

**Evaluating Comprehensive State Welfare Reforms: A Conference**

March 1997

The Joyce Foundation and the Charles Stewart Mott Foundation provided support for this project through a grant to the Institute for Research on Poverty. Opinions are those of the authors and not necessarily those of the sponsoring institutions.

# Evaluating Comprehensive State Welfare Reforms: A Conference

Institute for Research on Poverty  
University of Wisconsin–Madison  
November 21–22, 1996

## Table of Contents

List of Contributors .....	iii
Introduction: Evaluating State Welfare Reforms Thomas Kaplan .....	1
The Next Generation of Welfare Reforms: An Assessment of the Evaluation Challenge Thomas Corbett .....	7
Potentials and Problems of a Pre-Post Design for State-Based Evaluation of National Welfare Reform Robert Haveman .....	25
Outcomes of Interest, Evaluation Constituencies, and the Necessary Trade-Offs Maria Cancian and Barbara Wolfe .....	43
Controlled Experiments in Evaluating the New Welfare Programs Glen G. Cain .....	61
Prospects for Comparing Wisconsin Works to Welfare Reform Programs outside the State Irving Piliavin and Mark Courtney .....	73
Toward a Basic Impact Evaluation of Wisconsin Works Thomas Kaplan and Daniel R. Meyer .....	83
Wisconsin Works: A View from the Ground Thomas Corbett and Elisabeth Boehnen .....	117
Process Evaluation of W-2: What It Is, Why It Is Useful, and How to Do It Karen C. Holden and Arthur Reynolds .....	139
A Management Information System for Wisconsin Works Michael Wiseman .....	157
Monitoring Income for Social and Economic Development Martin H. David .....	183
Evaluating the Impacts of W-2 on Family Structure and Maternal and Child Health: A Brief Background Paper Gary Sandefur and Molly Martin .....	201
Evaluation of Child Care Services under the Wisconsin Works Program (W-2) Karen Fox Folk .....	211

Potential Effects of Wisconsin Welfare Reform (W-2) on Child Support Stuart Kipnis .....	223
Potential Effects of Wisconsin Welfare Reform (W-2) on Families Who Have Very Young Children with Disabilities and Special Health Care Needs George Jesien .....	231
Welfare Reform and Education John F. Witte .....	245
Welfare Reform and Child Welfare Services: Issues of Concern and Potential Evaluation Strategies Mark E. Courtney .....	253
Assessing the Impact of W-2 on Homelessness Irving Piliavin .....	267

## List of Contributors

Elisabeth Boehnen is Database Administrator, Institute for Research on Poverty, University of Wisconsin–Madison

Glen Cain is Professor Emeritus, Department of Economics, University of Wisconsin–Madison

Maria Cancian is Assistant Professor, School of Social Work and La Follette Institute of Public Affairs, University of Wisconsin–Madison

Thomas Corbett is Acting Director, Institute for Research on Poverty, and Assistant Professor, School of Social Work, University of Wisconsin–Madison

Mark Courtney is Assistant Professor, School of Social Work, University of Wisconsin–Madison

Martin David is Professor of Economics, University of Wisconsin–Madison

Karen Fox Folk is Associate Professor, School of Human Ecology and La Follette Institute of Public Affairs, University of Wisconsin–Madison

Robert Haveman is John Bascom Professor, Department of Economics, University of Wisconsin–Madison

Karen Holden is Associate Director, La Follette Institute of Public Affairs, and Professor, School of Human Ecology, University of Wisconsin–Madison

George Jesien is Director, Early Intervention Program, Waisman Center on Mental Retardation and Human Development, University of Wisconsin–Madison

Thomas Kaplan is Associate Scientist, Institute for Research on Poverty, University of Wisconsin–Madison

Stuart Kipnis is a Ph.D. candidate, School of Social Work, University of Wisconsin–Madison

Molly Martin is a Ph.D. candidate, Department of Sociology, University of Wisconsin–Madison

Daniel Meyer is Associate Professor, School of Social Work, University of Wisconsin–Madison

Irving Piliavin is Professor Emeritus, School of Social Work, University of Wisconsin–Madison

Arthur Reynolds is Assistant Professor, School of Social Work, University of Wisconsin–Madison

Gary Sandefur is Professor, Department of Sociology, University of Wisconsin–Madison

Michael Wiseman is Professor, Department of Regional and Urban Affairs and La Follette Institute of Public Affairs, University of Wisconsin–Madison

John Witte is Professor, Department of Political Science, University of Wisconsin–Madison

Barbara Wolfe is Director (on leave), Institute for Research on Poverty, and Professor, Department of Economics and Department of Preventive Medicine, University of Wisconsin–Madison



**Introduction:**  
**Evaluating State Welfare Reforms**

Thomas Kaplan

In August 1996, Congress and the president replaced the 60-year-old Aid to Families with Dependent Children (AFDC) program with a block grant, Temporary Assistance for Needy Families (TANF), permitting states to experiment with new forms of assistance to low-income families. States seem likely to try increasingly ambitious reforms as they gain experience under their TANF flexibility, at least if they respond as they did to federal permission to waive AFDC requirements. When President Bush first announced the new AFDC waiver options, states sought approval for only one or two proposals at a time. In the last months of AFDC, however, state waivers grew much more comprehensive. North Carolina, for example, secured a waiver that would allow the state to

- introduce an individualized personal responsibility contract;
- start a Learnfare initiative;
- require teen parents to live in a supervised setting;
- change the 100-hours rule in the Unemployed Parents program;
- increase the asset eligibility limits for AFDC recipients;
- expand the number of families required to participate in a JOBS program;
- increase sanctions for noncompliance with JOBS;
- impose time limits on receipt of AFDC benefits;
- require parents to obtain routine immunizations and medical examinations for their children.

Thomas Corbett's paper on "The Next Generation of Welfare Reforms" describes the nature of these reforms and some of their implications for evaluation. The comprehensiveness of such reforms and the potential for still broader change in the future create at least two challenges to evaluation. First, because components of a broad reform could be synergistic, the impact of a comprehensive reform may surpass the sum of the impact of each individual reform within it. If so, evaluators would understate the impact of a reform component *in a particular context* if they evaluate only the one component. Yet assessing a comprehensive reform in its entirety, especially if its elements change over time, can be harder than assessing components of it.

Second, comprehensive reforms create special difficulties for evaluations that utilize an experimental design, a standard way of determining cause and effect. The difficulties arise because comprehensive welfare reforms often bundle new requirements for school attendance, work, or a daily activity that leads to work, along with modified policies for adjusting family income according to family size, into packages designed to generate

social messages that discourage dependency on public assistance, promote work, and influence family-formation decisions. It is possible that such messages would be less effectively communicated in an environment of random assignment to control and treatment groups than in a more universal program. It is also possible that the control group would be influenced—and thus rendered invalid as a true control—by the messages conveyed to the treatment group.

Researchers have increasingly remarked on these and other problems of evaluating comprehensive welfare reform. But reaching consensus on evaluating such programs has proved elusive. In November 1996, IRP held a conference on the evaluation of the Wisconsin welfare reform program (Wisconsin Works, or W-2), in an effort to use one state welfare reform plan as a laboratory for evaluation. Although Wisconsin is not the only state with comprehensive reform, W-2 is among the most detailed and ambitious of the state plans. (Its basic features are summarized in Table 1.) IRP affiliates and staff believed that the program offered a worthy test of our ability to plan an evaluation of comprehensive welfare reform.

Intensive IRP work on the evaluation of W-2 started in the summer of 1995 and was supported by financial assistance from the Joyce Foundation and the Charles Stuart Mott Foundation. These were some of the issues that researchers confronted:

## THE DISTINCTION BETWEEN MONITORING AND EVALUATION

Monitoring involves tracking income, earnings, family composition, and other social and economic indicators as a new program is created. Evaluation goes beyond monitoring, to determine if a causal link exists between the reform and any changes observed in the indicators. Experimental evaluation designs, controlled so that the only difference between the experimental and control group is the new program, offer the standard method of determining how much the changed policies altered the relevant indicators. But if experimental design cannot be used, will any other design allow for evaluation, or is close monitoring of the indicators (a useful task in itself) all that will be possible?

Members of the IRP working group agreed that experimental design was inappropriate for evaluation of W-2, but differed over whether reasonable causal inferences could be made in the absence of experimental design. The majority maintained some optimism that at least limited causal attributions would be possible without experimental design. Comprehensive welfare reform, after all, is not the only policy intervention for which experimental evaluation designs are unsuitable. The effects on the economy of Federal Reserve Board policies, the deterrent effects on potential aggressors of new weapons systems, and the impacts of local



government efforts to encourage economic development are also unknowable through formal experimental design. Yet useful causal knowledge can accumulate concerning these and similarly complex interventions.

## NONEXPERIMENTAL COUNTERFACTUALS

In an experimental design, the causal link between the intervention and observed changes in indicators arises from the experimental structure: people or families are randomly assigned to the new treatment and to the status quo. If we have grounds for believing (1) that the people in each group did not differ in relevant characteristics upon their entry into the program, (2) that the treatment and control groups did not “contaminate” each other during the experiment, and (3) that the two groups were equally affected by other changes occurring at the same time, then any post-treatment differences between the two groups in the relevant indicators can be attributed to the interventions.

But if experimental designs cannot be used in the new welfare reform evaluations, a different “counterfactual” is required if we are to have any sense of what would have happened without the reforms. IRP researchers discussed seven other counterfactual possibilities, which we later determined to fall broadly into two approaches: *pre-post designs*, in which post-treatment indicators are compared to pretreatment indicators, and *cross-state designs*, in which Wisconsin indicators are compared to indicators in other states with different interventions. The advantages and disadvantages of these approaches are discussed in papers by Robert Haveman, Glen Cain, and Irving Piliavin and Mark Courtney. The paper by Thomas Kaplan and Daniel R. Meyer, “Toward a Basic Impact Evaluation of Wisconsin Works,” proposes a particular form of pre-post design.

## KEY INDICATORS

Regardless of whether or how causality is established, the populations and outcomes to be studied must be identified—issues discussed in the paper by Maria Cancian and Barbara Wolfe on “Outcomes of Interest, Evaluation Constituencies, and the Necessary Trade-Offs.” All members of the working group favored a concentration on the broad low-income population—those with incomes below 200 percent of the federal poverty line—not just on participants in the W-2 program. The reason for looking beyond program participants is the possibility that the program will have entry and exit effects—that is, the program itself may induce some people to enter, avoid, or drop out of it, in which case its full impact could not be inferred from observing only

program participants. There was less agreement on the possible outcomes of interest. Some members of our group favored concentrating on earnings, hours worked, and receipt of welfare. Others favored attention to a broader set of concerns, including the effects on family-formation decisions, health status, participation in health care programs, the incidence of child abuse and neglect, and the use of foster homes and other out-of-home placements. All of us wanted to give attention to “process indicators,” those indicators that would help program managers judge the quality of implementation and make timely policy adjustments based on early program experience. We did not reach consensus on what those process indicators might be, although the separate papers by Thomas Corbett, Michael Wiseman, and Karen Holden and Arthur Reynolds narrowed our differences.

## OTHER ISSUES

We discussed several other issues at less length during the months leading to production of the conference papers: how long the W-2 program should operate before it was subjected to impact evaluation; the dependability of various data sources; evolutions in the W-2 proposal itself; and whether—especially for pre-post analysis—W-2 changes should be deemed to include major policy modifications that the state of Wisconsin made within AFDC during 1996, or just the establishment in 1997 of W-2 itself. The papers by Martin David, Gary Sandefur and Molly Martin, Karen Fox Folk, Stuart Kipnis, George Jesien, John F. Witte, Mark E. Courtney, and Irving Piliavin discuss these and other issues in the context of assessing the potential impact of W-2 on incomes, family structure, child care services, child support, families containing young children with disabilities, education, child welfare services, and homelessness.

---

### Wisconsin Works (W-2): A Brief Description

The full W-2 program is scheduled to start on September 1, 1997, but transitional parts of the program have already begun. The two most significant early implementation phases of W-2 have been in place since March 1996:

- *Self-Sufficiency First.* This is a diversion program which requires applicants for AFDC to complete an interview with a financial planning resource specialist before applying for AFDC and then to participate in the state JOBS program during the 30-day AFDC application processing period. In these 30 days, the applicant must devote at least 60 hours to JOBS participation, including 30 hours of direct employer contact. If needed, child care is provided during the applicant's hours of JOBS participation.
- *Pay for Performance.* This is an intensive JOBS program requiring 20–40 hours per week of participation from AFDC case heads and imposing heavy sanctions on those who fail to comply. For each missed hour of required JOBS participation, a penalty equal to the federal hourly minimum wage is imposed, first on AFDC and then on Food Stamp benefits. Recipients who participate in JOBS for less than 25 percent of scheduled hours receive the full penalty, which reduces the AFDC grant for the next month to \$0 and the Food Stamp benefit to the federal minimum of \$10. Subsequent participation in JOBS restores benefits for future months.

The two transitional programs will phase into the full W-2 program, which will eliminate AFDC and replace it with cash assistance available only through work or participation in work-like activities. Under the full W-2 program, participation requirements will begin when the youngest child is 12 weeks. Families will be assigned to a Financial and Employment Planner, who places families on a level of a “self-sufficiency ladder” and helps them move up the ladder to greater independence, as indicated in the grid in Table 1. Small loans, which can be paid back in cash or community service, will be available in order to assist participants to find and keep work. Under W-2, there is no provision for subsidized formal education. Although some W-2 recipients will technically receive a grant, all cash income in the program will be disbursed on an hourly basis in return for each hour of work or program participation. Failure to participate will reduce income. Unlike the current AFDC program, the income will not depend on family size, but only on the case head's hours of participation and level on the W-2 self-sufficiency ladder. Also unlike AFDC, W-2 is not statutorily identified as an entitlement.

W-2 participants will also receive assistance with child care costs; this assistance will require copayments from recipients varying by recipient income and the licensing level of the care they select. All working families with incomes below 165 percent of the poverty line at program entry will be eligible for child care subsidies. Wisconsin has also requested but not at this writing received a federal waiver to replace the current Medicaid program for W-2 recipients with a W-2 medical program that would provide benefits generally similar to those available under Medicaid to a somewhat different population (some expansions and some contractions in eligibility) contingent on the payment of a premium. All W-2 services will be available to both single- and two-parent families with children under 18 and with incomes within 115 percent of the poverty line. W-2 will also eliminate the current practice under which child support income beyond the first \$50 in a month goes to public agencies to reimburse welfare expenditures. W-2 participants will be able to keep all the child support paid on their behalf.

---

**TABLE 1**

Level of W-2	Basic Income Package	Time Required of Recipients	Program Time Limits	Est. Child Care Copays (\$/mo.)	
				Licensed Care	Certified Care
Unsubsidized employment	Market wage + Food Stamps + EITC	40 hrs/wk standard	None	\$101–\$134	\$71–\$92
Trial Job (W-2 pays maximum of \$300/mo. to the employer)	At least minimum wage + Food Stamps + EITC	40 hrs/wk standard	Per job: 3 mo. with an option for one 3-mo. extension ; total 24 mo.	\$55	\$38
Community Service Job (CSJ)	\$673 per mo. + Food Stamps (no EITC)	30 hrs/wk standard; and up to 10 hrs/wk in education and training Total: 24 mo.	Per job: 6 mo. with an option for one 3-mo. extension	\$38	\$25
W-2 Transition (placement contingent on an assessment by the state Vocational Rehab. agency)	\$628 per mo. + Food Stamps (no EITC)	28 hrs/wk work activities standard; and up to 12 hr/wk in education and training	24-mo. limit, but extensions permitted on a case-by-case basis	\$38	\$25

**Sources:** Folk (1996) and presentation materials created by the Wisconsin Department of Workforce Development.

**Notes:** The income package and child care copayment are based on Governor Thompson's proposals in the 1997–99 biennial budget. Estimated child care copayments are for a three-person family with two children receiving no child support payments. Department of Workforce Development materials express child care copayments on a weekly basis; the monthly copayments shown in the table assume 4.2 weeks per month. For the purpose of estimating child care copayments, the Trial Jobs position is assumed to pay minimum wage, which, after October 1, 1997, will be \$5.15 per hour, or \$858 per month, and the unsubsidized-employment portion is assumed to range from \$6–\$7 per hour, or \$1,000–\$1,170 per month.

# The Next Generation of Welfare Reforms: An Assessment of the Evaluation Challenge

Thomas Corbett

## INTRODUCTION

Recent welfare reform legislation devolves a great deal of responsibility over the design and financing of welfare programs for poor families with children to the states. Based upon the welfare innovations carried out over the past decade under the waiver authority vested in the Secretary of the Department of Health and Human Services, there is reason to anticipate a fundamental change in welfare design and operations. The very mission and culture of welfare may well be transformed.<sup>1</sup>

There is every reason to believe that welfare devolution will profoundly affect how we evaluate policy and program innovations. In turn, our response to these challenges will dictate whether we exploit the opportunity to learn inherent in the explosion of experimentation throughout the states. Existing standards for doing *impact* evaluations based on classic experimental techniques may no longer be feasible nor warranted in all instances. Moreover, standards for conducting *implementation* and *process* analyses have never been well developed. They are increasingly critical given the complex character of recent reform initiatives. In short, the evaluation community must rethink its set of tools and strategies if it is to respond appropriately to the challenges laid out by devolution.

The papers in this volume focus on a dramatic new reform being introduced in Wisconsin called *Wisconsin Works* (W-2). Though W-2 is the object of attention throughout these papers, the purpose of the conference and this special report is broader and more fundamental. It is to begin the process of thinking through the challenges that the evaluation community must address and a dialogue through which appropriate responses might be crafted.

## THE ISSUES

Federal legislation enacted in August 1996, The Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA), transfers substantial responsibility over what had been the Aid to Families with Dependent Children (AFDC) program to the states. Key provisions of the Act convert public assistance to poor families with children from an open individual entitlement to cash support where the fiscal cost is

---

<sup>1</sup>The author is grateful for the contributions of Michael Wiseman.

shared by the federal and state governments to a state entitlement with a fixed or capped federal contribution. Since future assistance is to be time-limited, the AFDC program is now known as the *Temporary Assistance for Needy Families* (TANF) program.

Under the new legislation states theoretically will have greater flexibility in designing and managing their support programs for poor families with children. Greater flexibility brings increased responsibility and, given the nature of the new legislation, additional risk. There may be greater risk because, under PRWORA, the federal contribution is fixed and will certainly decline in value over time.<sup>2</sup> This suggests that states and local governments will bear the full fiscal risk of policy decisions *at the margin*, after the federal fiscal contribution is exhausted. Thus, states (or whatever the level of government program in which responsibility ultimately is vested) now will have a greater stake in obtaining a priori knowledge about the likely consequences of their policy decisions. If policy makers assume certain behavioral responses to a reform and guess wrong, they easily could incur substantial costs or be required somehow to ration services and benefits as budgets come under increased scrutiny.

In an environment of such change and challenge, program evaluations and other forms of analyses become critically important. Arguably, knowledge is more valuable in a devolved policy environment, since the decisions vested in local governments are both more complex and more consequential. But though the *value* of knowledge is greater, the *price* of obtaining knowledge also will increase. The federal government will neither mandate that evaluations be done, nor ensure that certain methodological standards be maintained, although it will continue to play a role (see below). Many jurisdictions will hope to be free riders, letting others incur the fiscal and other costs of doing evaluations while still accessing the results. Even if that were not the case, evaluations of the *next* generation of reforms would be challenging because of the character of emerging programs and policies.

As discussed in more detail below, many states emphasize varying combinations of what Lawrence Mead (1995) has termed “new paternalism” programs. These programs, which often require school attendance, work, or a daily activity that leads to work, are bundled into varying packages designed, in their totality, to discourage dependency on public assistance, influence family formation decisions, and promote work. State officials implementing these packages expect their full reform strategy to yield greater impacts than would be assumed from the sum of the impact of each individual reform, and they worry that separate evaluations of program components may understate overall impacts. They also fear that the social messages of the “new

---

<sup>2</sup>Even assuming modest inflation rates, the \$16.4 billion capped federal commitment may decline by 20 to 25 percent in inflation-adjusted terms over the course of the legislation.

paternalism” will be less effectively communicated in an environment of random assignment of individuals or agencies to control and experimental groups than would be possible under universal reform.

Simply put, the *next generation* of reforms are designed to change behavior rather than provide income support, and thus require a whole new relationship between agency and client. They are complex, subject to continuing change, and often call for radical alterations in agency *culture* which, among other things, decentralize decision making, rendering management control more problematic. Classic experimental methods, which work best in situations of limited change within stable program environments, may neither be feasible nor warranted in this emerging policy milieu.

These and related challenges to the evaluation community have come to dominate discussions about program evaluative strategies.<sup>3</sup> Over the past year, interest in developing a coherent response to what might be called *the evaluation challenge* has emerged and become more pressing.

## DEVOLUTION AND THE EVALUATION CHALLENGE

The most recent welfare reform debate focused on the “devolution revolution” theme. *Structural or legislated* devolution calls for the transfer of control over welfare to the states in the form of block grants. Proponents argued that the federal government could not solve the vexing problems of welfare, poverty, and teen pregnancy. This form of devolution encompassed the following themes: (1) terminate the individual entitlement status of selected income maintenance and service programs; (2) group related programs into broad program areas, defined either by common target populations or common service technologies; (3) convert federal contributions for programs from matching formulae into closed-ended block grants so that, on the margin at least, costs are no longer proportionately shared among local and federal governments; and (4) eliminate or reduce the federal role in rule making, evaluation, and technical assistance.

The advent of PRWORA partially realizes the vision of a “devolution revolution” and reveals the willingness of both the Congress and the president to endorse block grants as the means for federal cost sharing in public assistance. This extraordinary development is the result both of widespread support for reform and a public perception that states have the recipe for making welfare reform work. This belief is in part the product of state innovation in welfare program operation that began in the early 1980s and, encouraged by subsequent presidents, has continued at an accelerating tempo. President Clinton approved more state experiments

---

<sup>3</sup>For example, see *Focus*, vol. 18, no. 1 (special issue 1996), “Monitoring the Effects of the New Federalism.”

involving AFDC, Food Stamps, and Medicaid than were approved over the entire preceding 12 years of Republican administrations. Waiver-based welfare innovation thus constituted the second reform theme sometimes referred to as the *incremental reform strategy*.

Until PRWORA, federal approval was required when a proposed innovation deviated from the requirements of the Social Security Act of 1935 as amended. Section 1115 of the Act permitted such experiments when the Department of Health and Human Services (DHHS) judged them “likely to assist in promoting the objectives” of the AFDC or Medicaid programs. All approved state demonstrations were championed on this basis and all were expected both to contribute to our stock of knowledge and to serve as input into the development of *national* policy. These *waivers* were granted if the state could legitimately argue that the reform would not cost the federal government more than it would have otherwise spent (the cost-neutrality constraint) and that a rigorous evaluation would be done, rigor was virtually always defined as employing an experimental design with families as the unit of analysis.

The outpouring of investigation associated with waiver-based demonstrations promised significant contributions toward understanding the dynamics of public assistance and for identifying empirically supported programmatic and management improvements. Prior to PRWORA, both Republican and Democratic administrations supported incorporation of an explicit evaluation scheme as a requirement for waiver approval. The promised evaluation would seem to assure the social policy equivalent of a Consumer Reports special issue on techniques of welfare reform that will allow both program and policy comparison and identification of the welfare equivalent of the Consumer Union “best buy.”

This happy state of affairs may not transpire, we believe, for several reasons. First, many of the experiments in progress contain serious flaws in content and assessment strategy that, should they be completed, will diminish both the management utility and the internal and external validity of the outcomes. Second, PRWORA changes both federal and state incentives for experimentation in ways that diminish the likelihood that most ongoing experiments will be completed. Third, the character of future reforms may be so different from the changes examined in the past that the new demonstrations will not provide much proper instruction about how to proceed in the future—a true *discontinuity* in policy making.

#### POLICY CONTINUITY OR DISCONTINUITY?

The federal role in providing income assistance to poor families with children had been established under Title IV-A of the 1935 Social Security Act, which provided fiscal resources enabling states to assist



needy children without fathers. The states defined “need,” set their own benefit levels, established within federal limitations the income and resource limits for eligibility, and administered the program. Title IV-A was not particularly controversial, as it accorded fiscal relief to hard-pressed states while allowing substantial state flexibility and control over the program. But increased federal regulatory oversight eventually followed, particularly in the late 1960s and 1970s, when stronger rights were attached to beneficiaries and the benefits themselves came to be viewed as an entitlement. Though states always retained some program control—to set welfare guarantees, for example—the locus of control over policy development clearly had shifted to Washington by the early 1970s. Full federal control almost became a reality in the form of President Nixon’s Family Support Act (FAP) which in effect would have instituted a federally guaranteed income floor under all families.<sup>4</sup>

Within a decade, the debate over where to locate program control was engaged once again. Since 1962, the Secretary of the U.S. Department of Health and Human Services has had authority to waive statutory program requirements to permit states to experiment with program elements in order to inform national policy about how to better achieve the objective of aiding children. Until the late 1980s, these Section 1115 waivers were seldom sought by states and rarely granted by federal authorities. That situation began to change, first with the passage of the 1981 Omnibus Budget Reconciliation Act (OBRA), which gave states some flexibility in designing programs to move recipients from welfare to work, and then in 1986, when the Reagan administration gave encouragement to state interested in pursuing waiver experimentation.

Waiver authority use increasingly resulted in the devolution of program authority over welfare from Washington to the states. Rather than being granted primarily to add knowledge to national policy deliberations, by the 1990s waivers were being granted to accommodate state preferences concerning policy goals and program strategies. This transformation of purpose unleashed a torrent of state activity and experimentation. In addition to the historical effort directed at enhancing the labor supply of AFDC adult caretakers through work-related policies and programs, other recipient behaviors became the focus of attention. These included personal decisions about marriage and cohabitation; decisions affecting family stability (e.g., divorce and other family composition changes over time); fertility decisions; and the quality of parenting. The thrust of reform activity was increasingly directed toward a new strain of social engineering: employing welfare innovations as strategies for influencing those behaviors society deems important.

Ironically, devolution of program and fiscal responsibility was an established fact, accomplished by means of the waiver process, by the time that the welfare bill was signed by the president. Arguably, any

---

<sup>4</sup>FAP failed, but welfare for the blind, disabled, and aged was largely federalized and the Food Stamp program was extended to all jurisdictions, creating what some regarded as an in-kind negative income tax.

remaining pretense to a national welfare policy, other than the procedural requirements associated with the cost neutrality and evaluation requirements, disappeared sometime during the first two years of the Clinton administration. Spurred on by the rhetoric of the 1992 campaign to “end welfare as we know it,” states became bolder in their waiver requests and the federal government clearly signaled that it would seriously entertain dramatic proposals for reform. States were clearly calling the shots by the time *PWRORA* was enacted.

Since waiver-based activity may provide clues to future policy directions, salient trends from the pre-*PWRORA* period are reviewed. First, the *scope* and *pace* of waiver activity demonstrated just how extensive devolution was, even before it was formally legislated in August. Over 90 percent of all states and the District of Columbia had at least one approved waiver. The new legislation provides that waivers already granted continue in effect for the duration of the originally granted time. Moreover, the *pace* of state level activity has been increasing.<sup>5</sup>

The *complexity* of state-based welfare demonstrations has also increased. In the early days, a state would request permission to modify a few provisions of the Social Security Act in order to implement one or two new ideas. In recent years, the number of major changes to program parameters contained in a single waiver request would be in double figures. States increasingly would “borrow” ideas from other jurisdictions and *bundle* them together in complex reform packages. Delaware’s Better Chance initiative, implemented in 1995, is an example of the increasing complexity in welfare waiver demonstrations. This program contains at least 18 major provisions, including a family cap, a 100-hour rule change, time limits, income disregards, a contract for program participation, changes in JOBS requirements, and school attendance and immunization requirements. The Virginia Independence Program, also implemented in 1995, bundles similar provisions, including term limits, family caps, a contract for program participation as well as many other provisions intended to affect participant behavior.

It took a while for states to take advantage of the flexibility offered through a liberalized Section 1115 policy, but initially tentative experimentation became bolder as state officials gained greater confidence.<sup>6</sup> Moreover, the political popularity of reform could not be ignored by the nation’s governors. Whether justified

---

<sup>5</sup>For a discussion of welfare waivers, see Elisabeth Boehnen and Thomas Corbett, “Welfare Waivers: Some Salient Trends,” *Focus* 18, no. 1 (special issue 1996): 34–37.

<sup>6</sup>States may have *less* flexibility now than they did under prior welfare law and a liberalized waiver policy. *PRWORA* is a very inconsistent piece of legislation where devolution of authority over welfare to states is offset by many prescriptive provisions respecting either program objectives that must be met or expenditures of federal dollars which are prohibited.

or not, claims of state successes gained widespread media attention while national solutions, such as President Clinton's *Work and Responsibility Act of 1994*, sputtered in Congress and quickly disappeared from sight.<sup>7</sup>

Perhaps the most important trend evidenced in recent state demonstration activity involves a shift in program objectives. The demonstrations increasingly stress changes designed to alter critical personal and interpersonal behaviors with proposed welfare reform changes concentrating on helping (or obligating) people to play by society's rules: getting a job, getting married, making responsible fertility decisions, being a good parent, and obeying the law. For example, as late as 1992, only a couple of states has shown interest in the *family cap* concept where benefits would not be raised in cases of additional births while the mother was on AFDC. By the summer of 1996, about 40 percent of all states had actual or proposed family cap provisions and some 18 states had (or had proposed) provisions that required minor parents on assistance to live with their parents or in a supervised setting as a condition of eligibility. Finally, about two-thirds of all states were involved in reforms that explicitly linked welfare benefits with *good behaviors* such as students attending school regularly and parents ensuring that their children had all required vaccinations. Social engineering, using AFDC policy to inform and shape a variety of behaviors, had become a primary goal.

In the old welfare world (just a few years ago), change was limited and incremental and linear. The tools for learning from these experiments had been fairly well developed and standardized by a handful of the top evaluation firms.<sup>8</sup> The new world confronts us with a discontinuity in policy making and imposes new expectations upon the conventional methods found in the evaluation community.

## EVALUATION METHODS AND THE NEXT GENERATION OF REFORMS

There is considerable debate about whether program devolution will stimulate or inhibit innovation and knowledge building.<sup>9</sup> On the one hand, states might continue to experiment with new policies and program

---

<sup>7</sup>Clinton's proposed national welfare reform was submitted to several committees on June 21, 1994, after more than a year of development. The proposed legislation, including explanatory text, was 622 pages in length. It was too long and complicated to get a fair hearing in a Congress that was increasingly concerned with mid-term elections. The 1994 Congressional election results buried any hope for the President's proposal and dramatically shifted the debate.

<sup>8</sup>Basically the Manpower Demonstration Corporation, Mathematica Policy Research, Abt Associates, Urban Institute, and a handful of others. They refined basic approaches that had been developed partly at IRP in the late 1960s and early 1970s, when the methods for evaluating the Negative Income Tax and Supported Work proposals were developed.

<sup>9</sup>While devolving authority over AFDC and other welfare programs to the states has promising features, the notion raises legitimate concerns; these are discussed in Thomas Corbett, "The New Federalism: Monitoring

forms given that they are freed from most federal regulations, though the legislation passed in August remains quite prescriptive. On the other hand, states might respond with less risk taking behavior, given that additional fiscal risk is shifted to the states. A third possibility is that states will initially engage in innovative behavior, given that many will experience a fiscal windfall in the short term. After a while, however, they will become more conservative as the federal contribution declines in real terms.

There may be reasons to be less than optimistic about whether the explosion of welfare demonstration activity will lead to more theoretical and practical knowledge. The federal government will no longer mandate evaluations or be in a position to require rigorous evaluation designs. DHHS will be able to support evaluation activity to some extent, but resources will fall short of requirements.<sup>10</sup> In addition, the evaluation standards developed by major firms such as the Manpower Demonstration Research Corporation, ABT Associates, Mathematica Policy Research, or the Urban Institute may evaporate if those firms no longer dominate the market. They are expensive to hire and may become a luxury for states making difficult budget choices down the road.

At this time, we do not know how such factors will play out in the future. We do know, however, that trends in policy and program development already have imposed severe challenges on existing evaluation standards and practices.

Over the past two decades, a good evaluation was defined by the use of an experimental design. This approach proved particularly powerful when the interventions of interest were relatively simple, when counterfactuals were easily developed, and when the outcomes of interest were limited in number. That simple world no longer represents policy realities. Of course, the past may not be prologue, and state decisions under TANF may not be predicted by pre-TANF waiver activity. But if pre-TANF waiver activity is a guidepost to the future, the next generation of reforms just might look something like the following:<sup>11</sup>

- The emerging generation of policy/program interventions are *complex* and typically introduce numerous changes simultaneously or introduce changes outside of the welfare framework.<sup>12</sup> The new

---

Consequences,” *Focus* 18, no 1 (special issue 1996): 3–4.

<sup>10</sup>PRWORA sets aside \$7.5 million per year to support state evaluations of their welfare reforms. An RFP to help select recipients of these resources was issued in Fall 1996, with responses due January 15, 1997. Some 30 states submitted 43 separate proposals totaling about \$20 million.

<sup>11</sup>Some of the ideas are explored more specifically in the context of W-2 in a later paper in this Special Report, “Wisconsin Works: A View from the Ground.”

<sup>12</sup>Simply put, welfare-type programs target benefits on the income and asset poor and on vulnerable families (e.g., single-parent families with children). Such targeting inevitably creates perverse incentives and so many recent reforms, including W-2, attempt to lessen the extent to which targeting is a policy attribute (e.g., the attempts in W-2

interventions are also characterized by increased policy *volatility* (continuous change across time) and *variability* (significant differences across jurisdictions or agencies within a state).

- The conceptual and temporal bases for thinking about what is going on are dramatically altered. Traditional public assistance programs employ monthly accounting units, each month, for most purposes, represents a new temporal period of interest. The new generation of reforms tends to be change oriented: the participants are expected to evidence different behaviors as a function of their experience in the system. That is, the conceptual basis for looking at a case is transformed from a *point-in-time* perspective to a *point-in-process* perspective.<sup>13</sup>
- The new reforms intend to affect a more diverse set of *actors*. There are several target populations: adult caretakers in participant families (the usual suspects); children, and non-custodial parents. W-2, however, is designed to change the attitudes and behaviors of those operating the program (agency culture change) in addition to selected populations in the community (e.g., employers, child care providers, etc.). Thus, understanding what each of these groups “experiences” is important.

Most new-generation reforms are designed to alter a range of personal and interpersonal behaviors, in sharp contrast to traditional welfare programs, whose purpose was to provide some measure of income support and thus redress immediate economic need. The primary roles of front-line welfare personnel were to process information; to validate those data necessary to issue checks in the correct benefit amounts to eligible families. The time frame was the monthly accounting period, each month representing a new and distinct set of calculations. Particularly in states like Wisconsin, where much of the actual case decision making had been automated as far back as the 1970s, little discretion remained in the system.<sup>14</sup>

There are some qualifications to this characterization of welfare administration. Various welfare-to-work initiatives had been around virtually from the time that AFDC had been transformed into an entitlement system in the late 1960s and early 1970s.<sup>15</sup> These activities were, in theory at least, designed to change

---

to decouple child support, child care, and health care from welfare status).

<sup>13</sup>This distinction between these perspectives was, to my knowledge, first developed by Michael Wiseman while working with Kenosha County, Wisconsin, on the development of its innovative Welfare-to-Work Jobs Center.

<sup>14</sup>The Computer Reporting Network (CRN) was developed in the 1970s. Prior to that, the governing rules of welfare in Wisconsin permitted a great deal of discretion. CRN automated virtually all core decisions regarding the determination of eligibility and the calculation of benefits. Permissive language in the welfare manuals—the use of words such as *may*—were replaced by the use of nondiscretionary words such as *shall*.

<sup>15</sup>Many do not fully realize that welfare administration from the 1930s through the 1960s was highly discretionary and personalized. The mission was multi-dimensional, to ensure that a variety of behavioral conditions were met as well as to dispense assistance. Welfare was not an entitlement but conditioned on character and behavior.

behavior. However, they often were viewed as tangential to primary welfare functions and were contracted out to service providers located outside the welfare agency boundaries.<sup>16</sup>

In all likelihood, the next generation of welfare will not be an extension of what we had in the past. Rather, it will constitute a qualitative change from what had existed just a few years before.

- The new welfare systems will be *behavior-oriented* rather than *income-oriented*. By extension, the dominant treatment modalities will shift from income support strategies to service technologies. In this context, service technologies are broadly understood. What is actually delivered may not be traditional social work services as conventionally understood.<sup>17</sup> But there will likely be a shift toward case management, crisis management, and counseling activities that resemble the social work functions that marked the provision of public assistance some three decades ago.
- The new welfare systems will be *dynamic* rather than *static*. They are oriented toward change. They take a family at a baseline status and actively work toward changing behavior and attitudes. W-2, for example, explicitly talks about participants ascending the multiple tracks (or tiers) built into the system. The assumption is that participants will progress or move up the tiers before “graduating” into the labor market and mainstream society.
- The new welfare regimes will be *longitudinal* in character. That is, participants’ status in the program will not be totally dependent upon their current monthly financial and categorical circumstances. The nature of their involvement will be determined, in part at least, by their experience in the program and the objectives set for (or with) them. That is, participants are viewed as being in a *process*. All participants will be subject to time limits, both within program components and in an overall sense. The workers and the system must remain sensitive to where they are relative to these temporal constraints.
- The new welfare regimes will be *craft* oriented rather than *routinized*. Under the old welfare, some of the decisions were fairly complex, but the basic intent was to treat all participants alike—a rough justice. The new generation of reforms are designed to treat participating families as individuals, or in individualized ways. Many involve the negotiation of individualized “social contracts” or “individualized employment plans.”
- The new welfare systems will be *multidimensional* rather than *unidimensional*. An extension of the principle of individualization is the notion that all participants will not proceed through the welfare experience in lock step. They are likely to be tracked along different paths. Differential *tracking* suggests that important decision points exist where key *triage* choices are made—the assignment of

---

<sup>16</sup>The Kenosha County Jobs Center model achieved significant notoriety because it fully integrated welfare to work and traditional welfare functions. Among other things, the Kenosha model co-located welfare staff with JOBS and Job Service staff, establishing a common caseload.

<sup>17</sup>Social work services were an integral part of welfare systems when it was argued that recipients might be rehabilitated or counseled out of dependency. This was particularly true in the 1960s, though the rehabilitation theme can be traced back to the Scientific Charity movement of the 1880s. Income maintenance functions were formally separated from service functions in the early 1970s.

participants along distinct program trajectories that implicitly recognize diversity within the welfare population.<sup>18</sup>

- The new welfare systems will be characterized by complex decisions that will require a good deal of *professionalism* and involve a good deal of *discretion*. There are three ways to sort out the basic ways personnel are recruited, trained, and permitted to operate. One is to control front-line behavior through organizational structure (pyramid shape, vertical communications), management style (detailed manuals, strong supervision), and training methods (encourage professional standards). The other is to recruit professionals rather than technicians who bring in their own standards and expertise and are likely to resist highly controlling organizational regimes. The third is to move to performance-based management or to engage in privatization.
- The new welfare systems will be *labor intensive*. The old welfare involved repetitive, routine decision making. The emphasis was on efficiency and accuracy. Participants who wanted help were referred to other systems. Not surprisingly, administrative costs often were less than 10 cents on the dollar of benefits issued. W-2 and similar reforms will require intensive case management and a very active rather than passive participant-worker interaction.

This paradigm shift inevitably is accompanied by differences in management and oversight. Standardized, routinized rules and procedures lend themselves to stable policy environments and vertical, hierarchical management structures. The new welfare policy environments are neither very stable, nor do they lend themselves to top-down or vertical control.<sup>19</sup> It becomes more difficult for states to centrally manage and prescribe all aspects of agency operations. Consequently, *intrastate variation* can begin to rival interstate variation as a descriptor of the next generation of reform.

## WISCONSIN WORKS: THE CASE STUDY

The subsequent papers use *Wisconsin Works* (W-2) as the policy example through which to explore the evaluation issues posed above. Arguably, no other state reform captures the attributes and spirit of the next generation of welfare reform as does W-2. Because W-2 is so complex and ambitious, finding ways to rigorously evaluate process is worth the effort. Consequently, we also argue that W-2 is a suitable, though far

---

<sup>18</sup>The Wisconsin Works reform plan is based on four tracks or separate program paths participants might take. The origins of the four W-2 tracks might be traced back to the development of the Kenosha County (Wisconsin) Jobs Center in the late 1980s which developed a similar typology. See Corbett, "Child Poverty and Welfare Reform: Progress or Paralysis," *Focus* 15, no. 1 (Spring, 1993): 1–17.

<sup>19</sup>In the last decade, Wisconsin, for example, has obtained 15 relatively major waivers and has launched other changes not requiring waivers. Recently, it has begun upon two major reforms, Self-Sufficiency First and Pay for Performance. The evaluation contract has just been let to evaluate these reforms and the state is already considering how to fold these initiatives into the next generation of reforms, to be implemented in less than 18 months.

from perfect, laboratory for evaluating program implementation questions.<sup>20</sup> Although the nature of W-2 may preclude the use of conventional experimental evaluative designs, and its advanced state (a number of provisions are already in place) precludes the introduction of pre-post/cross site designs that will capture the full effects of the intervention, there remains much to learn about how well the initiative is working.

Wisconsin Works (W-2) symbolizes the logical consequence of the waiver dynamics evidenced in recent years and thus represents a classic example of the next generation of reform. W-2 purports to end AFDC by transforming virtually every aspect of the AFDC program and essentially proposes to develop supports for poor families with children outside of the welfare framework. It ends the individual entitlement to any cash assistance and replaces most cash benefits with time-limited, work-based supports. It decouples child support from AFDC in the sense that support payments will fully go to those families rather than to offset program costs. It decouples those portions of Medicaid and child care subsidies that were formerly linked to AFDC from the welfare program. It changes the whole mission and purpose of agencies and redefines the tasks of operational personnel. Finally, W-2 terminates the government monopoly on the management of welfare programs by, in some circumstances, selecting the local management teams on the basis of competitive bids open to for-profit and nonprofit firms.

We have no way of predicting in advance the net result of W-2 on the well-being of low-income families in Wisconsin and on their communities, labor markets, and the systems that provide critical services. The point of this project, and these papers, is not to evaluate W-2, which is merely emblematic of a fundamental transformation in the public response to the economic and related needs of low-income children and their families. This transformation has occurred so quickly, at least relative to the glacial pace of typical policy change, that our capacity to learn from that change and assess effects on vulnerable populations is severely tested. But it is crucial that we think through the methodological issues to inform and shape evaluations of the next generation of reforms.

## FUTURE DIRECTIONS FOR RESEARCH AND POLICY EVALUATION

The conventional approach to evaluation has been to change one or two parameters of the existing welfare program, randomly assign participating families into either an experimental or a control group, and examine “net” outcomes on a limited number of measures that virtually everyone agreed was important. None of that holds any longer. The changes are broad in scope and complex in character; the target groups of interest

---

<sup>20</sup>For an excellent treatment of administrative issues, see Lawrence Mead, “Welfare Policy: The Administrative Frontier,” IRP Discussion Paper 1093-96, University of Wisconsin–Madison, 1996.



numerous; the outcomes of interest even more numerous. Policy designers want to radically alter agency cultures and communicate a whole new set of messages to low-income communities. If anything, random assignment would mute the message and make changing agency culture more difficult. Randomly assigning agencies raises migration concerns. In effect, it is a new ball game.

The next generation of welfare reforms could well have its largest impacts through changing the social norms of society—by increasing the social stigma attached to out-of-wedlock childbearing, repeat childbearing among those already dependent on welfare, and childbearing among those unable to support their children. Indeed, the explicit arguments for many of the reforms relate to their power in promoting family values and work ethic. Thus from the parochial perspective of the states implementing the reforms, there is little reason to look beyond whether welfare rolls decline and whether fewer children being reared out-of-wedlock. From a broader social view we ought to adopt a more comprehensive and longer range research and evaluation plan involving multiple goals and assessment strategies.

In consequence, the evaluation challenges are daunting. How does one establish a *counterfactual*? How are the correct *criterion variables* selected? How does one agree upon which *target groups* to examine? What is the appropriate *unit of analysis*, individual or case or agency or county? How does one go about determining *overall* and *component* effects? Should the implementation analysis be used in a formative way, if that increases policy instability and confounds the impact analysis? Should local discretion and flexibility be curtailed, so that the character of the intervention might better be understood? There are many such puzzling questions that evaluators must confront in the future, not all of which lend themselves to simple technical solutions. As the need for good, empirical information increases, the cost and difficulty of obtaining those answers increase commensurately.

Any future evaluation agenda should cover not only an assessment of the success of the reforms in meeting their central objectives, but also other possible consequences and the mechanisms through which both intended and unintended consequences occurred. For example, while traditional reforms and their evaluations focused on program participants, the next generation are intended to affect communities. In consequence, population and entry effects are of considerable importance. We also need to understand the context for the findings, insofar as this affects both their interpretation and generalizability to other settings.

Clearly, the evaluation challenges raised in this “age of policy discontinuity” are several and serious. As a convenient way of viewing some of these, we can locate a number of the issues along three axes: (X) major methodological questions; (Y) substantive policy changes; and (Z) alternate foci of concern (or units of analysis).

- \* **X axis.** The major methodological questions are typically organized in three parts. First, we want to assess implementation and process. Carefully assessing how closely operations reflect policy intent is critical, particularly with complex interventions requiring complicated sequencing of tasks carried out by varied actors representing different institutions. Otherwise, we wind up with “black box” evaluations that are difficult to interpret. Second we want to assess benefit/cost or cost/effectiveness ratios. Do benefits exceed costs or, at a given cost, do alternative strategies produce differential outcomes? Third, we want to assess “net” outcomes or impacts. This requires comparing outcomes for those exposed to the intervention with an appropriate control or comparison group (the counterfactual), created through random assignment of individuals, sites, or through statistical procedures designed to account for heterogeneity between groups.
  
- \* **Y axis.** The new generation of complex, multidimensional social policy innovations are intentionally designed to affect several aspects (or domains) of the lives of those exposed to the program. This is a dramatic departure from the limited, linear framework within which earlier demonstrations might be understood. Then, one or two welfare parameters were modified to effect changes in a single domain—perhaps a modest sanction intervention to motivate changes in school attendance. W-2, for example, arguably is designed to effect changes in the following domains: labor supply, skill or human capital development (largely through increased labor force participation), health status, child care arrangements, fertility decisions, family formation and functioning, and child development. Each of these areas raises questions about what outcomes are critical, how the outcomes can be operationalized, what data sources exist and how good they are. Multiple domains also raise questions about interpreting results that may go in contrary directions.
  
- \* **Z axis.** The new generation of reforms raises questions about the appropriate unit of analysis. Are we most interested in the adult caretaker, the family as a whole, the child(ren), the community, the workplace, the service providers? Again, as the demonstrations become more ambitious and dramatic, the anticipated effects are likely to be felt at all of these levels. IRP (and its affiliates) have long been concerned about this, and organized a national conference some years ago to discuss the complexity of moving from one-generation evaluations (adult-only) to two-generation (adult and child) evaluations. Along this axis, we envision different issues and challenges associated with measuring effects on the following: the adult caretakers(s); the family as a whole; the child, community, and labor market effects; and institutional/service provider effects.

We thus have a set of evaluation issues and challenges organized in a three-dimensional perspective. This sounds more complicated than it really is. In short, the new generation of evaluations must know what is happening (process) that leads to what outcomes (impacts) at what level of investment (benefits costs). They want to know this at the individual, family, institutional, and community levels, and across several domains (economic, behavioral, developmental).

Within each of these, the list of issues is fairly extensive. Illustrations will have to suffice. Process (or implementation) evaluations have traditionally been the stepchild of impact evaluations. They have become more important as a host of “black box” welfare-to-work evaluations done in the 1970s and 1980s raised questions about what really was being tested (if anything). But the craft of doing these well is still in its

infancy. There are few accepted protocols for how to collect data on complex operational systems, or how to report out the findings. The objectivity and comparability of process evaluations must be enhanced.

Impact evaluations, on the other hand, raise important questions about how to establish a counterfactual (no exposure to the program) when the state wants to saturate the county or state. States often have good reasons for this (interest in changing community norms, or worries about administrative complexity, or migration effects, etc.). However, statewide programs do complicate the task of making causal inferences.

Recognizing that effects may be evidenced in different domains is one thing. Measuring changes in different domains is quite another. We have pretty good data on welfare utilization and wages earned (but not hours worked or fringe benefits) every three months. Data collection and quality, as well as the interpretability of data, become more problematic as we move toward broader measures. More thought must be given to how best to move beyond the conventional outcomes of welfare utilization, economic well-being, and labor force participation.

Conventional evaluations often have focused on one generation, the adult who is already a program participant (on welfare). Moreover, the focus is on the transition off assistance (or perhaps less reliance on welfare). The new reforms break many of the old conventions. They radically change the meaning of participation, alter the eligible population, expect to transform community norms, individual and family behaviors, and agency cultures. At a minimum, this means more attention to community effects, entry effects, institutional effects, and perhaps even long-term labor market effects.

Some evaluation issues would cut across any classification scheme. For example, at all levels of analysis, there is the data question. Critical data may not exist, or exist in the right form, or be credible and reliable. The data issue will be explored in relation to the questions within each axis above; revisions in extant data sources will be discussed; and new data collection recommended, where appropriate.

## WHAT NEXT?

Neither the papers prepared for the November conference, the conference itself, nor any subsequent products such as this Special Report answer the evaluation challenges raised here. That will take a long, multifaceted dialogue involving many actors and several strategies. Since the conference, a number of analysts from within and without IRP have been discussing a set of options for pursuing resolutions to the challenges

facing the evaluation community.<sup>21</sup> Below, we introduce some of the primary strategies that have surfaced so far.

### Developing Common Standards and Practices

In the absence of a strong federal role in ensuring the application of rigorous methods to program evaluations, it has been proposed that a series of conferences and workshops be conducted to work toward a convergence of methods and practices in the following areas.

- **Common terms and definitions.** Even where everyday terms are used, such as “case,” or “successful outcome,” or “full employment,” or “noncompliance,” great variation in meaning may occur in a devolved policy world. Without agreement on the meaning of terms, cross-state comparisons will be difficult.
- **Common research questions.** All evaluations start with research questions; management and theoretical issues about which empirical information is desired. The questions posed inform all other features of the evaluation—choice of outcomes, choice of methods, and so forth. If a consensus (or movement toward a consensus), could be developed around a set of management and research questions, more comparability across evaluations would be likely.
- **Common outcomes of interest.** Defining success in a devolved policy world will be a major challenge and a likely point of contention. Those who select the outcomes of interest and choose how to operationalize those outcomes greatly influence the political agenda of reform. Although local preferences should not be ignored, a core group of common outcomes, consistently operationalized, is needed to anchor state and local evaluations and permit analyses across jurisdictions.
- **Common reporting of findings.** Several evaluations conducted by the Manpower Demonstration Research Corporation were influential in part because they reported findings in a predictable and understandable manner. More varied evaluations by an increasing number of evaluators are likely in the future, and standards for reporting results would be helpful.
- **Consensus regarding acceptable methods.** Both *process* (formative) and *impact* (summative) evaluations are difficult to do and both are important. There have never been agreed-upon methods for doing process analyses; consequently, studies permitting cross-site comparisons are not common. Since the classic experimental design may no longer be the *sine qua non* for doing impact analyses, what alternatives are reasonable and acceptable for establishing causal inferences? Establishing minimal standards may not be a reasonable goal; moving toward such standards may be a necessary objective.

### Developing “Expert” Review Panels

---

<sup>21</sup>The key persons from outside IRP who are involved in these discussions are Larry Aber and Barbara Blum of the National Center for Children In Poverty; Howard Rolston, Ann Segal, Don Oellerich, and Matt Stagner of the U.S. Department of Health and Human Services; Rebecca Blank of the Joint Poverty Center at Northwestern/Chicago; and Barry Van Lare and Rikki Kramer of the Welfare Information network.

Program evaluations of reform initiatives will inevitably vary in quality and utility despite efforts toward some consensus. Consequently, it has been proposed that a pool of expert reviewers be established to examine the more important evaluations, for the following purposes:

- To isolate which results and findings from an evaluation might be accepted with some degree of confidence. This would be an effort to sort out credible findings from those where methods or management of an evaluation raise doubt.
- To identify and relate credible findings across diverse evaluations in order to build up a body of knowledge based on common findings in diverse program and situational contexts.
- To identify findings that seem contradictory and thus may constitute research and management issues that require further study. This is an attempt to identify the cutting edge issues requiring analytic attention.
- In general, articulate those issues that require additional study, and possibly suggest ways they might be examined. That is, the panel should push the envelope in terms of issue identification and resolution.

The composition of the panel is critical. Clearly, academics drawn from affiliates of the two Poverty Centers would be likely participants. But methodological input is only one dimension of interest. We also want the consumer's perspective, and perhaps even input from the most common producers of program evaluations—the top evaluation firms. Thus, we potentially see panels drawn from academics, state evaluation and policy units, evaluation firm, and the federal government. Mechanisms for ensuring cross-panel communication would be developed.

There may be some urgency to moving on this particular strategy. There are a number of evaluations of waiver-based reforms that will be issued over the next 12 months. A critical review of these evaluations would both enhance their utility and help in the process of improving the next round of PRWORA-motivated assessments.

### Information Technology Diffusion and Utilization

A general strategy involving the dissemination of information and “institutional” behavior modification should be developed. Some selected target populations and behaviors might include:

- *state officials*—develop increased appreciation for the necessity and value of evaluations, employ a common core of research questions and outcomes to increase the probability that cross-site comparisons are possible, use methods that permit causal inferences with some confidence.
- *evaluation firms*—move toward common terms, definitions, methods of reporting results. Move toward common ways of doing process and implementation analyses. Again, the issue is one of comparability and building up a stock of knowledge.

- *academics*—get more involved in policy and program evaluation activities. Develop increased ability to reporting findings in a way that allows researchers and their work to become more influential and useful to policy makers.
- *foundations*—move toward common standards that they will look toward in making awards, improve the technical knowledge of program officers, increase interest in research and evaluation activities.
- *federal officials*—increase technical assistance role and continue to play an active, though perhaps less official, role in enhancing the quality of evaluations. Increase the use by federal officials and organizations of neutral, third-party experts to perform review and monitoring functions that can no longer be performed internally.

Influencing these target populations is partly a matter of continuous communication and contact. Dissemination in this context should not be confused with one-way communication of “fact.” Rather, we are talking about a semicoordinated effort to keep the evaluation agenda alive, move toward a common understandings and tools, and communicate information that may be useful to policy audiences in ways they can use. Each of the steering group institutions is connected with the target group members in some way. Each of these organizations has, or is developing, venues for communicating with relevant parts of the policy and evaluation worlds.<sup>22</sup> We intend to use the various communication vehicles that each possesses to increase and maintain this contact. Our premise is that continuous dialogue and information dissemination through a variety of communication outlets will, over time, inform and shape the way selected individuals and institutions act, by constantly keeping evaluation issues at the forefront of the debate. We are not looking for miracles, but we do hope for progress.

---

<sup>22</sup>For example, preparing journalists to make sense of confusing claims and counterclaims made about program and policy “success” would be a valuable contribution.

## Potentials and Problems of a Pre-Post Design for State-Based Evaluation of National Welfare Reform

Robert Haveman

### I. THE LEGISLATION AND THE EVALUATION OF ITS EFFECTS

The 1996 federal welfare reform—the Temporary Assistance for Needy Families legislation—has important implications for the well-being of the entire low-income population, and in particular for those individuals and families that could have been eligible for assistance under the pre-reform law. Time limits will be in effect starting July 1, 1997, the entitlement to support is abandoned, some groups eligible for support under pre-reform law are no longer eligible, and states are required to tailor-make their own systems of support through the block-grant provisions of the new law.

While all of these changes have significant implications for the well-being of the low-income population, the new law has several features that make a reliable evaluation of its overall impact—and the particular impact of any state’s implementation of the law—difficult, and perhaps impossible. These include:

- A key provision of the new law is the block-granting of funds to individual states, together with the requirement that each **state design its own tailor-made system of support largely free of federal requirements regarding benefit eligibility and levels, and administration**. As a result, a national welfare system characterized by some degree of coherence and similarity<sup>1</sup> will be replaced by a system in which state-specific policies will show substantially greater disparities. Subsequent to its implementation, the pre-reform system will be replaced by 50 quite different systems, constrained only by the few mandates in the national legislation. As a result, the diversity across the nation of “systems” for the support of the low-income population will be far greater than existed under the old law. This fact has particular relevance for any effort to provide an overall, nationwide assessment of the new law, but it also has important implications for designing an evaluation of any particular state’s policy response to the federal legislation.
- Because the new law represents a national policy change with state-specific implementation, there are several different levels on which evaluation can focus. The broadest level is that of a national evaluation, in which the analyst would seek a picture of the overall impacts of the law, neglecting state or regionally specific effects. Alternatively, one might seek an assessment of the effects on a state’s

---

<sup>1</sup>Prior to the new law, each state had a support system consisting of a state-specific AFDC program that provided income support to primarily single-parent families, which program had to meet a detailed set of federal requirements and specifications; a federal Food Stamp program that provided food-based assistance on a national uniform basis to families supported by the AFDC program in a state and to other low-income families; and a state-specific Medicaid program which provided health care support to AFDC-supported families plus other low-income families that met certain criteria, which program (like AFDC) had to meet a detailed set of federal requirements and specifications. While this pre-reform set of programs varied by state in terms of benefit levels and accessibility, it retained a semblance of a coherent national system through both the uniform national Food Stamp program and the uniform federal requirements for the AFDC and Medicaid programs.

population of the particular version of the policy change implemented in that state. It is the latter form of assessment to which this paper is addressed.<sup>2</sup>

- The embedding of a state's policy change within a large-scale national policy shift poses a variety of research challenges for the analyst. On one hand, the analyst may be able to compare the behavior and administrative changes occurring over time in one location (state) to those occurring in another location. Such a simple comparison, however, is not equivalent to an evaluation of the impacts of the policy change, because underlying social and economic conditions in the different states may change in different ways and at different rates. Alternatively, the goal of the evaluation may be to assess the changes in variables of interest attributable to a particular state's policy change (for example, a shift from the prior-to-W-2 AFDC program to W-2), which is a question of "evaluation." Such an evaluation requires that change in underlying economic and social conditions in the state (apart from changes that are due to the policy shift) be known, and that the research is able to control for these changes. In both cases, research is hindered by the fact that, in a "general equilibrium" world, what happens in a particular state will be affected by the policy changes that are adopted in other states.
- The process and speed of the transition from the old welfare system to the new system will not be uniform across the states. Some states seem prepared to install and implement a new and radically different program generally consistent with the new law on short notice; others are likely to make changes far more slowly. This has clear implications for identifying the time horizon necessary for an evaluation of the entire national policy change, but it also has important implications for structuring the evaluation of the policy changes occurring in a particular state (e.g., evaluating the effects of W-2). Successful evaluation of the policy change adopted by a particular state is required to take account of the speed of implementation within that state, as well as the speed of implementation in neighboring states.
- The extent of policy change that will be undertaken by individual states in response to the new law is potentially enormous. The populations eligible for support under any state's old and new systems will be quite different. While the pre-reform system in all states had the objective of supporting the income of eligible people, the new system that will be adopted will seek to enforce work on a different pool of eligible citizens, and to condition assistance on work. The financial incentives offered by states to their low-income populations under the new law will be quite different from those that existed under the old law.

These considerations pose difficult problems for program evaluators concerned with evaluating either the effects of the new law on the nation or the effects of those changes in legislation that will be adopted by

---

<sup>2</sup>Assessment of the national impact of the policy change may well be more feasible than assessing the effects of a particular state's new law. At the national level, a number of data sets, such as the Survey of Income and Program Participation (SIPP), the National Longitudinal Surveys (NLS), and the Current Population Survey (CPS), exist to provide the basis for an assessment of the changes that the legislation has had over time on the populations of interest. Moreover, the Bureau of the Census has designed a nationally-representative, longitudinal survey, the Survey of Program Dynamics, that builds on SIPP, and that will have pre-reform period, reform implementation period, and medium-term post-reform period information. A wide variety of program eligibility, access and participation, income and in-kind benefits (by source), economic and demographic variables, as well as a variety of other information relevant to assessing the effect of the reform on family circumstances will be included in the survey (see Weinberg, 1996). The Urban Institute project "Evaluating the New Federalism" intends to make use of some of these databases in its impact evaluation of the reform, and to undertake special surveys in a selection of states.



particular states on the well-being of their citizens. Some of these involve large questions of evaluation design; others involve practical questions regarding which outcomes (variables of interest) are to be studied. In this paper, I will discuss the evaluation design issue (emphasizing the choice of design strategy for evaluation of a particular state's initiative), and comment on the difficulties confronted **by all feasible designs**. I will conclude that all of the available evaluation design options have serious fundamental problems, and that any overall decision regarding which design is the most effective will depend on considerations of both the cost of undertaking the evaluation and the relative weights assigned to the various problems associated with each of the designs.

## II. SOME BASIC PRINCIPLES FOR AN "IDEAL" EVALUATION

In this section, I attempt to describe the characteristics of an "ideal" evaluation of a state-based welfare reform measure and outline some of the principles that should guide such an evaluation. These basic principles must be kept in mind in thinking through both the ideal evaluation and any real-time, feasible evaluation.

### A. Some Assumptions

In discussing this "ideal" evaluation, I make a number of assumptions, most or all of which will be violated in any feasible real-time evaluation. In a subsequent section, I discuss feasible evaluation strategies in a complicated world in which these assumptions do not hold.

In my view, one question is central in designing a state-based welfare reform evaluation study:

**What is the impact of the policy change on the economic well-being and the economic activities (e.g., work effort, family structure changes, health and nutrition changes, changes in the care and nurturing of children) of those individuals and families that are the most likely to be affected by the policy change?**

A key word in this question is "impact," implying that the study seeks to isolate the changes in particular variables that are attributable to the policy from the changes that may be due to other factors. A second key point is the focus on individual people and their well-being. After all, a public policy is efficient only if the benefits that it conveys to people (its positive impact on their lives, living conditions, and well-being) exceed the costs that it imposes on them.

I would argue that if this fundamental question regarding net effects on well-being cannot be answered reliably, one should seriously question the wisdom of devoting substantial resources to studying the wide range

of other questions that might be asked regarding changes that might follow upon the policy shift. Consistent with this point of view, I presume in the following discussion that a well-designed, longitudinal sample survey of households, and the measurement of outcome variables observable in survey responses, must form the core of a reliable analysis. However, while I emphasize what can be learned about the impacts of a state's policy change from appropriately structured household surveys, I do not ignore the important evaluative information that can be obtained through both administrative data and time-series information on aggregate effects. This information can be studied apart from data obtained from the survey of households.<sup>3</sup>

Adopting this state-based evaluation posture avoids some of the problems that would confront an evaluator attempting to measure the impact of the policy change on the entire nation. However, this posture creates other problems. Perhaps the largest is that created by the potential migration of citizens in response to the policy change. In discussing the ideal evaluation, I ignore the difficulty for the evaluator posed by policy-induced migration.

Second, in discussing the characteristics of an "ideal" evaluation, I assume that the policy change from the pre- to the post-reform systems will be discrete, and that the characteristics of the pre-reform system and the post-reform system can be clearly identified and described. This neglects the fact that, in some states, major reforms have been undertaken prior to—and in anticipation of—passage of the new law, so that both defining and measuring the pre-reform system will be problematic. It also ignores the likelihood that real-time policy implementation will be slow and uneven, so that in the years following passage of a state's reform law all that the analyst will be able to observe will be some unknown combination of the pre-reform system and the post-reform system. In this situation, identifying the post-reform system may be impossible; any actual evaluator will have to assess how best to measure the effects of a "policy change process" as opposed to a discrete policy change.

Third, in discussing the ideal evaluation, I assume that the post-reform system will develop in response to the 1996 legislation, and that no major changes in federal law will be made in subsequent years, through additional legislation. In any real-time evaluation, this assumption too is problematic.

Fourth, I assume that the policy change undertaken by states in response the national law will be designed to "change the culture" in the state. That is, I assume that one of the objectives of the policy will be to change citizen perspectives regarding the responsibility of the public sector to provide income support, and of the need for individuals to accept responsibility for their own financial well-being. Hence, virtually all

---

<sup>3</sup>Indeed, the case for the "pre-post" evaluation design rests heavily on the superiority of the "with" and "without" comparison of administrative and aggregate data that is possible with it.

citizens with low permanent incomes will be affected by the legislation; evaluation should not be limited to the population of current program recipients, or those that would have been eligible for the current program.

Fifth, I assume that each of the designs considered is “feasible” in the sense that there is no prohibition in any state’s legislation that would forestall any particular approach, and that securing data necessary for implementing a specific design is possible, even though that might require data collection and household surveys in other states.

## B. Some Basic Principles

Several basic principles are relevant for securing a reliable evaluation of the policy change embodied in the new welfare law. These include the following:

- Precise specification of both the “factual” and the “counterfactual” is necessary. The analyst needs to have a clear description of both the nature of the public intervention **without** the new policy, and that existing **with** the new policy in place. More concretely, in considering the design of an evaluation of a particular state’s policy change, the analyst needs a clear description of the nature of the state’s welfare system as it existed prior to the new welfare reform law (that is, without the new law), and as it exists (or will exist) with the new law.
- Evaluation of the impact of introducing a new policy involves comparing the observed level of a variable of interest (e.g., the work hours and earnings of benefit recipients and other low-income individuals) with the new policy in effect,  $L^n$ , minus the observed level of the variable with the prior policy,  $L^p$ .
- Having made the assumptions indicated by these two principles, a sound evaluation of the impact of a replacement policy requires that both  $L^n$  (the factual) and  $L^p$  (the “without policy” counterfactual) be observed and measured reliably. Only in this way can  $\Delta$ , the impact of the policy change, be obtained. If the new policy is imposed as a replacement of the prior policy, the measurement of  $L^n$  is, in principle, straightforward; it is the level of the variable observed over time subsequent to the imposition of the new policy. Observation of  $L^p$ , however, is not so simple. It requires identification of a counterfactual which will yield behavioral incentives and state of the world conditions (and hence, observed levels of the variable of interest) identical to those that would have existed with the prior policy in effect. Hence, in a context in which a new policy replaces an existing policy (e.g., the replacement of the prior welfare system with the new system that will exist given the 1996 legislation), this basic measure of the impact of the policy change may be difficult to obtain.

In general, there are three ways of measuring the counterfactual level of the variable of interest,  $L^p$ . They are:

- a) establishing an Experimental design, such that a randomly assigned sample of those decision makers affected by the policy change continue to operate under the rules of the prior policy, P. In this case, the level of the variable of interest can be measured for this randomly selected control group, used as  $L^p$ , and compared with  $L^n$  observed and measured for the treatment group operating under the rules of the new policy, N.

b) defining a comparison group of individuals which is not randomly assigned, but which confronts behavioral incentives and state-of-the-world conditions that are identical to (or at least similar to) those of the prior policy, P. Here, the variable of interest can be measured for individuals in this “control site,” used as  $L^p$ , and compared with  $L^n$  observed and measured for the group confronting the rules of the new policy, N. This will be referred to as a comparison-site design.<sup>4</sup>

c) defining a comparison group of individuals who are not randomly assigned but who, in fact, confront the prior policy, P. Here, the variable of interest is measured as it exists (or existed) for the “control group” (taken to be those confronting the prior policy), used as  $L^p$ , and compared with  $L^n$  observed and measured for the group confronting the rules of the new policy. This design will be referred to as a pre-post design.

### III. THE PRINCIPLES, AND CHOOSING AN EVALUATION DESIGN: THE EXPERIMENTAL AND COMPARISON-SITE OPTIONS

These basic principles have implications for the decision regarding the choice of a design strategy for evaluating a state-specific welfare reform, in the context of a national law encouraging a wide variety of state-specific policy changes. In this section, I discuss the experimental and comparison-site design options, and indicate the implications of the basic principles for each of them.

#### A. An Experimental Design

The “social experimentation” technique has become the method of choice among many evaluators assessing the impact of proposed changes in social policy measures. Using this technique, individuals in the target group (say, a state’s citizens, or a state’s lower-income citizens) are randomly assigned into a treatment and a control group, with only the former group being subjected to a “treatment” in the form of a policy environment that differs in some well-specified way from the “without policy” environment. The standard policy proposal to which this design has been applied is, for example, a new job training program viewed as a replacement for the existing training environment. Simultaneous observation of the two groups over time will reveal a difference between them in the level of a variable of interest; that is, a difference between the level of the variable of the treatment group,  $L_t$ , and the level of the variable of the control group,  $L_c$ . This difference— $\Delta = (L_t - L_c) = (L^n - L^p)$ —is taken to be the impact of the policy. This procedure assumes that the new (or “treatment”) policy does not affect the underlying “state of the world” (or that the new policy is “small”).

---

<sup>4</sup>Most discussions of this possibility consider another, appropriately chosen, jurisdiction as a the “control group” to which a state’s residents would be compared.

Stated alternatively, this technique assumes that the level of the variable observed under the prior (or “without”) policy reflects the state of the world into which the new policy is introduced, and hence that only the change in the level of the variable from that observed with the prior policy to that observed with the new policy is relevant.

With this latter assumption, the experimentation technique can proceed if the randomly assigned control group can be isolated from the treatment group, such that the control group becomes the counterfactual against which the treatment group is compared. This requires that, in spite of the policy change, it is feasible to secure a control group of individuals who will be confronted with the “state of the world” represented by that which existed under the prior policy.

If an experimental design is chosen for studying the impact of the policy on households and individuals, a separate design for using administrative and aggregate data for studying the effect of the policy would have to be developed. By definition, administrative/aggregate information is available: (1) only for civil jurisdictions or administrative agencies and (2) in time-series form. Hence, the design of administrative and aggregate data collection for policy evaluation must employ a pre-post design format. The different conceptual bases for the evaluation—contemporaneous observation of samples of control and treatment group members in the experimental design and before-after administrative information for the pre-post design—could pose problems of interpretation.

#### B. A Comparison-Site Design

With this design, a group of individuals located in an environment with both “state of the world” and behavioral incentives identical to those existing under the prior policy in a particular state (e.g., Wisconsin), P, must be identified. Then, the level of the variable of interest must be measured for this group, and taken to be  $L^P$ . From this, and the similar measurement of the variable of interest for the group of individuals confronting the new policy in the state of interest, N, a reliable estimate of  $\Delta = (L^N - L^P)$  can be obtained.

As in the case of the experimental design, this counterfactual group facing the state of the world associated with P must both match and be isolated from the group of individuals who are subject to state of the world and the incentives of the new policy, N. However, while the random assignment procedure of the experimental design assures that the control group “matches” the characteristics of the treatment group, there

is no such assurance in forming the comparison site group. Hence, in this design, a “statistical match” of individuals in the two groups must be made.<sup>5</sup>

This design must also confront the difficulty of securing comparability of those elements of the “state of the world” that are not associated with the difference in policy between P and N, between the treatment and control (comparison-site) groups. In the context of welfare reform, statistically matched individuals located in a comparison site must face economic and social conditions which are identical to those that prevail in the site subject to N (e.g., Wisconsin). In sum, the only aspect of the environment that can differ between individuals in the new program site, N, and those in the comparison site is the incentives and constraints associated with the policy that is in place.

What is most important, the incentives and constraints of the policy in place in the comparison site must be identical to those of the prior policy, P, in the site for which the evaluation is being undertaken. The fact that pre-reform policy differences between any two states (or sites) are likely to be substantial is an unavoidable fact, and reduces the ability of an evaluator to secure a comparison state in which existing policy is identical (or even similar) to pre-reform policy in the state for which the evaluation is being made.

Finally, the implications of the comparison-site design for the use of administrative and aggregate information in evaluation must be recognized. Because the nature of the data available in the comparative site will inevitably differ in subject matter, coverage, and definition from that available in the evaluation site, serious difficulties will be encountered in securing comparable administrative or aggregate information from two geographic sites required for a reliable evaluation. As with the experimental design option, a separate strategy would have to be developed for securing reliable pre-post administrative and aggregate data from the evaluation site for studying the effect of the policy.

#### IV. THE PRINCIPLES, AND A PRE-POST DESIGN

In both the experimental and the comparison-site designs, assessing the impact of the policy change requires contemporaneous measurements taken over the groups representing the with-policy and the without-policy alternatives (respectively, the treatment/control and the comparison site/policy site groups). In the pre-post design, however, this assessment for the without-policy alternative requires measurement while the prior policy, P, is actually in effect (that is, prior to the implementation of the new policy, N, in this site); assessment

---

<sup>5</sup>The loss of comparability associated with a statistical match—as opposed to random assignment of a state’s permanently poor residents to a prior policy regime and the new policy regime—is an important limitation on the reliability of results based on the comparison-site design.

of the effect of the new policy, N, on relevant variables is made subsequent to implementation of the new policy.

In the pre-post design (as with the experimental and comparison-site designs) two groups of individuals must again be designated. In this case, one group must be subject to the prior policy, P, and the other subject to the new policy, N. As with the comparison-site design, this will require a statistical matching of individuals in the two groups, and with this an associated loss of comparability.

The problem of securing comparability of those elements of the “state of the world”—changes in demographic, social and economic variables apart from any effect of the policy—that are not associated with the difference in policy between P and N is also serious in this design. And, as in the case of the comparison-site design, the seriousness of this problem is not under the control of the evaluator. Because an evaluation using the pre-post design is based on measurements at two points in time, the analyst must attempt to secure identical underlying economic and social conditions during both the pre- and post- periods. Securing such identical (or even similar) “state of the world” circumstances is devilishly difficult. In the absence of such similarity, statistical control for differences in these environmental conditions will be required. The basis for securing such control is not clear.<sup>6</sup>

However, unlike the comparison-site design, securing an accurate representation of the prior policy, P, does not pose a problem in the pre-post design. By definition, the incentives and constraints confronting individuals in the without (or pre-reform) period are those of the pre-reform policy.

Finally, adopting this design for evaluating the impact of the policy change on households and individuals has the important advantage of being entirely “within site,” and hence consistent with the necessary “within site” design for the use of administrative and aggregate data for evaluation purposes. The impacts of the policy on families within the state (obtained through a longitudinal household survey) could be made consistent in both time and coverage with the impacts measured by administrative and aggregative data.

## V. TRADE-OFFS AMONG THE DESIGN OPTIONS: A SUMMARY

---

<sup>6</sup>The problem of statistically controlling for the effect of changed social and economic conditions on variables relevant to evaluating state changes in welfare policy confront the same challenge as that faced in studies that have attempted to model or “forecast” the change over time in state welfare caseloads. In these studies, the challenge has been to measure how underlying changes in demographic, social, and economic conditions will affect the number of welfare recipients, as reflected in entry and exit rates of welfare programs. Assessment of the forecasting success of the former studies is decidedly mixed. See U.S. Congressional Budget Office (1993), Garasky (1990), and Plotnick and Lidman (1987).

This discussion, then, has revealed problems associated with each of the potential designs, and hence the need for considering trade-offs among the difficulties and the costs of each. In terms of the basic principles, I conclude that:

a. The experimental design has a serious (and, perhaps, fatal) flaw: In the face of a policy change that seeks to change the “culture” of public income-support expectations within the state, it will be difficult or impossible to isolate a within-state control group from the incentives of the new policy. Implementing an experimental design also confronts the problem that a state is required to permit some of its citizens to continue to rely on the pre-reform system, and to mandate that program administrators work with these clients in a manner consistent with the operation of the program prior to the policy change. Conversely, the experimental design has the advantage of measuring the outcomes for the control and experimental groups contemporaneously, hence securing a constant “state of the world” (other than the changed incentives and constraints imposed by the policy change) across the two groups. This design for evaluating the effect of the policy on households and individuals, however, implies the need for a separate design for the use of administrative and aggregate data for program evaluation, given that these data are available within the site on only a pre-post basis.

b. The comparison-site design also encounters serious difficulties. The major difficulty is that of securing a comparison site that has a “current policy” in effect which is identical (or similar) to the pre-reform system in the state whose policy change is being evaluated. Given that all states have to respond to the policy mandate of the 1996 legislation, it is unclear that any state will meet this without-policy-change requirement.<sup>7</sup> A second serious problem is securing comparability in terms of the nonprogram “state of the world” characteristics (e.g., economic and social conditions) between the comparison state and the state whose policy change is being evaluated. Third, this design strategy requires statistical matching of the households in the comparison state with those in the state whose policy change is being evaluated. Fourth, while this approach avoids the costs and difficulties of implementing a random assignment strategy, it may require the permission of the authorities in the comparison state for undertaking a selective household survey for this purpose, and does require data collection in a separate and perhaps distant jurisdiction. Finally, like the experimental design,

---

<sup>7</sup>In principle, establishing a within-state comparison group design might be possible. This would require that some part of the state maintain the prior policy regime, and consciously administer that regime as if no state policy change had occurred. (For example, it has been suggested that, in Wisconsin, counties that border Minnesota, and which are dominated by Minnesota newspaper, radio and television media, should be administratively mandated to maintain the pre-W-2 AFDC program.) It seems unlikely that a segment of a state could, in fact, be kept immune from a state-based policy change with implications as fundamental and far-reaching as those required by the federal legislation.



adoption of this design requires development of a separate pre-post, within-site design for the use of administrative and aggregate data in evaluation.

c. The pre-post design faces the difficult problem of securing comparability in the “state of the world” characteristics (economic and social conditions) between the pre-reform period and the post-reform period. Unlike the experimental design and the comparison site design, this strategy does not measure impacts contemporaneously. Developing techniques for statistically adjusting for the effects of differing underlying economic and social conditions on those household variables and market changes that are central to the evaluation would be necessary. (As indicated in footnote 7, development of models that would enable reliable adjustments is not straightforward.) Unlike the other designs, however, the with-program and without-program groups can be rather clearly identified. While specifying the policy change in this case seems relatively straightforward, the real-time evolution of state policy may preclude a clear delineation of the with and without policy environments (see below). The pre-post design, however, does place the results from it on the same analytical footing as the evaluation results obtained from the use of within-site administrative and aggregate data.

## VI. PROBLEMS AND PROSPECTS FOR A PRE-POST EVALUATION DESIGN FOR WISCONSIN WORKS (W-2), AND NEXT STEPS

To assess the options for designing a pre-post evaluation of W-2, it will be helpful to define the constraints and the environment with which the evaluation will have to cope. The following are some facts and presumptions which are relevant for considering this issue:

- While successful evaluation of W-2 requires that the nature of the policy change be clearly defined, this requirement may not be attainable. Indeed, some of the elements of W-2 are currently being implemented in several counties, and all counties are being urged to seek employment for applicants prior to offering income support benefits.<sup>8</sup> This is so even though formal, statewide implementation of W-2 (though perhaps somewhat changed from the initial proposal) will not occur until the beginning of 1998. This fact will make evaluation of W-2, by any design, difficult.
- It seems unlikely that the state of Wisconsin will be willing to exempt some sites (counties) from the W-2 legislation, so as to make an experimental design feasible.
- Major welfare reforms in other states will also be pursued over the next few years.

---

<sup>8</sup>For example, the Pay for Performance program is effectively in effect in several counties in Wisconsin resulting in the sanctioning of rather large number of families. This program has more in common with W-2 than it does with the conventional, pre-policy-change AFDC program.

- Throughout the nation, but especially in Wisconsin, reform of the welfare system will be the primary instrument for “changing the culture,” a phrase meant to convey a major and discrete change in the social expectations regarding work and individual responsibility of low earnings capacity individuals (especially, poor single mothers). This culture change has already begun in Wisconsin.
- The economy of the state in 1997 and subsequent years will not be as robust as it is now, though there is little reason now to assume a major downturn. However, because of W-2 the low wage labor market will experience an increase in labor supply (and hence downward wage pressure), relative to conditions without W-2.

I conclude that a pre-post evaluation design is the most feasible of the options for securing a reliable assessment of W-2. I reach this conclusion not because of any inherent superiority of this design. Rather, the characteristics and requirements of the experimental and the comparison-site designs, together with the constraints imposed by the policy environment (e.g., reforms designed to change citizen expectations and the “culture,” simultaneous and differential policy changes in all states in response to the federal legislation), severely limit the feasibility of these designs, relative to that of the pre-post design.<sup>9</sup>

Given this conclusion, what steps that should be taken in order to implement a pre-post evaluation design for W-2?

1. As soon as possible, plans for a longitudinal household survey should be developed, including the specification of the variables on which information will be sought. Such a survey should be designed to elicit relevant information on behavior and well-being from the state’s citizens, both under the existing welfare system and, later, under W-2 when implemented. I presume that the longitudinal survey should be at least annual, although more frequent interviewing would be desirable. Moreover, the survey should be fielded for a large number, say, 6–8 years, in order to obtain reliable information over time on the changes in well-being and behavior that is sought. The variables included in this survey would emphasize the employment and work activities of the adults in the family, family income sources, information on children and various aspects of their well-being, location, family structure, housing arrangements, and health and health care for family members.

2. The sample of families to be included in the survey should be identified. I would suggest a sample of “low permanent income” families (perhaps, families in the bottom quintile of the permanent income

---

<sup>9</sup>In particular, an experimental design seems infeasible because of: 1) the absence of state exemption of sites or individuals from the provisions of the W-2 policy (and the maintenance of pre-reform AFDC service provision for these sites or individuals) and 2) the state-wide culture change which will affect all low income citizens, including existing recipients and potential applicants, rendering impossible the establishment of a reliable  $L^p$ , even were state exemptions to be granted. A Comparative Site design seems infeasible because: 1) no state with base economic and social (and other state of the world) characteristics similar to Wisconsin will have in place a welfare system similar to the pre-reform Wisconsin system, hence rendering impossible the establishment of a reliable  $L^p$ , and 2) the culture change posited for Wisconsin will also influence the behavior of low earnings capacity citizens in other states, even if these states have not implemented a change as drastic as W-2, again eliminating the possibility of establishing a reliable  $L^p$ .

distribution).<sup>10</sup> Such a sample would include all current welfare recipients as well as virtually all of those in the state who might, in the future, apply for support and be affected by labor market developments caused by implementation of W-2. By following such a sample from the present into the future, one should be able to secure a picture of changes in work, family structure, and well-being of families attributable to the W-2 program, and thereby discern how the change from the prior system to W-2 will influence the state's low-income population.

This suggested procedure assumes that observation of these families from the present (or the earliest date for fielding a survey) to a date at which W-2 is judged to be "implemented" would yield sufficient reliable information to form a picture of the counterfactual, pre-policy-change environment,  $L^p$ .<sup>11</sup> It seems clear that observing these families at a point in time after which W-2 has been in place can provide an accurate picture of their behavior and well-being in the post-W-2 period,  $L^n$ .

3. The site(s) that will be oversampled in the survey should be identified. For two reasons, I would argue that Milwaukee County be heavily oversampled in the survey. First, the conditions and behaviors to which the 1996 legislation was addressed are largely those present in large urban areas—that is, minority, inner-city welfare recipients. Hence, it is most important to secure reliable evaluation information on them. Second, it seems likely that implementation of W-2 will be slow in Milwaukee (and that information regarding the implications of W-2 is less well understood there); hence, the possibility of securing reliable pre-reform information on well-being and behavior may be the greatest there.

4. The survey should be implemented as soon as possible.

This discussion has rested upon a basic assumption, namely, that the evaluation of W-2 should focus on its effects on the well-being—as measured by the income, work, and structure—of low-income families in

---

<sup>10</sup>The New Jersey Income-Maintenance Experiment used such a "low permanent income" criterion as the basis for selecting the sample for this experiment. The volumes describing this experiment would be a guide in defining the sample for a W-2 evaluation.

<sup>11</sup>It must be recognized that such a survey may, in fact, not be able to yield an accurate picture of behavior and well-being under the prior policy, namely AFDC, given that implementation of changes within the AFDC program designed to move it toward W-2 are already being undertaken. A danger is that what such a survey would capture is the policy during an "interregnum" period in which the state's program is some undefined transition/combination of the prior system and W-2. In fact, this concern is as relevant to the experimental and comparison-site designs as is to the pre-post design. Irrespective of design, the evaluator needs a reliable estimate of well-being and behavior of the low permanent income population prior to implementation of welfare reform. To the extent that what can be observed in the state is **not** pre-reform AFDC but some combination of AFDC and elements of W-2, an intrastate Experimental design will also be unable to measure the impact of the policy change. A comparison-site design may be able to measure the impact but only if a comparable non-Wisconsin comparison site is available which would not have implemented welfare program changes in response to the federal legislation by the time the household survey is undertaken. As noted above, these requirements are daunting.

Wisconsin. While, in my view, this should be the primary focus of any evaluation, there are other important questions that are not addressed by this focus. These questions largely concern the provision and availability of public and private services to Wisconsin citizens and the accessibility of other sources of income to them (especially employment income). These provision and accessibility issues can be thought of as elements of the economic environment in which low income citizens live, which elements may be affected by W-2. Prominent examples of these elements could include:<sup>12</sup>

- the availability of quality day care services and the price of these services
- the availability of alternative sources of income, housing, food and education/training services from nonwelfare (including private) sources
- the availability of family planning and abortion services.

In addition, there are other changes in non-W-2, income-related program services that might reflect the impact of W-2. While some of these may be captured in a comprehensive and carefully designed longitudinal survey, program and administrative data might be better able to reflect the impact of W-2, including:

- the level of foster care placements in various parts of the state
- food pantry and emergency housing and other service provision by public and private agencies
- the incidence of eviction from public or subsidized housing due to inability to meet rental payments
- the prevalence of nutrition deficiencies and behavior problems in the local schools.

Finally, it might be of interest to assess the impact on a variety of policy-relevant variables not directly affected by the policy change, and which may not be picked up by the survey data, such as the level of child support collections from absent parents, the incidence of marriage (abortion, nonmarital childbearing) of young women and men, or the level of employment of young married males.

These considerations suggest the importance of simultaneously initiating efforts to design a plan for using administrative and aggregate data for evaluating W-2. As indicated above, adoption of the pre-post design for evaluating the impact of the policy change (from the prior Wisconsin AFDC-based welfare system to W-2) has the important advantage of being conceptually consistent with the necessarily pre-post nature of within-state administrative data. Hence, the design for a pre-post evaluation should be made at the same time as, and in coordination with, the design of a strategy for collecting administrative and aggregate socioeconomic data. A clear delineation of the nature of the findings from the two coordinated, but independent, evaluation

---

<sup>12</sup>Cancian and Wolfe (1997) discuss the program and aggregative data relevant to evaluating state welfare reform, and the trade-offs that must be confronted in designing an evaluation study using this information.

efforts should be attempted at the outset, so as to avoid duplication of efforts in gathering pre-post information.<sup>13</sup>

## VII. SOME FINAL THOUGHTS

In this discussion, I have not considered the alternative of a combined comparison-site/pre-post design. Such a design would allow both the cross-time and the cross-site effects of the policy change to be discerned. This would enable a “difference within differences” analysis framework which, by exploiting both the cross-time and the cross-site effects of the policy change, could provide additional observations and reliability.<sup>14</sup> Designing such an evaluation would be exceedingly complex, and would not eliminate the need to obtain statistical control for the effect of changes in demographic, social, and economic factors apart from the policy change on those variables relevant for evaluation. This combined strategy also has potentially large cost implications that should not be neglected.

If either a combined comparison-site/pre-post design—or one of either the comparison-site or pre-post approaches—are adopted, the evaluator will confront the difficult problem of distinguishing “state of the world” changes that are **not** due to the changed incentives and constraints imposed by the policy change from the effects of the policy itself. The most concrete example of this problem is the need to adjust pre-post results from both the household survey and from administrative/aggregate information for changes in the underlying state of the economy from the period prior to the adoption of the new policy to the period after (see footnote 7). The relevant question here is: “How would the work status (or some other relevant variable) of the individuals in the household survey have changed from the before to the after period, if underlying demographic, economic and social conditions changed without any policy change?” If this question could be answered, one could adjust observed changes in the relevant variables for the effects of changes in the state of the world apart from the policy shift.

Apart from the studies cited above, little thought has been given to this issue, in spite of major recent advances in statistical modeling of the determinants of changes in time series data. Given the importance of the question, and the nature of the policy change at issue, additional research efforts designed to develop reliable

---

<sup>13</sup>See Kaplan and Meyer (1997) for an evaluation design that rests on the use of administrative and aggregate data.

<sup>14</sup>See Piliavin and Courtney (1997).

statistical models for forecasting without-policy changes relevant to evaluating state welfare reform would seem worthwhile.

### References

- Cancian, Maria and Barbara Wolfe. 1997. "Outcomes of Interest: The Universe of Outcomes Likely to Interest Different Constituencies," Institute for Research on Poverty, University of Wisconsin–Madison, mimeo.
- Garasky, Steven. 1990. "Analyzing the Effect of Massachusetts' ET Choices Program on the State's AFDC-Basic Caseload," Evaluation Review (December), pp. 701–710.
- Kaplan, Thomas and Daniel Meyer. 1997. "The Structure of a Proposed Impact Evaluation of the Wisconsin Works Program," Institute for Research on Poverty, University of Wisconsin–Madison, mimeo.
- Piliavin, Irving and Mark Courtney. 1997. "Prospects for Comparing Wisconsin Works to Welfare Reform Programs Outside the State," Institute for Research on Poverty, University of Wisconsin–Madison, mimeo.
- Plotnick, Robert D., and Russell M. Lidman. 1987. "Forecasting Welfare Caseloads: A Tool to Improve Budgeting," Public Budgeting and Finance (Autumn), pp. 70–81.
- U. S. Congressional Budget Office. 1993. Forecasting AFDC Caseloads, with an Emphasis on Economic Factors. Washington, D.C.: Congressional Budget Office.
- Weinberg, Daniel, "A Survey of Program Dynamics for Assessing Welfare Reform," memo, U.S. Census Bureau, October 29, 1996.





## Outcomes of Interest, Evaluation Constituencies, and the Necessary Trade-Offs

Maria Cancian and Barbara Wolfe

Under Title I of the welfare reform bill of 1996, providing Temporary Assistance for Needy Families Block Grants, or TANF, broad-based welfare reform will take place in most, if not all, states. Will state-based, comprehensive welfare reform be a success? How will we know? In contrast to limited policy changes with clearly defined goals, the critical outcomes of interest for current reforms are a matter of debate. Even setting aside problems of measurement and data collection, no clear consensus exists on how to define success. This paper begins to identify the outcomes of interest for evaluating the reforms that will take place in the wake of TANF. We begin with a brief review of some of the factors that contribute to the need for a broadly focused evaluation. We then discuss alternative constituencies and the trade-offs that will be faced in collecting data and designing evaluations that address their interests. Finally, we outline the implications of these trade-offs for a representative set of outcomes.

Wisconsin has been a leader in welfare reform. The Wisconsin Works (or W-2) plan is substantially more developed than are the plans of most states. We use W-2 as a point of reference in the following discussion, but expect that most of the issues raised will apply to other states as well.<sup>1</sup>

### WHY A BROADLY FOCUSED EVALUATION?

For a number of reasons the outcomes of interest to evaluate current reforms will be broader than simply measures of impacts on welfare participants and their families. First, federal welfare reform legislation has moved primary responsibility for welfare design to the states. With block grants, states will have more freedom to design programs for needy families and will more fully absorb the financial consequences of their decisions. This raises a host of issues regarding the **fiscal and administrative responsibilities** of federal, state, and local governments. As the new system develops, evaluations will have to consider the impact at each of these levels of government of programs which replace AFDC, JOBS, and Emergency Assistance.

Under the new program there is no individual entitlement. States are free to determine which families receive assistance and under what circumstances. This raises questions of **accessibility**. There is no longer an assumption that state residents are all guaranteed equal benefits, given equal income and family size.

---

<sup>1</sup>As noted throughout, we draw extensively from the background issue papers on W-2 prepared for the IRP conference on Evaluation of Comprehensive State Welfare Reform.

Block grants mean greater state autonomy and variation in programs across states, raising issues of **interstate equity**. Historically, substantial cross-state variation in the level of AFDC benefits has occurred. Under the new legislation, program structure and eligibility standards are expected to vary even more substantially. In addition, because the 60-month lifetime cash benefit limit prevails across states, participants in a low-benefit state are not only receiving lower benefits, but are also reducing their opportunity to receive benefits in a higher-benefit state at a later time.

In Wisconsin in particular, the scale and timing of the reform effort increases the potential for important impacts on the state's **labor market and social institutions**. W-2 calls for almost all current recipients to be engaged in employment activities by September 1997. The vast majority of current AFDC participants are expected to be in community service jobs or private-sector jobs, subsidized or unsubsidized. Concerns have been raised about the potential impact of this initiative on the low-wage labor market as well as on the market for child care. Because almost all recipients are expected to make dramatic behavioral changes over a short period of time, the potential for impacts outside the participant community are substantial. Given the time limits and work requirements associated with TANF, other states are likely to face similar issues.

Finally, the goals of W-2 go beyond income support. W-2 was designed to "reinforce appropriate behavior." The philosophy and goals of W-2 emphasize work, support by both parents of their children, provision only of requested or needed services, and reliance on market and performance mechanisms for program implementation. The program states that W-2 "will demand more of participants, but in the long run it will provide independence and a future." Proponents of W-2 and TANF have argued that welfare reform will encourage **substantial behavioral changes**. This suggests that an evaluation should include an assessment of the extent to which reforms achieve these goals.

#### ALTERNATIVE CONSTITUENCIES FOR AN EVALUATION: EVALUATION QUESTIONS AND THE UNIT OF ANALYSIS

There are a very large number of potentially interesting outcomes. In part because the unit of analysis (e.g., individuals, households, or communities) and data needs differ according to the particular outcome, it will be necessary to prioritize. However, different constituencies have different questions, and hence are likely to be interested in alternative outcomes based on widely differing units of analysis. For example, analysts of welfare reform are interested in impacts on the population eligible (or formerly eligible) for AFDC, on all low-income persons, and on the general community. Questions concerning the fiscal and administrative aspects of welfare reform do not directly speak to the impacts on families, but are clearly of interest to citizens and to

government officeholders, especially in the context of devolution. Program participants also have an interest in evaluations of the impact of programs on their quality of life, while program administrators are likely to be interested in outcomes that capture the impact of the decisions they make at the local level. And there exist a number of other constituencies which have direct and indirect interests in the outcomes of reforms: private charities, employers, schools and teachers, doctors' offices and clinics are examples.

In sum, many constituencies are interested in the outcomes of welfare reform. However, they are not all interested in the same type of question. Their interests are sufficiently diverse that multiple units of analysis and approaches would be required to satisfy all constituencies.

### Trade-Off 1: Population

In assessing data collection priorities for the evaluation of welfare reform, it may be useful to recognize a number of trade-offs. It seems clear that in order to understand the consequences of reform for the eligible and potentially eligible population, we require data on a representative sample of families who were previously eligible. However, since past as well as future eligibles form only a small part of the state's total population, very few of them appear in existing national data sets—such as the Current Population Survey or the Survey of Income and Program Participation—or state samples, such as the Wisconsin Family Health Survey. Even a very large general household sample would yield only a small number of eligibles and their children. Therefore, for those interested in this particular population, a special survey sample, or substantial oversampling of current welfare users, with some addition of low-income families and young teens at high risk of pregnancy, would be preferable. This sample or oversampling would allow studies of the impact of welfare reform on those most directly affected—recipients or would-be recipients. Simple statistical power calculations could be used to learn the needed sample size to observe even a small change in this group.

The sampling issue is even more critical for the study of smaller populations—for example, children with disabilities, immigrants with low incomes, or infants. For these groups, a household survey will have very limited numbers of observations. The bottom line on this is simple—the smaller the group within the total population, the greater the need to oversample the particular subpopulations to learn about impacts on this group.

On the other hand, in order to learn what proportion of the population uses any services, the type of jobs they enter, the performance of the labor market, etc., it is necessary to have a sample of all households—participants and nonparticipants. This will also provide the ability to analyze unmet needs in the population (such as medical care for a chronic illness, or adequate child care) as well as permitting us to observe changes in the level of unmet needs. For analysis of the state's overall population, the sample sizes of

recipients and potential recipients are not very important. Hence there is a direct trade-off—the smaller the proportion of any group of interest (for example, disabled children), the greater the need to oversample, but the lower the value of the oversampled cases to the overall analysis. In considering this trade-off, it is important to recognize another complicating factor: it usually costs more per respondent to sample a targeted group than to conduct a general household survey.<sup>2</sup>

### Trade-Off 2: Individual vs. Community-Based Analysis

A second trade-off exists between using the local community as the basis of data collection versus a sample of individuals (households) across a state. (Both may oversample a targeted population or may focus on the general or only the low-income population.) Underlying conditions are likely to differ across communities in measurable ways—level of unemployment, type of firms, amount and quality of child care, quality and availability of low-cost housing, ratio of medical doctors to population—that may significantly influence the impacts of W-2. Resources allocated to W-2 and the implementation of the program may also differ across communities. Hence an evaluation that considers program components and community conditions is likely to be a far better and more comprehensive evaluation. But, because collecting such data is expensive, not all communities would be included. Thus there is a trade-off between collecting data that is representative of the state versus those that could inform us about impacts in a local, county, or any smaller area in which we might have an interest.

Some of the outcomes that analysts, policy makers, citizens, and participants are interested in will affect children or their families across the state—for example, is the low-income population of the state better off along a variety of dimensions under W-2 than it would have been under prior policies? To answer this question, we need statewide representative data. Many government officials and analysts are instead, or also, interested in the effects of policy implementation and service provision at the state and community level on individuals (households) and/or on firms, schools, and other institutions. Related questions of interest include variation in the impact of W-2 as a result of other environmental factors. For example, what is the impact in a community where there is near full employment? in rural areas or small towns? in communities in which there

---

<sup>2</sup>This occurs whenever prescreening is required to identify sample members. Thus if the population to be oversampled is potential eligibles and young teens, a household screen would be required to locate such individuals. If the targeted population is those of permanently low income, a screen is also required. If the targeted population is instead current recipients of services, this need for prescreening is removed, but locating the population may itself increase costs. In order to reduce survey costs, clustering is usually used, which raises another related issue. If the screening or selection of the targeted populations is done in clusters, costs are reduced, but so is the “usefulness” or representativeness of the sample. In effect, by clustering, the degrees of freedom are limited so that the “effective” sample is less than the number of observations.

is adequate child care versus those with waiting lists? To answer these questions and better understand differences across communities, data must be collected within communities or neighborhoods. Community data which would be useful might include not only state data on services and implementation (such as the proportion of the W-2 population placed in private sector jobs within two months, or the waiting list for positions in licensed child care centers), but also information from such diverse organizations as employers, community organizations, employment offices, the school system, and private charities. Because these data are expensive to collect, there is incentive to consider only a few communities. However, it is difficult to evaluate statewide outcomes on the basis of data collected in a limited set of communities. In addition, concentrating on a few communities clearly limits the degrees of freedom in the analysis of the impact of community differences on the population(s) of interest.<sup>3</sup>

### Trade-Off 3: Evaluation vs. Monitoring

A third trade-off exists between (1) evaluating the overall impact of the new program, (2) comparing implementation and impacts across communities, and (3) simply monitoring outcomes. To evaluate the overall impact, we need measures of the counterfactual (i.e., what the outcomes would have been if the program had not been changed). This means we require data that capture the earlier outcome, adjusted to account for other exogenous changes that would have affected outcomes even in the absence of the new program. In cases where we can measure outcomes pre- and post-reform by means of administrative data or existing survey or other data, we will have some ability to evaluate welfare reform. In other cases data to establish an appropriate counterfactual may not exist, and an evaluation of the program impact will be far more difficult than a comparison across sites or simple monitoring of outcomes. Cross-site comparisons and simple monitoring can give us important information on program implementation and effectiveness and on critical areas requiring further policy change. If we do not collect post-reform measures of outcomes because, for example, *pre-reform*

---

<sup>3</sup>One approach that has been used in a number of studies, particularly in the health area (the Rand Health Insurance Study, the Epidemiological Catchment Area Study), has been to select communities as the first basis for data collection. They can be chosen with a probability that is proportional to the size of their reported AFDC cases, for example. Within the community, agencies can be selected and within this, those who use the agencies can serve as the basis for the selection of a random sample. Or, a random sample from the community can be drawn (with oversampling as determined by the choices discussed above). Employers within the community can also be randomly selected or selected based on a probability of employing low-skilled workers. The use of data based in communities allows study of the impact of services provided, the labor market, etc., on outcomes of interest. The negative consequence of this design is in limiting understanding of the broader consequences throughout the state. And it is not clear if community studies can be pooled for analysis. If there are only a limited number of sites—say 4–6—there would be too few degrees of freedom to characterize the system. And, if data were pooled, we would need to have accurate weights on individuals in the community-based samples to represent the state, and it is not at all clear what would serve as the basis of these weights.

period information is unavailable, we may omit outcomes that are of significant importance for a variety of audiences. Hence we may wish to compromise and not require that only outcomes that can be evaluated be considered, even though evaluation remains our primary goal.

#### Trade-Off 4: Short vs. Long Time Frame

Finally, the fourth trade-off involves the desired time frame. Some outcome measures would be useful immediately, such as labor market and income measures. Some, however, are longer-term in nature, such as the impact of reduced parental time on young children. To assess that full outcome, many years will have to pass. (Which makes data collection and any causal inference even more difficult.) Moreover, for some outcomes, the short run and longer run may differ. In the very short run, with a full-employment economy the labor market outcomes may be quite positive; but if there is a subsequent downturn in the economy, these outcomes may change. Outcomes may also change over time because of program changes made to respond to unanticipated problems or changes in priorities of the state or local government. Even in the absence of program changes, outcomes may change over time. A sanction may be immediately deleterious to a family, but we may wish to know the medium- and longer-term outcomes. In any evaluation there are likely to be outcomes of immediate interest and those that must wait. In allocating resources for evaluation, a trade-off exists between funding an immediate evaluation versus a more comprehensive one which would require setting aside resources for future data collection and evaluation efforts.

## OUTCOMES OF INTEREST

We now turn to a representative set of outcomes and discuss issues common to a fuller set of outcomes. (The Appendix contains a comprehensive list of outcomes of interest.) In determining the focus of evaluation and monitoring efforts, we recommend considering the trade-offs identified above. We apply this approach to four representative areas: (1) work requirements, (2) child care, (3) child health, and (4) family formation. In each case we consider outcomes of interest to different constituencies and the potential importance of the trade-offs.

Much of the discussion that follows focuses on children. With the major changes demanded of parents, and resulting changes in parental resources, children in these low-income households are likely to be affected in significant ways. Children on average will spend less time with their parents, starting from young ages, and will spend more time in child care settings. Children of young mothers (<18 and not married) not currently

living with an adult or in a supervised setting will experience a change in living arrangements. Through added pressures, penalties, and uncertainties, more children may spend time with adults other than their parents, including foster parents. Locational mobility may increase, and with it, added uncertainties for the child. Requirements on parents, including increased working hours, may also affect a child's schooling as well as health status and medical care use. A broad-based evaluation would include outcome measures that attempt to capture these potential impacts on children.

### The Effects of Work Requirements

W-2 requires almost all participants to engage in full-time work in order to continue in the program. There are four tiers in the "Job Self-Sufficiency Ladder": unsubsidized employment, subsidized employment, community service jobs, and W-2 Transitions. Time limits are associated with all except the top tier, and with the overall program. Clients will be directed to the highest tier in which they can participate, as determined by their Financial and Employment Planner.

Outcomes associated with work requirements that are of interest to many constituencies include the distribution of participants across tiers; the probability and timing of movements toward (or away from) unsubsidized employment and the stability of such employment; earnings, benefits, and work-related expenses; administrative efficiency and cost of community service jobs and W-2 Transitions, changes in wages or job availability for low-wage workers not participating in W-2; and the availability, productivity, and cost of labor for firms.

*Labor market outcomes are of interest for all low-wage workers, for W-2 participants, and for selected groups.* Primary outcomes of interest for W-2 participants include the probability of initial success in securing a private sector job and the probability of maintaining such a job, as well as the timing of moves between community service jobs and subsidized and unsubsidized private-sector employment for those who begin their work experience in tier three, community service jobs. A variety of measures of compensation such as earnings, the availability of health insurance and other benefits, and work-related expenses are also of interest. The extent to which these outcomes differ across participant groups will be important in evaluating the adequacy of client evaluation, placement, and support services. For example, the new work requirements for mothers of young infants suggest that we may wish to pay close attention to outcomes for this group, which constitutes only a small fraction of the general population. At the same time, the timing and scale of work requirements raise the possibility of impacts on the general labor market, especially the market for low-wage workers. Concerns have also been raised about the potential for subsidized employment and/or community service jobs to displace current employees. To assess these outcomes will require data on nonparticipants,

especially but not exclusively low-wage workers. Thus, assessing work requirement outcomes will require consideration of the trade-off between collecting a broad sample and oversampling participants and/or particular groups of participants.

*Outcomes at the individual and community levels are of interest.* Some work outcomes, such as total earnings and benefits, and proportion of participants in each tier are of interest at the individual level and on a statewide basis. However, there are many outcomes, particularly those relating to the implementation of work requirements and the availability of planning and support services, where community-level outcomes may be of equal or greater interest. For example, both individual and community-level differences in the type of initial job placement and the probability and timing of transitions between tiers are likely to be influenced by local labor market conditions (openings, type of firms, etc.). Thus, not only individual but also community-level data will be important in assessing work requirement outcomes.

*Feasibility of impact evaluation, cross-community comparisons, and monitoring varies across outcomes.* By linking administrative data from the AFDC, W-2, and unemployment insurance (UI) programs, we can trace the earnings of program participants and other workers before and after implementation of W-2. UI records allow us to trace the earnings of all individuals who work in covered employment and to learn how these earnings vary over time and across employers. We might then use a pre-post design to evaluate the impact of W-2 on such outcomes as earnings and employment stability among W-2 participants and other workers. These data are available on a county basis, allowing cross-community comparisons as well. Nonetheless, evaluating the impact of work requirements on the well-being of the eligible population will be more difficult owing to the absence of information on individuals who may be discouraged from applying for W-2 services. This may be a particular problem for younger persons just finishing their schooling.

*Short-term and long-term outcomes are of interest.* W-2 requires a large and immediate transition to work. The feasibility of this approach and the nature of administrative and labor market incompatibilities are clearly of interest. It will be important to examine the extent to which wages and earnings increase over time and the potential for work experience alone to lead to improved economic status, especially in light of the retreat from long-term training and education and an increased focus on immediate job placement and experience. Employment patterns of continuing AFDC recipients and new entrants are expected to be substantially different, further suggesting the importance of long-term evaluation. Long-term changes in staffing and proprietorship of W-2 agencies (which ultimately may influence their cultures and functioning) are also important. Finally, the Wisconsin economy is currently very strong, with very low rates of unemployment. It will be important to evaluate the effectiveness of W-2 over the business cycle.



### The Effects of Child Care<sup>4</sup>

W-2 work requirements will increase the time children spend in child care. W-2 changes eligibility for state-subsidized child care, the types of child care providers that may be used, and the cost of care to participants. Because the state has announced that no mother with a child under age six will be required to take a job unless childcare is available, there is a clear state incentive to rapidly increase the supply of child care providers. To expand availability of care, the state has developed a new category of “provisionally certified” providers, who will not be required to undergo child care training and who will be reimbursed at a lower rate. Thus, there are reasons to expect a supply of lower-quality child care. W-2 requires copayments that are figured as a percentage of the family’s income. Because there is a differential in favor of certified as opposed to licensed care, there is an incentive to place children in lower-quality (lower cost) care.<sup>5</sup> In addition, any rush to increase the supply of formal or regulated care may lead to problems with the quality of the providers and/or facilities, at least in the short run.

Child care outcomes are of interest to diverse constituencies, including those concerned with effects on children’s well-being, with the effects of child care availability and cost on W-2 employment transitions and related costs, with the effects of availability and quality of child care on the general population, and with effects on the working conditions and wages of child care providers.

*Outcomes of interest include availability and quality of child care for all families, for W-2 participants, and for selected groups.* W-2 work requirements will substantially increase the demand for child care, especially for care of very young children (whose mothers were previously exempt, but must now work full time once the child is 12 weeks old). In addition, changes in the child care subsidy structure, which provides a subsidy up to 200 percent of the poverty line, will also increase demand for child care. Because of the timing and scale of the increased demand for child care, and given the incentives of the new copayment structure, W-2 may affect the quality and availability of child care for nonparticipants as well as participants. Thus, outcomes of interest include not only the availability and quality of child care for W-2 participants, but also for other low-income families and all families with children. Also of particular interest is the availability of child care for children with special needs. In assessing data collection priorities for the evaluation of child care, we will

---

<sup>4</sup>This section draws on the companion paper by Karen Fox Folk, “The Evaluation of Child Care Services under the Wisconsin Works Program (W-2).”

<sup>5</sup>There is a 30 percent differential, on average, between licensed and certified care. This means for example, that a family with an income level below 70 percent of the poverty line would pay \$40 to place two children in licensed child care but \$28 to place the same two children in certified care; at the poverty line, a family would be required to copay \$114 for licensed care compared to \$80 for two children in certified care.

face the trade-off between collecting a statewide representative sample of households (which might allow us to assess the impact on child care among all families) or oversampling particular groups (which might allow us to assess the impact on children with special needs, for example).

*Outcomes at the individual and community level are of interest.* A primary outcome of interest is the effect of care on child well-being, including such indicators as total time in care, ratio of providers to children at any one time, and stability over time and adjustments to child care. Also of interest is the use of informal care, and the probability and consequences of older children being left without supervision. The availability, cost, and convenience of child care also have important implications for the ability of parents to work and to have sufficient income to cover other expenses. Individual data will be necessary in order to evaluate these outcomes. However, community-level outcomes are also of great interest. These include the supply of child care by type, the success of W-2 agencies and community steering committees in coordinating child care services, and the interaction of public sector agencies, employers, community organizations and private charities. Community-level data will be necessary to evaluate these outcomes. Thus, in assessing data collection priorities for the evaluation of child care we will also face the trade-off between collecting individual or community data.

*The feasibility of impact evaluation, cross-community comparisons, and monitoring varies across outcomes.* Reliable baseline data are available on the regulated child care market, including enrollment, open slots, new facilities, and closings. Administrative data are also available on the use and cost of subsidized child care for low-income parents. These data are generally at the county level. A pre-post design might therefore be used to evaluate the impact of W-2 on the general regulated child care market and on the demand for subsidized care. Cross-county comparisons could also be made, comparing for example services in rural and urban areas. However, no systematic pre-implementation data exist for unregulated care. We also lack state baseline data on individual-level outcomes. This limits the ability to do an impact evaluation. Nonetheless, these outcomes may well be of sufficient importance to warrant data collection, if only to monitor effects.

*Short-term and long-term outcomes are of interest.* The availability of an adequate supply of child care in the short term is important, especially in light of time limits. Short-term outcomes related to child care administration, coordination, and the implementation of a new category of providers will also be important in assessing the need for change. However, many outcomes are long term in nature: the impact of child care on child well-being, school readiness, and later life outcomes can only be assessed over a long term. Other outcomes may change over time: the quantity and quality of child care may be expected to adjust over time in response to increased demand, the new structure of copayments, changes in child care workers' wages, and the new category of providers. The evaluation of child care will involve outcomes with a variety of time frames.

### Effects on Child Health

The changes resulting from welfare reform may also affect the health of children in a variety of ways. This discussion includes factors that may influence health directly and the measures of health that may capture other changes. Health-related outcomes of interest to alternative constituencies include child health status, nutrition, and ability to participate in school; parental availability to monitor health, take children to the doctor, and provide support to children with special needs; and the demand for health services and access to and costs of regular and emergency care.

*Outcomes of interest include measures of health of children directly affected, of children who are in contact with these children, and parents and the general public.* Studies of children in low-income environments find that they are more likely to be exposed to lead paint, to obstacles that can cause injury, to adults under stress who may engage in endangering acts, to poor nutrition, and to emotional stress. For children in families participating in W-2, other factors may come into play. Increased demands on the time of parents in W-2 may reduce their use of preventive medical care for their children, may increase the use of child care and of schools for ill children (as parents go to work instead of staying home with an ill child) and may reduce the provision of healthy food. One outcome that should therefore be monitored is the health status of children in W-2 families.<sup>6</sup> On the other hand, secondary health problems may also occur: higher rates of illness among W-2 children may lead to increased health problems among the children with whom they come in immediate contact, at child care centers or school. Finally, all of this may mean a change in demand on health care providers, including a demand for a change in hours of operation, an increase in the demand for services for acute illnesses and related expenditures by Medicaid and perhaps private insurers. Increased illness may also affect the operation of child care centers and of schools.

Among the *overall population* we might collect data on the proportion of persons by groups (especially children) who have health insurance coverage, who report excellent, good, fair, or poor health among children, the proportion of children 1–4 who are fully vaccinated, the proportion of low-weight births, and accident and injury reports from hospitals or police. These should be able to capture broader impacts of W-2 on children's health status.

*Outcomes at the individual and community level are of interest.* Health can be measured by anthropometric indicators such as height and weight-for-height, and by surveys of parents questioned about the general health of their child and more specifically about use of emergency rooms, etc. Some of these data

---

<sup>6</sup>The health outcome of specific groups may be of particular interest; for example, infants and children with disabilities may be especially vulnerable to the impact of new work requirements for their mothers.

are collected in existing surveys,<sup>7</sup> but these surveys do not contain a large sample of low-income children. Nonetheless, health status could be measured at the individual level if medical records on these children were made available.<sup>8</sup> Community-level data would be of interest because supply factors, such as the ratio of providers relative to the population, and the use of paraprofessionals, counselors, and other support personnel by provider groups (clinics), may all influence utilization of care and the follow-up use of medication. Responses to changes in demand, such as changes in hours of operation, are likely to affect changes in utilization, if not health itself. On the *community level* there exist such measures as a greater incidence of injury seen at hospitals, or greater or reduced congestion at medical facilities, which could be measured through average waiting times for care and to obtain appointments.

*The feasibility of impact evaluation, cross-community comparisons, and monitoring varies across outcomes.* Most of these measures of children's health could be collected for children in the pre-W-2 period by using medical records, vital statistics, existing surveys, and the like. Measures such as days of school missed, proportion of children with a diagnosed physical and/or mental disability, or proportion of children with a learning disability may be obtainable from administrative or school records in a selected set of communities and could be obtained for the pre-implementation period. Birth records could provide data on low-birth-weight infants; hospital records could provide data on the use of emergency rooms.<sup>9</sup> However, even if we lack

---

<sup>7</sup>Existing surveys include the SIPP, which has sampling units within the state, including Milwaukee, and the Wisconsin Family Health Survey, covering the whole state.

<sup>8</sup>Data on children covered by Medicaid could be obtained from the computerized collection of Medicaid records. Using these data or hospital data, we could analyze the proportion of children with categories of diagnosis usually related to accidents or violence. Medicaid data would also permit analysis of immunization rates among the covered population. Records of low-birth-weight infants in low-income areas or to mothers living in certain areas could be collected over time to see if changed policies affect the probability of low birth weight and hence the health of infants (and their future health). An indicator of emotional or mental health is a child's ability to engage in activities that are age-appropriate, such as playing with other children. Once children are in school or in preschools, teachers' records might provide such information. (In many cases these data may be at the community rather than the individual level.)

<sup>9</sup>For more on these options, see the conference paper by Sandefur (1997). A preferred strategy might focus on children enrolled in Medicaid, using computerized Medicaid data on a variety of indicators of health and medical care use among this population. Alternatively, a sample of children in W-2 families might be identified, their medical care provider approached, and permission requested to obtain medical records for a specified period of time, including some years prior to welfare reform as well as after. This might be much easier among a random sample of children covered by Medicaid or for cases in a hospital than if a large number of individual providers or HMOs have to be contacted. Nevertheless, it might be possible to do a case study with the cooperation of a limited number of HMOs serving the low-income population, or serving a sizable number of families covered by Medicaid. This would be complemented by data from the Wisconsin Family Health Survey.

resources to collect the data necessary for a pre-post evaluation, certain health outcomes, such as emergency room use, are of sufficient importance to warrant collecting these data for monitoring purposes.

*Short-term and long-term outcomes are of interest.* A number of short-term outcomes will give us information on the immediate effects of W-2 on child health. We might expect to see changes in days of school and/or child care missed, in the proportion of ill children in these settings, in the rate of injuries, and in the use of medical care in communities that serve a high proportion of W-2 families. Still other outcomes are of interest in both the short and longer run: for example, we should be interested in measures that would capture both short- and long-term effects of increased stress on children. These data could be gathered in surveys of parents, teachers, and child care workers. Other changes resulting from W-2 may take far longer to have an impact. Only over time should we expect to see effects that reflect nutritional changes, which are indicated by anthropometric measures.<sup>10</sup> Similarly, among children aged 1–4, the proportion without immunizations, or with an incomplete set, may gradually change over time. And the proportion of the low-income population with health insurance coverage may change as parents change employment tiers, or as children's or family's eligibility for Medicaid is lost.

#### Effects on Family Formation

One of the philosophical principles behind welfare reform is to encourage parents to be responsible for their children. Another principle, explicit in the federal reform bill, is to reduce teenage parenting, nonmarital parenting, and in particular nonmarital teenage parenting. The national legislation requires teenage mothers to live at home or with a responsible adult in order to receive any benefits and moves away from the restrictive criteria under AFDC that gave benefits primarily to single-parent households. Under a number of welfare reform plans, including W-2, both two-parent and one-parent families are eligible for assistance. Noncustodial parents whose children live in families eligible for TANF services may themselves qualify for a variety of services. TANF programs may also have a family cap, under which benefits (cash benefits in particular) do not increase with the birth of additional children. And a teen's parents may face requirements concerning the care of any children born to their underage child. All of these programmatic changes are expected to change incentives with regard to family formation.<sup>11</sup>

---

<sup>10</sup>Because anthropometric measures reflect nutrition over a length of time, collection of items such as height and height-for-weight at an early point in W-2 would serve as a measure of nutrition prior to W-2. Measures several years later would be post-W-2 measures. Unfortunately, however, this effort would be relatively expensive.

<sup>11</sup>The welfare reform programs, particularly W-2, change incentives in complicated ways. For example, the ability to get benefits (services) as a two-parent family may make it attractive to marry. However, a noncustodial parent may now get benefits, which would seem to make it more attractive to be a noncustodial parent.

Outcomes of interest with regard to family formation include living arrangements—whether married or not, whether cohabiting in a stable relationship, whether living with one’s parents or other adults. We would also wish to collect measures of subsequent fertility, such as delaying or forgoing subsequent births. Among noncustodial parents, we would like to know whether involvement with their children has been influenced by the change in welfare reform.

*Outcomes of interest include measures of family formation among the population most directly affected (single mothers with limited assets and income), teenagers who are potentially eligible for services, noncustodial parents, and selected subgroups of the general public.* Among the primary variables of interest under family formation is the rate of teen out-of-wedlock fertility among the “at risk” population. This population includes those who have already given birth as an unmarried teen as well as younger teens living in low-income communities. We might also wish to collect data on teen fertility, teen out-of-wedlock fertility, and out-of-wedlock fertility for the entire population and subpopulations of the state, since such data are readily available over a long period of time. They would also serve as a type of control for changes in fertility rates that might be due to other events and changing circumstances within the state.

Another measure of considerable interest is the time parents spend with their children. Of particular interest is involvement of fathers in their children’s lives. This might include measures of paying support for and of time spent with the children and other contacts (phone calls, attendance at school and sports activities, etc.). We would concentrate on the fathers of children who are eligible for W-2, but might also wish to include children who have lived in households on AFDC prior to the change in the law. If society is successful in increasing the involvement of fathers in the lives of their children we might expect to find this reflected in the broader population, and so would like measures of a noncustodial parent’s time spent with children who are growing up in single-parent households, and with children who began life in a single-parent household, as well as the proportion of children living in single-parent households.

*Outcomes at the individual and community level are of interest.* We may wish to compare individual data on marital status, childbearing, and living arrangements of W-2 participants with past AFDC recipients. In particular communities, or zip-code areas or census tracts, that have high rates of participation in welfare-related programs, it would be useful to compare rates of out-of-wedlock childbearing by women’s ages over time. TANF includes substantial funding for abstinence education programs as well as financial incentives for states that reduce nonmarital childbearing. To evaluate the effectiveness of education programs and other efforts we would need to collect indicators of education and attitude changes at the community level.

*The feasibility of impact evaluation, cross-community comparisons, and monitoring varies across outcomes.* The easiest data to collect, and for which historical data exist, are administrative data from birth

certificates, containing information on mother's age, whether the mother is married, race of mother, and whether she had previous live births. The mother's county of residence and city, village or township would allow aggregation of these data to the community level. Also easily collectable are rates of marriage by age and county. Similar information exists for divorce and annulments. The data on these certificates also include presence of children and award of custody. In Wisconsin, data from administrative records and surveys of parents are available from the Wisconsin Court Record Database (WCRD). These data have some limitations,<sup>12</sup> but might serve to construct a baseline. Collecting data on the amount of time absentee fathers spend with their children is more difficult. However, there are a number of data sets that might provide approximations of the time allocation before W-2. For example the WCRD parent surveys include questions on contact between children and noncustodial fathers in Wisconsin, and the PSID has a long history of asking such questions. For time allocation after W-2, survey data would be necessary.

*Short-term and long-term outcomes are of interest.* We can observe short-term effects on teen fertility, marriage rates, divorce rates, and the parenting role of the noncustodial parent. Over time, however, the patterns may change. In all of the family-formation measures, the short-run response may be smaller than, and even of different direction from, the long-run response. For example, if the teen nonmarital birth rate is reduced in the short run through a variety of targeted programs, they may change the culture as well as contraceptive knowledge, which will further reduce such births. If fathers allocate more time to their children, their relationship may be closer and would increase over time. However, if the father feels that too large a proportion of his earnings are allocated to the child, or the child's demands are too great, this allocation of time might decline in the longer run.

## CONCLUSIONS

Because resources are limited, choices will have to be made regarding the principal outcomes of interest. These cover a broad range, from the effects of employment requirements on participants, to effects on child well-being and family income, to child care quality and availability across the entire population. Some important outcomes may be evaluated using existing or easily available administrative or survey data; such evaluation should clearly be pursued. However, to gauge important effects of welfare reform, some new data will be required. In deciding what data to collect, all the trade-offs must be recognized. Should data be

---

<sup>12</sup>The database contains records of divorce and paternity cases from 21 Wisconsin counties. However, these data may be limited because of the small sample of AFDC-eligible cases.

restricted to eligibles, or also include near eligibles, or focus on special groups, such as families with a disabled child, or be spread to the entire low-income population? Should the data base attempt to link outcomes to community opportunities and structure, to the structure and personnel of the service programs, and to local nongovernmental organizations, making the community the organizing basis for data collection, or should the data base cover the state more generally? What can be learned about the state's programs if the community approach is followed? Should only those impacts for which we can construct a counterfactual be included, or are there outcomes so important that we wish to measure them regardless of their evaluation potential? Over what time period should we measure these outcomes?

Resources for evaluation are clearly limited. Prioritizing is essential but difficult. Different constituencies would make different choices. We would emphasize (1) those aspects that can be evaluated rather than simply monitored (but not to the absolute exclusion of important effects not available for the pre-reform period); (2) those impacts that can be evaluated using administrative data (which will of necessity emphasize the target population rather than the overall population or near eligibles); and (3) broad measures of well-being in the overall population that can be drawn from existing, large-scale data sets such as the CPS and the Wisconsin Family Health Survey. We view the major choice remaining to be made in evaluation design as the choice between a panel survey of the low-income population of the entire state versus a community-based panel survey of the low-income population; within these, choices involve focusing on a particular group of interest (e.g., children with disabilities) versus a larger sample of the low-income population. Because of the potential importance of differences in service availability, job opportunities, and other community-level factors, we would argue for a community-based panel survey.



**APPENDIX TABLE 1**

Outcome	Trade-Offs			
	Participants/ Population	Individual/ Community	Evaluation/ Monitoring	Short-term/ Long-term
I. Well-Being of Participants				
A. Economic				
1. Work experience				
2. Earnings/compensation				
3. Other sources of income & benefits in kind				
4. Total income relative to needs				
5. Dependency: transfers as a proportion of income				
B. Non-economic				
1. Health Status				
2. Education/training				
3. Stability of interpersonal ties				
II. Well-Being of Low-Income Families				
A. Child well-being				
1. Health				
2. Education				
3. Child care				
4. Nutrition				
5. Maltreatment/child well-being				
B. Family structure/formation				
1. Family Structure				
2. Time with parents				
3. Paternity establishment				
4. Birth rate; out-of-wedlock birth rate				
By age group				
C. Family Human Capital				
1. Parents' education				
2. Parental investments in children				
D. Housing				
1. Homelessness				
2. Density				
3. Geographical mobility				
III. Labor Market Outcomes				
A. Firm				
1. Vacancies				
2. Net Cost per low-skill employee				
3. Worker productivity				
B. Other low-wage workers				
IV. Externalities				
A. Crime				
B. School quality				
C. Child care				
D. Health care				
E. Migration				
F. Supply of Auxiliary Service Providers				
V. Fiscal Effects				
A. Total costs				
B. Allocation between federal, state and local				

# Controlled Experiments in Evaluating the New Welfare Programs

Glen G. Cain

## I. INTRODUCTION AND CONCLUSIONS

A conclusion of this paper is that a controlled experiment is not a practical method for evaluating Wisconsin's new welfare system, labeled Wisconsin Works (W-2), and is unlikely to be practical in any other state. Controlled experiments are seldom conducted to evaluate even small-scale social programs. None have been used for such large-scale economic programs as federal minimum wage laws, federal tax legislation, health care programs, or even for state-level versions of these programs.<sup>1</sup> This has not prevented economists from usefully employing a variety of observational studies to evaluate these and many other programs, even though such studies tend to be inconclusive about the quantitative magnitudes of the program effects and are often inconclusive about even the qualitative impacts of the programs. Another, and more upbeat, conclusion is that controlled experiments can be used to evaluate component parts of the new welfare programs. Finally, by analyzing the methods of controlled experiments we can better choose alternative methods for the majority of cases when controlled experiments are not feasible.

One problem that looms in the background of evaluating the new state welfare programs is the uncertainty about the appropriate counterfactual. The question of whether the new system is working well or poorly must be posed in conjunction with the question: Compared to what? The "what," the counterfactual, that is of greatest interest to policy makers is some feasible alternative program. In most evaluation studies that use controlled experiments, the alternative to the experimental program is the status quo. I doubt, however, that state officials view a return to the old system as an alternative to the new system, particularly since restoring the old is not under their full control, given the 1996 federal welfare legislation. Instead, comparisons between or among states are more policy relevant; that is, relevant to the question of what changes need to be made to improve the new welfare system of the individual state. In a controlled experiment, however, what would the experimental (alternative) welfare program be, what would it look like, if the old system is ruled out? The question has received no attention by state government officials because none (to my knowledge) has proposed

---

<sup>1</sup>One exception to this statement is the U.S. Department of Labor's evaluation of the Job Training Partnership program, which combined both experimental and nonexperimental designs, beginning in the early 1990s. How successful either of these evaluation designs were is still under debate.

using a controlled experiment. Wisconsin officials who are planning the state's evaluation have explicitly rejected this method.<sup>2</sup>

The choice of a counterfactual also arises in evaluating the new state programs by observational studies. Consider the distinction between a comparison that is appropriate for scientific analysis and one that is appropriate for policy analysis. Certainly, there is profound scientific interest in whether the new state programs are working well or poorly compared to the old system, and the interest is by no means confined to economic historians. A comparison to the old system has the virtue of specifying a well-defined counterfactual. Its weakness is its limited policy usefulness. In contrast, the counterfactual of some "viable alternative welfare program," although it is of greatest policy relevance, has not been defined. This paper gives attention to both controlled and observational evaluation methods.

Part II examines a conventional example of an experiment in welfare reform to illustrate both the potential power of the method and standards for achieving a good evaluation. Part III deals with specific problems in using controlled experiments to evaluate the new state welfare programs, with a focus on the particular difficulties posed by W-2 in Wisconsin. Part IV briefly discusses evaluating a state's new welfare program with observational data across states. Two other papers, one by Robert Haveman and one by Thomas Kaplan and Daniel Meyer, examine a within-state, pre/post design, where the old welfare system is the counterfactual.

## II. THE CONTROLLED EXPERIMENT AS THE EVALUATION DESIGN THAT SETS A STANDARD AGAINST WHICH OTHER EVALUATION DESIGNS CAN BE COMPARED AND JUDGED

In its conventional application the controlled experiment tests some new program to allow a comparison of its outcomes with the outcomes from the existing system. Consider experiments to determine whether a training program increases the employment, earnings, incomes, and affects other outcomes for welfare mothers. Typically, a sample of welfare mothers would be selected and randomly assigned to either a treatment group or a control group. The treatment group is eligible to receive the training, while the control group continues in the existing welfare system without access to the training program. The goal of the experiment is to measure what the outcomes would be if the experimental program were made a permanent part

---

<sup>2</sup>As indicated by the Research Plan contained in the Wisconsin Waiver Proposal, submitted to the federal government, May 28, 1996.

of the welfare program and were made available to all welfare mothers. The experiment is applied to the population of welfare mothers, including new entrants who would also be randomly assigned to the treatment and control groups.

Assume this experiment has the following three positive features. (1) The training program in the experiment is similar to what the program would be if it were extended to the entire population of welfare mothers. Similarity refers to the content and quality of the program, including the quality of the facilities, administrators, trainers, and so on. We may then assume that the treatment group's response to the experimental training is similar to the way welfare mothers would respond to a permanent training program applied to all welfare mothers. (2) The treatment and control families are followed for a sufficient length of time to determine the post-training and long-run outcomes under study, using household interviews and other sources of data, such as administrative records. Persons and families in the treatment or control groups who move would be followed and surveyed in their new locations. Both features, (1) and (2), are necessary to achieve unbiased estimates of the relevant outcomes of interest—the full (or long-run) benefits and costs of the permanent training program. They contribute to the validity of the experiment.

(3) The third feature is that the proposed training program is voluntary, and this is considered positive for a less obvious reason. A voluntary program refers to the welfare mother's choice to participate or not to participate in the training program. In the experiment, those who choose not to participate, as well as those who drop out of the program, would remain in the treatment group for purposes of comparison with the control group. The reason a voluntary training program is considered a positive feature is not because it makes the experiment more valid, but rather because being voluntary removes a possible ethical objection; namely, that mandatory training could make some of those randomly assigned to the experimental worse off compared to the control group. The control group and welfare recipients who are not in the experiment do not receive the training, but neither do they risk losing welfare benefits if they do not participate in the training program. If the permanent program were mandatory—requiring participation in the training program to receive welfare benefits—then the experiment would also have to have that feature; otherwise the experiment would not test the proposed program. The ethical objection to an experiment involving mandatory training may be overruled, but I mention the issue because it arises in discussing controlled experiments of the new state welfare reforms.

Now consider two potentially important weaknesses in the training experiment. One is the possibility that if the experimental training program were adopted and made a permanent part of the welfare system, the population of new entrants would change. If the training program were notably rewarding, some women who would not have entered the old welfare system might decide to enter the new system. Conversely, if the training

program were, say, mandatory and onerous, some women who would have entered the old welfare system may not enter the new system. The experiment would not detect these outcomes because the experimental training was not imposed on all new entrants to the welfare program.

A second weakness is that a full-scale training program, unlike the relatively small experimental program, is likely to have an adverse effect on the overall low-wage labor market by substantially increasing the supply of low-wage workers. Without measures to increase the demand for low-wage workers, the expected market response to the increased supply is a decline in wages and other conditions of employment of low-wage workers and an increase in their unemployment. Almost all low-wage workers are not on welfare, so this population was not, and could not have been, studied to determine its response to the experiment.

Neither of the two weaknesses just discussed implies that the experiment would be a futile endeavor. Rather, the message is that an experiment can answer some, but not all, important questions. As described above, the experiment offers persuasive evidence of the ability of a replicable training program to affect the earnings capacities of welfare mothers. The experiment's central question and its intended policy decision are well defined. How do the outcomes from the training program compare with those of the existing system, and should the existing system be changed to include the training program? The first question is fundamentally scientific; the second question is fundamentally political. The experiment permits the measured change in outcomes to be attributable to the training program rather than to changes in the economy or to various characteristics of the women in the treatment group. The experiment also provides most of the information needed for a benefit-cost analysis of the training program, which is essential for the policy decision. The experiment's measure of gains from the training would probably require a downward adjustment to allow for the smaller expected gains in a permanent program, because of the marketwide supply effects from the full-scale program. The costs of the full-scale program, on the other hand, might be adjusted downward in anticipation of economies of scale.

Note that the experiment could also provide information about the child care arrangements of the children of welfare mothers who undertake training and obtain jobs. Indeed, a child care program might have been usefully included as another component of the experiment, but note that in this case the experiment would not be able to measure how a full-scale child care program would affect marketwide prices and other conditions of the child care industry. Similarly, the training experiment could provide information about whether and how women in the treatment group obtained health insurance. Alternatively, health insurance could have been another component of the experiment. If the experiment contained three components—training, child care, and health insurance—then all the outcomes being studied should, in general, be attributable to the package of the

three components, and assigning the causal effect of any particular outcome of the experiment to one of the components may be unwarranted. But such individual attribution of causality is not necessary if the full-scale permanent program matches that which was carried out in the experiment. The method of controlled experiments in evaluating changes in welfare programs is appealing in many settings.

### III. CAN AN ONGOING COMPLETE CHANGE IN A STATE'S WELFARE SYSTEM BE EVALUATED BY A CONTROLLED EXPERIMENT? A PESSIMISTIC APPRAISAL

In describing the evaluation of the hypothetical training program, emphasis was placed on the clear policy decision at issue: whether or not to adopt the training program in the welfare system. What is the comparable decision at stake in evaluating a new state welfare system; specifically, W-2? Has anyone said, "If W-2 does not work out, we will go back to the old system?" The question answers itself. The old system was part of a national program that has been repealed and replaced, so the closest Wisconsin could come to this alternative and still operate within its legal boundaries would be to use its block grant money to install a system with as many features of the old system as current legislation permits. I do not know of anyone who has given any thought to what such an alternative state program would be, and designing one and implementing it as a controlled experiment seems hopelessly unrealistic, even if the technical difficulties (discussed below) of carrying out such an experiment could be overcome.

Another justification for using the existing welfare system as a basis for comparison with W-2 is simply that comparing the social impact of the new program with the old is an interesting scientific question with findings that would be useful to making changes in the new program. On these grounds Wisconsin could request permission—a waiver—to conduct an experimental program that consisted of the old system. But unless returning to the old is a realistic policy option, it is difficult to see much public support, let alone enthusiasm, for such an experiment. I discuss below comparisons for W-2 other than the old system, but first let us look at the technical and logistic problems facing a controlled experiment, with the old system as the treatment program and the new system providing the control group. I see five important problems, which I order from least serious to most serious, using my imperfect judgment about this ranking.

1. Using a controlled experiment to evaluate the new welfare reform may raise ethical objections if the random assignment to the treatment group would make the families worse off than participating in the ongoing

new welfare program.<sup>3</sup> The ethical issue does not arise if the new welfare program in a given state is unambiguously less desirable to welfare mothers than the old system. If this were true, then the experiment and its random assignment of some families to the old welfare system would be benign. W-2, however, makes some changes in the old system that are, at least on paper, beneficial; improved child care arrangements and Medicaid extensions to poor husband-wife families are examples. There may be ways of dealing with the ethical problem, perhaps by providing a lump-sum compensation that overcomes a family's opposition to participating in the treatment group, but such devices are complications and interfere with the integrity of the experiment.

2. A second problem is that the experiment would have to be relatively long, say five or six years, to represent behavior that matches the behavior of those who would be in the old welfare program if that program had continued to be the state's program or if it were reestablished. Long-duration experiments are more difficult to carry out and more expensive than short-duration experiments, but it is the reason for a long-duration experiment that highlights the importance of the problem. To see this, consider the example of the training experiment. Here, the welfare mothers in the treatment group are in a training program that is expected to be completed in a year or so, and the effects of this on the participants' employment and earnings are expected to be revealed in, say, one or two more years. Finally, the training program is to be adopted if the experiment shows it to be successful.

In sharp contrast, if the experiment consists of the old program the relevant time duration for its participants varies from one to 20 years. Thus, to view the welfare mothers' behavior in the treatment group as representing what their behavior would be if the treatment (the old system) were in place, we would have to assume that they believed, or acted as if, their treatment were part of a permanent program. Perhaps the average time horizon for welfare mothers is five or six years, with some women expecting to be in the welfare system only a year or less and others expecting to be in it for 15 to 20 years. If the experiment lasted only a year or two, many women in the treatment group would not behave as if they were in the old welfare system but rather in some temporary program, with the new welfare program as their soon-to-be destination. A time period of five to six years for a controlled experiment in which the old program is the treatment group is not an insurmountable problem, but it is a drawback, and a drawback that is probably exacerbated by the fact that the old program is not seriously considered as an alternative to the new program.

---

<sup>3</sup>Clearly, if persons assigned to the treatment group that consisted of the old welfare system were permitted to refuse and to enter the new program, then the validity of the controlled experiment would be fatally impaired. The assignment process would no longer be random, but would instead depend on various unobserved factors that would prevent the attribution of causality to the treatment.

3. The third problem is that the new state welfare systems will not be in their steady-state mode of operation for some time, the length of time depending on many factors: the complexity of their systems, how radical the new changes are, the quality of the state agencies, the stability of their economic environments, and so on. Perhaps two years or so will have to elapse before the new welfare system settles into its normal, long-run operational status. A comparison between the assigned control group in the new system, which is starting up, and the treatment group in the old welfare system, which has a 60-year history, could be criticized as being not only unfair to the new system but meaningless. Again, contrast this evaluation of the two welfare systems with the evaluation of the training program, where the latter was the new reform and contained the treatment group, and the old system of welfare provided the control group.

The fact that the new state welfare programs will not be operating smoothly for some time is a challenge to any evaluation strategy, but the problem seems more serious if the evaluation is a controlled experiment that compares the old and new systems. Consider an alternative evaluation of W-2 based on an observational study of other state programs. Although this evaluation design has many problems, it does compare programs that are in similar stages of development.

4. The fourth problem is that components of W-2 have been under way in Wisconsin for a year or more, which implies that the old welfare system no longer exists at the time when the experiment would begin. Thus, assigning a treatment group of families to the old system will involve families that are already altered (“contaminated”) by exposure to and participation in the new program.<sup>4</sup> The empirical question is whether there is a sufficiently large group of current welfare recipients who have been unaffected by the new programs and who, therefore, can be assigned to a treatment group in the old system and to a control group that remains in W-2. Note that this obstacle to experimenting with the old system in Wisconsin does not apply to all states. Some states have made no changes in their current welfare system and others have made minor changes that have affected (and “contaminated”) only small numbers of families.

5. A fifth problem with a controlled experiment that has welfare families randomly assigned to the old welfare system is that the state’s economic environment is expected to change under the new system in ways that it would not have under the old system. Therefore, the outcomes of the treatment families in the old welfare plan are unlikely to represent what their employment, housing, child care, and other outcomes would be if the old system had been maintained. A clear example is the outcome of finding a job. Under the old welfare system, a substantial fraction, say 25 percent, of welfare mothers find a job that takes them off welfare during a period

---

<sup>4</sup>This criticism of the controlled experiment design is emphasized by Haveman in his paper.



of, say, one to three years. Thus, if the labor market for the (generally) low-skilled women in the treatment group did not change, then 25 percent of them would be expected to find jobs over the given period. However, we know that the pressure for welfare mothers to find jobs under the new welfare system will be great, making employment for welfare mothers in the treatment group (under the old system) more difficult and less rewarding. Assume only 10 percent of these treatment-group women find jobs and leave welfare. Assume further that 40 percent of welfare mothers in the new program become employed and leave welfare. The new system would look good compared to the old system by this criterion of success—a 40 percent success rate compared to a 10 percent success rate. But the new system's superior outcome is upwardly biased; the (hypothetically) true difference is 40 percent compared to 25 percent. If the old welfare system had been maintained, which is the counterfactual under study, then its welfare mothers would not have faced the expanded labor supply, and they would not have had the sharp drop in employment.

Although a similar weakness in the controlled experiment for evaluating the training program was pointed out, the problem is more severe here. The treatment of training was not influenced by the behavior of women in the control group and the existing welfare system. Also, the crucial outcome was the test of whether the earnings capacities of the women in the treatment group could be increased, and the difference between the employment and earnings outcomes observed during the experiment and those that would occur in the full-scale implementation of the program seems minor and amenable to a modest adjustment.

In summary, the obstacles to using controlled experiments to test new welfare programs against the old programs seem insurmountable. If the technical problems could be overcome, there is, indeed, intense scientific interest in the comparison between the two systems, but the fact remains that the old programs are not the relevant counterfactual in the sense of being realistic alternatives to the new programs.

The best chance for a useful controlled experiment to test the new program against the old program exists in states that (1) have new programs that are decidedly less generous than the old program, (2) can carry out an experiment for at least five years, (3) have new programs that are relatively simple and can be operating at their steady-state level quickly, and (4) can begin the experiment before the new program gets under way. The value of controlled experiments that test some reform or component of an ongoing program that has reached a state of normal operation remains, and there will surely be ample opportunities for these experiments in the future.<sup>5</sup>

---

<sup>5</sup>Kaplan and Meyer mention the example of a controlled experiment to test the effects of different copayment schedules for child care. This would not only affect the quantity and quality of child care purchases, but would be expected to affect employment and earnings of the parent (or parents)—directly because of the effect of child care on the parent's employment and indirectly because the copayment schedule affects the net wage and the net earnings of

It is useful to consider briefly the arguments against a controlled experiment of W-2 raised in two other conference papers, by Haveman (see footnote 4 above) and by Kaplan and Meyer, who mention four objections. Their first is that so many “program elements are undergoing change” that administration of the experiment is difficult and perhaps not feasible. Using the old welfare system as the experimental group, however, does not appear administratively complex, because it merely requires that state officials continue doing for the treatment group what they have been doing. The problem of attributing causality to the individual program elements has been discussed above and does not seem necessary in the evaluation. Second, the state agency (in Wisconsin) will be uncooperative. This is a problem, of course, but if the state officials have valid reasons for opposing the program, then these reasons are the immediate issue of contention.

The third argument of Kaplan and Meyer, which is the most complicated, is that there would be “contamination” between the control and treatment groups. The precise nature of such contamination needs to be examined before judging this criticism. One version, argued by Haveman in his paper, is that the new welfare program promises to “change the culture,” which implies that the experimental group under the old welfare system will behave differently than they would if the new welfare system had not been implemented. Presumably, the same claim and the same criticism could be made of the past experiments of negative income tax programs. Would the cultural changes that would occur if the negative income tax replaced welfare be so extensive that, say, the labor-supply outcomes of either the treatment or control groups were incorrectly estimated? Economists would view this form of cultural change as a change in workers’ tastes (or preferences) for market work. Is that likely? Perhaps the cultural change in the wake of the new state welfare system takes the form of certain institutional changes; for example, an expanded role by private charities. Again, we would need to examine how this change would bias the experiment’s estimates of specific outcomes. Still another type of “contamination” is that people in the randomly assigned treatment and control groups will behave differently just because they are aware of the existence of each other. Again, this problem was not raised, as such, in the negative income tax experiments. Was this an oversight in the analysis? The famous “Hawthorne effect” was considered in the negative income tax experiments, but the program was considered sufficiently unobtrusive to avoid this problem.

Finally, Kaplan and Meyer refer to “community feedback effects” as a problem with the controlled experiments. One type of such feedback might be cultural change. Another is the feedback in the form of a market-wide change that biases the experiment’s estimate of the change in outcome. This problem has been

---

the employed mother (or father) in the family. A higher copayment imposed when earnings rise acts as an effective tax on the rise in earnings.

discussed above (see point 5) and considered quite serious in the case of a controlled experiment in evaluating the new state program; less serious in the case of the training program.

#### IV. BRIEF REMARKS ON AN ALTERNATIVE EVALUATION STRATEGY OF COMPARISONS AMONG STATE PROGRAMS

Five summary points follow.

1. In evaluating a state's new welfare program, the programs in other states appear to have more policy relevance—although less scientific interest—than the old system in the given state. Cross-state evaluations will not produce a rigorous evaluation, in the sense of obtaining unbiased estimates of the causal effects of even the state program as a whole, let alone isolating the causal effects of the various components of the state program.<sup>6</sup> “Ball park” estimates are more likely. Nevertheless, the idea that the programs in other states offer alternatives, in whole or in part, to a given state's program is realistic and is in the spirit of the old tradition of viewing variation in state programs as “experimental laboratories.” Note that virtually all states would have programs in similar stages of development, so the comparison studies would not be biased by the factor of time-in-operation.

2. A preferred strategy for gathering the data for the evaluation requires surveys across states before the new programs get under way (or soon after) and a subsequent longitudinal survey design. With these data the analytic model commonly called “difference-in-differences” can be used. An important point is that surveys, compared to administrative data, are uniform and easily comparable across states, cover a wider range of behavioral outcomes, and cover groups who are not participating in, or in the records of, the welfare system. Repeating a point made above, welfare changes will affect persons who are not “on welfare,” especially persons with low wages and families with low incomes. I should acknowledge, however, that a more optimistic view of the scope and coverage of administrative data is discussed in the Kaplan-Meyer paper, where a number of favorable features of the administrative sources of data in Wisconsin are described.

3. The longitudinal surveys must continue over a relatively long period, say six years. The long duration is necessary for several reasons. Many outcomes of interest occur after several years elapse, such as the effects of the mandatory terminations of welfare after a period of up to five years. Other behavioral outcomes, like family composition and fertility, may take several years to appear. Another important reason for a long-duration study is to avoid several biases that are likely to appear in the short run. I suggest that the

---

<sup>6</sup>The problem in attributing causality to component parts of a program also exists with controlled experiments, as noted in Part II.

new programs will, on the one hand, improve over time in their performance as administrators learn what works. On the other hand, the program's performance will appear to decline because the administrators have every incentive to "cream" in the early stage of the new program. That is, they will tend to find jobs first for those on welfare who are the easiest to place, and they will place their clients in jobs that are easiest to fill, such as low-level public jobs or jobs in subsidized nonprofit agencies. Moreover, jobs will be easier to find early on, because as the number of entrants to the labor market increases over time, wages and job opportunities will decrease. All this will make the program look good in its first or second year and help the program gain early political favor. (These are not necessarily objectionable motives, but they do affect the evaluation.)

4. Cross-state comparisons are natural, because the programs are legislated to vary across states, and under normal circumstances all the state's inhabitants are covered by the same. In contrast, if the evaluation design is based on within-state variation, such as variation across counties, these variations must be imposed as exceptions to the law. Or, more generally, the reasons there are differences in the county programs would have to be known and taken into account in the models used to explain variation in the outcomes among counties. Also, the comparisons across counties that have variation in their welfare programs will be biased by intercounty migration that is selective of families on the basis of unobservable characteristics that are themselves causal to the outcomes of interest. Interstate migration will also complicate the interpretations of interstate comparisons, but this should be less frequent, because such migration is more costly than within-state moves.

5. The diversity in state programs has good and bad consequences for evaluation. Wide variation in the causal program variables is useful in predicting and explaining outcomes. But we know that some program variation is an effect of underlying variation in the states' economic and social conditions. Unless these underlying conditions are measured and controlled for, biases in the estimates of program effects can be severe. For example, the large stock and flow of immigrants in California will partly determine that state's welfare program, so any comparison of outcomes between California and other states will have to take into account not only immigration as a feature of California's economic conditions but as a cause of certain components of the welfare program in California. Depending on the context, immigration can be viewed as both a cause and an effect of the California welfare system. Interestingly, in Wisconsin no one talks about immigrants as a group that will be affected by W-2, although there has been considerable attention to welfare-induced migration from Illinois to Wisconsin.

A persuasive program evaluation based on cross-state comparisons is basically difficult because it ultimately depends on statistical controls for not only good measures of program differences but for the levels

and changes of the states' economic and social environments. Picking similar states to compare, controlling for state differences in observable environmental factors, and the "difference-in-differences" model are ways that can only lessen, but not eliminate, the basic methodological problem. Some degree of reliance on theoretical reasoning is needed. For examples, if a state provides higher benefits and longer time limits for its welfare recipients, its job placements are likely to be lower; if the state offers a more generous child care program, then its job placements are expected to be higher, and so on. Finally, the state programs may be ranked by the level of their generosity, and even if specific explanations for the differences in outcomes cannot be attributed to specific components of the state programs, there is a value and use in knowing the relation between various outcomes and the overall levels of program generosity. I end, therefore, with the modest proposal of evaluation by the traditional means of a theoretically based observational study.

## **Prospects for Comparing Wisconsin Works to Welfare Reform Programs outside the State**

Irving Piliavin and Mark Courtney

The issues which need to be studied as part of the evaluation of welfare reform efforts are numerous and quite complex. The research approaches that might be used to study the issues are also numerous and at times quite complex. While we hope eventually to address these complexities with the thoroughness they deserve, we have not done so as yet, a flaw which we hope will not seriously detract from the paper's intent.

The evaluation of a state's new welfare reform effort obviously requires one or more counterfactuals—what would have happened in the absence of the new program. One counterfactual is to reintroduce, or maintain, among a portion of program eligibles the elements of the welfare program previously implemented. These elements would presumably constitute the state's version of AFDC. In some states (e.g., Wisconsin), recent and numerous changes introduced in AFDC programs may make it necessary to decide on some version of the prior AFDC program. As an experiment, this would involve assigning individuals within an area (say, a county) by some probabilistic method to either the AFDC alternative or the reform alternative. Another design would be to assign the program alternatives to entire counties, with all eligible persons in a given county assigned to whatever program is introduced in that county. The latter design is preferred by some analysts, who believe it better approximates the conditions and consequences of permanent programs. The variants of this experimental approach as well as their strengths and weaknesses are taken up elsewhere in this conference.

Another approach for evaluating welfare reform efforts, to which we devote most of our attention here, involves comparing the experiences of applicants and/or participants of welfare reform efforts across states. Lacking the controls of an experimental design, this approach obviously does not permit an explicit examination of whether families' responses to welfare reform programs differ from those that might have been the case under AFDC. On the other hand, assuming that welfare reform efforts can be ordered in terms of some important dimensions (e.g., financial generosity, work requirements, employment opportunities) it will be possible, given sufficient program sites, to examine, under relevant statistical controls, the effects of variation on a number of applicant/participant experiences.

We distinguish cross-state comparison approaches along two dimensions: sources of information and type of design. We restrict our discussion to two sources of information and three forms of design. The two information sources consist of (a) merged agency administrative data, and (b) interviews with agency staff

members and program participants.<sup>1</sup> The designs include single-wave cross-sectional studies, multi-wave cross-sectional studies, and multi-wave panel (longitudinal) studies. While we assume that data dealing with pre-welfare-reform experiences may be available from agency data bases, we assume this will not be true for information supplied through program participant interviews.<sup>2</sup>

Various forms of investigation that are possible by crossing the data source and design possibilities noted above are represented in the cells of Table 1. After taking up some issues concerning the quality of the data sources, we discuss selected combinations of design and data source, indicating relative strengths and weaknesses. Then we briefly take up broader issues dealing with program conceptualization, costs and, finally, data retrieval limitations inherent in the cross-state investigations represented in Table 1.

**Table 1: Potential Designs for Cross-State Evaluation of Welfare Reform Based on Administrative and/or Interview Data**

DATA SOURCE	DESIGN		
	One-Wave Cross Section	Multi-Wave Cross Section	Multi-Wave Panel
Merged administrative data	1	4a Post-only administrative data	7a Post-only administrative data
		4b Pre and post administrative data	7b Pre and post administrative data
Interviews	2	5	8
Merged administrative data and interviews	3	6a Post-only administrative data	9a Post-only administrative data
		6b Pre and post administrative data	9b Pre and post administrative data

---

<sup>1</sup>For the sake of consistency we use the term “interview” throughout the paper when referring to data collected directly from individuals for the purpose of a cross-state evaluation. Nevertheless, data collection from program participants might involve in-person or telephone interviews, observation, etc.

<sup>2</sup>While it is possible that data dealing with pre-reform experiences may be obtained from post-reform interviews, recall problems and program exposure effects may limit their usefulness.

## STRENGTHS AND WEAKNESSES OF DATA SOURCES

The public institutions that make use of administrative data bases (e.g., work programs, child welfare services, corrections, schools, housing services, mental health programs) generally use administrative data to monitor program participation, including reasons for beginning and ending participation. Such information is often useful to evaluators, particularly when program participation or service provision are important “inputs” or “outcomes” of interest. Furthermore, when data on program participation for individuals can be merged across data systems, it becomes possible to study the timing of transitions that are of interest to evaluators (e.g., movement of public aid recipients into jobs, movement of children from one child care arrangement to another or from families receiving aid into out-of-home care). Administrative data are also relatively inexpensive to use and often facilitate the use of large-sample statistical methods.

The disadvantages of administrative data for cross-state evaluation of welfare reform have to do with the range and reliability of data collected via management information systems and the comparability of data across states. As has already been mentioned, these data are primarily used for monitoring program participation. Although this can be very useful information, analysis of administrative data often leads to questions about *why* or *how* individuals came to be involved with a program, or moved from one program to another, information that is seldom available from administrative data. Moreover, the low utility of much administrative data to those who generate it (i.e., “line workers” in public and private agencies), and the lack of performance monitoring of data entry in many social welfare agencies, means that many data elements are of questionable reliability. Perhaps the most significant problem with using administrative data for cross-state comparison is the fact that there may be very little comparability across states. There is scant reason for states to collect data in the same manner for similarly defined populations unless there is a federal mandate to do so and federal funding to pay for the data collection. Considerable detailed study of data systems across states will be required prior to any use of cross-state administrative data for evaluation purposes.

At any rate, administrative data will be useful when comparable data are available across states on program participation and transitions of interest. Their utility will be enhanced by combining administrative data with interview data. Unless stated otherwise, we assume in our discussion below that evaluators have access in each state of interest to administrative data that can be merged across relevant service systems.

The importance of agency staff (job developers, social workers, child care providers) and program participant interviews lies in part in their being a source of much detail on family experiences, transitions, and interpersonal relations during the course of participation in the programs of interest. Many of these phenomena,



we believe, will be relevant to and enable better interpretation of changes recorded, but not elaborated upon, in agency records. The occurrence of and reasons for job placements and losses, school withdrawals, child placement, and marital dissolutions illustrate such changes. Perhaps most important, interview data will be the only source of information about family members' experiences following their withdrawal from participation in welfare reform programs.

The primary disadvantages of surveys/interviews are practical and financial. Before designing data collection instruments and working out data collection procedures, important design decisions will need to be made concerning whom to interview, and when. The sample needed in each state to make meaningful inferences about program impacts across states will be very large. The logistical problems involved in contacting and interviewing family members and other informants will be formidable. Needless to say, costs will be high for any cross-state evaluation that relies significantly on data collected directly from panel members.

#### SELECTED CROSS-STATE EVALUATION DESIGNS

The most detailed class of studies among those represented in Table 1, that located in cell 9, involves the use of panel studies, employs individual families as the unit of analysis, and generates data through both interviews and administrative records. The most common investigation within this class is that in which a sample of families who apply for participation in the welfare reform program in each of several states is followed over a common period of time. Policy-relevant family changes and experiences can be recorded and compared across samples from states that implement distinct welfare reform programs. Relationships can be assessed, conditioned on family attributes and selected state attributes, between changes in family circumstances and *specific* program characteristics. The approach also enables analysis as to whether families with different demographic attributes receive different forms of service and experience different family outcomes. Finally, to the extent our assumption is correct that in some states pre-welfare-reform implementation data are available on AFDC recipients, comparisons can be made across states of family experiences during AFDC participation with those subsequent to implementation of whatever programs succeed AFDC.

A second example within this class of investigations expands on that described above by following more than one sample entry cohort in each state. This approach provides opportunity to assess the relative consequences of different welfare reform efforts as they mature over time. Planning for study of multiple entry cohorts over time is appealing, given the uncertainty surrounding the implementation complications and

potential consequences of many program changes being contemplated by states. A possible shortcoming with this design is that its use lessens the relevance of data from pre-welfare-reform agency data bases, since families in later cohorts are less likely to have been served by pre-reform programs. On the other hand, because program maturity implies program change, the failure to employ a multiple-entry cohort design raises the question as to when, following implementation, the impacts of welfare reform programs should be assessed, a question we raise but cannot answer.

The above comments refer only in passing to what we believe to be the major advantage of the class of investigations in Cell 9 relative to most others represented in Table 1. That is, other than the class of investigations represented in Cell 8, none is able to provide data on *any* experiences of sample families when they are no longer served by agencies for which administrative data are available. This obviously includes events and transitions whose occurrence might be critical in assessing the consequences of welfare reform efforts (moves, family composition changes, receipt of services from agencies not included among those for which data base linkages are possible).

The class of panel studies represented in Cell 8 of Table 1 is similar to that represented in Cell 9 except that the data are limited to interview responses from family members and/or other informants. Its clear weakness relative to the class of studies in Cell 9 is that information on agency transactions may be subject to the unknown biases of informants. Its strength relative to the investigations represented in the remaining cells of Table 1 is that it enables the collection of data on changes in family composition and experiences that go beyond what might be collected through panel designs using agency data bases (Cell 7) or which are simply not available due to the cross-sectional nature of the study design (Cells 1 through 6).

The class of studies included in Cell 6 of Table 1 comprises investigations based on multi-wave cross-section designs, utilizing data from interviews as well as agency administrative data bases.<sup>3</sup> Because the studies provide detailed information on the present status of current participants of (or applicants to) welfare reform programs, they enable comparison of the circumstances of families in different programs. They also permit the study of how these circumstances change across cross-sectional samples over time, and whether these changes vary across programs that can be distinguished from each other by important attributes. However, the data from these investigations do not permit analysis of within-family change over time as a function of welfare

---

<sup>3</sup>It may seem odd to discuss “cross-sectional” administrative data, but in many cases administrative data bases only retain snapshots of program participation (e.g., whether a person received a particular service in a given month) and are not suited to the construction of actual event histories.

reform program experiences. Nor does this class of studies provide information about the circumstances of families who formerly participated in, or applied to, aid programs.

The class of studies represented in Cell 4 of Table 1 provides a limited amount of information about program impacts of pre- and post-welfare-reform efforts. The available information is that contained in linked agency data bases over time, as well as information on the substance of agency programs and statewide demographics. This will make possible, using some area-wide controls, comparisons regarding the *concomitant* agency status of individuals who participate in different versions of welfare reform and AFDC programs. Over time this information will provide data time on the volume and attributes of participants in different welfare reform efforts as well as data on their involvement in programs of other social service and control agencies. These data could be interpreted, depending on the sources of information, as indicative of the relative success of different reform efforts across states and times. No information would be available on people not served by these agencies.

Finally, we would like to comment briefly on a form of investigation which, although not included in Table 1, has occurred sufficiently often in the past that it deserves mention. Here we refer to the comparison across states, perhaps over time, of *unlinked* agency administrative data. The data consist simply of rates of people served by various social agencies across states and time. Possible comparisons include the relative rate of service use in jurisdictions whose welfare reform or AFDC programs vary along certain relevant dimensions. Presumably, depending on type of agency, these numbers may be used to argue for welfare reform program failures and/or successes. A simple case in point is the current finding among agencies in Milwaukee and Madison, Wisconsin, that the number of individuals and families requesting help from homeless shelters has dramatically increased since the early spring of 1996. On the one hand, the administrators of these agencies argue that the cause for the increase, at a time when requests have traditionally decreased, is the gradual implementation within Wisconsin of elements of what will become the W-2 welfare reform effort. However, lacking any systematic data that those seeking help were (1) individuals who might have been or were participants in AFDC, and (2) were rejected from or not qualified for the W-2 program, the argument is merely that. On the other hand, supporters of W-2 may argue that the increase in Madison and Milwaukee of homelessness may reflect the fact that because of its low unemployment rate, Wisconsin has become attractive to former residents of other states and that the spurt in homelessness reflects what might be expected when there is a sudden and recent increase in the immigration to an area of relatively poor people who are seeking but have not yet obtained employment. Lacking any systematic data suggesting that the upsurge in requests for shelter care in Madison and Milwaukee is attributable to recent state immigrants, the supporters' argument is merely

that. It is likely that except under very unusual circumstances investigations based on this design will fail to provide the information necessary for even minimally convincing evaluations.

## OVERARCHING ISSUES

Before closing our discussion we note four general issues that must be addressed in considering implementation of the evaluation approaches discussed here. These concern program conceptualization, scope of data collection, costs, and the generalization of findings.

### Program Conceptualization

Welfare reform programs are likely to differ on a variety of characteristics, including conditions for and generosity of benefits, time limits within which participants will be dropped, and availability of services. Because costs are likely to limit the number of state programs that can be compared, the number of program characteristics that can be independently studied must also be limited. The task of identifying characteristics warranting study is complicated by a general absence of work seeking to identify and conceptualize the contents of most welfare reform efforts and, within these efforts, the degree to which program elements follow comparable orderings. To the extent that substantively different elements of each state's program can be included within more broadly conceptualized dimensions (e.g., "generosity," "service orientation," "time constraints"), the problem of incorporating program elements for study becomes more easily handled. Although we have noted above that there has been a general absence of work on this problem, there are at least two exceptions, the IRP project given the acronym SLIP (State and Local Initiatives Program), and the project of the Urban Institute titled *Assessing the New Federalism*. These efforts are still in an early stage, a situation not entirely independent of the fact that the contents of most welfare reform efforts have yet to be well identified.

### Scope of Data Collection

The scope of evaluating the projected consequences of welfare reform efforts is exceptionally broad, so broad as to suggest that no one study can hope to assess them in their entirety. This possibility must be addressed and, if accepted, some priorities need to be made among the impacts to be studied. This issue far transcends the scope of this brief paper.

### Costs

The costs of the various approaches identified in Table 1 vary significantly. In general, designs that rely solely on administrative data are much less expensive than designs that require direct collection of data from informants. In some cases, those who are responsible for administering management information systems at the state level will be able with relatively little effort to generate reports on program participation of cross-sectional samples. In states with this capacity, the administrative data necessary for designs represented by Cells 1, 3, 4, and 6 of Table 1 may be able to be generated by state personnel at relatively low cost. In contrast, state data managers generally have less experience in merging data on program participation across service systems and little if any experience in analyzing event histories across service systems. Thus, assuming that the appropriate data are available, designs represented by Cells 7 and 9 of Table 1 will certainly require the active and ongoing participation of experts in complex data management, matching, and analysis.

The above and other complications notwithstanding, we have received some estimates from staff of the University of Wisconsin Survey Research Center regarding the direct costs of undertaking panel surveys of program participants. Costs of interviews with program operatives are not estimated, in part because we have not yet articulated the character of the information these officials might usefully provide. In any event we believe the costs associated with operatives' surveys will be much less than those associated with applicant/participant surveys. We emphasize that the estimates are not the product of detailed analysis but rather represent "reasonable" ballpark estimates. First, we assume initial interviews will be 1.5 hours in duration and that subsequent interviews will be .75 hours in duration. Second, we assume that the initial sample of applicant/participants will be 1,000 in each state and that the sample attrition will be 10 percent between consecutive waves. Third, we assume that the panel in each state will receive a total of 6 interviews over a period of 2½ years. The costs of collecting data from the panel in each state, given these conditions, will be slightly more than \$1 million. Costs of overhead, data preparation, and analyses must be added to this sum. If it is also desired that brief phone interviews be undertaken to obtain details of major changes in family circumstances as these occur, and, assuming that the incidence of such changes will average 3 per family, an additional \$125,000 should be added to the total of each state panel costs.

### Generalization of Findings

The projects we have represented in Table 1 assume, within each state, samples from the population of families which have had some form of systematically recorded contact with welfare reform programs and possibly their AFDC forerunners. The samples, depending on the interests of evaluators, may be composed of

program participants, or for more thorough analyses, program applicants. An important population not tapped by these samples is that of families who, despite being potentially eligible, fail to apply for or participate in welfare reform or AFDC programs. Rebecca Blank and Patricia Ruggles<sup>4</sup> have recently estimated that single mothers fail to enroll in AFDC during 30 to 38 percent of the months for which they are program eligible. The authors' analyses indicate that this failure occurs among sample members who are less needy (e.g., higher proportion employed, smaller proportion receiving other forms of assistance), are more advantaged (higher education, less likelihood of disabilities), and who anticipate benefits insufficient in size or duration to warrant their program participation. The possibility that the relevance of these factors is affected by attributes of programs other than their anticipated benefits,<sup>5</sup> could not be assessed with the data available to these analysts. The required analyses must be based on state-based probability samples of nonparticipating program eligibles. A possible strategy, now being considered, for obtaining these samples is to substantially supplement the size of low-income family samples to be obtained by the Urban Institute in its study of welfare reform. We do not know whether the sample supplementation will be carried out among a sufficient number of states or if the yield of nonparticipant welfare reform program eligibles in each state is will be of sufficient size to enable the required analyses.

### Comment

As is probably clear from the preceding comments, it is our view that the cross-state evaluations represented in Cells 8 and 9 of Table 1 offer distinct advantages over the alternatives represented in the table, particularly over evaluations based solely on administrative data. The data gaps inhering in the latter are, we believe, sufficiently great that valid program evaluations based solely on them seem impossible. On the other hand, since reliable estimates regarding consequences of program parameters may require comparisons across at least 8–10 states, it might appear that the costs of the evaluations represented in Cells 8 and 9 are prohibitive. Yet the fact is that in the not-too-distant past some multimillion-dollar evaluations were undertaken of programs whose scope and potential impact appear to be much more limited than what will be the case under welfare reform efforts. One example is the National Supported Work Demonstration, an important and high-

---

<sup>4</sup>“When Do Women Use Aid to Families with Dependent Children and Food Stamps?” *Journal of Human Resources* 31(1, 1996): 57–89.

<sup>5</sup>Anticipated benefits were not assessed directly. They were inferred by mothers' age, education, and disability status. It is of interest that the maximum AFDC benefit available, perhaps a measure of program generosity, was not a significant predictor of program “take up”.

cost evaluation of a jobs program targeted at adult offenders, drug addicts, school dropouts, and AFDC recipients. The projected benefits and possible problems of Supported Work as a demonstration pale to insignificance as compared to those projected by protagonists and antagonists of the various versions of welfare reform to be introduced throughout the nation. The magnitude of these outcomes, positive as well as negative, suggest that costs to be incurred in an 8–10 state implementation of the most elaborate evaluation efforts outlined in Table 1 are relatively small.

# Toward a Basic Impact Evaluation of the Wisconsin Works Program

Thomas Kaplan and Daniel R. Meyer

## I. INTRODUCTION

The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 overturns 60 years of federal welfare policy, eliminating the federal structure of AFDC and replacing it with a block grant which states can use to design and operate programs for the poor. Wisconsin has received federal approval for its planned use of block grant funds, called Wisconsin Works (W-2). In this paper, we present a proposed impact evaluation of the Wisconsin program. Because our proposal does not cover all domains and groups potentially affected by W-2, and discusses only a few indicators within some domains, the proposal should be read as a “basic evaluation plan,” containing our view of the minimal elements of an impact evaluation of W-2. Additional impact evaluations focusing on other potential consequences of W-2 should supplement this basic evaluation.

## II. THE W-2 INTERVENTION: A BRIEF DESCRIPTION

The full W-2 program is scheduled to start on September 1, 1997, but transitional parts of the program have already begun. Because the phased implementation of W-2 complicates some pre-post comparisons, it is useful to understand the previously implemented parts of W-2. Two primary transitional programs have been in place since March 1996:

- Self-Sufficiency First. This is a diversion program which requires applicants for AFDC to complete an interview with a financial planning resource specialist before applying for AFDC and then to participate in the state JOBS program during the 30-day AFDC application processing period. In these 30 days, the applicant must devote at least 60 hours to JOBS participation, including 30 hours of direct employer contact. If needed, child care is provided during the applicant’s hours of JOBS participation.
- Pay for Performance. This is an intensive JOBS program requiring 20–40 hours per week of participation from AFDC case heads and imposing heavy sanctions on those who fail to comply. For each missed hour of required JOBS participation, a penalty equal to the federal hourly minimum wage is imposed, first on AFDC and then on Food Stamp benefits. Recipients who participate in JOBS for less than 25 percent of scheduled hours receive the full penalty, which reduces the AFDC grant for the next month to \$0 and the Food Stamp benefit to the federal minimum of \$10. Subsequent participation in JOBS restores benefits for future months.



These transitional programs, which will be part of the broader changes created under W-2, have already had an impact. In September 1996, a total of 5,182 AFDC recipients in Milwaukee County, out of a caseload of 31,000, received initial sanctions. For the full state in the same month, 8,000 AFDC recipients out of a caseload of 55,000 received initial sanctions (Schuldt, 1996).

In addition to the transitional programs, the full W-2 program will eliminate AFDC and replace it with cash assistance available only through work or participation in work-like activities. Under the full W-2 program, participation requirements will begin when the youngest child is 12 weeks, compared to the threshold of one year under current regulations. Also, the limited ability to meet current JOBS requirements through participation in formal educational programs will not be available under W-2. Although some W-2 recipients will technically receive a grant, all cash income in the program will be disbursed on an hourly basis in return for each hour of work or program participation. Failure to participate will reduce income. Unlike the current AFDC program, the income will not depend on family size, but only on the case head's hours of participation and level on the W-2 "self-sufficiency ladder" (see Table 1, on p. 6 of this report). Also unlike AFDC, W-2 is not statutorily identified as an entitlement; a participant with remaining eligibility could conceivably receive no activity assignment and no income if public funds fall short (although state government officials have stated that would never happen).

W-2 participants will also receive assistance with child care costs beyond a mandatory participant copayment, the level of which depends on family income, number of children in care, and the licensing level of the provider. Wisconsin has also requested but not (at this writing) received a federal waiver to replace the current Medicaid program for W-2 recipients with a W-2 medical program, which would provide benefits generally similar to those available under Medicaid to a somewhat different population (some expansions and some contractions in eligibility) contingent on the payment of a premium. All W-2 services will be available to both single- and two-parent families. W-2 will also eliminate the current practice under which child support income beyond the first \$50 in a month goes to public agencies to reimburse welfare expenditures; W-2 participants will be able to keep all child support paid on their behalf. Table 1 (on p. 6 in this report) summarizes key features of W-2.

The W-2 program marks a significant change from the Wisconsin AFDC program that existed through 1995. W-2 ends the entitlement for a cash benefit for most recipients; places time limitations on most benefits (as required under federal block grant legislation); proposes to decouple Medicaid and child care from cash payments, making them independent programs with income-tested copayments or premiums; alters the

relationship between child support and cash assistance; and imposes a wide array of changes in management purposes and procedures.

Such a major change is worth careful evaluation, but a useful evaluation will not be easy to perform. Measuring the impact of one discrete change when related policies remain constant is difficult enough in social programs. Measuring the impact of many major changes that occur continuously over a long period—with no easy ability to randomly assign participants to treatment and control groups or determine in other ways how the program would have evolved without the policy changes—is still more difficult.

### III. OVERVIEW OF THE EVALUATION PLAN

#### A. The Scope of the Evaluation

Because W-2 is an attempt to completely replace the previous welfare system, it could influence many dimensions of the everyday life of AFDC recipients, as pointed out in the accompanying paper by Maria Cancian and Barbara Wolfe. Indeed, hopes of influencing several dimensions of the daily lives of AFDC or potential AFDC recipients—ranging from their work life to the way they relate to their children—are a distinguishing feature of emerging versions of comprehensive welfare reform. To reflect the aspirations of program designers and the fears of program critics, evaluations must cover multiple dimensions, but they must also be small enough to be manageable. We view our task here as limited: we propose a “basic” impact evaluation, an evaluation built around selected critical consequences that we believe *must* be evaluated. We have selected six “primary” impact domains (income, dependency, child care, child welfare, health status, and living arrangements and family structure) based on judgments that they (a) are central to the purposes of the reform and to the well-being of the families affected, and (b) offer a high likelihood of finding a measurable effect with data that are now available or could be available in the future. We also identify a second tier of impact domains that we consider important but not as central.

Just as W-2 could affect many dimensions of everyday life, the program could affect many groups, including W-2 participants, those who would be eligible but do not participate, those not eligible but likely to be affected (primarily adults in the low-wage labor market, including those with no minor children), employers, child care providers, schools, social service agencies, governmental units, and taxpayers. Again, to keep this proposed evaluation manageable, we focus our attention on individuals rather than employers, institutions, or entire markets. We do not propose a benefit-cost analysis, which would require a major and complex effort.

Determining direct costs in a benefit-cost analysis is fairly straightforward, although the analysis would have to be encompassing enough to incorporate expenditures on Food Stamps and the Earned Income Tax Credit as well as child care, cash assistance, and medical assistance. Indirect costs and benefits are quite problematic, particularly a comparison of the indirect costs of W-2 and AFDC. Examples of costs and benefits that would be difficult to estimate include increased (or decreased) costs borne by the educational system if children are less (or more) prepared for school, increased human costs and costs in the social service and criminal justice systems if individuals are afraid to leave situations of domestic violence because they perceive fewer financial resources available to make it on their own, the potential for increased institutionalization of children with disabilities, long-term increases or decreases in juvenile crime, increases or decreases in the general readiness of the labor market, etc.

We propose a *limited* evaluation approach, one that we believe has the potential to answer the most critical questions, even though more complex or comprehensive approaches are possible. We therefore recommend a single counterfactual and a similar approach to evaluating each potential impact. We envision that other evaluations would also be done, perhaps on other impacts, perhaps with other approaches, and we periodically comment on areas that we tentatively view as most promising. We believe that our contribution is not to identify an idealized evaluation, but to think about the difficult decisions one must make if one is to conduct a limited evaluation, with a limited budget. In summary, we view our task as identifying key impacts—but not all impacts—covering key concerns—but not all concerns, with a limited approach.

## B. The Counterfactual

We propose that the primary comparison be between outcomes under W-2 and outcomes in Wisconsin before the implementation of W-2 (a pre-post design). Part of our reasoning stems from the weaknesses or unsuitability in this setting of other options. (For a more comprehensive review of these other options, see the papers by Haveman, Piliavin/Courtney, and Cain.) One method of evaluating W-2 would be to randomly assign some applicants to W-2 and others to AFDC. But experimental designs work best when (a) only a few program elements are undergoing change (because it is then administratively feasible to operate an experimental and control program), (b) the agencies that would implement random assignment are cooperative (Hotz, 1992), (c) the control group can easily be isolated from “contamination” introduced by the experiment (Cook and Campbell, 1979), and (d) the intervention is not expected to generate community feedback effects (Garfinkel, Manski and Michalopoulos, 1992). W-2 meets none of these conditions: it is an attempt to create an entirely new, radically different program; it is unlikely that agencies implementing W-2 (especially, but not only, the

state Department of Workforce Development) would be willing to operate a parallel AFDC system in some locations; a control group cannot be isolated; and the designers of W-2 hope that it causes a fundamental restructuring of norms in low-income communities. Further, federal welfare reform complicates the possibility of leaving a part of the AFDC program still in operation.

A second possibility would be to compare results under W-2 to results from jurisdictions outside the state. This may be particularly useful to policymakers because they are unlikely to consider the pre-W-2 system in Wisconsin as a policy-relevant alternative, whereas another state's approach to assisting low-income families might be more feasible. Nonetheless, a comparison of W-2 to AFDC is of significant scientific interest, and it is also politically relevant in the sense that proponents of welfare reform make the claim that their proposals will be better than AFDC. Comparisons of W-2 to the system in another state would be most helpful if the case can be made that the other jurisdictions would have evolved identically to Wisconsin had it not been for their varying public assistance strategies: differences in the well-being of low-income households or the status of the low-wage labor force could then reasonably be attributed to the different public assistance strategies. But many other policy and economic variables could influence observed differences among states, and it would be hard to show that the divergence in public assistance programs was the critical variable. Some forces affecting low-income families affect those in different states differently, like a recession in a particular industry, a natural disaster, or state policy toward related public programs (elimination of General Assistance, removal of public schools from the property tax, etc.), so a cross-state comparison does not ensure that the only relevant difference between two states is whether low-skill parents face W-2 or an alternate public assistance regime. A final difficulty with this approach is that states have very different administrative record-keeping systems, and, as we outline below, we propose relying heavily on administrative data in the basic evaluation.<sup>1</sup>

A pre-post time-series design is not without serious problems of its own. The main problem is the possibility that something else changed at about the same time that AFDC was changed to W-2 (Cook and Campbell, 1979). Clearly, some major factors *did* change: for example, the minimum wage increased from \$4.25 to \$4.75 per hour in 1996 and will increase again to \$5.15 per hour in 1997, making an evaluation of W-2's effect on wages in the low-skill labor market problematic. Similarly, benefits under the Earned Income Tax Credit (EITC) were altered significantly in 1990 and in 1993 through phased increases, so an increase in employment among low-skilled workers may be the effect of EITC changes rather than the replacement of AFDC with W-2. Finally, the ending of AFDC, changes in the Food Stamp program, and changes in the

---

<sup>1</sup>In particular, different states have different levels of income at which taxes must be filed, and the extent to which low-income families file tax returns probably differs substantially across states.

Supplemental Security Income (SSI) program—all components of the federal welfare reform passed in 1996—will take effect contemporaneously with W-2, further complicating the simple comparison of W-2 and AFDC. Below we identify some strategies for dealing with these changes, recognizing that none are perfect. We selected a pre-post design not because it is flawless, but because it is feasible, cost-effective, and seems no more flawed than other designs.

Our reliance on a pre-post design does not negate possibilities for meta-analysis of pre-post studies in several states, and, indeed, this is in some ways a superior strategy if feasible. With its limited evaluation funds, it may be possible for the federal government to encourage enough common variables and definitions so that meta-analysis of pre-post evaluations could investigate the independent effect of selected variables. By comparing the post-reform with the pre-reform in each state, and evaluating the “difference in differences,” the evaluator could state with more confidence that observed results were caused by W-2 (or perhaps even by particular features of W-2). While we are intrigued with this approach, we do not propose it here because we are concerned about its feasibility and cost-effectiveness, and because our one-state, pre-post model is consistent with a basic, limited evaluation.

Our reliance on a pre-post design also does not negate possibilities for randomized evaluations of particular program components. For example, the state could still decide that families in a few counties would be randomly assigned to different child care copayment schedules. This would enable policy makers to examine the effects of alternate program elements, while not compromising the main thrust of the reform. In this limited evaluation, however, we are more concerned with determining the overall impacts of W-2 than with determining the program mechanisms responsible for these impacts. Identifying with reasonable certainty the key impacts of W-2 will be a large enough challenge; determining the independent impact of particular features within W-2 seems, in general, too daunting.

### C. Basic Data Source

We propose to depend primarily on administrative data to examine effects. Our preference for pre-post administrative data over surveys stems in part from the expense of obtaining data through surveys over time. Capturing the effects of an intervention on income growth, for example, would require panel studies measuring the same people over a period of years, and panel studies are the most expensive kinds of surveys to operate. In contrast, it is possible to follow the same family’s income tax records much more efficiently over time.

Panel surveys are especially difficult to conceptualize as a way of evaluating the impact of W-2 in Wisconsin because the first wave of the survey could at best be mounted in early 1997, well after much of W-2

has already gone into effect. Because so much of W-2 has already been implemented, the best way to assess its full impact is somehow to go back in time to capture the situation that existed before W-2 began. Administrative data in both electronic and paper form do extend back into the past and offer some reliability, and our proposal rests heavily on such data. Further, it may take some time before W-2 is operating smoothly, and thus W-2 may have different effects in its first year of operation than the second or third year. If these effects are of interest, this makes a panel survey even more costly because either a very large initial sample would have to be selected (in order to ensure that there would be a sufficient number of new W-2 recipients in later years) or a new sample would have to be drawn each year. Drawing new samples each year and following them through administrative data, however, is less expensive. Finally, the implementation analysis may reveal that W-2 is implemented very differently in some counties than others. Using administrative records as a sampling frame would enable researchers to oversample in some counties several years *after* the program differences had been detected. In contrast, a desire to oversample from particular counties not even identified until a substantial period of time after the program has been in operation would be quite difficult in an evaluation relying on survey data.

Even though we rely on administrative data for the basic evaluation we propose, we believe that surveys of low-income families could provide valuable additional information. Because administrative data typically have very limited demographic information, an intriguing possibility is to conduct a simple survey gathering demographic information on individuals in the administrative data bases. Analyses using the Urban Institute's New Federalism Household Survey could also supplement the analyses proposed here; we do not rely on it as a primary data source because we believe annual data on the same families over a period of more than five years is desirable. A comprehensive panel survey of the low-income population, while much more expensive to gather than the administrative data we propose using, could provide a wealth of important data. Furthermore, while we propose primarily quantitative analyses in the basic evaluation, analyses using ethnographic data would also add an important untapped dimension to our knowledge about the effects of W-2.

#### D. Time Period

All evaluations face a difficult decision regarding the length of time a program will be evaluated, as discussed by Cancian and Wolfe.<sup>2</sup> If the time period is too long, the evaluation will be expensive, the results may be too late to influence program development, and the counterfactual becomes more difficult (a pre-post

---

<sup>2</sup>The time frame is one of four trade-offs identified by Cancian and Wolfe. The others are targeted vs. broad-based samples, community vs. individual data, and monitoring vs. evaluating.

evaluation has to face increased likelihood of other significant events occurring during the evaluation period that could create their own effects; similarly, more time increases the likelihood that significant reform will occur in the “control” site in a cross-site design). If the period is too short, however, other difficulties arise: a program evaluated too early for too short a period may not tell us anything about the way the program would work when it is more mature and some of the “debugging” has occurred (e.g., Campbell, 1984); further, programs like W-2 seek to affect community norms, and this process takes some time. Complicating this decision further is the fact that W-2 has time limits built into some of its components: no one can receive subsidized job services for more than five years unless granted an exception, and thus any evaluation that is concerned about the effect of this five-year limit must span a long enough period that some will exhaust their eligibility and outcomes can then be examined. Similarly, participants can only take part in each of the three subsidized components (Transitions, Community Service Jobs [CSJs], and trial jobs) for 24 months. Finally, W-2’s proponents recognize that in the short term, income may decline, but the hope is for longer-term increases in income. Again the evaluator needs to allow enough time to elapse that the individual who received valuable work experience in a trial job could be in the unsubsidized workforce long enough to show wage gains resulting from that experience.

Therefore, we propose a two-part evaluation, a three-year examination (focusing on short-term effects) and a seven-year examination (focusing on longer-term effects; in seven years some families will have exhausted benefits and the time period should be long enough to capture longer-term wage increases). Further, we propose that no impact evaluations be done for the first year, only process evaluations, allowing at least some time for the program to change, develop, and reach a steady state.

#### IV. IMPACT DOMAINS TO EVALUATE IN A BASIC EVALUATION

We propose that an evaluation that sought to answer only the most important questions would focus on six impact domains: income, dependency, child care, child welfare, health status, and living arrangements and family structure. We select these six because we believe all are central to the purpose of W-2 and to the well-being of affected families. We also believe that data sources for income, dependency, health status, and family structure are sufficient to offer the prospect of finding a measurable effect. The data sources for child care and child welfare are less certain, but could be improved with minimal future investment.

Primary Potential Impact 1: Income and Poverty

Hypothesis 1.1: W-2 will have an impact on incomes among low-income families with children. Because some factors are likely to result in increased incomes and others in decreased incomes, we do not specify the predicted direction of effect. In this domain, we propose that it is important to identify whether incomes of low-income parents as a whole increase or decrease, as well as identifying the types of families experiencing increases and the types experiencing decreases.

A. Rationales

Income increases could occur through five primary mechanisms:

1. Some low-income families do not currently have the job skills to compete in the labor force and would receive few unsubsidized job offers if they were to search. Because W-2 provides CSJs and/or trial jobs for these individuals, they may eventually obtain the job skills (and/or the work experience) that will enable them to earn more than they could have earned in the past and also be able to earn more than they would have received on AFDC.
2. Closely related to the first group is a group of recipients who have some job skills and could receive some job offers, but the wage offers they would receive would be too low to make working worthwhile. This could occur either because work expenses like child care and transportation are so high or because the job offers they would get would come without health insurance. For these individuals, W-2's provisions of child care and (perhaps) health insurance, combined with the requirements to be working, may shift the incentives and move them into employment; as a result they will eventually earn more than they would have had the old AFDC system been in existence.
3. Adults in some low-income families do not currently work outside the home because AFDC allowed them to receive benefits with minimal effort. By eliminating the possibility of receiving cash assistance without effort, by limiting the amount of time a family can receive any assistance, and by requiring employment, W-2 may push these individuals into the labor force, which will eventually lead to higher incomes than they would have received under AFDC.
4. Many husband-wife low-income families are not currently eligible for assistance because of well-known limitations in the AFDC-U program and severe funding restrictions for child care assistance. Because W-2 opens up child care assistance and (perhaps) health assistance to husband-wife families, disposable income will automatically increase for some families. Further, the provision of subsidies for child care may enable some families that were single-earner families to become dual-earner families, increasing their incomes. Finally, if AFDC encouraged family breakup by primarily helping only single-parent families, under W-2 these families may stay together, and family incomes in husband-wife families are generally higher than in single-parent families.
5. Under AFDC, a recipient family was able to keep only \$50/month of child support, with the remainder going to offset AFDC costs. W-2 allows families receiving assistance to keep the entire amount of child support, which directly increases incomes. Further, some ethnographic evidence suggests some nonresident parents only paid \$50/month of child support because additional amounts did not benefit the family (Edin, 1995). Thus W-2's elimination of this disincentive to pay may increase the total amount of child support paid, which would further increase the resident parent family's income.



On the other hand, W-2 could reduce income. W-2 will not provide any benefits to three groups of families who used to receive AFDC. First, because W-2 has a time limit for receiving benefits, some families will exhaust this time limit and stop receiving benefits. Second, because W-2 is not an entitlement, it is possible that during periods of recession the budget will be exhausted, and then some families who would have been participants will not receive benefits. Third, the sanctions for noncompliance are stricter under W-2 than they were under AFDC before 1996, and thus more families will become ineligible for benefits for not complying with various provisions of W-2. In each of these cases in which a family no longer receives benefits, whether their total income increases or decreases depends on their response to this new regime.

In addition to the families who receive nothing from W-2, for some families W-2 will provide some assistance, but at a level below what they received under AFDC. One key group is families with two or more children. Under AFDC, families with one adult and two children received \$517/month; under W-2-Transitions they receive \$519/month but may have to pay around \$50/month for their child care co-payment and (perhaps) \$20/month for health coverage, making their net income about \$449/month. If the same family is working at a CSJ, its net income under W-2 would total about \$485 (\$555 for the CSJ minus the \$50 and \$20), and could be even less if there were other job-related expenses. For larger families, the disparity between W-2 and AFDC is greater. Again, eventual total income depends on a family's response to the lowered benefits.

#### B. Data, Comparison Group, and Analysis Approach

This is a domain in which we have fairly reasonable data on the AFDC period, with the ability to examine earnings of AFDC recipients before, during, and after receipt (through the quarterly earnings records of the Department of Industry, Labor, and Human Resources, or DILHR) and family income (through an extract from the tax records of the Department of Revenue, or DOR).<sup>3</sup> A companion paper for this conference by Martin David discusses a related approach in more detail.

Our proposed approach is built around three main comparisons:

- comparing income changes of W-2 and AFDC recipients
- comparing pre- and post-W-2 income changes of a broader sample of low-income families
- comparing pre- and post-W-2 income changes among higher-income families.

---

<sup>3</sup>While federal income tax data have historically been very difficult for researchers to access because of confidentiality concerns, researchers in Wisconsin have been given access to these data after appropriate safeguards have been negotiated. For analyses using Wisconsin tax data, see Phillips and Garfinkel (1993) or Meyer (1995).

For the W-2-AFDC comparison, we propose selecting two samples of W-2 recipients in 1998 (or another year after full implementation of W-2), a sample of those who entered the program during the year and a sample of those who were participating at the beginning of the year. Using the Social Security numbers of the adults, we propose merging tax, earnings, W-2, and food stamp records to calculate the total income of these families in each year from 1999 through 2005. For the comparison AFDC group, we would select two parallel samples of recipients drawn from the administrative AFDC records for 1988.<sup>4</sup> Again we calculate their family incomes in each year from 1989 through 1995 using Social Security numbers, tax records, earnings records, AFDC records, and food stamp records. (Note that the W-2 group would look different from the AFDC group on several dimensions because of policy differences between the programs; for example, the W-2 sample would contain more husband-wife families.) We propose drawing separate samples of entrants and current recipients because many analyses show that these two groups are quite different (e.g., Cancian and Meyer, 1995; Friedlander and Burtless, 1995).

The primary longer-term evaluation is a comparison of the later family incomes (2005 and 1995) for these samples of entrants and recipients. The analysis includes estimating a multivariate equation in which later family income is the dependent variable and the key independent variable is whether this is a family who was an AFDC recipient (the early period) or a W-2 recipient (the later period).<sup>5</sup> A variety of variables will be incorporated as control variables, including features of the local labor market, individuals' educational levels, family size and structure, etc. We will also explore interacting the variable indicating the period with a variety of sociodemographic characteristics to determine if W-2 had different effects for different groups of people.

The main short-term evaluation is a comparison of the 2001 incomes of the 1998 W-2 recipients with the 1991 incomes of the 1988 AFDC recipients, using a similar analysis. The three-year time frame also enables the researcher to examine whether a difference between W-2 and AFDC is a result of the program

---

<sup>4</sup>The year of initiation for the seven-year AFDC sample is difficult to determine. 1996 as an endpoint for the sample would be too late to capture the longer-term effects of AFDC, because many components of W-2 would already have been implemented. 1993 seems too early an endpoint because that would require the selection of cases from 1985, a year before earnings records were kept by DILHR. However, 1994 and 1995 are difficult years for data collection because the state was in the process of changing administrative records systems, and data may not be accurate in some counties. If this is discovered to be the case, we propose focusing only on counties in which we are sure of the data.

<sup>5</sup>A simple test is whether later incomes during W-2 are higher (or lower) than later incomes under AFDC, ignoring incomes during the year in which the sample was drawn and in all intervening years. This will suffice for most analyses because it compares the incomes of people three (or seven) years after entry into AFDC (or receiving AFDC) with the incomes of people three (or seven) years after entry into W-2 (or receiving W-2). Other tests would incorporate income during other years, enabling the researcher to estimate whether the *trend* in individual incomes had been affected by W-2.

change or merely the result of long-term time trend.<sup>6</sup> The analyst could calculate income three years later for several samples of AFDC recipients, perhaps drawn from each year between 1988 and 1992. If time trends are detected, it may be possible to then remove this trend from the comparison of W-2 and AFDC.

Another analysis could test whether W-2's effect on incomes differs for those experiencing W-2 when it is first established from when it is a more mature program. By using administrative data, it is relatively simple to draw other W-2 samples, perhaps from 1999 and 2000, to follow them three years after their W-2 experience as well, and to test whether W-2's effect differs as the program develops.

Another analysis could be a test of whether some types of W-2 programs lead to different effects, as suggested by Holden and Reynolds in their paper. Agencies administering W-2 have considerable latitude in several program features, and it is therefore possible that the "intervention" will look very different in different counties. If the implementation evaluation identifies key program differences across sites, we propose using this information to test whether different types of W-2 programs lead to different effects on incomes.<sup>7</sup>

All of the analyses above are tests of whether the later incomes of W-2 families are different from the later incomes of AFDC families. But W-2 may also have entry effects, affecting whether low-income families become recipients, and thus our second major comparison is low-income families with children during the W-2 and AFDC periods. To make this comparison, we propose drawing two additional samples of low-income families with children, defined as families with income less than twice the poverty line. We propose drawing these samples based on Wisconsin tax records from 1998 and 1988, and then examine income in 2001 and 1991 (short-term effects) and 2005 and 1995 (longer-term effects). Perhaps the tax data would have to be supplemented with survey data to gain a full sample of low-income nonrecipients. Once the sample was selected, an analysis parallel to the analysis of recipients would be conducted.

One method of increasing our confidence that any observed change in incomes was due to W-2, rather than to other factors, is to include a variety of control variables. Another method is to compare incomes over time of higher-income families with children, defined as having taxable income above 300 percent of poverty (taken from the DOR records of the same years), with analyses parallel to those described above. Because W-2 is likely to have little effect on the incomes of upper-income families, this will provide one estimate of general

---

<sup>6</sup>Hoynes and MaCurdy (1994) have recently documented that, by some measures, AFDC spells have been getting shorter since the late 1960s and early 1970s. If this time trend is a result of changes in the low-wage labor market, and if these trends continue, incomes in the W-2 period may differ from incomes in the AFDC period without being caused by W-2.

<sup>7</sup>Of course it will be difficult in this analysis to determine whether different effects in different counties are due to different program features or different labor market opportunities.

changes in income over these periods, and this could strengthen our confidence in whether observed changes between AFDC and W-2 were due to W-2 or to other changes in the economy. For example, if the family incomes of the low-income sample in 2001 are higher than in 1991, but the family incomes of the higher-income sample did not change, this could increase our confidence that the effect was the result of W-2 (or other changes affecting only the low-income population) rather than economywide changes.<sup>8</sup>

### C. Issues and Limitations

One difficult issue for evaluating the effect of W-2 on income is the definition of “income.” The traditional definition employed by the Census Bureau has included only pretax cash income. Many analysts prefer a concept of disposable income, which includes near-cash in-kind benefits and also deducts taxes. While assigning the cash value of food stamps to a household’s income is straightforward and generally noncontroversial, there is less consensus over whether and how to value public health insurance, employer-provided health insurance, housing assistance, or the value of home ownership.<sup>9</sup> In the basic evaluation we propose here, which relies only on administrative data, we will not have measures of employer-provided health insurance, housing assistance, or a good measure of home ownership, limiting the measure of income we can construct. For those filing tax returns, we propose to calculate income by summing taxable income, AFDC/W-2 benefits, food stamps, and the federal Earned Income Tax Credit, and subtracting federal and state income taxes and payroll taxes. For those not filing tax returns but with in-state AFDC/W-2 income or in-state earnings, we will estimate income by summing earnings, AFDC/W-2 benefits and food stamps, and subtracting estimated payroll taxes. We will also explore subtracting estimated out-of-pocket child care expenditures; unfortunately, these will be more available under W-2 than they were under AFDC.

Another issue facing this analysis and the ones that follow is defining who is a W-2 recipient. We propose that individuals receiving trial jobs, CSJs, or W-2-transitions are clearly recipients. Although families that receive child care subsidies only could be considered recipients, we propose to not count them as W-2 recipients in these analyses because we generally lack information on families that only received child care assistance from the AFDC period. Similarly, if there is accessible information on individuals receiving

---

<sup>8</sup>Note that this analysis only enables the researcher to control for factors that affect individuals of all income levels; the researcher could still not distinguish between effects of W-2 and effects of other changes that affected only low-income individuals.

<sup>9</sup>Still another difficulty is the possibility that the private not-for-profit sector will provide increased benefits to some families, replacing public benefits. If public benefits are valued, probably private ones should be too if we want to know whether total income for low-income families increased or decreased. However, these figures are generally not available administratively, and the accuracy of self-reports is unknown.

Medicaid-only from the AFDC period, we could include both them and the corresponding group from the W-2 period (those receiving health insurance subsidies only, if a W-2 health plan is implemented). If the data are available, analyses could be conducted both with and without this group. Another difficult group includes those who come for help, are referred to a private sector unsubsidized job, begin working, and receive no other services. While W-2 considers these individuals clients, and appropriately will count them as successes, we believe that they cannot be included in the sample of “W-2 clients” because a comparable person was not included as an AFDC client (and some individuals did receive this type of help from AFDC, although presumably a much smaller number). The analysis of income changes using the entire low-income sample may enable us to estimate effects on this “referral-only” population.

A final difficult issue concerns what variables to hold constant in this analysis. This is a well-known problem of quasi-experimental designs; we discuss two examples, family structure and the local unemployment rates, but the issue relates to several other variables as well. In evaluating W-2’s effect on income, one perspective is that the analyst should examine effects on single-parent families and husband-wife families separately. The argument for doing so is that because husband-wife families typically have higher incomes than single-parent families, and because a higher proportion of W-2 recipients are husband-wife families, a finding that W-2 participants have higher incomes than AFDC participants may merely reflect the composition of the caseload rather than an effect of W-2. On the other hand, W-2 could affect family structure decisions, and under this perspective, an evaluation of W-2’s effect on income should not control for family structure. Similarly, the analyst would want to control for the local unemployment rate, because a change in the unemployment rate could confound an analysis of the effect of W-2. On the other hand, if W-2 affects the local unemployment rate, the analyst should not hold this constant. Both perspectives have limitations; our general proposal would be to conduct both types of analyses.

The proposed evaluation strategy has several limitations related to gaining information on income. First, relying on Wisconsin’s administrative records of income misses some individuals; for example, no information is available on those who move out of state. Second, tax records also miss information on those who do not file taxes; individuals with taxable incomes below about \$6,000 are not *required* to file forms. This is less of a problem in Wisconsin than it would be in other states because Wisconsin offers a refundable credit for low-income renters and homeowners, increasing the likelihood that a low-income family will file a tax form.<sup>10</sup> Further, incorporating information from the earnings records and welfare records increases the

---

<sup>10</sup>In 1995, a family with one child with earnings of less than \$8250 and rent of \$300/month that does not include heat would receive \$724 by filling out the one-page form. If this family owned a home and paid \$1500 in

coverage, because individuals who do not file tax forms frequently have income in these other records. Analyses conducted with data from the Wisconsin child support project suggest that income estimates should be available for about 90 percent of a low-income sample and perhaps over 95 percent. Nonetheless, the 5-10 percent for whom we lack income information are likely to be the most vulnerable group, and they deserve more attention. One approach in the comparison of W-2 and AFDC recipients would be to identify those who were initially W-2/AFDC recipients who do not appear in later administrative records, attempt to locate them, and survey them.

The measures of family income are also limited. Taxable income does not include various income sources, including transfers and child support.<sup>11</sup> For those not filing tax forms, earnings records from DILHR/DWD are limited in that they only report earnings in *covered* employment, which does not include earnings in the informal sector. Expenses associated with working are not fully available: the amount of federal and state taxes is straightforward, but child care costs may not be, and work-related transportation expenses will definitely not be.

Tax records provide only limited information on family composition, and family composition is needed to determine family income. Tax filing status can tell us whether the taxpayer is married or not on December 31, but cannot tell at what point the marriage or divorce occurred. Second, there is no way to identify other adults in the home (cohabitators, roommates, relatives); to the extent that these individuals contribute to the family's income, we will underestimate income.<sup>12</sup>

A related concern is that administrative records contain very limited information on these families. For families who have never received AFDC or W-2, the tax record includes only the age of the adults, filing status, the number and ages of children, and address. For families who do not file taxes and have never received AFDC or W-2, and appear in the sample only as part of the DILHR/DWD earnings file, we would not even know this information. Thus for many families we will not know basic data like race and educational level. This information will be available for any families who ever received AFDC or W-2, but even for these families we will not know whether they speak English nor anything about their family of origin, and we will have only

---

property taxes, they would receive \$1160. Full-year AFDC recipients are not eligible for this credit.

<sup>11</sup>Administrative records of child support income for most families will exist for the W-2 period. Unfortunately the child support administrative data system that existed during the AFDC period had several problems, limiting the comparability of these data across the two periods.

<sup>12</sup>An additional difficulty is that there is occasionally not an exact correspondence between where children live and where they are claimed as dependents for tax purposes.

sketchy information on work experience. We consider this limitation potentially quite problematic, and believe that the development of a short survey gathering basic data on individuals who are in the administrative records should be explored.

The degree to which we can be certain that W-2 caused differences in income under W-2 and AFDC is limited by other changes that occurred between the AFDC and W-2 regimes, as discussed earlier. One key change is the increase in the minimum wage. There are several ways to try to account for this, none of them entirely satisfactory. One way would be to use the results of other studies on the effect of the increase in the minimum wage on family income. One could then subtract this effect from any observed difference in income, assuming that the remainder was due to W-2. A second approach would be to simulate what income in the AFDC situation would have been like had the new minimum wage been in effect,<sup>13</sup> and to compare income under W-2 and the new minimum wage with simulated income under AFDC and the higher minimum wage. A third approach is to redefine the “treatment” to include not just W-2 but W-2 and a new minimum wage. In this type of issue in which alternative approaches seem justifiable, the basic approach we favor is that the researcher conduct multiple analyses, report the results of different analyses, and make the data available for others who would prefer another approach.<sup>14</sup>

Evaluations of any policy change of this magnitude in a complex and dynamic environment will be imperfect. Although this approach contains several limitations, alternatives also present serious problems, and we believe this approach has the potential to provide reasonably clear information on the impacts of W-2 at a reasonable cost.

Hypothesis 1.2: W-2 will decrease incomes among families *without* children.

A. Rationale

Three provisions of W-2 may have a direct impact on the earnings (and therefore income) of adults without children. First, W-2 will force many low-skilled parents into the labor market, so the general supply of low-skilled people wanting jobs will increase. This increased supply will decrease wages for low-skilled jobs.

---

<sup>13</sup>The approach is conceptually simple but complicated in practice. Presumably there are three effects of an increase in the minimum wage: low-wage individuals gain higher wages, some low-wage individuals lose their jobs, and some non-working individuals begin working now that work is more worthwhile. By combining empirical estimates of the magnitude of these effects with several plausible scenarios, one can estimate what income might have been had the new minimum wage been in effect.

<sup>14</sup>Another potential limitation is that the analysis attempts to compare two policy regimes that are ten years apart, and any attempt to put incomes in constant dollars is limited by variations in prices over regions, over urban versus rural areas, and over time.

The increased supply will also cause employers to be less likely to retain marginal current employees. Second, W-2's provision of employer subsidies for hiring from a targeted class may make the hiring and retention of nontargeted employees less likely, decreasing their incomes. Finally, W-2 creates many community service jobs in nonprofit organizations and local governments. The possibility of having access to a large number of potential employees at little or no cost may cause these organizations to forgo hiring individuals that they would have hired and may make them less likely to retain marginal employees, both of which lead to lower incomes. (Because public employee unions will be paying close attention to employment trends in local government, any potential impact in this direction may be more pronounced in nonprofit organizations than in local government.)

B. Data, Comparison Group, and Analysis Approach

The approach to evaluating this question follows the approach to evaluate the first hypothesis: draw a sample of low-income families without children in 1998 and 1988, calculate their incomes in 2001/1991 and 2005/1995; examine whether the level of income in 2005/2001 is lower than in 1995/1991; and examine whether individual changes in income from 1998 to 2005/2001 are lower than they were from 1988 to 1995/1991.

C. Issues and Limitations

Difficulties affecting the definition of income and limitations regarding available data (discussed above) are also relevant here. An additional difficulty is that the DILHR/DWD data include only information on quarterly earnings, not information on hours worked or weeks worked. Thus an evaluator can identify those with low earnings and determine whether the proportion with low earnings increased, but cannot differentiate those who changed the number of hours they worked to keep total earnings roughly comparable. For example, assume that W-2 causes hourly wages to fall. Some individuals who were working 250 hours quarterly at \$10/hour (total quarterly earnings of \$2500) under AFDC may decide under W-2 to keep total earnings constant even though their hourly wage has declined to \$5/hour, working 500 hours/quarter. This important effect of W-2 could not be detected under this basic evaluation scheme.

Hypothesis 1.3: W-2 will change the proportion of families with children who are poor, and will change the poverty gap (the aggregate amount needed to bring everyone up to the poverty line). Conventional measures of poverty in the United States are merely measures in which family income is compared to a threshold that varies by family size. Because of this, the rationales, data, comparison group, basic approach, issues, and limitations provided under Hypothesis 1.1 are relevant here. The additional issues are in determining an appropriate poverty threshold and identifying what income is to be compared to this threshold. The official



governmental approach has well-known limitations (see Citro and Michael, 1995, for a recent critique). We propose to use a variety of measures of poverty, including the official measure and a measure as close to that proposed by the Panel on Poverty and Family Assistance of the National Research Council (Citro and Michael, 1995) as data will allow.

Although the limitations noted above are also relevant for this hypothesis, an assessment of poverty status makes the need for accurate information on family composition more critical, since the number of people in the family affects the poverty threshold. Further, the lack of information on child care and other work expenses is particularly troubling for measures of poverty status that rely heavily on the assumption that the critical concept is disposable income rather than gross cash income. Still, there are no better alternatives in that there is no information on work expenses during the AFDC period.

### Primary Potential Impact 2: Dependency

Hypothesis 2.1: W-2 will lead to a decreased reliance on means-tested transfers among low-income families with children. Although there are several possible ways to conceptualize “dependency” and “reliance” (Meyer, 1990; Smiley, 1996), we focus here on the proportion of income that comes from means-tested transfers (programs specifically for the poor) (Meyer, 1990; Gottschalk and Moffitt, 1994; Meyer and Cancian, 1996).

#### A. Rationale

First, the rationale under Hypothesis 1.1 above suggests several mechanisms through which earnings would eventually increase; because all of these are increases to the denominator (income) without increasing the numerator (means-tested transfers), the share of income from means-tested transfers is predicted to decline automatically. Second, some recipients will immediately move into unsubsidized employment, which obviously decreases their reliance on transfers. Third, some recipients will move immediately into trial jobs; among these people, at least the portion of wages that is not part of the subsidy to the employer is nontransfer income, and they too should thus show decreased reliance on transfers. Fourth, as we noted above, W-2 could lead to increased child support, which, as an addition to income that is not a means-tested transfer, would also decrease the transfer ratio. Finally, recipients who lose transfer income due to time limits, lack of entitlement, or sanctioning will also see their transfer ratio decrease.

### B. Data, Comparison Group, and Analysis Approach

The basic structure of the evaluation of this question parallels that of the evaluation of incomes. We begin with the two basic recipient samples: W-2 recipients in 1998 and AFDC recipients in 1988. For the short-term evaluation, we suggest calculating the “dependency ratio” three years later (2001 and 1991), using a combination of tax records, earnings records, and benefit records. For the longer-term evaluation, we propose calculating dependency ratios for 2005 and 1995. We would estimate a multivariate equation in which the dependency ratio is the dependent variable, the key independent variable is whether this is a person who was an AFDC recipient (the early period) or a W-2 recipient (the later period), and we would incorporate other control variables.

Thus the main evaluation in this domain has to do with whether *recipients* are able to become less dependent. But W-2 may also affect whether low-income families become recipients; in fact, by opening eligibility to two-parent families, more people may receive benefits. To evaluate this possibility, we will use the sample of low-income families with children described above, and conduct an analysis in parallel fashion to the analysis of recipients.

### C. Issues and Limitations

All of the issues about the difficulty of knowing what to include in income mentioned above are also applicable here. But looking at the percentage of income derived from means-tested transfers also highlights some new difficulties in deciding what is a means-tested transfer. Some benefits are clear: food stamps, the AFDC benefit, SSI and W-2-Transitions are all means-tested transfers. Presumably the Earned Income Tax Credit and public subsidies for child care and health care (both Medicaid and the public portion of the premium under W-2) are also transfers. But within the W-2 program, how should we view trial jobs? From a taxpayer perspective, the subsidy going to the employer is a benefit available only to the poor, and any earnings above the subsidy counts as earnings (non-benefit). But from the perspective of the recipient, they are *working* for the entire amount, and none of it should count as an “unearned” benefit. Community service jobs create a similar problem, although it is clear that the state treats these as grants, rather than earnings. Finally, what if one effect of W-2 is that private not-for-profit organizations provide more vouchers or cash to some individuals who have exhausted their time limit for benefits or who have been sanctioned under W-2? These “gifts” do not count as transfers from the taxpayers’ perspective, but they represent a transfer of dependency from public assistance to private charitable sources. (Note, however, that public administrative records do not keep track

of these private-sector transfers, so in practice this would not be evaluated in a basic evaluation even if it were thought to be conceptually important.)

### Primary Potential Impact 3: Child Care

The quality of child care arrangements for low-income families could be affected by W-2. Karen Folk identified in a paper for this conference several potential effects of W-2 in the domain of child care; we focus here on what we view as the most important, child care quality. We select this outcome because it has the potential to have long-term effects on child development and because questions of the availability of child care for W-2 participants will be addressed in the companion paper on implementation evaluation. Unfortunately, reaching agreement concerning an appropriate measure of child care quality is difficult. We focus on two indirect measures: formal parental complaints lodged against child care providers and substantiations of abuse or neglect by child care providers.

Hypothesis 3.1: W-2 will cause (a) an increase in formal complaints against child care providers made to the office that licenses the providers, and (b) an increase in substantiated cases of child abuse and neglect by child care providers.

#### A. Rationale

W-2 requires copayments and sets the copayment amount as a percentage of income with adjustments for the level (certified, licensed, etc.) of child care; these provisions may lead families to select lower-cost child care options. Further, W-2 establishes a new category of child care providers who have limited training requirements and for whom there is limited regulation, and these providers may provide lower quality care. Finally, by requiring work of all recipients, demand for child care is likely to outstrip supply (at least in the short run), increasing the general cost and causing some to accept child care arrangements of lower quality than they would have liked.

#### B. Data, Comparison Group, and Analysis Approach

Complaints can be made to the office that licenses child care providers, and it is not clear whether this office has kept, and will keep, track of complaints, particularly complaints about those providers with provisional licenses. If these data are kept, analyses of the effects of W-2 on substantiated parental complaints might be possible.

The most serious allegations, those for abuse or neglect, however, *are* part of the child welfare record-keeping system. Standard data processing forms used throughout the state indicate whether perpetrators are

child care providers and the type of child care setting—center, family child care, or care provided in the child’s own home. (It is not always clear, however, whether county staff investigating abuse and neglect cases enter these data reliably.)

We propose to examine records in counties that have adequate records to determine the annual number of substantiated abuse/neglect reports in which the perpetrator was a child care provider. Presumably the entire child care market is affected by W-2, so we propose looking at incidents of abuse/neglect by child care providers among the entire potential population of child care users, not merely among welfare recipients. Counties are the proposed unit of analysis. The analysis strategy is fairly straightforward: identify the counties in which there are adequate records, count substantiated incidents over a several-year period, and, through a multivariate time-series analysis, determine if the number of substantiated incidents after the implementation of W-2 is greater than the number prior to W-2.

Both complaints and substantiations of abuse and neglect will have to be expressed as rates. Otherwise, any increase in raw numbers could be the effect of more children in child care and not an indicator of child care quality. Unfortunately, so far as we know, the number of children in Wisconsin in child care is unknown; perhaps the number could be broadly estimated from data on child care use in the Survey of Income and Program Participation (SIPP), although that survey is not designed to be representative of a state.

### C. Issues and Limitations

Even if the rate of substantiated parental complaints could be adequately estimated, this is a quite limited measure of quality. The number of complaints can be affected by the level of information about who to complain to, the perceived threshold of dissatisfaction under which one is justified in making a complaint, and the likelihood of corrective action following complaint, all of which could change substantially over a several-year period.

The number of substantiated abuse/neglect incidents is also less than perfect. A child care provider could negatively affect a child’s development and emotional well-being without creating an incident that parents felt was clear enough to warrant calling it abuse or neglect. In addition, parents with other child care options may remove their child without lodging an abuse/neglect complaint. Parents with few other child care options may not want to risk losing the only arrangement they have been able to find by taking such serious action. Nonetheless, substantiated incidents of abuse and neglect are so serious that we believe they are important to track and to try to evaluate.

It is unfortunate that better measures of child care quality are unavailable. The generally accepted measures of child care quality are child/staff ratios, the size of child groups (regardless of ratios), the child care

training and formal education of providers, and the frequency of turnover of child care providers (Gormley, 1995). We know little about these measures in Wisconsin, but it would be appropriate to start to collect this data for licensed, certified, and provisionally certified providers. Even if formal pre-post analysis and causal inference are impossible, a careful start at monitoring these measures would be desirable.

#### Primary Potential Impact 4: Child Welfare

In this section, we refer not to child welfare in the sense of overall child well-being, but to the official concerns of the formal child welfare system in Wisconsin—child abuse and neglect and the placement of children in substitute care (primarily foster care) when parental care is unavailable or considered inadequate. Another paper written for this conference by Mark Courtney considers these issues in greater detail.<sup>15</sup>

The relationship of reports and substantiations of child abuse and neglect to actual abuse and neglect is not well known. Reports can increase because of greater public attention to the issue or because of legal changes concerning who is required to report suspected cases, even if the underlying rate of abuse or neglect remains unchanged. Nevertheless, some relationship presumably exists between the underlying phenomenon and reports and substantiations of it. Because actual abuse and neglect is unambiguously bad, it is useful to try to assess whether major social policies such as welfare reform influence the direction of the imperfect indicators of the problem available to us—reports and substantiations of abuse and neglect. The W-2 program could influence child abuse and neglect reports and substantiations in several ways:

- If parents do not participate in the required W-2 programs and are sanctioned, they lose a source of income that would have been available to them under W-2. The loss of income could lead parents to neglect the basic health and physical needs of their children, which could in turn increase the reports and substantiations of neglect. Although Wisconsin statutes officially prohibit findings of neglect solely for reasons of poverty, a finding of neglect because a parent did not take full advantage of all opportunities available in W-2 would be possible.
- Parents who would under AFDC have worked in the labor force for few or no hours each week but who participate for 40 hours per week under W-2 may find the pressure of full-time work combined with single parenting to be overwhelming, and they may behave inappropriately to their children as a result.
- Reports and substantiations could rise simply because new officials—the Financial and Employment Planner or the Supportive Services Planner—are playing close attention to the family.

---

<sup>15</sup>We consider the health status of children below. Researchers are only beginning to conceptualize impacts of welfare reform on measures of child well-being that go beyond the traditional measures of health status, child abuse and neglect, substitute care, and educational achievement (Collins and Aber, 1996).

- Child abuse and neglect could fall because parents who formally did not conform to broad social values of independence do so under W-2. The parents may feel better about themselves generally and thus more competently fulfill their parenting roles.

The placement of children in foster care could increase or decrease for largely the same reasons. Parents who do not conform to the requirements of W-2 may, if they have no income to support their child, leave child welfare authorities with no choice but to place their children in alternative arrangements and, eventually, even to terminate parental rights. On the other hand, it is possible that some parents forced by W-2 to take control of their work lives would improve their general functioning enough so that W-2 could reduce the need for out-of-home care.

For reasons identified in the Courtney paper, a panel study following low-income cases starting before and continuing well after the imposition of W-2 would be desirable. However, as that paper points out, the panel size would have to be quite large to detect statistically significant differences over time, because the percentage of any population in the formal child welfare system owing to abuse, neglect, or out-of-home placement is small. In addition, a survey approach is complicated by the fact that some of W-2 has already been imposed, and retrospective questions eliciting accurate information about abuse, neglect, or out-of-home placement would be hard to construct. We thus turn again to administrative data for the basic evaluation design, although an additional targeted evaluation of child welfare issues, probably using survey data, also has merit.

#### A. Data, Comparison Group, and Analysis Approach

This approach would be demanding, but it would be possible to use the basic pre-post samples described above to obtain some information about the direction of change in child welfare indicators. The tax information provides the county of residence and a name or Social Security number, and it would be necessary to ask selected counties if, in the particular year at issue, any child in the family was reported or substantiated as abused or neglected or placed in out-of-home care. Once these matches have been accomplished, the basic comparison could be similar to the income/dependency comparisons described above: is there a change in the low-income samples before and after W-2, and is that change different than in the middle class sample?

#### B. Issues and Limitations

One likely difficulty with this analysis is that only small numbers of families are affected by child welfare programs. As a result, any change may not be statistically significant even if it would have enormous practical significance. Also, we expect that data from the AFDC period would be available in only a limited number of counties, which may or may not be representative of the state. The major difficulty is likely to be

separating the impact of W-2 on child welfare from the impact of other changes in the child welfare environment. There has been a strong trend over time toward increased use of out-of-home care and greater reporting and substantiation of child abuse and neglect. Any increase in these phenomena after W-2 is implemented may be greater than the increase under AFDC solely because of these other factors. In this case, comparing increases among low-income samples potentially affected by W-2 with increases among middle-income samples may not be very informative, because the child welfare system predominantly affects low-income people.

#### Primary Potential Impact 5: Health Status

The original plan for W-2 included the replacement of the Medicaid program for W-2 participants with a new W-2 health plan. That change would require a federal Medicaid waiver, which has not yet been granted. Evaluation issues surrounding such a change are discussed below in the “Health Insurance Coverage” section. Even without the change to a W-2 health care program, however, other provisions of W-2 could have health care implications, in at least the following ways:

- a. Low income is at least a predictor, and possibly a cause, of poor health among children (Geltman et al., 1996), and family incomes could rise or fall under W-2.
- b. The future administrative relationship between W-2 and Medicaid remains to be fully developed. It is not certain that eligibility for W-2 will confer automatic eligibility for Medicaid, or that a joint application process will exist. Health status could be affected if the process of achieving eligibility for Medicaid becomes more time-consuming and complex.
- c. Parents who would have worked in the labor force for fewer hours under AFDC than they are required to work under W-2 may reduce utilization of routine primary health care, because the parent might not be able to fit routine health care into a busier schedule. On the other hand, it is conceivable that parents who feel better about their less dependent social role could feel empowered with respect to use of the health care system. At any rate, it would be desirable to know if W-2 changes the utilization of routine health care by low-income children.
- d. W-2 could influence the level of prenatal care received by women expecting their first child. Under the Wisconsin AFDC program, women who are pregnant and have no other children and who fall below income and asset thresholds gain eligibility for Medicaid upon confirmation of pregnancy. This will still be the case under W-2, whether Medicaid or the W-2 health plan is operating. Under the AFDC program, however, women with no other children become eligible for AFDC in the seventh month of pregnancy, whereas under W-2 women in this status will not be eligible for W-2 work programs until after their child is born. Since some women delay obtaining their Medicaid card until they become eligible for AFDC benefits in the seventh month, the lack of cash benefits throughout pregnancy may also reduce the use of prenatal care.
- e. W-2 could also affect the prenatal care of women who are enrolled in W-2 and have other children. Women who have a child and subsequently become pregnant must remain in a W-2 work program;

the only time they can obtain benefits without participating in a W-2 job or activity is during the first 12 weeks after birth. Although a medically difficult pregnancy might qualify the mother for the W-2 Transitions program and activity compatible with her medical condition, it may still be the case that the requirement to work through pregnancies could impair the health of some pregnant women and their newborn children, at least in comparison to what would have been the case under AFDC. It is also possible, however, that the greater self-reliance imposed by W-2 could positively influence the care a pregnant woman takes during her pregnancy.

A. Data, Comparison Group, and Analysis Approach

The companion paper by Maria Cancian and Barbara Wolfe offers many useful suggestions for assessing utilization of routine care. Perhaps the most direct indicator, if the basic Medicaid program is not converted to a W-2 health plan, would be to measure the percentage of children below age 6 on Medicaid (children below age 6 are eligible for Medicaid up to 165 percent of the poverty line) who received Healthcheck screens (the Wisconsin equivalent of the Early and Periodic Screening, Diagnostic, and Treatment program) and standard immunizations. Both indicators are available from Medicaid administrative records, and both should be sensitive gauges of the use of routine health care.

We propose limiting the evaluation here to Medicaid recipients before and after implementation of W-2, because we assume that Medicaid eligibility and “take-up” among low-income children will not change as a result of W-2. This assumption of no change in eligibility owing to W-2 does not apply to pregnant women, as we say below. But among young children, assuming a continuation of the Medicaid program and no conversion to the proposed W-2 health care plan, we believe our assumption of little change in Medicaid eligibility or take-up is warranted, at least for a basic evaluation plan. Medicaid was not tightly linked to AFDC in the early 1990s because of expanded Medicaid eligibility features, and Medicaid will similarly not be linked to W-2 when that program takes effect. Case heads who do not wish to subject themselves to W-2 might conceivably avoid obtaining eligibility for Medicaid because the same agency may confer eligibility for both programs. But any visit to any health care provider is likely to lead to Medicaid enrollment, because providers want to be reimbursed.

Assessing the impact of W-2 on the health status of newborns requires a broader approach, discussed in the companion paper by Gary Sandefur. The most widely accepted indicator of the health of a newborn is birthweight, measured by discrete categories: normal birthweight (5.5 pounds and above); low birthweight (3.3–5.4 pounds); and very low birthweight (under 3.3 pounds). Because both birthweight and the mother’s Social Security number appear on birth certificates, it should be possible to match maternal Social Security numbers with income tax records to compare the proportions of births to low-income women (with incomes



below 200 percent of the poverty line) in each birthweight category before and after the creation of W-2. If desired, this could be augmented with a measure—either the Kessner or Kotelchuck index (Kotelchuck, 1994)—of the adequacy of prenatal care. (Such an index could be constructed from data available on birth certificates for each birth.)

#### B. Issues and Limitations

One significant limitation of a pre-post comparison of health care utilization among Medicaid children is that the state's administrative emphasis on routine health care has been increasing, and health care providers are under more pressure to offer it. W-2 is thus not the only relevant change, and it will be hard to disentangle the effects of W-2 from the new administrative emphases. However, much of that increase in emphasis occurred in the early 1990s and may not affect a pre-post study if the "pre" period is the mid-1990s. A lesser concern stems from the way these data are reported to the state. Children on Medicaid are increasingly served by HMOs, which generate no direct billing record when a Healthcheck screen or immunization is provided. The source of data on the provision of these services thus comes from surveys filled out by HMO administrative staff. Because HMOs are judged in part on whether they provide these services, it would probably be desirable to check on a random basis the survey reports against patient charts.

A pre-post comparison of birth weights is reasonably straightforward, although it may be necessary to perform the comparisons separately for the major racial/ethnic groups identified on the birth certificate, since a change in the racial/ethnic composition of mothers in the state could influence birth weights independently of W-2.

#### Primary Potential Impact 6: Living Arrangements and Family Structure of Low-Income Households

As noted in a companion paper by Gary Sandefur, W-2 could conceivably influence marriage and divorce and both nonmarital and marital childbearing patterns in Wisconsin. If AFDC encouraged nonmarital births by providing financial support to parents without any requiring any work or work-like activities, W-2-required work activities may decrease nonmarital births. Similarly, if AFDC encouraged divorce or separation or discouraged marriage because eligibility was easier for those who were single parents, W-2's offer of services contingent only on the income and assets of the family, without regard to whether the family contains one or two adults, could decrease divorces and increase the rate of marriage. Moreover, by eliminating the practice of providing benefits based on the number of people in the assistance unit, W-2 could reduce the

fertility of welfare recipients. Finally, if W-2 is perceived as offering attractive assistance into the labor force, the program could encourage women to have one child, whether in or out of wedlock, to obtain W-2 eligibility.

Determining the actual influence of W-2 on these and related possibilities will be difficult. W-2 administrative records will need to contain data on the number of people in the household, because child care copayments and (if implemented) premiums for the W-2 health care program will depend on household income relative to the poverty line. But because these calculations do not require information on the actual composition of the assistance unit (number of adults and their relationship to each other, number of children, and the relationship of the children to the adults), it is not certain that administrative records will contain such data.

In contrast, the likelihood of good administrative data on births is much higher than the likelihood of good administrative data on family composition. About 70,000 babies are born each year in Wisconsin, and the birth order of the child and the marital status, zip code and Social Security number of the mother are recorded on the birth certificate or a worksheet attached to it. The availability of the Social Security number allows for matches to income tax and wage records. Thus one analysis could be a comparison of the percentage of births to low-income women that were nonmarital in the AFDC and W-2 periods. A second analysis could use the zip code data to identify low-income areas and compare the percentage of nonmarital births in these areas pre- and post-W-2. A third type of analysis could examine the later fertility of recipients of AFDC and W-2. By merging AFDC, W-2, and birth data, the analyst could compare whether recipients had another birth within a given time period.

Because administrative records on marital status are not likely to be available, a brief survey could be done to collect these data. Respondents could be asked about their marriage and divorce histories (to obtain “pre” information for the AFDC period) and their recent and current marital status. Questions for collecting both current and retrospective information could be adopted from national surveys. Alternatively, the Social Security numbers of AFDC and W-2 case heads could be matched against Social Security numbers in divorce records to gain a sense of the relative impact of AFDC and W-2 on divorce.

Because W-2 could affect childbearing and marital decisions by women not on the program, the full impact of W-2 on living arrangements and family structure is discernible only through a consideration of the broader low-income population, not just of those on AFDC or W-2. Whatever type of data is used, the sample would need to consist of families with minor children and with incomes below, say, 200 percent of the poverty line. Attributing any observed changes to W-2, rather than to broader cultural and environmental factors that influence marital and fertility decisions, will be challenging.

## V. SECONDARY POTENTIAL IMPACTS

We view several potential consequences of W-2 as of slightly less evaluative significance, generally because they are less central to the purposes of the reform, likely to affect fewer people, or (as is the case for health insurance coverage) may change little from the status quo after federal waiver possibilities are clarified. These impacts include homelessness, residential mobility, health coverage, and the living arrangements of children with disabilities.

### Secondary Potential Impact 1: Homelessness

W-2 could influence the level of homelessness among low-income families with children, as discussed in the companion paper by Irving Piliavin. W-2 could increase homelessness because families that would have received AFDC may not receive W-2 funds, either because they do not want to subject themselves to the work requirement of W-2 or because they exhaust their time-limited benefits. With no other source of income, they may lose the ability to pay rent and thus face eviction. On the other hand, W-2 could conceivably decrease homelessness because, if W-2 succeeds as its advocates hope, income will rise among program participants. Some participants may move to higher-income neighborhoods, perhaps closer to their jobs. Rental prices may decline in their former neighborhoods because of reduced demand.

Both the Milwaukee and Madison homeless shelter programs have one intake point for families seeking shelter. When families reach that intake point, at least if a prospect of providing services exists, a common form is filled out. The form notes family status (single person, children, etc.), Social Security number, and a staff person's selection from a list of possible "events precipitating homelessness." At least in Milwaukee, these forms have remained roughly constant since 1990, so that limited comparisons before and after W-2 are possible. Evaluators could:

- a. Note the number of families cited as homeless with "loss of benefits" cited as the reason for homelessness before and after W-2. This could, however, include families losing SSI or other program benefits, not just AFDC. Evaluators would probably have to examine AFDC administrative data to determine whether someone with that Social Security number was on the program and lost benefits at about the date of the homeless report.
- b. Investigate the extent to which people become homeless not for loss of benefits but because they did not apply for benefits: some people may, in effect, choose homelessness over what they believe to be the rigors of the W-2 program. No simple approach to discovering this is available. Qualitative research in homeless shelters would probably be the best approach.

Attributing any change in the level of homelessness to W-2 will be difficult for at least two reasons: (a) federal policies have been changed to make it easier for landlords to evict Section 8 tenants, which presumably increases homelessness independent of W-2, and (b) the growth in supply of homeless shelters that has occurred over the last several years presumably increases the number of people served by them, again independent of any changes in W-2. A comparison of the number of people who are homeless (according to the form filled out by homeless program staff, augmented by checks of AFDC administrative data) because of loss of AFDC benefits should be less influenced by the Section 8 change. Still, it will be hard to control for the possibility that more people who lose benefits may present themselves to homeless shelters if they believe that capacity exists to serve them. For families (as opposed to individuals), the alternative to application to a homeless program is probably not so much actual homelessness—living in a parked car or on the street—but rather efforts to move in with relatives or friends. The perceived availability of a homeless shelter for families then could reduce doubling up and increase apparent homelessness among families independent of W-2.

Some define homeless families as including those whose financial emergency causes them to live with others when they would prefer to live separately (Jencks, 1994). We know of no way to determine if W-2 influences the level of this kind of “near-homelessness.” It would be possible, however, to follow a sample of people sanctioned from W-2 or who reached the time limit. Although we could not be certain whether any problems observed were greater than what would have been the case under AFDC, carefully documenting the extent of any problems would be important.

### Secondary Potential Impact 2: Residential Mobility

Short of homelessness, frequent moves from one household to another may be another response to economic pressure. Frequent moves among families with children (whether for a better job, or a worse job, or to move in with others) can complicate school performance and reduce the ability of parents to draw easily on informal resources in the community that might help them with their parenting tasks, such as friends who can provide occasional child care or help their children obtain summer jobs (McLanahan and Sandefur, 1994).

W-2 could increase or decrease the level of residential mobility for the same reasons it could affect the level of homelessness. Low-income families may lose income because of W-2, causing them to be evicted or to have to move to avoid eviction. W-2 participants may also move to find a job, or they may move to better quarters because W-2 helps them obtain increased income. We would not propose in this part of the evaluation trying to ascertain whether any change in geographic mobility attributable to W-2 stemmed from higher or lower income, but only whether W-2 has had an effect on geographic mobility.

A standard pre-post design could be applied to this issue, relying on tax returns, which indicate place of residence as of December 31. The basic question would address whether the number of moves made by low-income families increased after the implementation of W-2. A corresponding analysis of the number of moves made by middle-income families before and after W-2 may increase confidence that any change among the low-income sample was indeed an effect of W-2. If some reason other than W-2 would lead low-income people to change their housing at a rate different from that of middle-income people, this analysis would overestimate the impact of W-2. We can identify no such factor at this time, but we cannot discount its possible existence. A final complication is an analysis based on tax data cannot identify multiple moves within a calendar year, only changes from one year to the next, thus underestimating actual mobility.

### Secondary Potential Impact 3: Health Coverage

This is the part of W-2 containing the most policy uncertainty. The original W-2 proposal was to create a new W-2 health care program with benefits similar to those under Medicaid, to be operated (as will the existing Medicaid program) through contracts with managed care organizations—primarily HMOs. All families containing minor children—whether headed by a single or two parents—would be eligible for the program, so long as their income and assets fall below a threshold and they have no access to employer-provided health insurance. The income threshold was proposed to be 165 percent of the poverty level for entrance and 200 percent of the poverty level for continuation in the program. Since these are generally higher income thresholds than those available in Medicaid, some expansion in coverage could have resulted from this policy change.

On the other hand, the W-2 health care program would have required payment of monthly premiums (\$20 for families with incomes below 160 percent of the poverty line and rising to \$143 for families at 200 percent of the poverty line). In addition, the W-2 health care program would not have been available to families with access to employer coverage, even if the employer plan was not feasible for the worker because of high premiums and deductibles. Some W-2 participants eligible for the W-2 health care plan would presumably not have joined it because of the premiums, and some W-2 participants not eligible for the W-2 health plan because of access to employer coverage would not have elected the employer coverage. These factors would act to decrease the number of people with some form of coverage.

All these changes in coverage require a federal Medicaid waiver, and negotiations around such a waiver are ongoing. If the waiver is granted, a panel study following the health insurance status of low-income families would be possible. Large change has not yet occurred in this area, and a panel study could be initiated in time to pick up the effects of policy change. Because the W-2 changes could increase as well as decrease health

coverage, it would not be appropriate to follow only families who were on the Medicaid program before the new policy went into effect. A better sampling frame would probably be families with young elementary children enrolled in the free and reduced-price school lunch program (which has an eligibility level up to 185 percent of the poverty line) or low-income families selected from tax records.

Because of the uncertain policy direction in this area, we say no more about it here.

#### Secondary Potential Impact 4: Living Arrangements of Children with Disabilities

As with the W-2 health care plan, some policies under W-2 concerning people with disabilities remain unclear at this writing. The basic premise of W-2 is that the adult in public assistance cases should work, and the basic premise of SSI is that an adult qualifying for it cannot work. Under AFDC, children of parents on SSI receive AFDC benefits based on the number of children in the family; the adult on SSI is not counted for purposes of calculating the AFDC benefit. The original plan under W-2 was to provide each child in this situation a flat \$77 per month (a decline from the level provided under AFDC), but the governor used his item-veto authority to delete this provision, apparently out of concern for violating federal SSI regulations. The policy outcome for children of parents on SSI remains unclear.

For parents who do not qualify for SSI but who have a condition which makes an unsubsidized job in the regular labor force unlikely, the W-2 program offers the W-2 Transition category. Some of these adults function at a fairly low level by conventional standards: in cases in which low IQ is the primary test for qualifying for SSI, a person functioning acceptably as a parent must often have an IQ score below 60 to qualify for SSI. The character and quality of the W-2 Transitions placement for those adults who score in the 60s on IQ tests but do not qualify for SSI is perhaps more a concern of implementation than of impact evaluation, although some monitoring of adults who reach their five-year limit may be appropriate. We assume that many will be exempted from the five-year limit and continue in the W-2 Transitions program.

For children with disabilities, as the companion paper by George Jesien notes, the primary impact of W-2 will be through new demands on their parents. Parents who under AFDC could have stayed home and attended to their children may no longer have that option under W-2, although it is possible that many parents in this situation will be placed in the Transitions program and given the assignment of caring for their disabled child. Again, whether that happens is probably more a concern of implementation than of impact evaluation, although what happens when a parent in this situation reaches the five-year limit is of interest.

From the standpoint of an impact evaluation, the most important concern is probably a potential increase in the placement of children in foster or institutional care because the parent under W-2 cannot stay

home with the child. This could occur either because the parent reaches the five-year W-2 limit or because the parent (especially if other children reside in the family) does not wish to restrict herself to the income available under the Transitions program. A parent seeking more income may move to higher levels of the W-2 program that would not allow the parent to remain home with the disabled child. A special, and affordable, child care setting for a child with disabilities—especially after school and during school vacations—would then have to be available. Employers of these parents may also have to allow the parents a more flexible schedule than is usual for employees, so that parents can go to their child's school or child care settings to address occasional crises. If such child care is unavailable or unaffordable, or if employer flexibility is not forthcoming, the result may be a higher number of out-of-home placements.

One evaluation approach would be to select school districts in Wisconsin enrolling large numbers of children with two of the disabilities likely to present the greatest challenge for a working parent—physical and cognitive. (Those with emotional disabilities should probably be omitted because changes in federal SSI provisions especially complicate a pre-post comparison for these children. For the same reason, legal immigrants could be omitted from the sample.) School district records indicate the child's living situation—with a parent, a relative, in a foster home, etc. The basic question would be whether children with physical and cognitive disabilities were more, less, or about equally likely to live with a parent under AFDC than under W-2.

The changes in the federal SSI program that were part of the welfare reform act of 1996 will complicate the ability to attribute changes to W-2 through a pre-post analysis. Children who lose SSI benefits may for that reason alone be more likely to be placed in out-of-home care. We try to control for this change by limiting the comparison to children likely to qualify for SSI both before and after the SSI changes, but this is unlikely to be a perfect control.

## VI. OTHER POTENTIAL DOMAINS AND POPULATIONS

Because W-2 is such a broad intervention, it could potentially affect a wide variety of other domains and populations as well. To close, we simply list a few other possibilities for careful evaluation below:

- a potential for increased asset accumulation, because W-2 has different asset provisions;
- a potential for decreased employment stability, because employers may be less likely to retain marginal employees;
- a potential for increased dependency among low-income families without children, because if their wages and employment stability decline, their food stamp usage might increase;

- school readiness among entering kindergartners could be affected;
- the educational achievement of older children could be affected;
- the rate and severity of juvenile crime could be affected.



### References

- Campbell, Donald T. 1984. "Can We Be Scientific in Applied Social Science?" in Roos F. Connor, David G. Attman, and Christin Jackson, eds. Evaluation Studies Review Annual 9: 26–48.
- Cancian, Maria and Daniel R. Meyer. 1995. "A Profile of the AFDC Caseload in Wisconsin: Implications for a Work-Based Welfare Reform Strategy." Institute for Research on Poverty, Special Report no. 67, University of Wisconsin–Madison.
- Citro, Constance F. and Robert T. Michael, eds. 1995. Measuring Poverty: A New Approach. Washington, D.C.: National Academy Press.
- Collins, Ann and J. Lawrence Aber. 1996. "State Welfare Waiver Evaluations: Will They Increase Our Understanding of the Impact of Welfare Reform on Children?" Working Paper, National Center for Children in Poverty, Columbia University School of Public Health, New York.
- Cook, Thomas D. and Donald T. Campbell. 1979. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton Mifflin.
- Edin, Kathryn. 1995. "Single Mother and Child Support: The Possibilities and Limits of Child Support Policy." Children and Youth Services Review. 17: 203–30.
- Folk, Karen Fox. 1996. "Welfare Reform under Construction: Wisconsin Works (W-2)." Focus 18(1): 55–57.
- Friedlander, Daniel and Gary Burtless. 1995. Five Years After: The Long-Term Effects of Welfare-to-Work Programs. New York: Russell Sage Foundation.
- Garfinkel, Irwin, Charles F. Manski, and Charles Michalopoulos. 1992. "Micro Experiments and Macro Effects." In Charles F. Manski and Irwin Garfinkel, eds., Evaluating Welfare and Training Programs. Cambridge, Mass.: Harvard University Press.
- Geltman, Paul L., Alan F. Meyers, Joshua Greenberg, and Barry Zuckerman. 1996. "Commentary: Welfare Reform and Children's Health." George Washington University Center for Child Health. Health Policy and Child Health 3(2): Special Report, 1–6.
- Gormley, William T., Jr. 1995. Everybody's Children: Child Care as a Public Problem. Washington, D.C.: Brookings Institution.
- Gottschalk, Peter and Robert A. Moffitt. 1994. "Welfare Dependence: Concepts, Measures, and Trends." American Economic Review 84(2): 38–42.
- Hotz, V. Joseph. 1992. "Designing an Evaluation of the Job Training Partnership Act," in Charles F. Manski and Irwin Garfinkel, Eds. Evaluating Welfare and Training Programs. Cambridge, Mass.: Harvard University Press.
- Hoynes, Hilary and Thomas MaCurdy. 1994. "Has the Decline in Benefits Shortened Welfare Spells?" American Economic Review. 84(2): 43–48.

- Jencks, Christopher. 1994. The Homeless. Cambridge, Mass.: Harvard University Press.
- Kotelchuck, Milton. 1994. "An Evaluation of the Kessner Adequacy of Prenatal Care Index and a Proposed Adequacy of Prenatal Care Utilization Index." American Journal of Public Health 84(9): 1414–1420.
- McLanahan, Sara and Gary Sandefur. 1994. Growing Up with a Single Parent: What Hurts, What Helps. Cambridge, Mass.: Harvard University Press.
- Meyer, Daniel R. 1990. Child Support and Welfare Dependency in Wisconsin. Unpublished doctoral dissertation, School of Social Work, University of Wisconsin–Madison.
- Meyer, Daniel R. 1995. "Supporting Children Born Outside of Marriage: Do Child Support Awards Keep Pace with Changes in Fathers' Incomes?" Social Science Quarterly 76: 577–93.
- Meyer, Daniel R. and Maria Cancian. 1996. "Economic Well-Being Following an Exit from AFDC." Paper presented at the Association for Public Policy Analysis and Management Research Conference, October 1996, Pittsburgh.
- Phillips, Elizabeth and Irwin Garfinkel. 1993. "Income Growth among Nonresident Fathers: Evidence from Wisconsin." Demography 30: 227–41.
- Schuldt, Gretchen. 1996. "AFDC Sanctions Rise in County." Milwaukee Journal Sentinel, September 12, 1996, p. 3b.
- Smiley, Marion. 1996. "Private Needs and Public Welfare: Rethinking the Idea of Dependency in a Democratic Culture." Manuscript, Department of Political Science, University of Wisconsin–Madison.

## Wisconsin Works: A View from the Ground

Thomas Corbett and Elisabeth Boehnen

This paper examines methods for assessing the implementation of the new generation of welfare reforms. Assessing the quality of program implementation and functioning is increasingly a critical component of a comprehensive evaluation, particularly given the dramatic changes now being contemplated. Donald Kettl of the La Follette Institute of Public Affairs at the University of Wisconsin–Madison argues that the initiatives being generated by

the new block grant proposals create two kinds of impossibilities: Asking states to do things (like creating one-stop shopping and comprehensive strategic planning) that no government has done before; and asking the states to do things that can be done but which cannot be effectively accomplished in the very short time frame the programs provide.<sup>1</sup>

The next generation of welfare reforms will constitute a sharp contrast with the past and break new ground in program design and management. Understanding how they are being operated will be crucial.

### DIMENSIONS OF A COMPREHENSIVE EVALUATION

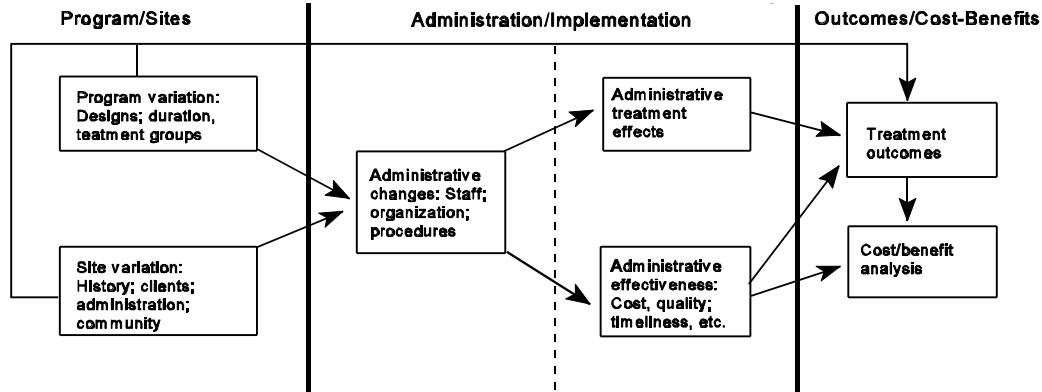
A comprehensive evaluation does not merely focus on the question “does X work,” the classic **summative** question. A serious evaluation requires a serious political dialogue accompanied by a rigorous analytic process to determine *what would constitute success* in the new program. Equally as important, one must establish acceptable methods for determining *whether or not program managers are running a program that reflects original intent* and *whether the experience at the operational level reflects the intentions of program designers*.

Summative evaluations constitute only one dimension of a comprehensive evaluation. In the short run, ultimate success is not the critical question in a dramatic innovation such as Wisconsin Works. Since it is virtually impossible to “get it right” the first time around when dealing with the multidimensional changes encompassed in many new-generation reforms, including W-2, it is reasonable to argue that evaluators should initially focus on whether program administrators are actually doing what they say they intended to do. In addition, evaluators might well focus on providing feedback necessary to make mid-course corrections—what is normally referred to as a **formative** evaluation.

---

<sup>1</sup>“The Devolution Revolution: Shifting the Politics of Pain,” a paper presented at the conference “The New Federalism and the Emerging Role of the States,” University of Illinois, Sept. 1996, p. 17.

Figure 1 captures a portion of the complexity associated with comprehensive reforms.<sup>2</sup> Both anticipated and unanticipated variations in the intervention can occur (column 1) because of differences in the way the program is implemented or because of observed or unobserved differences in the sites (local agencies and counties). The middle two columns summarize factors that can influence the way in which the intervention is done, thereby affecting the treatment outcome of interest or the cost-benefit analysis. The efficiency of management and administrative systems and variations in their procedures may or may not directly affect treatment outcomes. Variation in administration approaches and procedures will have different costs and should therefore be included in any cost-benefit analysis.



**Figure 1. Sources of variation in the evaluation of social programs.**

**Source:** John F. Witte and Thomas Corbett, "Research and Evaluation Issues Relating to W-2," *Focus* 18, no. 1 (Special Issue 1996): 75.

Comprehensive evaluations typically would encompass the following:

- Observers would first want to know whether the program has been *implemented* in a way consistent with original intent. Because critical differences between intent and operational reality may exist, this is often referred to as a **discrepancy** analysis.
- Observers would require that a **process** analysis be planned which captures key indicators of program operations after they have become somewhat stable and institutionalized (in addition to the evaluation of the implementation process described above). Moreover, the analysis would consider selected temporal factors inherent in new-generation reforms. For example, the early experience may not reflect how the program will be run over time. Two theoretical possibilities exist. First, there may be a "learning curve," where operations improve as early lessons are incorporated into ongoing processes,

<sup>2</sup>The author is grateful to John Witte for this schematic.

staff is better trained or selected, or ill-designed features are corrected. Second, there may be some deterioration of performance as the early energy often associated with a new initiative wanes.

- Observers would also want an **outcome** analysis. What changes in individuals, families, systems, and communities are important to the definition of success? Outcomes are not necessarily process measures, as defined above, nor impacts, as defined below. Outcomes are merely phenomena of interest that observers wish to track and monitor as reform takes place. Where feasible, comparisons still need to be made to make sense of these “outcomes”—perhaps with comparable jurisdictions or with historical data.
- Finally, observers would want an **impact** analysis—an analysis of *net* outcomes. The outcome analysis may provide evidence that things are getting better or worse, but only this analysis can causally link the intervention with key outcomes. It helps sort out false negatives (gross outcomes that are going in the wrong direction, but that would have gotten worse even if the new program did not exist) and false positives (gross outcomes getting better irrespective of the existence of the program). Getting at net outcomes, and being able to make *causal attributions*, requires a more rigorous evaluative design.

These levels of evaluation are not mutually exclusive; they do not represent choices that administrators or program evaluators are required to make. They are best thought of as complementing one another. The first two are formative: the intent is to determine *how well* the program is working from an operations perspective; this involves identifying those components either functioning or not functioning well, and providing the kind of information necessary to improve operations. The latter two are summative: their intent is to determine whether program objectives are being achieved or not.<sup>3</sup>

Until we know something about the former, however, it is difficult to say much about the latter. Let us assume we find outcomes or impacts that are very modest in scale, or whose effects are not measurable. Do we assume the intervention failed? Not necessarily. The first question is whether it was tried in the first place or, perhaps more accurately, whether it was tried well. Disappointing impacts can be attributed to an intervention that was tried and found wanting (a “bad” idea that was tried); to an intervention that was never tried (the “good” idea that was never implemented); or to an idea that was so changed in the way it was operationalized that the sponsors would not recognize it (a “good” or “bad” idea that replaced the original “good” or “bad” idea).

This is the dilemma of the black box. We too often believe we know what is being implemented and evaluated and so wind up assessing labels rather than realities and drawing totally inappropriate conclusions about the policies under consideration.

---

<sup>3</sup>We gloss over a contradiction between formative and summative evaluations. To the extent that formative evaluations lead to policy, program, and management changes, additional instability is introduced that makes many summative analyses more problematic.

Process Analyses—The Neglected Stepchild of Evaluations

Process or formative-type evaluations typically have been the neglected element of the evaluation industry. Traditionally, the policy world has obsessed over impacts. That narrow view of evaluation has been changing. In recent years, evaluation Requests-for-Proposals (RFPs) have demanded submissions that generally include plans for process, impact, and benefit-cost analyses. A state-of-the-art process evaluation would focus on the following types of questions and issues:

- A *descriptive* analysis, in which the following types of data are presented: *contextual* data on the environment within which the program(s) is operating, including demographic information, labor market information, measures of social disorganization (e.g., economic distress); *population* data on the demographics and circumstances of actual and potential participants; and *agency* data—general information on organizational structure and personnel.
- An analysis of the *discrepancy* between the intended sequence of activities and services in a program model and the actual sequence. Program models as articulated in legislation, planning documents, and management interviews are compared with reality as measured through direct observations, case records, and interviews with program operators and participants.
- Some form of *participation* analysis. The sophistication of these analyses will vary. Many focus on cross-section pictures (point-in-time analyses, described further below) which depict the proportion of the target population in various participation statuses in a given month or on a given day. More sophisticated analyses are moving toward longitudinal participation analyses, where participation is calibrated on the basis of the proportion of the target population that has been engaged in the program over a period of time since initial program contact—for example, the proportion of a given target group participating in X (whatever the innovation calls for) within six months after initial program contact.
- Process analyses often include some form of *dosage* or *intensity* analyses. Such analyses attempt to calibrate the extent to which typical or modal participants are engaged in and by the program. The number calculated typically is those who participate over those eligible to participate.
- Good process analyses attempt to measure *continuity* of program engagement. This includes measures of “leakage” (falling through the cracks), down time, queuing up for services, etc. Observers might want to know how quickly the system responds to noncompliance by the participant and to significant changes in program status (from a job search component to a community service component). In the most sophisticated of analyses, a description of modal or typical participant experience is undertaken to determine the sequence of activities and events experienced by participants.
- Process analyses sometimes attempt to assess program *coherence*. That is, do all key actors view the system as a whole, or key components of the program, in the same way? Do program managers, operations personnel, and participants have similar understandings about program intent. If a particular message is to be communicated at a particular point, is that what the participant hears; is that what operations people understand respecting their role?
- In complex and multisite initiatives, some process analyses include a *comparative* component presenting cross-site data in ways that highlight similarities or critical differences. This may include comparisons of many of the features noted above: population and agency attributes, program design

features, and various performance features. This is particularly helpful when variation in circumstances or approach is likely to be significant.

- Finally, some wish to assess overall program *quality*, or at least the quality of selected program components. This is very difficult to do reliably and well; techniques vary, from relying on observers' judgments, to describing and assessing inputs, or to assessing participants' satisfaction.

Most process analyses are not so comprehensive. What passes for a process analysis is often little more than a perfunctory review of planning documents and a few quasi-structured interviews with key informants. For the new generation of reforms, we need more rigorous methods.

## THE NEW GENERATION OF REFORMS

Over the past two decades, a good evaluation was defined by the use of an experimental design. This approach proved particularly powerful when the interventions of interest were relatively simple, when counterfactuals were easily developed, and when the outcomes of interest were limited in number. That simple world no longer represents reality. The salient attributes of the new generation of reforms are described in the background paper to the conference, "Reform and Evaluation: An Overview," and will only be briefly reviewed here.<sup>4</sup>

The emerging generation of reforms are characterized by the following: by their *complexity*, the simultaneous introduction of multiple changes; by their *boldness*, the attempt to influence outcomes in multiple domains among several target populations; by their *volatility*, variation and change in policy direction or delivery over time, and by their *diversity*, the extent to which intrastate variation exists. The new reforms require a whole new way of looking at the *core technology* of welfare agencies—what they actually do.<sup>5</sup> For example, since new programs are designed to change behavior rather than merely offset immediate income deficits, they need to be conceptualized quite differently. They require a *point-in-process* rather than a *point-in-time* perspective.

What sets the new generation of reforms apart from prior reform eras is their focus on changing a broad range of personal, interpersonal, and community behaviors. Their emphasis on behavioral change, well

---

<sup>4</sup>We use the term *new* generation of reforms, but the more appropriate phrase may be the *next* generation of reforms. *New* implies an old and a new, whereas *next* implies continuing evolution. The latter may better represent what is going on.

<sup>5</sup>Of course, many of the new generation of agencies will not be called welfare, social service, or income maintenance agencies. They are likely to be called jobs programs or workforce development programs.

beyond the goal of merely enhancing labor market attachment, is the most recent in an historical sequence of shifting and cumulating goals.<sup>6</sup>

When the Aid to Dependent Children's program was enacted as part of the Social Security Act, the focus was on children. It was assumed that benefits paid for their well-being would enable them to stay with their liable caretaker (in the overwhelming majority of cases a widowed female), an arrangement assumed better for the development of the child than substitute care (e.g., in an orphanage) or living with a working mother. By the mid-1960s, the reduction of income poverty became an explicit policy goal. Income strategies (e.g., various Negative Income Tax schemes) emerged; welfare was transformed into more of an entitlement. By the late 1960s, the focus had shifted to the caretaker, by now mostly women with children whose economic woes typically stemmed from a failed marriage or the failure to marry. The primary concerns were dependency and strategies for moving women into the labor market. The unit of analysis and concern broadened to the family, first in the form of child support enforcement reforms and later in a concern about adopting promarriage and profamily policies. Most recently, there has been a subtle shift back to the children.<sup>7</sup> Though reform goals concerning the mother (or caretaker) and family remain, it is argued that what really counts are the longer-term prospects for healthy child development.<sup>8</sup> Emerging reform objectives do not fully replace prior goals. Often, they become part of the multiplicity of ends important to reformers.<sup>9</sup>

---

<sup>6</sup>Of course, the more traditional question has been the relationship between welfare and work, mothers and the labor force. The relationship between mothers and work also changed dramatically over time. At the inception of AFDC, less than one-quarter of all women were in the labor force, and mothers were disproportionately underrepresented. Societal norms did not expect them to work and many firms as a matter of policy encouraged mothers (or women who married) to leave. Over the post-war decades, the labor market participation of women in general, mothers in particular, began to rise, at some periods increasing by one percentage point a year. By the early 1960s, the question of whether welfare mothers ought to work had been engaged. By the 1990s, there was little debate on that point. Societal norms now accepted that mothers, even with young children, were obliged to be self-sufficient. Further, a host of welfare-to-work experiments had apparently demonstrated that such programs were feasible and at least modestly effective.

<sup>7</sup>This shift was evidenced by the inclusion of child outcomes in the evaluation of the JOBS provisions included in the Family Support Act of 1988. Child Trends, under contract to the Manpower Demonstration Research Corporation (MDRC), is focusing on this aspect of the evaluation. Also see the recently published working paper by the National Center for Children in Poverty, "State Welfare Waiver Evaluations: Will They Increase Our Understanding of the Impact of Welfare Reform on Children?" by Ann Collins and Larry Aber (1995). The Board on Families and Children has also held a workshop on Welfare and Child Development.

<sup>8</sup>For example, see Deborah A. Phillips and Anne Bridgman, eds., *New Findings on Children, Families, and Economic Self-Sufficiency: Summary of a Research Briefing* (Washington, D.C.: National Academy Press, 1995).

<sup>9</sup>These subtle transitions are important to the reform debate and are not always fully appreciated. The debate has responded to the demographic earthquake evidenced in this country. Both divorce and nonmarital births have increased sharply over the last thirty years. In 1960, only 5 percent of births occurred outside of marriage. In 1993, the



This is in sharp contrast to traditional welfare programs, whose purpose was to provide some measure of income support. The nominal objective was to redress economic need in the short run. The primary roles of operational personnel were to process information and to validate those data necessary to issue accurate checks to eligible families. The time frame was the monthly accounting period, each month representing a new and distinct set of calculations. Particularly in Wisconsin, where much of the actual case decision making had been automated as far back as the 1970s, little discretion remained in the system.<sup>10</sup>

There are some exceptions to this characterization of welfare administration. Various welfare to work initiatives had been around virtually from the time that AFDC had been transformed into an entitlement-type system in the late 1960s and early 1970s.<sup>11</sup> These activities were, in theory at least, designed to change behavior. However, they often were viewed as tangential to primary welfare functions and often were contracted out to service providers located outside the welfare agency boundaries.<sup>12</sup>

Resulting from the ambitious expectations regarding what these reforms might accomplish, the next generation of welfare constitutes a qualitative change from those of just a few years before. Again, drawing from the conference background paper:

- The new welfare regimes will be *longitudinal* in character; participants are viewed as being in a *change* process.
- The new welfare systems will be *behavior-oriented* rather than *income-oriented*, where the dominant objective is geared toward enabling participants to be self-sufficient rather than guaranteeing certain economic outcomes.

---

comparable number was 31 percent. In 1960, the divorce rate was 9.0 per thousand married women; that had grown to 21.0 in 1990. In 1970, less than 12 percent of all families with children were headed by a single parent (3.4 million); that percentage had risen to nearly 27 percent by 1994 (9.8 million). The dramatic change in the demographic landscape sharpened the call for change.

<sup>10</sup>The Computer Reporting Network (CRN) was developed in the 1970s. Prior to that, the governing rules of welfare in Wisconsin permitted a great deal of discretion. CRN automated virtually all core decisions regarding the determination of eligibility and the calculation of benefits. Permissive language in the welfare manuals—the use of words such as *may*—were replaced by non-discretionary words such as *shall*.

<sup>11</sup>Many do not fully realize that welfare administration from the 1930s through the 1960s was highly discretionary and personalized. The mission was multidimensional, to ensure that a variety of behavioral conditions were met as well as to dispense assistance. Welfare was not an entitlement but conditional on character and behavior. Welfare was transformed into more of a rights-based entitlement between the mid-1960s and mid-1970s; the entitlement-based concept of AFDC was under full attack by the mid-1980s, when reforms bringing back earlier behavioral conditions (e.g., Learnfare) made a comeback.

<sup>12</sup>The Kenosha County Jobs Center model drew significant attention because it fully integrated welfare to work and traditional welfare functions. Among other things, the Kenosha model colocated welfare staff with JOBS and Job Service staff, establishing a common caseload.

- The new welfare systems will be *multidimensional* rather than *unidimensional*; participants will be tracked along varying program paths and thus experience the program quite differently.
- The new welfare systems will be *dynamic* rather than *static*, continually changing for a variety of reasons from political instability to significant transformations in the labor market.
- The new welfare regimes will be *craft-oriented* rather than *standardized*; each case will be approached and treated in an individualized manner relative to the former system.
- The new welfare systems will be characterized by complex, nonroutine decisions that will require a good deal of *professional judgment* and involve a good deal of *discretion*.
- The new welfare systems will be *labor-intensive and management-intensive*; efficiency and low overhead may be less important than achieving certain outcomes.

The old welfare was routinized, repetitive, limited, static, and unidimensional. The new welfare is dynamic, complex, proactive, change-oriented, longitudinal, and labor-intensive.

Among all the state-initiated reforms on the horizon, Wisconsin Works (W-2) represents a very sharp departure from past practice. Because W-2 is so complex and ambitious, finding ways to rigorously evaluate its process is worth the effort. Consequently, we also argue that W-2 is a suitable, though far from perfect, laboratory for evaluating program implementation and process questions.<sup>13</sup> The same may not necessarily be said for evaluating W-2 outcomes. Many conventional evaluation strategies may not work well, as is discussed in other papers. Moreover, the advanced state of W-2 (a number of provisions are already in place) precludes the introduction of pre-post/cross site designs that will capture the full effects of the intervention. It is in developing new standards for assessing process that W-2 might well make its most important methodological contributions.

### Change in Agency Culture

The change in organizational culture that W-2 requires can hardly be more pronounced and would probably be catastrophic had the alterations not been going on for a decade now. Still, the remaining transitions are dramatic and the administrative challenges awesome. Brodkyn has argued that discretion has always been a part of welfare administration, even during the height of the entitlement period, and that control of operations at the street level is difficult at best though this has been the thrust of national and state policy over the past

---

<sup>13</sup>For an excellent treatment of administrative issues, see Lawrence Mead, "Welfare Policy: The Administrative Frontier," IRP Discussion Paper no.1093-96, University of Wisconsin: Institute for Research on Poverty, 1996.

several decades.<sup>14</sup> Despite, or perhaps because of this observation, knowing what is happening at the operations level assumes greater importance at this juncture and raises serious methodological questions about how to conceptualize and measure cultural change in organizations.

One approach to understanding the “meaning” of a new initiative requires that the evaluator obtain information on what the participants and the program operators actually *experience*. And since many of the new reforms are dynamic, change-oriented, longitudinal, etc., the evaluative strategy must take these qualities into account. If participants do not experience something quite different after a reform, serious questions can be raised about whether we can anticipate any effects on participants’ behavior.

Below, we try to describe the character of W-2 even though such an exercise is surrounded with uncertainty. Developers of the initiative have sketched in only a modest amount of the operational detail. In fact, there has been some reluctance to be over-prescriptive, on the philosophical and managerial principle that it is better to focus on performance rather than process, on what is accomplished rather than how it is accomplished. The most radical extension of this principle would be to move in the direction of performance contracting and privatization. At this writing, the state appears to be adopting a middle ground, moving cautiously in the direction of opening up the selection of providers to a limited competitive process, yet struggling with the complex task of describing (e.g., flow charting) what the new system will look like in practice.<sup>15</sup>

W-2 and similar reforms are best conceptualized as a series of consecutive interactions (or episodes) between the participant and the agency that can be characterized as events, activities, and decisions. They are not independent; the outcome of earlier episodes determines (or should determine) the consequent participant experience in the program. If problems arise in the execution of events or in the appropriateness of the decisions being made, the effects can ripple throughout the system. Queues (waiting lists) can develop, participants can be allocated to the wrong tracks, services not delivered or coordinated. This is just the surface of what can go wrong.

What holds a W-2 program together are the decisions made about (or with) the participant. These decisions are the glue that ties together the complex sequence of activities and events that constitute the

---

<sup>14</sup>See Evelyn Brodkyn, “The State Side of the ‘Welfare Contract’: Discretion and Accountability in Policy Delivery,” SSA Working Paper no. 6, University of Chicago: School of Social Service Administration, November 1995.

<sup>15</sup>Counties meeting certain performance standards (caseload reduction) during a baseline period were given the right of first refusal to operate W-2; otherwise, selection of the provider was through a competitive bidding process. Also, see attached for an example of an early attempt by the state to describe W-2 operations.

substance of the W-2 experience. Who makes the decisions, how they are made, and what information they are based on are critical to the operational success of the undertaking. Decisions can be routinized and made according to rigid protocols. Expert systems can be developed to guide, but not determine, some of the more complex decisions. Conversely, full and professional discretion can be accorded the decision makers.

## THE GROUND-LEVEL VIEW OF W-2

W-2 is very complex and is still evolving. Space considerations make it impossible to capture the full scope of the change. We first describe some of the key actors responsible for operating W-2 and then attempt to lay out the critical interactions between the participant and those actors.

### Key Institutional Actors

The number of actors and institutional systems involved in W-2 is substantial. There are several populations of interest: the adult caretakers and their children, noncustodial parents, and members of special groups such as minor parents, kinship care providers, etc. Interest also exists in such ancillary systems as child care, health care, child support, child welfare, community supports, the business community, and a host of other critical institutional contributors to W-2. For the purposes of this paper, however, we will focus only on the institutional actors who operate the system and the program participants with whom they interact.

If one were to examine the typical income maintenance agency of a decade ago, understanding the structure and organizational roles would have been a relatively simple matter. Since the governing organizational technology was routine and standardized, the participant-agency interface could easily be understood and specified. Personnel could be arranged in a classic hierarchical structure. The character of the participant-worker relationship was specified at the top (sometimes decision rules were built right into the agency's computer system) and organizational communication flows were essentially vertical.

There was little variation (at least theoretically) at the point where participants actually interacted with the agency. The test for the welfare agency of a decade ago was the efficiency and accuracy with which it processed data and made routine calculations. Discretion and professional judgment were discouraged. Role parameters were prescribed.

This vision of routine and limited worker-client interactions was never the actual norm, of course. Over time, new and more complex responsibilities were added to the income maintenance worker's duties, among

them responsibilities attached to carrying out the requirements associated with new child support and welfare-to-work program initiatives. Over time, welfare agencies became more organizationally complex, relying upon contracting for certain services or developing team approaches to more effectively complete increasingly demanding operational responsibilities. W-2 takes this growing complexity to qualitatively new levels.

The key roles (activities grouped into logical functions which would be performed by the same person) include the following:

- **Receptionist.** Performs the initial gatekeeping function; receives and processes initial public contacts. In many cases, will refer visitor to resources outside the W-2 agency, such as Job Net (a computer-assisted method for linking job seekers to available jobs).
- **Resource specialist.** Performs some key gatekeeping and triage functions; for example, performs initial referrals, attempts to divert potential participants away from W-2 to other resources and opportunities as appropriate, and conducts preliminary evaluation for W-2 services.
- **Financial and Employment Planner.** Performs key case management functions such as determining eligibility for work-supporting services, monitoring performance, and making tracking decisions at the vital process points.<sup>16</sup>
- **Supportive Services Planner.** Determines eligibility and provides for the delivery of work-supporting services for W-2 participants who are not receiving employment case management.
- **Other agency personnel.** Local grievance officer, child support and child care specialists, job developers, motivation trainers, soft-skills trainers, etc.
- **Selected community and service providers.** Child care and health care providers; education and training specialists; counseling, child welfare, law enforcement, and other personnel; housing, food, energy, and others who provide critical supports; employers, job market experts, temporary services agencies; and many, many more.

We stress that a comprehensive process evaluation would encompass the views and perceptions of nonoperations personnel, those not on the front lines who actually have contact with program participants. This includes program managers and those in the community affected by the program.

### Key Procedural Steps

The complexity of W-2<sup>17</sup> can be reduced to several macro-activities or modules. These are:

---

<sup>16</sup>We cannot stress too strongly that this is a simplistic and generic view of W-2. The actual program is likely to be much more complex with an even greater elaboration of these roles. For example, the plan for Dane County (where Madison is located), calls for the development of four distinct FEP roles.

<sup>17</sup>We try to capture some of the complexity in the crude flow chart that is attached (Figure 2).

- **Gatekeeping.** This is a simple module relative to others, but quite important, because it encompasses those activities and decisions that the participant first experiences. The character of this experience may say a lot about whether clients continue and about their subsequent behavior and attitude.
- **Triage/plan development.** Next is a very complex set of activities and decisions related to determining eligibility, moving the participant into the appropriate track or group, and organizing a set of appropriate activities and experiences.
- **Participation/ monitoring.** The basic participant experience in W-2 will be the activities encompassed in one of several tracks—Trial Jobs, Community Service Jobs, and W-2 Transitions. However, tangential involvement with supportive service systems such as child support, child care, health care, and food stamps may also be going on. Monitoring is central, since benefits are tied to work or activity compliance, and other supportive services such as child care and, potentially, health care may depend on continuing copayments by participants.
- **Review/adjustment.** Everything is time-limited, including exposure to the experiences associated with participating in specific tiers (five-year overall limit, two years in a specific tier; six to nine months in a given subsidized job within a tier). Participants might have their status reviewed a number of times and be retraced through triage (where the participant may be reassigned to a different program track) if they are complying with all the rules. They might face more reviews (and intensive case management) if problems occur or if they do not comply with program rules.
- **Exit/follow up.** Actual or potential participants can exit at any time. In fact, diversion from the program is to be encouraged. The kinds of activities that will be developed to minimize recidivism and to otherwise support the family in the labor market (such as access to health care) are still unclear.

#### Generic W-2 Agency Case Management Flow

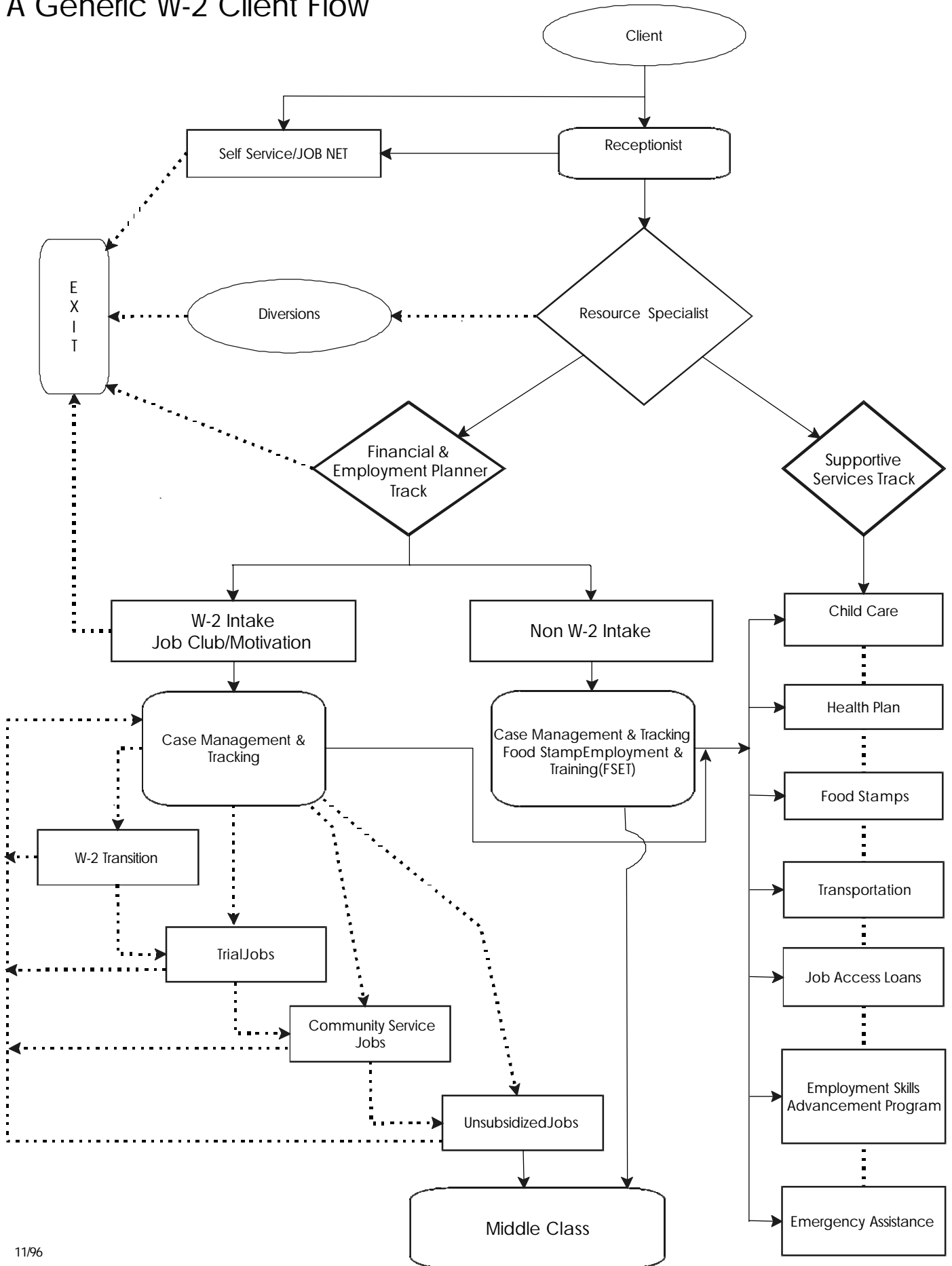
These modules are not discrete enough to develop a detailed implementation or process analysis. Although there will be an extraordinary amount of variation in local practices and operations, a generic flow model for W-2 agency case management is evolving (see Figures 2 and 3). Whatever the final model, we begin developing an implementation/process analysis plan by identifying *those points in the sequence of activities where things can go wrong, or at least can go in directions not intended by program designers and operators*. Good candidates for closer examination are those points where critical decisions are made by or about the participant: whether to proceed to another step, which track is to be assigned, and whether or not a participant is in compliance. The examples below point to the complexity of these decisions:

- **Screen initial contact.** The *receptionist* identifies the reason for potential participant contact (employment or employment service request),<sup>18</sup> determines appropriate referrals to be made, ascertains whether special accommodations need to be made (e.g., whether hearing or physically

---

<sup>18</sup>In case there is any confusion, W-2 assumes that agency cultures will be radically reoriented at this point. Potential participants will be walking into a Job Center, not a welfare office. They are likely to be greeted by signs and messages that are job-oriented, not income-maintenance-oriented. This is all part of the cultural transformation of welfare that started with the development of the Kenosha Job Center almost a decade ago.

# A Generic W-2 Client Flow





# W-2 Agency Case Management Flow

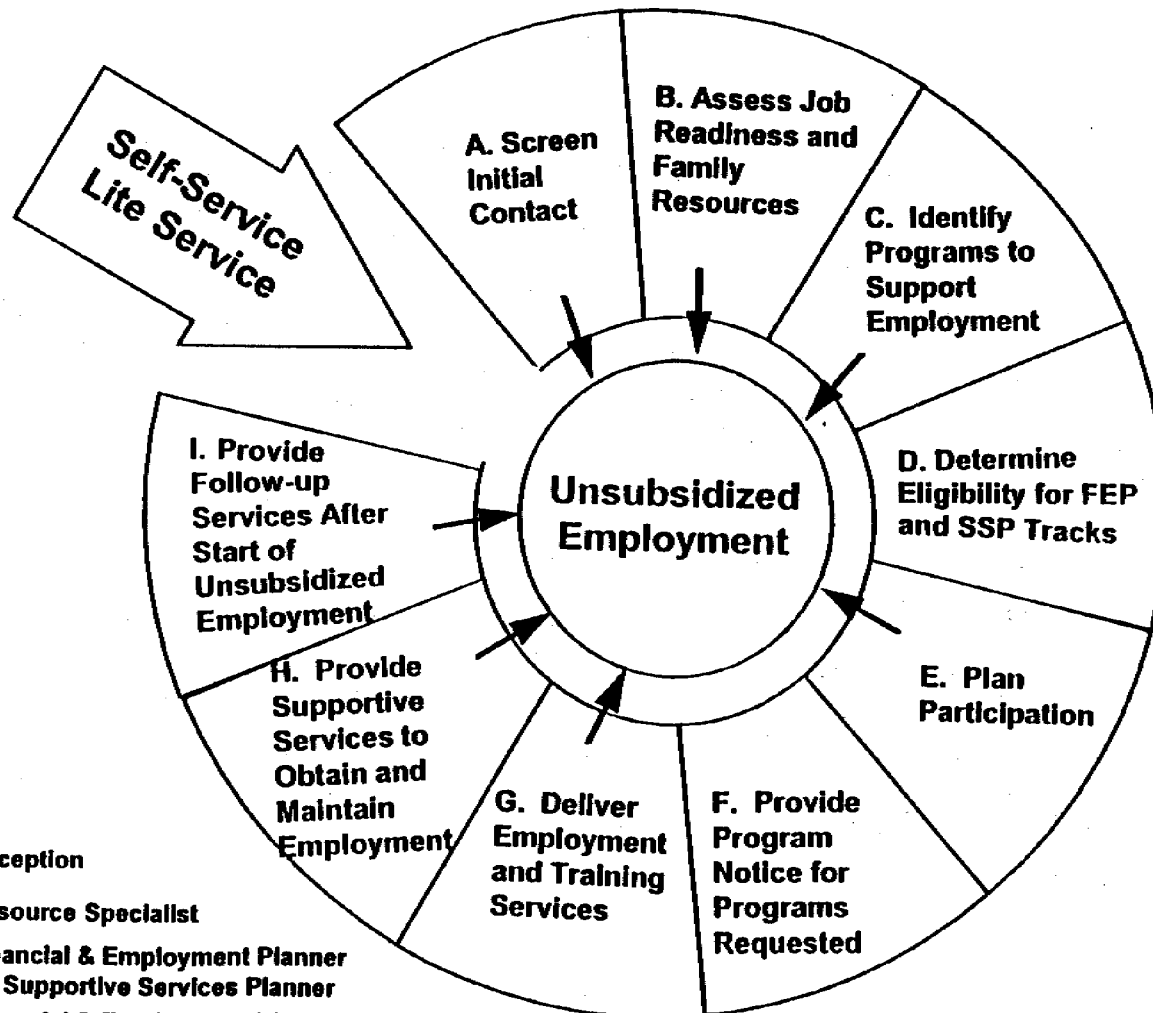


Figure 3

impaired, non-English-speaking, etc.), collects demographic information, and checks current and prior contacts.

- **Assess job readiness and family resources.** The *resource specialist* determines the current employment situation and assesses basic human capital, determines family composition as it affects work, explores self-sufficiency alternatives,<sup>19</sup> documents the contact and screens for priority (emergency) service needs.
- **Identify programs to support employment.** At this point, participants encounter a major sorting process. The program expectation is that a majority of clients will “self-sort.” The hope is that participants will be able to indicate fairly clearly what their minimum level of need is. Some will then continue on the main W-2 track,<sup>20</sup> others will move to what is called the supportive services track (needing only supplemental work services). The *resource specialist* also develops a preliminary self-sufficiency plan with the participant and refers the participant to child support and other potential benefits (SSI, OASDI, UC, VA, etc).
- **Determine eligibility for FEP and SSP tracks.** The *Financial and Employment Planner (FEP)* or the *Supportive Services Planner (SSP)* processes the formal application for assistance. This includes verifying program compliance (child support cooperation, two-week job search, application for other benefits); gathering, recording, and verifying data and making appropriate eligibility decisions; explain all program requirements and philosophy; and checking for fraud or abuse (including national lifetime time-limit).
- **Develop a participation plan.** The *FEP* will then identify employment strengths through various assessment tools; negotiate and document a mutually acceptable employability/self-sufficiency plan with the participant to determine the highest appropriate level of work activity for the participant, schedule and assign participants to planned activities (motivational training, employment positions, Food Stamp Employment and Training programs); and explain all results of the eligibility determination as well as all rights and responsibilities.
- **Deliver employment and training services.** The *FEP* now performs a whole range of sophisticated services, including expanding and enhancing the self-sufficiency plan to what is called the “detail” level; assigning the participant ongoing, specific work or training activities (including matching person to openings, negotiating with vendors or employers, establishing feedback mechanisms, and providing or arranging for “soft-skill” services); arranging for rehabilitation services if needed; tracking participant compliance; performing a range of monitoring and documentation services; negotiating all contracts and trouble shooting problems; processing subsidized payments to employers where applicable (Trial Jobs—level 2); processing payments to participants where applicable (Community

---

<sup>19</sup>Exploring self-sufficiency alternatives essentially is the client-diversion strategy. Key tasks include identifying extended family resources; identifying transportation and child care resources; exploring health care and nutrition options; exploring various service, training, and employment possibilities. All is discussed with the potential participant to determine if an alternative to W-2 participation is feasible and warranted.

<sup>20</sup>This track involves preparation for various W-2 employment positions and the delivery of various support services and assistance including child care, the healthplan, food stamps, Job Access loans, and emergency assistance. The supportive services track provides all the ancillary services and assistance.

Service Jobs and Transition Benefits); providing services to advance participants on the self-sufficiency ladder.<sup>21</sup>

- **Provide supportive services to obtain and maintain employment.** The *FEP* (*SSP* in some instances) will determine the need for child care and refer as needed for services; reassess need on an ongoing basis and assess noncompliance with rules; identify need for health care and administer the program (issue authorizations, monitor use, process all payments and copayments, repayments, and overpayments, etc.); provide and administer food stamps, transportation services, job access loans, services under the Employment Skills Access Program (ESAP), and Emergency Assistance (EA) benefits.<sup>22</sup>
- **Review and manage participant transitions.** At a number of points the case must be reviewed and decisions made. Within a track, time limits exist. The *FEP* must determine if another such job should be offered or if the participant should be moved up, or down, the ladder. Time limits in a given track are generally two years. Decisions must be made to extend that time limit, if appropriate. Additionally, there is a five-year lifetime limit. Therefore there are many reviews imposed by state W-2 regulations or required by federal laws.
- **Provide follow-up services after start of unsubsidized employment.** Services continue after the participant has achieved unsubsidized employment. The *FEP* (*SSP* in some cases) will follow up the participant and/or employer; determine whether continuing services are needed to maintain successful employment; deliver (or arrange) for such services; identify when services may end; close case when activity ceases and document closure.

Many process or flow problems might be identified when W-2 is assessed at the ground level.<sup>23</sup> We name a few examples at this point:

- **Queuing.** As participants flow through the system, do bottlenecks develop at critical points; at eligibility determination, when child care or community service jobs are to be arranged, or where the remediation of a barrier to employment must be addressed? A bottleneck occurs when key events do not take place or important decisions are not made in a prompt and seamless fashion.

---

<sup>21</sup>Note that complaints and grievances can occur at any time. The initial steps will be handled by a *Local Grievance Officer (LGO)* and subsequent steps by the *Division of Hearings and Appeals (DHA)*. It is not clear how this will really work, since there presumably is no individual entitlement to cash benefits under federal law or W-2.

<sup>22</sup>All of these services and benefits from Food Stamps to EA involve a complex set of activities: identifying need, issuing authorizations, monitoring use and assessing any abuse, processing payments or overseeing service delivery, reassessing need or level of benefits and services, and implementing sanctions where appropriate.

<sup>23</sup>As a reminder, we are looking only at the participant/worker set of interactions. A number of macro-level or administrative functions and tasks would also be part of any comprehensive process evaluation. For example, key W-2 administrative responsibilities include maintaining all internal operations (e.g., hiring, training, providing management supports, etc.); establishing and maintaining a working relationship with the state, including meeting all contractual agreements; developing and maintaining service partner relationships; establishing and maintaining relationships with the community; and establishing and maintaining positive relationships with employers.

- **Leakage.** Do participants disappear, get lost in the system, fail to move from one worker to another or one component to the next? This has been a major issue in early welfare-to-work programs and has a greater potential for being a problem here, given the complexity of W-2.
- **Misallocation.** Participants are sorted and tracked along different paths at a number of points in the process. Are these allocation decisions made according to accepted policy guidelines? Are participants proceeding along the various program tracks according to a priori expectations?
- **Conflict.** All kinds of disputes can arise, given the number of complex decisions to be made and the extent to which discretion is afforded front-line personnel. Are there more disputes than expected? Where are they occurring, and what are the issues?

The simplest tracking devices can be of great assistance in understanding what actually happens to participants as opposed to what is supposed to happen. Measuring time between significant events and collecting minimal data on the perceptions of key operations personnel and participants can be a useful management tool. Expectation can be compared to reality, problem areas pinpointed, and responses formulated and delivered.

*The intersection of actors and functions—an example.* It will be very difficult for the state to prescribe a given administrative approach; considerable diversity exists at the local or agency level for several reasons. First, nonroutine, discretionary core activities are not easily managed and controlled by centralized authorities. Second, the state has increasingly wanted to move in the direction of performance-based contracting, in which what an agency accomplishes is viewed as more important than how it accomplishes it. Third, the state has taken some steps to privatize the management of W-2 in some families, a step beyond performance contracting.

Both the character of the new reforms and the desire for decentralization increase the prospects for local variation in program design and operations. Some observers believe that several distinguishable models are emerging as counties, welfare agencies, and other potential W-2 providers develop their service delivery approaches.<sup>24</sup> The designers of this model believe that the FEP role is so complex that four distinct positions (or roles) will be required:

- **FEP I: Initial assessment and referral specialist.** Conducts the individualized assessment with Job Center *customers* seeking specialized W-2 assistance. The goal is to explore with the potential applicant the options available to meet personal and employment goals using family and community resources; to make a preliminary determination of whether or not W-2 services are appropriate and, if they are, to facilitate the maximum degree of self-sufficiency as the applicant interacts with W-2.
- **FEP II: Applicant specialist.** Meets with applicants to determine their eligibility for W-2 and non-W-2 programs and services and refers to the appropriate services. This person performs the gatekeeper function.

---

<sup>24</sup>According to Dan Kittel, operations manager of the Dane County Jobs center.

- ***FEP III: Employability planning specialist.*** Applicants who have not found employment are referred to *FEP III* for orientation and enrollment into the appropriate W-2 employment service tier. They are responsible for assigning participants to the appropriate tier or level and for managing their progress through W-2.
- ***FEP IV: Employment retention specialist.*** Interviews persons found eligible for services in the W-2 unsubsidized employment tier. *FEP IV* is responsible for seeking out and providing or for referring participants to employment retention services such as W-2 child care, W-2 health plan, advanced EITC payments, child support enforcement activities, transportation services, etc. *FEP IV* will also provide feedback to the program design and coordination teams within the Job Center that are addressing labor market shortages and retention issues to identify opportunities for program development and specialized employer retention services.

Other important roles include:

- the *Job Developer*, responsible for developing relationships and partnerships with area businesses, both public and private, to promote W-2 services and participants and for identifying subsidized and unsubsidized employment opportunities for participants and negotiating contracts with businesses;
- the *Employment Skills Development Specialist*, responsible for the development and implementation both in the classroom and at the training site; and
- *Support Service Workers and Child Care Specialists*, responsible for the management and distribution of child care and transportation services for W-2 participants. Their activities include determining eligibility, issuing vouchers, preparing and reviewing payrolls, tracking expenditures, resolving billing issues with providers, serving as a link between participants and community service providers, and analyzing the support needs of W-2 participants.

Such role differentiation reflects the complexity and diversity among W-2 providers. This local diversity argues that there is not a W-2 program, but a host of W-2 programs.<sup>25</sup> Over time there may exist as much structural and procedural variation within Wisconsin as there will exist between Wisconsin and various other jurisdictions. Thus, there is not likely to be a single *implementation* or *process* analysis that will capture the diversity of approaches and strategies. We must think about how to group models over time as agencies find some institutional stability and select those features and arrangements best suited to their philosophy and context.

---

<sup>25</sup>We do not yet know the final number, but it may run to more than 80 agencies. There are 72 counties and several Indian tribes in Wisconsin. Milwaukee County has been divided into six districts; the W-2 management entity in each was decided by a competitive process that produced an eclectic array of providers, from Goodwill Industries to a for-profit firm such as Maximus, Inc.

## ADDRESSING THE MEASUREMENT CHALLENGE

I keep the discussion of measurement issues simple and generic at this point, preferring to indicate a general strategy rather than a detailed plan.

Doing an implementation or process analysis essentially means capturing participants' experience throughout the program. If participants do not encounter an experience qualitatively different than they would have under the old program (counterfactual #1) or different than they would have under alternative policy and program structures (counterfactual #2), then the program has not succeeded in doing something different and we might be a bit less sanguine about significant impacts. Most new-generation reforms purport to alter the essential *culture* of welfare offices and the way those delivering actual services function. There are three general areas of concern in considering what constitutes this experience and how we might capture it. First are those attributes of the participant's experience that shape the character of that experience in the program (1); second are those points in the flow of participant-agency contact that should be most closely examined (2); and finally there is the issue of data collection methods (3).

1. Attributes of the Participant's Experience (What)

- **Clarity.** Altering the culture of offices means communicating a different message to potential applicants and actual participants. Is the message communicated clearly and concisely? Is there ambiguity, or inconsistency, or role confusion among personnel? Recall the number of FEP roles identified above. Are the benefits, obligations, and mechanics of the new system communicated well?
- **Celerity.** Changing the culture and the message means delivering services quickly and efficiently. If there is excessive "down time," the meaning of the new program may be unclear, the seriousness questioned. This may be particularly true at the front end (contact, gatekeeping, triage, and tracking phases).
- **Consequences.** There must be consequences for failure to perform on both sides (agency and participant). W-2 would seem to have consequences built in for noncompliance, but administrative failures could still undermine that dimension of the program, particularly in tiers 3 and 4, where the agency issues checks.
- **Seamlessness.** Transitions are critical in W-2. When one W-2 placement ends (or is supposed to end), when does the next begin?
- **Accuracy.** A number of decisions are made in W-2 about where participants belong and how much they are to pay for child and health care and whether or not they are in compliance with the rules. Some of these decisions are prescribed, others involve discretion. Most can be assessed for accuracy in some way or another.
- **Quality.** At least minimal levels of quality can be assessed in areas such as the provision of child care, job search assistance, or job development and matching.

## 2. Points of Concern in the Participant Flow (Where)

- **Contact.** What does the potential applicant experience when first approaching the program? What is the substance of the initial message (welfare or employment); what is the quality of the message (hostile or supportive)?
- **Gatekeeping.** How is the complex set of services, expectations, data gathering and data verification at the front end carried out? This is an extremely hectic and complex set of procedures, with much information being exchanged and very difficult decisions being made.
- **Diversions.** Diverting potential participants from the program is an important goal but can be abused unintentionally or intentionally (e.g., budget pressures may create perverse incentives). Getting inside this delicate procedure may be critical.
- **Tracking.** Making the decision about where a participant will start in W-2 is another of those seminal decision points with great consequences. Pegging participants too high (tier 2, for example) may lead to failure and employer dissatisfaction. Putting them in too low a track may not set high enough expectations for participants.
- **Monitoring.** Monitoring participant performance and administering benefits and payments is a huge responsibility. Participants can fall by the wayside; disruptions in benefits or copayments can occur for appropriate or inappropriate reasons. Monitoring may reveal noncompliance, which raises other complex issues, such as how to decide culpability where the facts are unclear.<sup>26</sup>
- **Transitions.** W-2 is full of key transitions: from one subsidized job to another; from one tier to another; from one form of time limit to another. Procedures must be in place regarding how and when these transitions will occur and who will be empowered to make these important decisions.
- **Exit/recidivism.** There are decisions and issues related to exits from the program (though when a person really exits from W-2 is a bit unclear). Some exits may be more desirable than others (employment, as opposed to giving up), and the availability and quality of any work support services delivered may determine recidivism rates or well-being over time.

## 3. The Intersection of What, Where, and How

- **Administrative/management data.** A good automated case management system capable of tracking individuals and families over time and across programs will be essential. Probably no state has such a capability in hand, although California, Oregon, Massachusetts, Indiana, Illinois, and a few others have some components in place.
- **Special case-tracking efforts.** While automated systems are being developed, generic tracking surveys may have to be developed to track cases over time and keep track of key dates, outcomes, decisions, and service delivery. This may have to be done on a sample basis, for reasons of cost and feasibility.

---

<sup>26</sup>For example, a participant in a community service position may claim sexual or racial harassment while the employer claims negligence and insubordination. How to resolve these matters is difficult.

- **Surveys.** Some decision points in the process may be so important that we suggest constructing questionnaires (or some data-gathering device) around those points. If the distribution of new participants across the W-2 tiers radically deviates from prior expectations, we may want to find out why.
- **Interviews (structured).** Interviews with operations personnel (those working directly with clients) would seem to be essential to document, among other things, how they view the program, what problems they perceive, and how they see their specific role.
- **Interviews (unstructured).** Replicate the above for management, with the unstructured format permitting exploration of problems unforeseen by the evaluators.

Evaluating process in the new reforms builds on the tools developed over the past several years in evaluating welfare reform initiatives, but it takes the enterprise more seriously. At the core, the evaluators must get inside the experience of both the potential and actual participants; identify the key decision points and other points where things can depart from plans; and capture data that will shed light on what happens as opposed to what is supposed to happen. The evaluators (or program designers) should set expectations about what is supposed to happen (e.g., how quickly certain events are supposed to occur, such as finding a child care slot) or how it is supposed to happen. This at least creates a set of norms for desired performance against which actual performance can be compared and/or adjusted.

## BEYOND W-2

Ultimately, neither this paper nor this conference is really about W-2. The Wisconsin reform is merely the laboratory for exploring issues and strategies associated with devolution (waiver-based and law-based) and the consequent emergence of a new generation of policy, program, and management innovation. As devolution plays itself out and the actual practice of running programs for disadvantaged families becomes more diverse, both the need to accurately describe what a program does (and what participants experience) and the challenge of completing such descriptions will increase. I suggest the following steps:

- That we work toward establishing general principles and methods for describing administrative and operations models in ways that really communicate what is going on to interested observers. In part, this means clearly conceptualizing the audience for such evaluations and determining what they really need.
- That states and evaluators begin a dialogue over common objectives to guide future work on implementation and process analyses and over common standards for conducting acceptable process analyses.



- That states begin to work together toward common definitions of terms and concepts so that communications among states and cross-state comparisons become more feasible.
- That evaluators begin a similar form of communication and exchange in order to develop and improve the practice of doing implementation/process studies.
- That program designers and evaluators begin thinking about how to better integrate formative with summative evaluations. Formative evaluations are designed to alter program parameters or operations. This creates instability that renders causal interpretation of impacts unclear. This tension needs to be resolved.
- That we all begin to think about making administrative outcomes a distinct matter of investigation. One could examine systematic (experimental) or natural variation in administrative strategies and see how they relate to administrative outcomes of interest—measures of satisfaction, efficiency, or accuracy.

I argue that implementation/process analyses should receive the same attention traditionally accorded to impact analyses.

## **Process Evaluation of W-2: What It Is, Why It Is Useful, and How to Do It**

Karen C. Holden and Arthur Reynolds

A process evaluation attempts to illuminate program administration and client behavior from the start of program implementation to the observation of program outcomes. Scriven (1991) describes process evaluation as an activity that “refers to an evaluation of the treatment . . . that focuses entirely on the variables between input and output . . . [and that] refer(s) to the process component of a full evaluation” (pp. 277 ff.). A process evaluation measures program processes, enabling evaluators to draw direct links between program activities and ultimate outcomes. It documents delivery of the hypothesized treatment, measuring dosage and duration of that treatment, examining variations in delivery across service agencies and client groups, identifying intervening and confounding activities, and explicitly linking variations in service mobilization and delivery to variations in measured outcomes.

An evaluation’s process component provides policy makers with guidance on whether and how to extend or modify programs in time, space, or coverage. What in fact took place between the initiation of an innovative policy and program outcomes must be measured in order for policy makers to make decisions about continuing programs over time (e.g., whether, after evaluation results are produced, W-2 as a whole or individual components should be continued as is, be modified, or be terminated), and about extending programs to different places (e.g., whether other states should adopt welfare program components similar to those of W-2 in expectation of similar results) or covering different populations (e.g., whether W-2 as a whole or specific components should be extended to families with somewhat higher incomes than the current program cutoff or to noncustodial parents, and whether health benefits should be extended to poor nonparents). To measure only outcomes—simply assuming that the full program was implemented as intended, regarding implementation as an unobserved “black box”—provides no guidance on program modification and extensions that might alter outcomes in desired ways.

The value of understanding process is widely accepted in the evaluation literature. In part because of the time-intensive nature of gathering process data, examples have been found more often in evaluations of relatively small demonstrations of social service programs, in which gathering information on program process was central to both program and evaluation design. Examples include substance abuse education programs (Ellickson and Bell, 1990; Pentz et al., 1989), several spousal abuse response experiments (Sherman et al., 1992; 1984), and juvenile justice experiments (Land, McCall, and Williams, 1990). In these small and fairly well-monitored experiments, variations in the method of delivery and intensity of the treatments were important

in explaining net outcomes. Despite the widely accepted importance of program implementation evaluations for understanding program causality, Lipsey et al. (1985) observed that only about 10 percent of the large number of evaluations they reviewed provided any documentation on whether programs were implemented as assumed by evaluators and on how program components may have affected observed outcomes. Large-scale employment training and adult education programs have devoted little attention to evaluation of program implementation (Riggin, 1990). In part this results from the challenges of gathering data on organizational expectations and administrative practices across many large program sites. In one attempt to do so, Riccio and Friedlander (1996) explore the relationship in California's GAIN program of implementation practices to earnings outcomes in selected sites.

It is important that Wisconsin begin now to develop plans to chart the implementation of Wisconsin Works and to consider how to follow the different entry cohorts of participants as they move through the job tiers. This information is likely be useful for short-run program management purposes; more important, whatever outcome evaluation design is adopted, understanding program implementation and provision of services to early participants will be important in the causal analysis of eventual outcomes among those same participants and subsequent cohorts.

## EXAMINING PROCESS IN WISCONSIN WORKS

Wisconsin Works (W-2) is a far more complex program in terms of program administration and services received by clients than was Wisconsin's AFDC program. (The basic features of the program are outlined in other papers and are not described here.) In contrast to AFDC, in which eligibility is based on income and assets and leads to entitlement for cash benefits, W-2 offers a bundle of potential services: the assistance of financial and employment planners (FEPs), referrals to community service agencies, subsidies for paid work, child care, and health insurance. The complexity of this statewide program will lead to variation across counties in services received by similar participants, even in the fully implemented program.

Also in contrast to AFDC is the intent of current welfare reform efforts in general, and W-2 in particular, to change the culture of the welfare population and, as Mead (1996) describes it, to "control the lifestyle of adult participants." W-2 does this by setting conditions for the receipt of aid and imposing strong administrative suasion and sanctions on participant behavior. Wisconsin Works is not an entitlement. While eligibility includes meeting an income and asset test, continued eligibility also requires, for example, that W-2 clients cooperate in efforts directed at establishing paternity of the dependent child, furnish the W-2 agency

with any relevant information the agency determines is necessary, and make a good faith effort, as determined by the agency on a case-by-case basis, to obtain employment. W-2 agencies are likely to operationalize these requirements in different ways, with different behavioral consequences for clients and different program outcomes.

The latitude allowed W-2 agencies in the administration of the program is likely to result in wide variation in program attributes and sanctioning policies across service delivery areas. This variation is both a challenge and an opportunity to evaluators who will be trying to identify the role of W-2 in moving participants toward greater economic independence. Gradual implementation makes it difficult to define the “program” to which individuals entering at different times were subject. Nonuniform program implementation across counties means that individuals subject to W-2 over the same calendar period will in fact be subject to different programs. At the same time, geographic and time-related variation in the W-2 program provides a laboratory in which to evaluate the relative importance to program outcomes of individual program components.

## KEY ISSUES IN PROCESS EVALUATION

### Implementation

In a process evaluation, the primary task is to assess program coverage and the delivery of program services. Three main questions are addressed:

- Are the administrative services and resources that are key to the success of the program in place when implementation begins?
- To what extent are program services being delivered to the target population as specified in the program design?
- What is the variation in program services and administration across program sites, and how do these relate to differences in the delivery and coverage of services?

For at least three reasons, evaluating implementation is essential to assess social programs and services:

1. *To validate a fundamental assumption of impact (outcomes) evaluation.* A key requirement for conducting an outcome evaluation is to learn whether the program has been delivered as intended to the target population, and a process evaluation allows one to test the accuracy of this basic assumption. Much previous research has indicated substantial deviations between the intended and actual implementation of program

treatments (Blalock, 1990; Weiss, 1987). The assumption of program implementation must be tested lest the magnitude of program outcomes be erroneously attributed to diminished program effects rather than accurately attributed to deficient implementation.

Because implementation problems usually result from incomplete delivery of programs or the implementation of the wrong program for the targeted population, outcome evaluations may often and unknowingly underestimate effectiveness. For example, evaluation of the initial Head Start Program for preschoolers and their parents greatly overestimated, owing to a lack of process documentation, the extent to which health services, family services, and staff training were available and provided. The outcome evaluation found no effects of the summer program and limited effects of the full-year program, in part because of undelivered services. As Cicirelli et al. (1969) indicated in their report, how the program was implemented “was not a basic question of the study, and the time schedule did not permit intensive investigation of the program.” They further explained that “it was impossible to know in detail the actual program that these children experienced” (see also Wu, 1991).

Evaluations of the implementation of Title I educational block grants illustrate a complementary problem—that even if resources are available and services delivered, they may be delivered nonuniformly and fail to reach the target population that is most in need (Doernberger & Zigler, 1993). If services are not available to the population for which program impacts are estimated, impact evaluation findings are misleading, because they are not evaluations of the impact of the intended program but of some other social forces or of some other social programs in place prior to the assumed implementation of the program.

2. *To help explain why a program worked or did not work.* Evaluating program implementation helps investigators clarify the precise program elements and features that were implemented well or not so well. This is especially the case for complex, multiple-component programs in which implementation may vary across geographic areas. Consequently, information on the elements that are well implemented can be used to detail the source of the effects of the program.

Alternatively, implementation information helps explain why the program may not have been as effective as expected or why effects varied across program locations. There are three general reasons why programs do not show their intended effects: inadequate program design or theory, poor program implementation, and inadequacies in the research design or measures. Process evaluations help distinguish among these three reasons for no-effect or small-effect findings. Even if services are delivered, variation in types and targets of service delivery implies that treatment and comparison groups may be imprecisely defined, reducing the ability of the impact evaluation to find effects that may exist. Because many employment training

and adult education programs devote little or no attention to evaluation of program implementation, it is impossible to assess whether small impacts are due to lack of programs implementation among the study sample, or to imprecise definition of the “program” variable, or to actual small program effects.

As noted by Cook & Shadish (1986), evaluators have learned that “analyses of implementation and causal mediation are crucial in evaluation, for they promote explanation, and explanation may be crucial for the transfer of evaluation findings to new settings and populations” (p. 226).

3. *To promote program replication and evaluation utilization.* By understanding and documenting the workings of a program, its essential operative features can be identified and disseminated for use in other settings, thus promoting program diffusion and replication.

Since a major goal in evaluation is the utilization of findings, implementation evaluations also encourage collaboration between evaluators, program designers, and other stakeholders in understanding the essential features of programs and making them effective. Program delivery systems typically consist of a number of separate elements that may not be uniformly administered across program sites. W-2 allows considerable discretion in program implementation and administration across sites, resulting in treatment and outcome variation. Utilization of the results of an outcome evaluation for program expansion or continuation requires that the delivery system(s) be described and the contribution of particular elements assessed.

## EVALUATING INNOVATIVE, LARGE-SCALE SOCIAL SERVICE PROGRAMS

Evaluating programs such as Wisconsin Works poses special challenges. One is the innovative nature of the program. Being untried, it may take many years to fully quantify program implementation and effectiveness for families and children. It is very difficult to conduct comprehensive outcome evaluations of large-scale programs without a careful and continual mapping of the character of the treatment. There will not be a single treatment regime introduced at one moment in time, but many different treatments introduced statewide, begun simultaneously and then gradually modified.

Gradual implementation raises questions about evaluation design. Should the emphasis be on process evaluation alone, impact evaluation alone, or some combination of the two? On the one hand, program administrators would like quick and early feedback on program implementation and service delivery in order to alter program services as soon as possible to address unexpected results and unwanted side effects. On the other hand, particular interest groups would like evaluation resources devoted to a “true” test of a fully implemented program in order to discover the effect of the full delivery of the program as intended. We propose

a middle-ground position combining these two views: to conduct a long-term outcome evaluation and give attention to documenting program implementation and program services over the period in order to properly specify the timing and type of treatment services received by program clients.

### Nonuniform Treatment Delivery

W-2 is a block-grant program whose effectiveness will depend in large part on the coordination and availability of resources at the local level. The type of agencies administering the program and the extent and quality of program implementation will differ by locality and geography. Agencies may be private or public, and may be those traditionally engaged in job search (e.g., temp agencies) or those traditionally engaged in educational or family services (e.g., Head Start agencies). These different agencies may organize administrative tasks differently, implement program components in different ways and on different schedules over the early months of the program, and have greater expertise in providing certain types of job counseling and placement services. Resource differences across geographic units may include the availability and quality of employment opportunities and training, child care, health benefits, schools, etc. In such treatments, a close tracking of service availability and provision may be especially important.

Evaluations of federal block grant programs such as Title I education programs and community development programs indicate that global impact evaluations (collapsing across sites) are not only less likely to measure a significant program impact but are often misleading, because the treatment is administered differently across sites (often legitimately, due to local needs). Nonuniform programs are really multiple treatments and should be evaluated as such.

One approach to evaluating nonuniform treatments like W-2 is to study variation in program implementation as well as response to the program by site (e.g., county, service delivery system). For example, across 20 sites there could be several different programs. Evaluations would be conducted at each site, then pooled to provide an overall picture of the implementation and effects of the program, while identifying key differences across sites in the effect of different populations and of different service components. The advantage of within-site evaluations is that site-specific and service-system characteristics are largely controlled.

### Attributions of Success of Wisconsin Works

Another special challenge of evaluating W-2 is that even under the optimistic assumptions of successful and full program implementation and appropriate identification of comparison groups, changes in the health, employment, or family status of program participants cannot be attributed to particular program components

without some understanding of the organization, coverage, and extent of those components. W-2 is largely a program that mobilizes existing community resources, but may also stimulate their expansion to meet this larger demand. Many of these resources are not under the direct control of program managers. Thus, changes that are attributed to W-2 will be the result in part of local resource availability, the particular resources mobilized for particular client groups, and the success of clients in using those resources to move through the W-2 job tiers toward greater self-sufficiency. Process evaluation can provide insight into whether differential success in moving through the job tiers across sites and client groups and over time among different client cohorts was due to W-2 per se, to particular local resource availability, or to the expansion in local resources provided in response to the program (e.g., training, child care, etc.).

## PROCESS AND OUTCOME EVALUATION

Other papers discuss the feasibility of evaluating W-2 using a randomized experimental design (Glen Cain), cross-state comparisons (Mark Courtney and Irving Piliavin) and before-after comparisons (Robert Haveman). Drawing causal conclusions from any of these designs implies an understanding of program implementation and process. While few of the papers explicitly link outcome and process evaluation issues, virtually all recognize the importance of understanding the causal link between the identified outcomes and the delivery of W-2 services. Table 1 lists key outcomes described in these papers and the explicit or implied process links. While recognizing the importance of understanding program process, these papers do not attempt to describe how such data can be collected. This is in part due to the absence of a widely accepted theoretical basis for choosing program components to be evaluated and standard methods for acquiring process data (but see a recent attempt by Scheirer, 1996). The ultimate goal of a process evaluation is to be able to link outcomes with program implementation, coverage, and dosage. Data should enable assessment of (1) the extent to which program administrative components are implemented, (2) the coverage and types of services provided to clients, and (3) key components of difference from a particular counterfactual, such as the previous policy or other comparison groups.

In the classic randomized experiment, program elements to which individuals are assigned are generally well defined, individuals are assigned to treatment or control groups at a specific point in time, and their participation is monitored. An experimental design integrates process and outcome evaluation, since the delivery of program services and client selection is designed explicitly to meet evaluation needs. Causal inference is based on measured outcomes and known treatment differences between the randomly assigned



Table 1 here

Table 1 continued

groups. Because W-2 is mandated as a statewide program replacing AFDC, assigning individuals randomly to W-2 and to some other non-W-2 treatment will be programmatically difficult, politically unfeasible, and probably not of immediate relevance to policy makers seeking advice on effective W-2 implementation. Using an experimental design to evaluate the effect of variations in W-2 program components is feasible, and could provide extremely valuable information on differential program effects (e.g., by randomly assigning participants to different child care copayment groups).

The cross-state comparison proposes one or a set of states as the counterfactual. This design seeks to attribute differences in an outcome measure to differences between Wisconsin's W-2 and the counterfactual states' welfare programs. To attribute the cause of differences in outcomes requires information that will distinguish the programmatic differences. This design demands that, in addition to the issues discussed in the Courtney and Piliavin paper, cross-state agency data must be obtained to describe and compare substantive program elements.

Robert Haveman argues for a pre-post (before-after) evaluation design, and Thomas Kaplan and Daniel Meyer describe one possible pre-post evaluation plan. A pre-post evaluation would measure W-2 program effects as differences in the value of selected outcome variables between a period defined as pre-W-2 and another defined as post-W-2. The power of this design (the probability of finding statistically significant effects if they exist) depends on selecting the pre- and post-comparison periods such that they are truly representative of the intended counterfactual "pre-program" years and of the "post-program" years during which W-2 is sufficiently operational to have the hypothesized effect on targeted populations and institutions. This design requires, perhaps more than others, a great deal of information on the implementation process. To the degree the design takes account of the process by looking at outcomes over several periods, the greater is the demand for program process data.

Defining pre- and post-program years for a W-2 evaluation is not straightforward. Though officially effective on a legislatively specified date, W-2 cannot be characterized as a discrete point-in-time shift from AFDC to full W-2 implementation. Preliminary demonstration programs have already altered the "pre-W-2" climate in some counties. In all counties, the development of a mature W-2 system will be gradual, some counties contracting with private agencies inexperienced in administering programs for low-income clients, and others with public agencies required to incorporate new W-2 responsibilities into their traditional operations. In this evaluation design, and in the others as well, failure to identify counties that either do not or more slowly deliver necessary services to some eligible families (either because they were not available or not offered, or

because families were uncertain about program eligibility during these early months) may lead to misleading inferences about program impact akin to those cited earlier in Head Start.

#### Data for Management Purposes

At the simplest level a process evaluation can provide useful descriptive data on W-2 implementation across selected sites. As a mechanism for providing feedback to program managers, this descriptive information should be acquired early in the program, with periodic data collection efforts as the program matures. Gathering comparable program data across service delivery areas can help answer questions about effects of administrative variation. Programs may change in response to feedback from these initial collection efforts, a process which itself must be measured in order to understand subsequent program outcomes. An early, fairly intensive effort to collect process data in a few program sites would inform program administrators about the management of this innovative program and provide insight into the type and means of collecting process data.

#### Data for Cross-Agency Comparisons

Gathering data on individual W-2 service agencies can provide information with which to measure strategies adopted to provide services to target populations, to compare services for consistency with program intent, to understand participant flow through and termination from the program, to understand the process by which participants are matched with appropriate services, to understand the shifts by participants among program services and components, and to chart differential selection into, persistence in, and attrition from program services.

Answering questions across service delivery areas is, we assert, the most valuable purpose of outcome and process evaluation of W-2. Even in the long run, W-2 service agencies are likely to adopt different programmatic arrangements and impose sanctions differently, providing the opportunity to test, as was done for the GAIN program, how these differences are associated with program outcomes. These comparisons are the essential comparison for welfare policy-making in a world in which federal and state legislation has eliminated the former AFDC regime as an alternative policy option.

Process data can also illuminate unintended program consequences. Wisconsin Works is a time-limited transition-to-work program. Although the threat of loss of program services may better facilitate a client's entry into the labor force than time-unlimited programs, several unintended consequences may occur, as illustrated by the following questions.

Do children receive poor-quality child care? The supply of good-quality child care providers is very limited. The program incentive for children to enroll in the least expensive care available could place many children at significant risk for developmental problems. Thus, the quality of care that children receive under Wisconsin Works should be extensively investigated.

Does the rapid transition into full-time work discourage parents from enrolling their children in compensatory programs that have proven effectiveness (e.g., Head Start, special education interventions)? Several of these programs, for example, are part-day programs.

Does program participation affect the amount and quality of time parents spend with their children? Does participation affect the amount and quality of parent participation in school?

Will employers provide sufficient training for program participants even if the amount of training necessary for an individual employee goes beyond the tax incentives provided to the employer?

A possibility to consider in investigating unintended consequences of W-2 is that participants will likely have different rates of compliance. Program success, for example, may depend on an individual's psychological readiness for transition activities and how responsive case workers are to the needs of participants. Level of motivation to participate in the program will probably vary substantially, and this may affect program outcomes. These attributes could be measured in survey questionnaires.

### Evaluability Assessment

One approach to evaluating the implementation of Wisconsin Works would be to conduct an evaluability assessment, developed by Wholey (1987) as a pre-evaluation effort designed to clarify program intent, the stages and feasibility of implementation, and the likelihood of improving program performance from the point of view of policy makers and interest groups. Although such assessment can be used during any phase of program development, it is especially appropriate at the beginning of new programs, to identify key attributes and the services that managers are both most interested in evaluating as well as those that will be provided by the program. Another way to view evaluability assessment is as a delineation of key factors of program implementation that may determine program effectiveness.

Some of these key factors and resources will, in part, determine the effects of Wisconsin Works, and are consistent with factors described in background papers for this conference as important to understanding program impacts. Three key points should be considered. First, the social context in which the program and resources are administered will likely affect both implementation and program effectiveness. This argues for site-by-site investigations of program effectiveness. Of course, social context factors may operate differently

in each county and may vary by service provider. Second, the identification and coordination of services provided by resource specialists and FEPs are critical to successful program participation. This includes the amount and quality of involvement by the FEP as well as the timing of job placement (e.g., length of service until first job placement). Finally, a focus on the operative resources that will affect short- and long-term outcomes are defined through employment training, child care, health care, and other community supports. Although the extent to which the target population is appropriately referred to services is a key evaluation question, the availability and quality of services and resources will directly influence the effects of program participation.

### Data Sources

Although administrative data are likely to be a main source of data in any process evaluation design, survey data will be essential to understanding the perceived and actual roles of administrative staff over time and as the program develops. In order to monitor the role of these staff, annual or semi-annual survey questionnaires could be administered.

To measure components of the process, instruments such as the Service Delivery System Questionnaire (from the Center for Health Policy and Program Evaluation, University of Wisconsin–Madison) could be utilized by program staff during the implementation of Wisconsin Works. The attached page from the questionnaire displays several dimensions that may significantly affect the learning experiences of program participants. As indicated, community services such as adult education, health care, and child care may differ across sites in their availability, coverage, and quality. Since program staff (e.g., FEPs) may be familiar with and will invariably refer participants to these and other services, their ratings may help to reveal the match between participant needs and community resources. Various indicators of interorganization coordination may affect program implementation across communities. Among the items are the extent to which “agencies share information and resources,” “interagency meetings occur,” and “children and families face barriers to obtaining services.” Again, such ratings help to define the level of coordination among service agencies that could be utilized in a process evaluation of welfare reform.

The ideal evaluation data system would allow linkage of individual client data, required by the outcome evaluation, to process data that summarize the character and services of the W-2 agency. This link would enable identification not only of the timing of program entry by each client, a standard item in any outcome evaluation, but of the administrative environment in which each client was served. These data would be matched through unique identifiers for the W-2 site responsible for coordinating the program services for the

W-2 participant at the particular time. Some process data may come from administrative records. However, observation of selected service sites and administrator surveys must be used to gather other data. Program administrators, adjusting initially to a completely new welfare system, may be reluctant to fully cooperate with a survey effort outside the normal demands of their administrative duties. For this reason the

Process Evaluation

154

Survey Questionnaire here



outcome and process evaluations should be designed to depend heavily on administrative data (as is proposed by Meyer and Kaplan in their paper) with only periodic and focused site surveys. For this reason, careful thought must be given to process evaluation needs in designing intake and client flow forms. Administrative cooperation in gathering these data can be increased by including evaluation plans (e.g., need for this effort, feedback to administrative staff) in staff training, and by convincing staff of the usefulness to them of these efforts. Contracts written with W-2 service delivery agencies could require periodic reporting of specific site and community resource characteristics.

W-2 administrative data from each linked site (see Michael Wiseman's paper on aggregate client flow data) and site surveys would provide information on administrative structure and services in the aggregate. Selecting a cohort of W-2 participants every 6–12 months for an outcomes study would provide information on how new cohorts moved through the system and, linked to additional program and periodically updated site-provided data, would allow a comparison of the effects of program implementation and service systems on clients over time.

We should note that, if, as suggested by John Witte, an evaluation of W-2 (or at least of its education effects) takes place in a limited number of communities, rather than on a statewide basis, so should the process evaluation. A statewide outcome evaluation must be accompanied by statewide process data, and process evaluations for selected-site outcome evaluations should take advantage of this more intensive data collection design. Some mix of methods may be valuable, with a few intensive case studies conducted early in the program in order to discover issues that may be particularly important to highlight in broader survey efforts. In either case, however, the explicit linkage process and outcome data need to be pursued.

## CONCLUSIONS

The consensus in these papers appears to center on the greater feasibility of a pre-post program impact evaluation design, assuming that the primary question to be addressed concerns changes in outcomes under W-2 as compared to the AFDC era. This design requires correctly identifying the period before W-2 went into effect and a period after it took effect. Pre-post designs can most accurately link outcomes to programs when programs are quickly implemented. When implementation is gradual, some studies eliminate the period of implementation, but even then it is necessary to be able to clearly identify the without-program and with-program period, and the time in between may be long.

Service delivery across W-2 agencies as contracted by the counties is likely to vary, both temporarily during implementation, but also in the long run. Public and private contract agencies may operate quite differently and mobilize resources differently. Clients may flow through these agencies at different rates. Data on basic agency characteristics, service provider characteristics, and agency practices will allow an assessment of how the different impacts of W-2 across agencies may be due to different types of agencies and their practices. It is this type of cross-county comparison that will be most valuable for management purposes, for policy makers in proposing legislative and regulatory changes, and for analysts trying to understand the socioeconomic impact of the complex W-2 structure.

Effects of cross-program variations in W-2 can be assessed with the alternative outcome evaluation designs. An experimental design can be used to assess the effects of different program provisions within W-2. Cross-county comparisons can examine effects of different organization of services by different agencies. Pre-post design can capture how changing the W-2 program in some substantive way alters program effects. All these, however, require good process data to causally relate outcome to program input.

A process evaluation must begin early in the program implementation period, especially if a pre-post program design is adopted for an outcome evaluation. A process evaluation of W-2 would be one of the first such efforts for a large-scale welfare and employment program. Beginning now would increase the chances that it would be done right and that it would provide the feedback to administrators that is necessary for early program modification and later explanations of program outcomes.

### References

- Blalock, A. B., ed. 1990. Evaluating Social Programs at the State and Local Level. Kalamazoo, Mich.: W. E. Upjohn Institute.
- Cicirelli, V. G., et al. 1969. The Impact of Head Start. Athens, Ohio; Westinghouse Learning Corporation and Ohio University.
- Cook, T. D., and Shadish, W. R. 1986. "Program Evaluation: The Worldly Science." Annual Review of Psychology, 37: 193–232.
- Doernberger, C., and Zigler, E. 1993. "America's Title I/Chapter I Programs: Why the Promise Has Not Been Met." In E. Zigler and S. J. Styfco, eds., Head Start and Beyond. New Haven, Conn.: Yale University Press.
- Ellickson, Phyllis, and R. M. Bell. 1990. "Drug Prevention in Junior High: A Multi-Site Longitudinal Test." Science 247: 1299–1305.
- Land, K. C., McCall, P. L., and Williams, J. R. 1990. "Something That Works in Juvenile Justice: An Evaluation of the North Carolina Court Counselors' Intensive Protective Supervision Randomized Experimental Project, 1987–1989." Evaluation Review, 14(6).
- Lipsey, M. W., et al. 1985. "Evaluation: The State of the Art and the Sorry State of the Science." New Directions for Program Evaluation, 27: 7–28.
- McGraw, S. A. 1996. "Using Process Data to Explain Outcomes." Evaluation Review, 20(3): 291–312.
- Mead, L. M. 1996. "Welfare Policy: The Administrative Frontier." Journal of Policy Analysis and Management, 15(4).
- Pentz, M. A., et al. 1989. "Multicommunity Trial for Primary Prevention of Adolescent Drug Abuse." Journal of the American Medical Association, 261(22).
- Riccio, James A., and Friedlander, D. 1996. "Understanding Best Practices of Operating Welfare-to-Work Programs." Evaluation Review, 20(1): 3–28.
- Riccio James A., and Friedlander, D. 1992. GAIN: Program Strategies, Participation, Patterns and First-Year Impacts in Six Counties. New York: Manpower Demonstration Research Corporation.
- Riggin, L. J. C. 1990. "Linking Program Theory and Social Science Theory." In L. Bickman, ed., Advances in Program Theory. New Directions in Program Evaluation, No. 47. San Francisco: Jossey-Bass.
- Rossi, P. H., and Freeman, H. E. 1993. Evaluation: A Systematic Approach. 5th edition. Newbury Park, Calif.: Sage Publications.

- Scheirer, M. A., ed. 1996. A User's Guide to Program Templates: A New Tool for Evaluating Program Content. New Directions for Evaluations, No. 72. Newbury Park, Calif.: Sage Publications.
- Scriven, M. 1991. Evaluation Thesaurus. 4th edition. Newbury Park, Calif.: Sage Publications.
- Sherman, L. W., et al. 1992. "Crime, Punishment, and Stake in Conformity: Legal and Informal Control of Domestic Violence." American Sociological Review, 57 (October): 680-690.
- Sherman, L. W., et al. 1984. "The Specific Deterrent Effects of Arrest for Domestic Assault." American Sociological Review, 49 (April): 261-272.
- Weiss, C. H. 1987. "Evaluating Social Programs: What Have We Learned?" Society (November/December), pp. 40-45.
- Wholey, J. S. 1987. "Evaluability Assessment: Developing Program Theory." New Directions for Program Evaluation, 33 (Spring): 77-92.
- Wu, P. 1991. "Structural Equation Models in the Analysis of Data from a Nonequivalent Group Design: A Reanalysis of the Westinghouse Head Start Evaluation." Department of Social Relations, Lehigh University.

**Table 1: Process Components of Outcomes Identified by Other Paper Authors**

Author	Outcome	Process Component
Jesien	Effect on families and services provided (to families with disabled children) Out of home placements Hospitalization Immunization rates Child abuse and neglect Disabled children in child care settings Waiting lists for established providers	Manner in which program implemented & degree of flexibility given to parents in work & hours; Changing interaction of parents with Birth to Three programs. Availability of parents to attend to needs Increased demand on community services.
Sandefur	Effect On family structure Out-wed. births Maternal & child health Marriage & divorce Involvement of noncustodial parents in lives of children.	Delivery of health care to low income families Treatment of 2 v 1-parent families Requirements imposed on parents of teen-age parents. Services provided to and services provided to noncustodial parents.
Folk	Impact of W-2 subsidies, regulations, and demand on child care market. Demand may exceed supply in the short run Relative supply for licensed care. Impact of services in enabling W-2 parents to work. Impact of child care on child development Quality of child care may be lower with consequences for child development	Relationship between child care services and parents' employment. Timely availability of child care services. Location of child care services Administrative delays & bottlenecks Availability of weekend & eve care Stability and quality Search for and satisfaction with child care services. Determination of eligibility for child care. Questions: child care, cc changes, reasons for cc changes, staff/child ratios, warmth of caregiver's relationships to child.
Witte	Effect on education attainment of children Effect on schools and school districts Differential across school districts Segregation effects	Affect of W-2 work requirements on value of education Magnitude of adjustment in total and share of resources (time and money) devoted to education Change in parental expectation Change in parents' supervision of child Relationship to mobility
Courtney	On Children and families; On Welfare services Child maltreatment Out of home placements	Cut off in benefits if families do not meet work requirements or unable to obtain adequate child care. Changes in kinship regulation and reimbursement. Reduced supervision of children. Must pay close attention to children welfare system

Author	Outcome	Process Component
Cancian and Wolfe	Impact on AFDC eligible population, on low income population and on the general community. income sources employment and duration Impact of child care on child well-being school performance health status delinquency Availability and quality of child care Family formation	Redistribution of types of care used. Availability of child care for special needs. Supply of child care providers Time spent in child care Success of W-2 agencies in coordinating child care providers. Supply of health providers
Cain	Experimental design Cross-state comparison	Other states' programs offer "natural experiment" as alternatives, in whole or in part. Long-term evaluations necessary to avoid short-run biases (e.g., creaming in early years),
Piliavin and Courtney	Cross-state comparison	Design implies program implemented Design assumes ability of order programs in some important dimension (e.g., financial generosity, work requirements). Identification of program elements that warrant study.
Haveman Views of W-2 agencies of child welfare agency role in removing children from parents who have poor prospects for work.	Economic well-being and economic activities (e.g., work effort, family structure, health and nutrition, care and nurturing of children) of families	Evaluators must determine how to measure effects of a policy change process as opposed to discrete policy change.
Wiseman	MIS requirements of an "active" welfare system which is history oriented and emphasizes case management.	Information systems must include history at level of case manager, site manager, state or federal agencies Services treated as different tiers Status and progression through tiers Client distributed across process
All authors  Participants & kids  Characteristics of child care services	Implementation/data system delays Concurrent changes in other systems, leading to potential support of null hypothesis Changes in program over time Conceptualization of program elements.	Linking child care and W-2 records (Folk) Linking OOHC and W-2 (Courtney) Policy will be implemented is slow and uneven (Haveman) Welfare reform simultaneous with changes in educational system (Witte) Outcomes change as program responds to problems (Cancian & Wolfe) Components already underway, new system not in steady-state mode for some time (Cain). Program changes require multiple entry cohort comparisons (Courtney and Piliavin)



**SECTION 3**  
**Perception of System Interorganizational Characteristics**

16. Listed below are statements that may apply to services for children and families within this school/school district's geographic area. Please select the response that best describes the situation.

In this geographic area:	Always 4	Usually 3	Sometimes 2	Rarely 1	Never 0	Don't Know 8
Children and families receive adequate services as needed.						
Programs reach all children and families who need them.						
Services are easily accessible to families						
Services for children and families are well-coordinated.						
Agencies share information and resources.						
Referrals are shared between agencies.						
Interagency meetings occur.						
There is joint planning across agencies.						
There is problem-solving across agencies to fill gaps in services.						
Agencies make children and families a very high priority in terms of providing services.						
Agencies are often in conflict with one another.						
Agencies have conflicting rules and eligibility requirements.						
Duplication of services is a problem.						
Children and families face barriers to obtaining services.						



## **A Management Information System for Wisconsin Works**

Michael Wiseman<sup>1</sup>

### INTRODUCTION

Wisconsin Works (W-2) is sufficiently radical to justify description as ending Aid to Families with Dependent Children (AFDC). But it would be a mistake to infer that the new system thereby circumvents all problems that have plagued public assistance system management in the past, or that experience gained in implementing Wisconsin Works will not be of interest or benefit to states that adopt different reform strategies. Many of the building blocks of W-2—for example child care, health care, case management, and wage subsidies—will be used in one form or another in construction efforts elsewhere. And experience has already taught us a great deal about these components that should be useful to reformers in Wisconsin and other states.

This paper considers an important building block for program administration: the management information system, or MIS. The MIS for a public assistance program encompasses procedures for collection, storage, and retrieval of information essential for operating and improving the program. Management information systems are themselves constructed from components, and it is easy, in discussing them, to become quickly entangled in detail. In this discussion I skirt detail to focus on the key information needs of active intervention programs like W-2 and on certain strategic problems that must be addressed in designing and implementing management information systems to support all such programs.

A caveat is in order: The Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996 includes extensive requirements for public assistance utilization data. These requirements exceed the capacity of all known public assistance management information systems. Thus states are presented, broadly speaking, with two options. One is to concentrate first on planning for their own needs and then turn to the necessary additional instruments to meet federal requirements. The alternative is to begin by designing a system to meet federal requirements and then working backward to the system needed at the ground level to produce what Washington requests. A case for adopting the start-at-home strategy is made in the conclusion.

---

<sup>1</sup>Michael Wiseman is a Visiting Scholar at the Russell Sage Foundation for the 1996–97 academic year and is Vice Chair, Wisconsin Works Management and Evaluation Project Steering Committee, Wisconsin Department of Workforce Development. I have benefited from discussions with members of the Steering Committee of the Wisconsin Works Management and Evaluation Project, Paul Saeman of the Wisconsin Department of Workforce Development, and Tom Corbett of IRP. Opinions expressed here are those of the author and do not represent the opinion or policy of the Department of Workforce Development or the W-2 MEP Steering Committee.

The paper is organized as follows. I begin with a discussion of the relationship between the design of information systems and the ideology or mission of public assistance policy. I then summarize the W-2 program and model it in terms of a hierarchy of responsible entities, or agents. Each level has particular information needs, which I describe in general terms. I next contrast the information requirements of the W-2 system with those required for operation of AFDC and JOBS, and I point out five major problems for the MIS engineers. I discuss briefly the PRWORA mandate and how to meet it. I close by posing some important research questions that fall in the domain of public management.

I assume in writing this paper that the state is committed to delivering, for those who are in need and for the citizens who will pay, the employment-based system of public assistance that W-2 promises. Such a commitment creates a need to know the things that a good MIS will provide. Every seasoned administrator has learned from experience that simply needing numbers does not assure that they will be available; demand is not a sufficient condition for assuring supply. Nevertheless, wanting the numbers is surely a necessary condition for getting them.<sup>2</sup>

#### MANAGEMENT INFORMATION AND PUBLIC ASSISTANCE IDEOLOGY

The term “public assistance” refers here to government-sponsored programs of aid for people in need. Such programs require methods of determining need and calibrating and delivering aid. How all three things are accomplished is very much influenced by the program’s ideology and mission. It is important to begin with the ideology or mission of public assistance, because the character of the information required for program management follows from it. When ideology and mission change, so should the MIS. Many management problems arise from disjunction between mission requirements and information available.

The evolution of public assistance ideology can be thought of in terms of two models: “passive” public assistance and “active” public assistance. By *passive* public assistance I refer to systems in which need is a circumstance that aid in and of itself alleviates. (Think, for example, of “hunger” as a “circumstance” and “food” as the obvious prescription for relief.) By *active* public assistance I refer to systems in which need is

---

<sup>2</sup>In this regard Governor Tommy Thompson’s statement to the heads of various foundations is encouraging: “There are many issues concerning [W-2] design, management, and potential effects that can only be resolved as we do the job,” he wrote. “With these issues in mind, I have encouraged those working on the W-2 project to consider evaluation, feedback, and improvement as essential elements of W-2 strategy” (letter to Deborah Leff, Joyce Foundation President, August 21, 1996).

treated as a situation that aid alters. The difference between “circumstance” (used in the definition of a passive system) and “situation” (used in regard to an active one) is narrow but significant. Circumstance is generally cast in terms of surroundings, the external factors that lead to a person’s condition. Situation, in contrast, is more action oriented. The nuance is apparent in application: I have never heard anyone refer to the *circumstances* of an army, or of a team, or of an athlete, only the *situation*. One doesn’t get into a circumstance.

The two models translate into different styles of delivering help. The approaches to determining need and calibrating and delivering aid are presented in Table 1. The idea here is that passive assistance systems respond to circumstances with alleviation. If circumstances can be readily assessed, program operation is a matter of delivering the goods (or the checks). “Entitlement” means that it is circumstances in such systems that qualify a person, or a family, as eligible. (It is possible to operate a passive system without delivering benefits to all who are qualified, so entitlement does not necessarily guarantee receipt, in the sense of establishing a legally enforceable claim to aid. What’s important to this story is that it’s not response that counts.) In contrast, active systems attempt to alter the situation of recipients and imply some movement or action by those seeking aid. The linguistic nuance in this case is that response *warrants* assistance but need not create entitlement.

The differences between passive and active systems translate into differences in required management information systems. A passive public assistance system emphasizes current transactions. Most of what needs to be recorded occurs in a point in time, although the “point” may be extended to cover an accounting period. Thus in AFDC as we knew it, eligibility and treatment were conducted on a monthly basis, and any month’s assessment was in principle independent of what had gone on before. Active systems are history oriented and emphasize case management: situations are diagnosed, treatments are prescribed, and outcomes, that is, changes in situation, are observed in real time. What happens next is very much a function of what came before. The implication is that management information must include transactions history and provide access not only to the current action but also to those which preceded. These differences have concrete, or at least electromagnetic, implications. A passive, current-transaction system has no memory. At each point at which eligibility is reassessed or payments are redetermined, the necessary data fields may be overwritten to preserve storage. The file for each case can have a fixed format. In contrast, an action-oriented MIS must store history. Since one life differs from another, the dimensions of information associated with each case will vary. The file will be event oriented, and the more events that have occurred over the history of the case, the bigger the file will be.

<b>TABLE 1</b>		
<b>Passive versus Active Public Assistance</b>		
<b>Activity</b>	<b>Passive Model</b>	<b>Active Model</b>
Eligibility Determination		
Standard of Need	Circumstance	Situation
Reporting	Fill out form; interview confirms information	Fill out form; interview establishes action
Basis for Claim	Entitlement	Warrant
Treatment		
Conditional upon:	Circumstances	Response
Variation	Small	Large
Operation	Centralized	Decentralized
Objective	Alter circumstances	Alter situation

Researchers may be interested in event histories even if they are irrelevant to program operation. Event-history data are occasionally cobbled together from a series of point-in-time snapshots drawn from administrative records.<sup>3</sup> The fact that such costly strategies are required to study poverty dynamics is telling. Most existing management information systems are not history oriented, and as a result they are inadequate for the new generation of active policies. From another perspective, the character of existing systems testifies that their management has not required much process-related information. Management of the more active strategies in prospect may require data more consistent with the interests of researchers. But what do managers need?

#### MANAGEMENT INFORMATION FOR ACTIVE INTERVENTION

In addressing the issue of “what managers need” we encounter a problem of methodology. One approach is to investigate what managers do and to develop theories of the factors that work to produce this outcome. From this perspective, information obtained is an outcome: it is determined by management decision and affects the decisions management can or is willing to make. An alternative is to begin with what managers ideally would do, work backwards to the information needed to do this job, and then to ask why such information is or is not on the manager’s desk. Since the object of a conference on evaluation strategies for state welfare reforms is to develop practical models, it seems appropriate to adopt the latter strategy and to focus on what the good manager should be doing and what she or he needs to know to do it. Required actions are quite program specific, so I shall discuss the issues with reference to a particular program, Wisconsin Works. My approach would be the same for any other state.

I begin with a description.

#### Wisconsin Works: A Summary

W-2 is outlined in detail in various state publications (see, for example, Wisconsin Department of Health and Social Services, 1995; the program is also briefly summarized in Folk, 1996, and in Wiseman, 1996). W-2 is a strategy realized in a program. The strategy has five major features (Wisconsin Department

---

<sup>3</sup>See, for example, Cancian and Meyer (1995). The nation’s most ambitious effort at creation of longitudinal histories of public assistance receipt is going on in California under sponsorship of the California Department of Social Services (see Brady and Snow, 1996).

of Health and Social Services, 1995; page references that follow are to this source). First, the program is fundamentally work oriented. Virtually all cash assistance is linked to some form of employment. Second, the variety of situations and capabilities of persons seeking public assistance is addressed by an ambitious program of tracking and case management. Third, the plan attempts to reduce the connection between benefits and dependence by decoupling cash assistance from provision of health insurance and child care assistance. Fourth, the “end of welfare” is taken seriously: the Thompson Administration is committed to rapid and complete implementation. Finally, the critical importance of administration is dramatized by a shift in responsibility for its operation from what was the Department of Health and Social Services to the state’s employment service agency, the Department of Industry, Labor, and Human Relations, now reorganized as the Department of Workforce Development (DWD). DWD will, in turn, be responsible for competitive subcontracting with various public and private organizations for W-2 operation.

The W-2 program has four tiers of support for adults with children. Persons seeking assistance will first meet with a Financial and Employment Planner (FEP). It is the job of the FEP to help needy applicants “to think through their best options to provide for the economic security of their families”(p. 34). Those meeting eligibility standards will be directed to one of the tiers in the self-sufficiency ladder. Some applicants will move into unsubsidized employment—the highest ladder rung. Applicants initially unable to find unsubsidized employment will be either accommodated in “trial” subsidized jobs in private or public organizations or placed in community services jobs (CSJs). The bottom rung, W-2 Transitions, is “for those legitimately unable to perform independent self-sustaining work even in a community service job” (p. 9). The program is structured so that movement upward on the ladder raises income, and duration of tenure in each category save W-2 Transitions is strictly limited. The overall lifetime limit on cumulative participation in any W-2 activity is five years, but the plan includes some options for extension.

Persons in each tier are eligible for subsidized health and child care, with copayments dependent upon income. More generally, Wisconsin Works commits the state to provide child care and health care subsidies to working parents with children on the basis of income and assets only. Eligibility for some subsidies is to be extended to families with incomes as great as 165 percent of the federal poverty line. Child care costs are to be constrained by some relaxation of standards for caretaker eligibility; health care costs are to be constrained by use of managed care, benefit reductions, monthly premiums, and restrictions on the ability of persons previously participating in employer-paid health insurance plans to transfer to the W-2 health program. The state’s intention is that public assistance be cash earned through work or delivered as services needed to support work; for some families, assistance may involve only subsidization of health insurance and child care.

In sum, the program is dramatic in scope and ambition. It offers an unparalleled administrative challenge. Some building blocks in child care and health insurance are made more passive in orientation than the current system. For example, welfare as we knew it offered extended Medicaid access to families leaving public assistance because of employment. This access was therefore contingent on a history of welfare receipt. In contrast, access to health insurance under W-2 is strictly income based. Nevertheless, the core component of work-based cash assistance is explicitly structured around a service progression intended to lead in a fixed amount of time to unsubsidized employment.

### The W-2 Management Hierarchy

W-2 MIS requirements may be derived from a review of what the “agent,” or responsible entity, at each level of the W-2 organization hierarchy is expected to do. W-2 presents four levels of information demand: (1) The case manager, (2) the site manager, (3) the system manager or state agency, and (4) the federal agency.

The front line of an active system is the *Case Manager*, or, in the case of W-2, the “Financial and Employment Planner”.<sup>4</sup> The FEP is the person ultimately responsible for the allocation of persons to services and, implicitly, to program tiers. This diagnosis must be based on an assessment, which in turn will call for information on an individual’s situation. The assessment will be followed by prescription and an outcome.

The *Site Manager* is the agent responsible for allocating cases among FEPs, coordinating resources for FEP use, contracting with providing agencies, coordinating activities with community agencies, and reporting to the state. The site manager may play a role in disseminating information among FEPs. The emphasis placed in W-2 literature on the use of nongovernmental organizations for site management is an important feature of the program, although at this writing it appears that initially site managers in virtually all counties will be the same agencies that managed welfare operation under AFDC.

The *State Manager* or *State Agency* is the agent responsible for designating agencies for site operation, evaluating agency performance, and financial management. The state may play a role in disseminating information on methods. Under W-2, state oversight is exercised through the Department of Workforce Development.

---

<sup>4</sup>The “Business Flow” plan for W-2 divides the case management function into four jobs: Reception, Resource Specialist, Financial and Employment Planner, and Supportive Services Planner (Department of Workforce Development, 1996, page 4). A FEP can perform all these roles, and it is unnecessary to make the distinction for this paper.

The *Federal Agency* is in principle responsible for assuring that federal funds are used in a manner consistent with the enabling legislation. In addition, the federal government may play a role in disseminating information on methods. Under current law, responsibility for public assistance remains distributed across a number of agencies, but for this paper federal interests will be treated as if they were unified.

These four levels have been described in relation to Wisconsin Works. Certain factors make it likely that similar structures will be observed in other state programs as well. An active system must involve a casework function that includes direct interaction with participants and a serial process of assessment, prescription, and evaluation. Since recommendations (prescriptions) related to strategies for movement to self-support are conditioned by features of local labor markets and services available, the casework function must be managed locally. Resources for public assistance are generally collected from statewide sources and are used with state administrative and legislative oversight. This requires a state agency. And finally, over half of all public assistance costs are borne by federal taxpayers. Stewardship over these funds requires federal agency oversight as well.

#### INFORMATION REQUIREMENTS: THE CASE MANAGER

Now, consider the information required for each of these four jobs. I will pay the greatest attention to requirements at the front line, or case manager level. If the information is not collected here, nothing worthwhile will rise to the upper reaches of the administrative hierarchy. At each level I comment on order of importance—that is, which information elements are most important to sustaining a genuinely active system.

The discussion that follows emphasizes movement to and from categories that include cash payments, either as wages (in subsidized employment) or as grants conditioned on activities (community service jobs and W-2 Preparation). The W-2 program integrates these activities with statuses that involve only receipt of employment-related services such as child care and health insurance subsidies and a special loan program for employment-related expenses. A general MIS would probably create a separate status for persons in this group.

#### The Case File

At the front line, the case manager needs an information system that provides four reports, one of which is essential to individual case management and three of which provide summary information on the manager's entire case portfolio. The core report for the case manager is a transactions history for each case



for which he or she is responsible. This is the electronic version of the standard welfare case file. It is here that assessment results and program prescriptions are recorded. Given that in W-2 the “clocks” relate to the individual experience of adults, the case record must be *relational*, linking all transactions for a family to action history for the adults.

However supported, the case file required by W-2 should allow aggregation to four summary FEP portfolio reports. The first covers allocation by activity, the second case flows, and the third and fourth provide two measures of time in status.

### The Participant Activity (PAR) Report

Official literature pays considerable attention to the division of W-2 services into tiers. From a case management standpoint, however, it is probably better to treat the tiers as simply different classes of service. The W-2 operations plan envisions a progression of cases from intake upward through the hierarchy to unsubsidized placement. Thus, at least in abstract, cases could be tabulated by status and progression. The *Participant Activity Report* for a caseload will thus present cases cross-classified by status and stage.

A sample activity report is sketched in Figure 1. In the activity report the categories identified on the horizontal axis are steps in the system process. The first step is intake, and thereafter come various assignments. These assignments could be continued in the sequence assignment-transition-assignment indefinitely, but in practice the number of assignments is limited. For example, the earliest welfare-to-work programs sometimes included only two, administered job search followed by community work experience (Greenberg and Wiseman, 1992).

The vertical axis of the PAR identifies the various statuses created by the program plus a few additional classifications that are common to most welfare-to-work programs and that are likely to show up, sooner or later and in one form or another, in W-2 case counts. In order to avoid being too abstract, I have included categories for missing information, both with regard to status and point in process.

The first three rows in Figure 1 cover the three W-2 statuses that involve active oversight: Subsidized Placement, Community Service Employment, and W-2 Preparation. (Unsubsidized placement is an outcome; it is discussed further below.) When made operational, the chart would be full of numbers. The number in cell A tells how many people in this FEP’s case portfolio began the current month in a second W-2 assignment to a subsidized job placement. The number in cell B shows how many of the caseworker’s entire portfolio of clients began the month in the subsidized employment status. Some persons presumably begin the month after filing their initial application but before completion of the registration process and

Figure 1  
**Participant Activity Report**  
 (Point in Time; probably beginning of month)

Participant Status	Activity						Status Totals
	Intake	Assignment One	Assignment Two	Assignment Three	Assignment Four	Assignment unknown	
Subsidized Placement			A				B
Community Service Job							
W2 Preparation							
Hold							C
Intake, in process	D						
Exemption							
Status Unknown						E	F
Totals						G	H
<b>Totals By Point in Process</b>							

Note: Shaded blocks identify impossible participant status/ activity combinations. Letters identify cells discussed in the text. The transition/ assignment process need not be limited (as done here) to four episodes.

assignment to an activity; the number of such persons appears in cell D. The chart includes a transition category for persons who have passed through intake but are, at the beginning of the month, unassigned or between assignments; this is “Hold,” and C is the total number of persons in this group.

The PAR yields many useful numbers. Cell H is the total number of persons for which the case manager has responsibility. The ratio E/H, that is the proportion of these people for whom the case manager knows neither status or assignment, is surely an indicator of loss of control. The complement of the ratio F/H, that is, of unassigned participants to total participants, is a type of participation rate. The larger the share of the caseload that falls in the upper rows, that is, in the most work-ready categories, the greater the likely turnover in the subsequent month, since these people are actively involved in welfare-to-work activity.

Readers with experience in public assistance administration will at this point be begging for a reality check. The Participant Activity Report calls for a great deal of information that may be cumbersome to collect and time consuming to record. Looking at the chart performs a useful function, however, in that it assists in identifying minimum information requirements. It would be difficult to claim that cases were really being managed if the number in cell F, that is, cases for which status is unknown, is large or if nothing like the counts in the “Status Totals” column is available to the case manager. I draw an analogy to parenting. You can’t be a good parent if you don’t know where your 14-year-old is at 10 p.m. You can’t be a good case manager if you don’t know what your clients are doing at the beginning of the month.

The PAR also helps in thinking about an “information expansion path,” that is, the order of importance of numbers. The choice here is in part a matter of policy, and stakeholders will differ about what counts. My own assessment is that it is most important to know the number of cases in intake process (cell D), since processing intake is essential to delivering aid to people in need. Beyond this information, in order of importance, are (2) total case count (H), since knowing this signals an environment of well-defined case management responsibility, (3) count of cases for which status is unknown (see cell F), since this is a measure of oversight detail, and (4) count of cases for which information on stage is unknown (see cell G). In W-2 it may suffice simply to establish a low incidence of “status unknown,” for once we know current status, we know what is supposed to come next. Note that full information does not an active program make. Our MIS may record the number of persons in “Hold” with remarkable precision, but if this number is large relative to the total number of cases the worker is managing, the program fails a key test of “active” policy.

#### The Participant Transitions Report (PTR)

The Participant Activity Report just discussed covers states and must be defined at a point in time. The Participant Transitions Report covers an interval and summarizes movement of individuals through the W-2 case management progression. A prototype PTR appears in Figure 2. The rows of the chart identify case status at the beginning of the month. The columns identify case status at the end. Letters carried over from Figure 1 reflect the same numbers; for example, D is the number of cases involved in the intake process at the beginning of the month, B is the number in subsidized placement, and H the total beginning-of-month caseload. For convenience, new intakes during the month are included here as J; of this group, K are still unassigned at the end of the month. L, M, and N represent various possible dispositions of the B participants who began the month in subsidized placement.

Once again, we can identify various matters of interest. The sum  $T = P + Q + R$  is terminations, and  $T/(H+J)$  is the termination rate. Comparison of new intake (J) with terminations (T) reveals the relative importance of accessions and terminations as sources of caseload change. The details of the chart can be manipulated to reveal the extent to which individuals change status. As was true for the Participant Activity Report, the share of the caseload for which Beginning-of-Month (BOM) or End-of-Month (EOM) status is unknown is one indicator of the efficiency of program oversight.

Here, as in the PAR report, the appropriate information expansion path is in significant part a matter of program objective. Beyond the importance of getting BOM and EOM status straight, it seems likely that an activist policy would focus on the disposition of recipients upon intake and on the nature of terminations. Wisconsin's emphasis on diversion would show up in this table as a sizable incidence of closures for persons listed in the left-hand columns as either new intakes or intake in process. Success in placement would be identified by a high ratio R/T.

### The Time in Status Report (TSR)

Participants in activist transfer programs carry two clocks. The first covers elapsed time for which the participant has been receiving assistance. In W-2, this clock is set for each adult on first entry. The second clock records hours required for assistance-to-work activities. The time-in-status report covers the first clock; it summarizes the duration of participants in each status and in the program overall. Time in status is assessed at a point (presumably the same reference date as is used for the activity report). A sample TSR is presented in Figure 3.

The setup for this chart is complicated because time limits differ by status. For simplicity here they are all listed with the same format. The output includes distribution of participants by time in status plus

information on the number of participants at risk of reaching a within-status time limit. The time limit for each status (indexed by  $I$ ) is  $T^*(I)$ . Risk is signaled when duration—elapsed time in status, or  $ET(I)$ —crosses a trigger denoted  $t^*(I)$ . These “action triggers” look algebraically the same for most of the activities, but the

Figure 2  
**Participant Transitions Report**  
 (Monthly Case Flow)

Beginning of Month		End of Month (EOM) Participant Status									
(BOM) Participant Status	(BOM) Participant Count*	Voluntary Termination	Administrative Termination	Unsubsidized Placement	Intake, in process	Subsidized Placement	Community Service Job	W2 Preparation	Hold	Exemption	EOM Status Unknown
New Intake*	J				K						
Intake, in process	D										
Subsidized Placement	B	L	M			N					
Community Service Job											
W2 Preparations											
Hold	C										
Exemption											
BOM Status Unknown	F										
<b>Total, BOM</b>	H	P	Q	R							
		<b>Terminations</b>			<b>Total, EOM Participant Count by Status Category</b>						

\*New intake count for month (J) is not included in beginning-of-month participant totals. Letters identify cells discussed in the text.

Figure 3  
**Participant Time in Status Report**  
 (Beginning of Month)

Beginning of Month (BOM) Participant Status		Action Trigger, Status(i) [t*(i)]	Elapsed Time in Status (Months)											At risk [t > t*]
			<1	1	2	.....	t*(i)	.....	T* -2	T* -1	T*(i)	Ex-ten-sion	ET Un-known	
Intake, in process		ET > 1												
Subsidized Placement		ET > t*(i)												
Community Service Job		ET > t*(i)						U						
W2 Preparations		ET > t*(i)												
Hold		ET > t*(i)												
Exemption		ET > t*(i)												
Status Unknown		ET > t*(i)												
<b>All Clocked Activities</b>		ET >54 months						V					W	
<b>Total, BOM Participant Count by Status Category</b>														

Note: Shaded cells are not tabulated. Status categories are indexed by (i). T\*(i) is the participation time limit for status (i); t\*(i) is the elapsed duration that places participant in category (i) "at risk" of reaching time limit. Both the limits and action triggers may vary by status. The action trigger for intake is assumed to be one month. The action trigger for all activities is assumed to be 54 months, i.e. 6 months before reaching the overall time limit.

actual values,  $t^*$ , will vary. For example, the action trigger for intake could be a single month, for subsidized placement ten months, and for community service jobs four months.

In addition to the distribution of participants individually, by time in each status, the TSR should include a summary of the caseload by time in program overall. This is the “All Clocked Categories” item. The cell marked U shows the number of persons two months away from maximum duration in Community Service Jobs; the cell marked V reveals the number two months away from the overall program time limit. In the figure, it is arbitrarily assumed that the action trigger for all clocked activities is 54 months; this means that once a case begins the last six months of the five-year duration maximum, it is flagged.<sup>5</sup>

In addition to the information already cited as most important from the participant activity and transitions reports, the minimum requirement in the TSR is the bottom line—the overall clock. Second in order of importance might be the “At Risk” column. However, construction of the “At Risk” column requires the information in the remainder of the table, so it may be difficult to sustain any time awareness without maintaining the data needed for the entire chart.

#### The Hours in Activity Report (HAR)

The TSR records participant situations at a point in time. But many programs also devote attention to the number of hours of activity that each status involves. An objective of assistance-to-work programs is often to raise this activity level so that time spent in activities amounts to or approximates full-time work. In Wisconsin’s “Pay for Performance” initiative, the sanction applied for nonparticipation is proportional to the scheduled hours of activity that participants fail to achieve without just cause. Effective application of such a sanction requires that hours missed be deducted from full time or from hours scheduled, whichever is less. Once again, if this sort of sanction is to be applied, actual hours of participation must be recorded in some systematic fashion.

A sample hours report appears in Figure 4. The left-hand side of this chart is by now familiar. The report is established at a point in time (again assumed to be beginning of the month), but the chart refers to activity over some reference interval. This could be, for example, the preceding week, but for each person enumerated, the count must be limited to the days for which the client was actually eligible for assignment. This feature of the chart is retrospective, since W-2 attempts to make payments (and to impose sanctions) on the basis of previous activity.

---

<sup>5</sup>Both triggers and time limits could vary by individual. In Iowa’s Family Investment Plan, the end dates are joint decisions made after assessments by the recipients and their caseworkers.



Figure 4  
**Hours in Activity Report**  
 (Beginning of Month)

Beginning of Month (BOM) Partici- pant Status		(BOM) Partici- pant Count		Scheduled Daily Hours in Reference Time Period*									Sanction in Reference Time Period					
				0	1	2	3	4	5	6	7	8	Sche- duled, Hours Unknown	Not Appli- cable	Schedule Status Unknown	Sanctioned	Not Sanctioned	Not Appli- cable
Intake, in process																		
Subsi- dized Place- ment																		
Com- munity Service Job																		
W2 Pre- para- tions																		
Hold																		
Exemp- tion	AF																	
Status Unknown																		
<b>Totals</b>	H	X									Y	Z	AA	AB	AC	AD	AE	

Note: Reference time period precedes BOM. Calculations cover activities in reference time period regardless of status under which hours accumulated. Letters identify cells discussed in the text. "Not Applicable" category arises in cases (e.g. new openings) for which scheduling in preceding period not possible.

Whatever the reference time interval, persons can be scheduled for an average of from 0 to 8 hours per day. A zero-hour schedule might be applied to a person in W-2 Preparation and awaiting the initiation of some particular service. Subsidized placements, on the other hand, are generally intended to be full time. Persons who are still in intake are unlikely to have been eligible for scheduling during any prior reference period; most such persons would be counted under "Not Applicable." As usual, we acknowledge that the system may lose people in "Schedule Status Unknown."

The right-hand side of the HAR report summarizes the incidence of sanctions. In W-2, sanctions take the form of reduction of grants or earnings in proportion to the scheduled hours missed. I have cast this figure simply to count sanction occurrences and to avoid tabulating the amount. The logic of including scheduled hours and excluding an hourly breakdown of sanctions is that the incidence of sanctions rather than the absolute amount is of primary interest, whereas for activity levels the principal concern is with the quantity.

The order of importance here seems to be the (1) the number of persons for which scheduling was appropriate but did not occur:  $(Y+AA)/(H-Z)$ ; (2) the incidence of sanctions among persons eligible for scheduling  $AB/(H-AD)$ ; (3) mean scheduled hours by status; and (4) the index of ignorance, that is  $(Y+AA)/H$  and  $AE/H$ .

In sum, the focus of activist public assistance policy is case management, and management information begins at the ground level. In principle an activist assistance-to-work system requires case managers who keep track of participants. Keeping track involves to some extent (a) retaining history, (b) knowing current activity, (c) following changes, and (d) watching the clock. The number of things that might be recorded about the assistance process is potentially very large, and information is not costless to assemble. Therefore it is important to develop a clear "information expansion path" to assist in focusing effort on the information likely to be of greatest value in assessing performance.

## MOVING UP THE LADDER

The ideas developed for thinking about the ground level carry upward in the administrative hierarchy. For this discussion I emphasize the approach rather than the content. The question to be asked before developing the information system is "What does this person need?"

*Site Managers* are responsible for supervising caseworkers, contracting for services, managing funds, responding to the community, and relations with the *State Agency*. The key is knowing what is going on. For sure, part of this information comes from the fabled managerial activity of "walking around," but numbers

count as well. The place to begin is sitewide aggregations of the Activity, Case Flows, Time in Status, and Hours in Status Reports, at least in their minimalist versions. Possession of such data allows comparison of the achievements of individual case managers to agency norms. This comparison is meaningful at the agency level, however, only if participants are assigned to case managers at random. Otherwise, some differentiation is needed among participants by background, for example, distinguishing those without work experience from those who report it. The choice of categories for such differentiation is a management problem for which state assistance and coordination would be useful.

As well as comparing caseworkers' performance, the site manager must also monitor the performance of contractors or contributing agencies that provide training, community service employment, and other program components. The site manager has the best vantage point for assembling information on service providers and evaluating their productivity. This information must then be shared with line staff. The site manager is also responsible for budget. Each activity has an associated cost, so at the site level the various case reports are complemented with corresponding data on outlays. Finally, the site manager may be responsible for some review of post-program experience, including the duration of placements and the extent of recidivism.

Under W-2, the *State Manager* or *State Agency* is responsible for site oversight. In principle, the state agency is unconcerned about the performance of the contractors used by each site in operating W-2<sup>6</sup>. Achievement is its concern, but just as caseworkers' achievements must be "normalized" on the basis of participants' characteristics, so must comparisons of agency performance. Moreover, the site aggregations of the Participant Activity, Participant Transitions, Time in Status, and Hours in Status reports are by definition of interest to the state if they are of interest to the site manager. Here again, some normalization by participant characteristics is needed for meaningful intersite comparisons.

The *Federal Agency* is responsible for assuring that federal funds are used in a manner consistent with the enabling legislation. In addition to fiscal audits, the required reports will feature whatever performance measures are consistent with the legislative mandate. The federal government may also play a role in disseminating information on methods. In general, methods-related studies are not likely to involve data aggregated over multiple sites to the state or regional level, but rather to consist of evaluations of programs conducted in a collection of sites. The design of information systems for such evaluations is beyond the scope of this paper.

---

<sup>6</sup>Auditor concerns are another matter.

## CHALLENGES

Wisconsin's current management information system is named Client Assistance for Reemployment and Economic Support, or CARES. Like systems sold by contractors to most other states, CARES is primarily oriented to supporting transactions linked to check writing and federal reimbursement. CARES is inadequate to the task of operating W-2, and major effort is being devoted to modifying the system to support the new program. As is true for many other features of W-2, designing and implementing the W-2 management information system present substantial challenges. Among them are what I term the issues of peaceful transition, graceful degradation, policy feedback, planned flexibility, system linkage, and meeting federal mandates.

*Peaceful Transition.* As the state's struggle with implementing its current case management system demonstrates, an operation as complex as public assistance does not change from one system to another overnight. Rather, a path of adjustment must be planned and carried out. The logic of W-2 suggests that the transition should involve gradually closing the AFDC payments system while simultaneously expanding JOBS to include all recipients as "mandatory," reclassifying JOBS participants by W-2 status, and paying for JOBS participation. This is what the state's "Pay for Performance" initiative is about.

*Graceful Degradation.* This paper has emphasized the importance of setting priorities and establishing a data quality expansion path. Just as it is important to decide the path to be followed in enhancing the accumulation of and access to information, so too it is important to assure that all is not lost should it prove impossible to sustain the MIS at the levels intended. Over the long run, the capability for orderly retreat may be as important as the capacity for innovation.

*Policy Feedback.* Wisconsin's CARES case data system does not automatically produce data pertinent to policy analysis. As a result, developing information relevant to management decisions almost always takes major programming effort; staff time in the Bureau of Welfare Initiatives is consumed by data extraction and programming issues at the expense of analysis. There exists nothing like the Participant Activity, Transitions, Time in Status, or Hours in Activity reports for either case workers or site managers, even in high-profile demonstration sites such as the Work Not Welfare operations in Pierce and Fond du Lac Counties (Bohnen and Corbett, 1996). The immediate consequence for management is that these demonstrations have failed to provide as much information for W-2 planning as they might have done, if a more capable information system had been in place. These failings also diminish the utility and accessibility of the state's administrative data for scholarly research.

The reports portfolio developed in this paper is intended to provide templates for consideration in ongoing negotiations between the state and its information systems contractor. Since the reports are founded on an event history for individual participants, it should be possible to create a recipient-oriented analysis file as a system byproduct. Such information provides the basis for more detailed study of a variety of poverty-related issues.

*Planned Flexibility.* The state's problems with implementing the CARES system occurred in the context of operation of a well-established transfer system. Neither AFDC nor JOBS were changing significantly at the time; it was the operating system that was being modified. Nevertheless, the change proved disruptive and has yet to be fully accomplished.

In contrast, the W-2 MIS will be implemented simultaneously with a new program. Past experience makes it highly likely that W-2 will change, and change significantly, as a result of problems encountered or opportunities discovered in the context of operation. The MIS must be planned with the expectation that the underlying program will change. The system of reports described in this paper is deliberately generic, in order to allow for variation in the underlying scheme.

*System Linkage.* In the present system both Medicaid and certain types of child care assistance are closely linked to AFDC, and service utilization is recorded within CARES. This connection should end with W-2, when in principle access to child care, health insurance, and child support assistance will be disconnected from participation in direct employment assistance. In practice, the systems will have to be linked in order to meet other program needs. For example, the current plan calls for paying the copayment for health insurance for persons in W-2 Transitions and Community Service Jobs by direct deduction from grants. Analysis of many policy questions will require linking data both across the services network and to other systems, including social security and tax withholding systems.

## THE NEW FEDERAL MANDATE

The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 is schizophrenic in its treatment of states. On the one hand, the legislation establishes block grants for "temporary assistance for needy families" to "increase the flexibility of states." On the other, the law creates remarkably rigid requirements for information collection, requirements that cannot be met at the present time by the management information systems available in any state. Then, as something of a mid-text afterthought, Congress instructs the Secretary of the Department of Health and Human Services to report on the status of state data processing

systems and to find out “what would be required to establish a system capable . . . of tracking participants in public programs over time.”<sup>7</sup>

While PRWORA may increase flexibility, the object is to make temporary assistance for needy families conditional on work. The TANF work requirement is summarized in Table 2.

To make sure that states are meeting the requirements, PRWORA also requires quarterly reports. As indicated by Table 3, the quarterly reports cover the information needed to assess the participation rates cited in Table 2 as well as other data. One set of information must be drawn from the experience of closed cases. The reports must be delivered quickly, and penalties are specified for laggards. The information may be derived from “the use of scientifically acceptable sampling methods approved by the Secretary.”

How should states respond? History suggests respectful skepticism. The PRWORA reporting requirements are yet another example of a long congressional history of dealing with nagging concerns by requiring data that nominally address the problem. Little, if any, consideration seems to be given to analysis strategies or the consistency of data requirements with other administrative or legislative goals. The participation rate requirements imposed by the Family Support Act of 1988 are a case in point (Wiseman, 1991). The object of the law was to raise the incidence of participation in welfare-to-work activities under the JOBS program. The measure of participation adopted by Congress could not be calculated using data from any available management information system and was useless for any administrative purpose other than satisfying the congressional requirement. Nevertheless, many meetings were held, many teeth were gnashed, and a great deal of money was passed to data management firms before it became clear that everyone would report something called participation rates and all sides would agree to “don’t ask, don’t tell” as a *modus vivendi*.

That was welfare reform as we knew it. The point here is that one step in the direction of ending welfare reform as we knew it is to return to the ground level. Looking back at Table 3, we note that a major part of the information required by the new law could be derived from the case management system used to develop the four basic case manager reports and the site manager outcomes report. The implication is that if employment is the objective of the state’s own system, and if development of an effective management information system for support of that system is part of the state plan, federal requirements may, for the most part, take care of themselves. The additional information that is required can be added through “scientifically acceptable” surveys, if it is useful. If it is not, it might be more appropriate to seek changes at the federal level.

---

<sup>7</sup>Public Law 104-193, Title I, Section 106(a).

**TABLE 2**  
**Temporary Assistance for Needy Families: Required Levels of Involvement in “Work Activities”**

Fiscal Year	Two-Parent Families			Others		
	Participation Rate	Hours per Week in Work Activities	Hours per Week in Specific Activities	Participation Rate	Hours per week in Work Activities	Hours per week in Specific Activities
1997	75	35	30	25	20	20
1998	75	35	30	30	20	20
1999	90	35	30	35	25	20
2000	90	35	30	40	30	20
2001	90	35	30	45	30	20
2002	90	35	30	50	30	20

**Source:** Public Law 104-193, Section 407; format from Brady and Snow (1996).

**Note:**

**Rate of Participation:** Fraction of families of indicated type receiving TANF assistance with adult in “work activities” for the required number of hours.

**“Work Activities”** include:

- \*1. Unsubsidized employment.
- \*2. Subsidized private sector employment.
- \*3. Subsidized public sector employment.
- \*4. Work experience if sufficient private sector employment is not available.
- \*5. On-the-job training.
- \*6. Job search and job readiness assistance.
- \*7. Community service programs.
- \*8. Vocational educational training (for no more than 12 months).
- 9. Job skills training directly related to employment.
- 10. Education directly related to employment for those without a high school diploma or equivalency.
- 11. Satisfactory attendance at secondary school or its equivalent.
- \*12. The provision of child care services to an individual who is participating in a community service program.

**“Specific Work Activities”** are those work activities marked with an asterisk.





**Table 3**  
**Temporary Assistance for Needy Families: Required Quarterly Reports Data**

---

**For Active Cases**

Demographic information:

- County of residence of family
- Number of individuals in the family and relationship of each family member to youngest child in the family
- Ages of family members
- Marital status of the adults in the family
- Race of each adult and child in the family
- Educational status of each adult and child in the family
- Citizenship

Employment Status:

- Employment status and earnings of the employed adult in the family including information to calculate participation rates
- Participation of adults in education, subsidized private sector employment, unsubsidized employment, public sector employment, job search, job skills training or on-the-job training, vocational education
- Unearned income received by any member of the family

Social Program Use and Status:

- Type and amount of assistance received under TANF
- Whether family received subsidized housing, Medicaid, food stamps, subsidized child care, and amount received for latter two and number of months receiving assistance under each program
- Disability of child or adult

**For Closed Cases**

- Whether the family left the program
  - If the family left the program, whether the family left due to employment, marriage, five-year time limit, sanction, or state policy
- 

**Source:** Public Law 104-193, Section 411; format from Brady and Snow (1996).

## RESEARCH CONCERNS

Development of the W-2 MIS is the focus of an information systems working group within the Department of Workforce Development. Given the importance of the information system for evaluation, the W-2 Management and Evaluation Project has identified MIS development as a priority, and an effort is being made to include an MIS expert in the national advisory committee. Apart from the fundamental issues of system design, a number of problems related to MIS use need attention, since understanding the problems may help assure that the MIS provides appropriate information. Sample problems include: (1) calibrating the triggers, (2) building the incentives, and (c) managing intake.

*Calibrating the Triggers.* Costs vary substantially across the various W-2 service tiers. Current forecasts of W-2 costs are based on estimates of the characteristics of incoming clients and transition flows from tier to tier that have little empirical support. Once W-2 is under way, experience will accumulate, but to date the resources have not been available to permit study of the possible “triggers” for directing management concern. At some point, some number will be pleasing. At some point, some number will be alarming. What are the numbers? What are the alarms?

*Building the Incentives.* An information system is of no use if no incentives exist for data entry. The system outlined in this paper is dependent upon appropriate incentives for data collection at intake (see below) and at each transaction. The system developed here is based very much on what is believed to be the essential information required to do certain jobs, in particular the job of the FEP and the job of the site manager. Consideration needs to be given to the actual operations of these agents and the extent to which incentives are adequate to ensure that the information is properly recorded. A big step in this direction is to assure that the information is made available promptly to those who need it. Much of the data entered by eligibility technicians and case managers in the current system is never returned in usable form.

*Managing Intake.* A common theme in both state and federal welfare reform is the “end of entitlement.” At first consideration, it is not clear why entitlement should be an issue for MIS design. But the meaning of loss of entitlement is unclear. The state views it as a way to avoid legal entanglements; critics see it as a sure indicator that many in need will fall through the cracks.

Much of the concern about the meaning of loss of entitlement involves intake. Under AFDC as operated in Wisconsin, families in need had the right to receive benefits within 30 days of application, and failure to deliver gave grounds for legal action. This recourse diminished the importance of collecting administrative data on time-to-assistance, since the system was to a significant extent self-policing. Without

entitlement, the opportunity for legal recourse is diminished, and it is possible that contractors will be tempted to discourage entry, especially by problem cases. The state's commitment to timely aid would be confirmed by appropriate steps to measure the elapsed time of persons in intake for W-2 (see Figure 3). If such data are to be used to evaluate performance, they must be collected and included in the MIS.

Given the state's general interest in assuring access and initiating the accumulation of information, a case can be made for the separate operation of intake on site by persons who are not employees of the site contractor. Such people could be responsible for initiating the case record and for dealing with child care and health insurance services that are operated independently of the employment track. How might such a system work?

In sum, within every state, as well as across all states, the Management Information System developed for TANF is important as an indicator of the nature of the program being implemented, a source of information on operations and consequences for families, and an object for study. Design of the W-2 MIS is an excellent example of the intersection of concerns of public management, policy analysis, and social science research in the study of welfare reform. Errors in design and failures of implementation will complicate program management, impede policy analysis, and diminish the utility of the administrative data for social science research. Thus there exists a common interest in design and in creating the necessary incentives to assure both implementation and ongoing operation.

**References**

- Boehnen, Elisabeth, and Thomas Corbett. 1996. "Work-Not-Welfare: Time limits in Fond du Lac County, Wisconsin." *Focus* 18(1): 77-81.
- Brady, Henry E. and Barbara West Snow. 1996. *Data Systems and Statistical Requirements for the Personal Responsibility and Work Opportunity Act of 1996*. Berkeley, California: University of California Data Archive and Technical Assistance.
- Cancian, Maria, and Daniel R. Meyer. *A Profile of the AFDC Caseload in Wisconsin*. IRP Special Report no. 67, Madison, Wisconsin, 1995.
- Folk, Karen. 1996. "Welfare Reform Under Construction: Wisconsin Works (W-2)." *Focus*, 18(1): 55-57.
- Greenberg, David, and Michael Wiseman. 1992. "What Did the OBRA Demonstrations Do?" in Charles F. Manski and Irwin Garfinkel, eds., *Evaluating Welfare and Training Programs*. Cambridge, Massachusetts: Harvard University Press.
- Wisconsin Department of Health and Social Services. 1995. *Wisconsin Works*. Madison, Wis.: The Department.
- Wisconsin Department of Workforce Development. 1996. *W-2 Business Flow Development Project Final Report*. Madison, Wis.: The Department.
- Wiseman, Michael. 1991. "Research and Policy: A Symposium on the Family Support Act of 1988." *Journal of Policy Analysis and Management*, 10(4): 588-666.
- Wiseman, Michael. 1996. "State Strategies for Welfare Reform: The Wisconsin Story." *Journal of Policy Analysis and Management*, 15(4): 515-546.

## **Monitoring Income for Social and Economic Development**

Martin H. David

### UNDERSTANDING THE IMPACT OF 1996 WELFARE REFORMS

The Personal Responsibility and Work Opportunity Act Reconciliation of 1996 (PRWO) mandates an end to the social assistance provided by Aid to Families with Dependent Children (AFDC), which is regulated and partially financed by the federal government. In general, the act requires that persons in need of social assistance must work to receive benefits. The act also limits the total benefit that can be received by any individual over a lifetime.

Because each state must respond to this new statute, more than fifty institutional structures will characterize social assistance in the United States. States will probably follow the lead of Wisconsin and offer wage subsidies and public service employment in order to place individuals in need of assistance in work slots. Understanding the change in family well-being and labor market outcomes for individuals induced by this complex change in social policy requires assembly of dynamic data on families and individuals. Furthermore, understanding requires samples within each administering unit (probably as small as a labor market consisting of a small group of counties). Samples with that level of geographical detail will permit comparisons of social services and subsidized work opportunities across administrative units; some comparisons can also be made with the social welfare system prior to 1997.

These requirements for data are monumental. To some extent they are mandated by the PRWO. The act requires an entirely new administrative record-keeping system in which the lifetime use of Temporary Assistance for Needy Families (TANF) can be monitored among the U.S. population. An individual's eligibility for TANF benefits will depend on her or his own past use of the program and the benefits received by other family members at households in which she or he lives. This system is more extensive than the current systems for tracking unemployment insurance eligibility and social insurance eligibility in the U.S. population.

The massive institutional change triggered by the PRWO act poses the question: How can existing administrative record systems be enhanced to produce data required to understand the outcomes of new social welfare policies? A partial answer for the state of Wisconsin is offered below.<sup>1</sup>

---

<sup>1</sup>Assistance from staff of the Wisconsin Departments of Revenue and Workforce Development is greatly appreciated. Dr. James Eisner retrieved many elusive documents. The document reflects my views; errors are my responsibility.

## BACKGROUND

Wisconsin has a long history of anticipating problems of economic and social development by forward-looking administration and fact-finding. For three decades the Department of Revenue (DOR) has collected samples of tax returns for revenue estimation. The Department of Health and Family Services (DHFS, formerly the Department of Health and Social Services) pioneered in the creation of a computer system, CRN, now CARES, to integrate program information on income and services provided to families and individuals eligible for Medicaid and food stamps. That agency has also funded the Wisconsin family Health Survey, which collects data on use of medical services and health of the Wisconsin population. The Department of Workforce Development (DWD, formerly the Department of Labor, Industry, and Human Relations), in conjunction with the University of Wisconsin Center for Community Economic Development, pioneered the use of employer unemployment tax returns to generate information about local labor markets that can be used to spur development and redevelopment.

Change in technology and in the environment in which government services respond require new ways of informing policy makers and citizens about the health of our economy and the well-being of our citizens. A potential exists to combine past efforts of these three agencies into an administrative record database. Such a database would be dedicated to statistical use and would provide information that is needed for understanding the consequences of Wisconsin job development programs, programs supporting small businesses, and programs that reach out to bring potentially dependent adults into the world of work. The database would simultaneously increase understanding of the economic environment in which these programs operate.

## CONCEPTS

Five concepts are required to enhance the use of administrative records for monitoring income and setting policy for social and economic development: (a) longitudinal data that follow a representative sample of individuals and business entities; (b) refreshing samples that represent new residents of the state and new business enterprises; (c) program beneficiary samples that identify users of safety net and work-oriented programs; (d) dual-frame samples that integrate the approach to a, b, and c; and (e) relational database technology that makes it possible to combine and extract information from these samples over time.

These concepts will create information that is not now available to administrators, policy makers, and the public from any source. Good data for distributing Title I funds to school districts with high poverty rates are unavailable. Information that demonstrates the success of many beneficiaries of welfare is unavailable. Information that documents the high level of tax compliance and good tax administration in the state is meager and poorly distributed.

The mission of several agencies—DOR, DWD, and DHFS—can be enhanced by better use of administrative data. Policy makers, legislators, and citizens can be better informed about: how government activities in several agencies relate to each other, what government services cost, what groups are covered by benefits, and what groups are excluded from benefits.

Implementing these concepts builds on a foundation of administrative information systems that is already in place. For example, the DOR already samples individual tax returns every two years to assist revenue estimation. The DHFS maintains a database on its population of beneficiaries that integrates information on Medicaid, food stamps, SSI, AFDC, and some additional social services. The DWD already keeps an on-line system of job applicants and a real-time capability for employers to announce vacancies. Additional effort required to implement the five concepts can be judged from an understanding of present record systems. This discussion seeks to understand how well present administrative records count the Wisconsin population and how record systems currently in use relate to each other. It seeks to understand how ongoing administrative records can be linked over time. The discussion proceeds by addressing four questions: (1) How do existing administrative records of the DOR, DWD, and the DHFS relate to the Wisconsin population and each other? (2) How can DOR records on individuals be linked over time? (3) How can DHFS and DWD records be linked over time? (4) How can DOR records be used to understand entry and exit from the W-2 program?

## APPLYING CONCEPTS TO WISCONSIN

### The Population Counted by DOR and Records on the Receipt of Social Assistance

The population of Wisconsin is partially represented in the records of the DOR, which include persons who file tax returns and persons who do not. We also have information on persons who are beneficiaries of the safety-net programs, including W-2, who are represented in the administrative records of DWD and DHFS. Other persons, including some who are in need of assistance, do not benefit from state programs and will not

appear in state administrative records. The extent to which administrative records reflect the population of Wisconsin is displayed conceptually in Table 1. Only the group  $q$  is not reflected in administrative records. The proportion of the population covered by tax returns,  $\underline{x} + \underline{y}$ , is extremely high because Wisconsin has a refundable Homestead Tax Credit. That credit is paid to many whose total income is too small to entail tax liability and whose shelter expenses are relatively large. Filing returns by many who owe no tax is what makes a sample of tax returns useful in studying the state's population. Most of the nonfiling population will be beneficiaries of the safety net and will be identifiable because of their use of SSI, food stamps, Medicaid, W-2, WIC, or other means-tested programs. At any point in time  $q$  will be small.

Historically the DOR has sampled at least one-quarter of 1 percent of the forms relating to individual returns every two years. The forms are then matched to the corresponding federal income tax returns through the data exchange program of the Internal Revenue Service. Among other things, this match yields information on use of the earned income tax credit (EITC). The sample is stratified to yield information on aspects of tax-return filing that are important for estimating revenue and analyzing the consequences of proposed legislation. Therefore, returns with lower amounts of taxable income are sampled at lower rates than returns with more significance for the state's tax policy and revenues. This feature is essential to good revenue estimating.

### Longitudinal Data

Tax Returns. Panel data for tax returns were pioneered in the state of Wisconsin in the 1960s by Harold Groves, Roger Miller, and Martin David (David, Gates, and Miller, 1974). This effort established the feasibility of tax return panels and provides unique historical information about the Wisconsin population for the years 1947–1964.

Longitudinal data require two capabilities: each individual involved in a return must be followed; and individuals must be allowed to enter the sample universe. Exits must be accounted for. The first capability requires that individuals must have unique identifiers which persist despite changes in names, and that all the identifiers that are associated with a return, either as filers or as dependents, must be captured so that members of the filing unit can be followed.<sup>2</sup> These capabilities are assured in present returns by the requirements of the US Tax Reform Act of 1986. All taxpayers and their dependents are now assigned taxpayer identification numbers.

The second capability relates to the concept of refreshing the sample, discussed below.

---

<sup>2</sup>The IRS has demonstrated this capability in its Individual Tax return panel (see Czajka and Schirm, 1990; Sailer and West, 1996).



**Table 1****Tax Filing and Beneficiary Status for the Wisconsin Population in a Particular Year**

Tax return filers?	<u>Beneficiaries of safety net?</u>		Total
	Yes	No	
Yes	x	y	x + y
No	p	q	p + q
<b>Total</b>	<b>x + p</b>	<b>y + q</b>	<b>x + y + p + q</b>

Safety-Net Programs. CRN established unique identifiers for each case. CRN also identified eligible units that left the welfare rolls and returned at a later point in time. Recommendations by Robbin and Hedstrom (1981) to implement a panel and to maintain a dynamic database of the caseload were not adopted. The successor to CRN, CARES, still does not routinely maintain individual case histories.<sup>3</sup> The PRWO will require that Wisconsin agencies administering food stamps and W-2 create this capability.

### Refreshing Samples

The DOR biennial sample assures a representative selection of taxpayers at two-year intervals. Similarly, a monthly readout from the CARES database assures a representation of beneficiaries at any point in time. Neither the DOR nor CARES samples automatically generate histories of individuals. Those histories can be attained by panel samples based on fixed identifiers, such as the social security number. The histories can be readily summarized by updating tables that show “spells” of pertinent activity. For example, DWD will be interested in the number and length of spells of Temporary Assistance for Needy Families. The DOR will be interested in the spells of filing proprietor income (Schedule C).

When individuals are followed, e.g. when a panel of data is collected, care must be taken to represent the population at each point in time. Newcomers must be added to the panel sample. Disappearances must be reconciled with known deaths and persons leaving the state. The extent to which nonfilers are eligible nonparticipants in safety-net and work programs must also be ascertained.

Representativeness for two years is illustrated by Table 2, which ignores immigrants and emigrants. Most individuals fall into group a. Extremely few fall into group b. Work with Canadian income tax records indicates that nearly all of group b can be accounted for: persons who died could be found by matching the sample to vital records; most of the remaining living persons filed tax returns in one of the three years prior to year 1 (Rubin-Bleuer, 1996). The ability to identify the proportion of individuals who do not file in consecutive years (groups b and c) is valuable administrative tool for designing audit procedures.

In Wisconsin tax records, b is larger than in the United States, because some persons move out of state and are not obligated to file nonresident returns. Similarly, c is greater, because more people move into the state. To see this more clearly, a row and column are added for non-residents of Wisconsin who are not required to file Wisconsin returns in Table 3.

---

<sup>3</sup>Almost no additional cost is entailed in this historical record, but the capability must be designed into the management information system for safety-net programs.

**Table 2****Tax Filing by the U.S. Population Over Two Years**

Tax return filers, Year 1?	<u>Tax return filers, Year 2?</u>		Total
	Yes	No	
Yes	a	b	a + b
No	c	d	c + d
<b>Total</b>	a + c	b + d	a + b + c + d

**Table 3****Tax Filing by the Wisconsin Population, Distinguishing Missing Nonresidents, Over Two Years**

Tax return filers, Year 1?	<u>Tax return filers, Year 2?</u>		Nonresident, nonfiler	Total
	Yes	No, WI resident		
Yes	a	b	e	a + b + e
No, WI resident	c	d	f	c + d + f
Nonresident, nonfiler	g	h	i	g + h + i
<b>Total</b>	a + c + g	b + d + h	e + f + i	sum a-i

The population groups g, h, and i are not at risk for Wisconsin safety-net programs in year 1; the population groups e, f, and j are not at risk in year 2.

To assure that group c is represented in each successive year of a panel, new tax returns must be sampled and added to the returns of individuals being followed. These returns can be identified in administrative processing. Returns of individuals who did not file in the prior year need to be sampled. Some of these people will already be known, because they were claimed as dependents on returns filed previously by others.

### Program Beneficiary Samples

Relationship to the Population. The procedures that generate a panel of tax returns are pertinent to generate a panel of beneficiaries of safety-net and work-oriented programs. Two additional considerations are important: citizens of the state have a concern for the poor and homeless; safety-net and work-oriented programs may not reach their target populations. To visualize these problems, we adapt Table 1 to create Table 4.

The first row of the table identifies a group that is presumed, by an officially sanctioned criterion, to have unmet economic needs. Group II includes persons who are presumed to need assistance and who are not reached by safety-net and work-oriented programs. They cannot always be identified from tax return data, because they may not file a return. Probability samples of Wisconsin families must be used. Several sources are available: the “long-form” sample of the U.S. Decennial Census, the Survey of Income and Program Participation, and the Wisconsin Family Health Survey. None of these sources provides high-quality, continuous annual measures of the distribution of family income by county in Wisconsin. As we shall see below, that deficiency justifies the expansion of panel data collected from administrative records.

The Decennial Census is the only source of information that describes poverty families by county in Wisconsin.<sup>4</sup> (Results are not available until 3–4 years after the census day). Annual samples of families in Wisconsin either do not represent the state’s population, fail to provide a precise measure of poverty status, or fail to give estimates at the county level where many programs are administered. For example, the Survey of Income and Program Participation represents only the Milwaukee metropolitan area, where precise measures of poverty by month are obtained for small samples that include few poor persons. The Wisconsin Family Health Survey collects information on annual income from a single question that is less inclusive than

---

<sup>4</sup>Congress has mandated that the Census Bureau produce poverty estimates for small areas. Initial work indicates that it is extremely difficult to estimate poverty counts for counties (Seigel and Coder, 1996).

**Table 4**

**Family Poverty Status of Individuals and Their Beneficiary Status in Some Period  
(Month or Year)**

Family poor?	<u>Beneficiaries of safety net?</u>		Total
	Yes	No	
Yes	I	II	I + II
No	III	IV	III + IV
<b>Total</b>	I + III	II + IV	Sum I-IV

income measured in the Survey of Income and Program Participation. The Health Survey gives estimates of the number of households below the poverty line, above twice the poverty line, or in between, for five Wisconsin regions.

Value of a Tax Return Panel. Some of the poor will file tax returns to receive earned income tax credits or homestead credits. They can be located in the continuing panel of tax returns. The number of poor who do not apply for these credits can be estimated by forecasts from the panel data for prior years. The panel provides historical information from which the number of persons moving into and out of poverty in successive years can be estimated. The number of poor migrating into or out of the state can also be estimated. Once a decade, these estimates can be checked against county estimates of the poor from the census. Once a year, these estimates can be checked against other survey data for Milwaukee or the regions of the state.

Preparing estimates of the poor requires that the tax return panel be organized into family units. Until 1993, combining returns into families and households could not be done precisely (Cilke and Wyscarver, 1990). Multiple tax returns from a given household (including students away at college, spouses filing separate returns) can now be aggregated to approximate a family. Sailer and Weber (1996) have demonstrated that this aggregation comes sufficiently close to the Census Bureau's counts of families to indicate that the difference may result from the failure of the bureau to identify cohabiting adults, as distinctly from legally married persons. The ability to aggregate tax returns to an approximation of families enhances use of tax data for understanding the well-being of the entire Wisconsin population. It creates the foundation for forecasts of family income distribution in the state.

Beneficiary Samples. Mandates for efficient administration and measures of program performance make it likely that safety-net programs will continue to keep data on their beneficiaries. The PRWO of 1996 mandates that food stamp and welfare assistance records be sufficient to limit lifetime receipt of benefits to five years. Following recipients of SSI, food stamps, and W-2 in Wisconsin entails tracking about half a million persons, some 10 percent of the State population. The continuing entry of persons into these programs and the exit of others implies that a ten-year history of beneficiaries could include three and a half times as many persons (see Duncan, 1984; Table 2 here).

Apart from the mandate of the PRWO, a history of past program use is important to understand the success of programs as measured by time elapsed since the last program benefit was received. History is also necessary to devise appropriate interventions for persons for who return to safety-net programs. Both of these uses of beneficiary data dictate a continuing "vital record" of program use, the parity of the current episode

of use, and the years since the last use. Such records would need to be kept for a long period following program use, ten or twenty years, perhaps a lifetime.

Taken alone, a panel of beneficiaries tells us little about the success of those who exit. When the beneficiary panel is matched to tax record information, a continuing picture of the economic activity of past beneficiaries can be generated. Data on “life after welfare” (Meyer and Cancian, 1996) can be produced to show those who cross the threshold of median family income, and their numbers in comparison to others who never achieve an income greater than the poverty threshold.

### Dual Frame Samples

Matching tax return and beneficiary panels cannot be achieved without an integrated sample design that crosses the record systems of all agencies involved. A few examples indicate why.

- Persons most likely to be at risk for safety-net programs are sampled at a rate of about 1/4 of 1 percent by the DOR. If all safety-net programs keep records for the entire panel of persons who enter, perhaps 0.05 percent of the income experience of the population will eventually be reflected in matched records. This 2,500 person sample will be insufficient to give information about income for many counties outside the metropolitan statistical areas defined by Green Bay, Kenosha, Madison, Milwaukee, and Racine.
- Persons currently benefitting from the W-2 program or food stamps probably constitute no more than 3 percent of the Wisconsin population. The match with the DOR sample will reveal substantially less than 0.008 percent of the population, about 400 persons, since many of these persons will not be required to file tax returns.

A common sampling plan that meets the mission of DWD, DHFS, and DOR can produce matches across a high proportion of records maintained in each agency (Czajka and Schirm, 1990; Czajka and Walker, 1989). This can be achieved because the DOR can choose a sample from the poverty population that includes the families who receive DWD or DHFS benefits. Those persons can be traced in future years by the continuing panel of tax returns. On the other side, persons in a DWD or DHFS sample who do not file tax returns alert the DOR to the extent to which a nonfiling population exists and its importance for enforcement.

### Relational Database Technology

Relational database technology assures several capabilities that are essential for panel samples. Individuals can be located in long lists at minimal cost. Sampling can be integrated into the administrative processing of cases, without adding additional steps or routines for the caseworkers and returns processors. Matches of records can be made across agencies to provide a basis for integrated sampling, without revealing



any other information about the sample. The relational principle assures that data from different months or years will have a unique name. Similarly, individuals must have a unique identity, making it possible to link an individual's records over time. The systems afford logical checking that can identify inconsistencies at the point of data capture and save resources that are wasted in incorrect payments, or separate data entry and verification procedures. Routine reports can be created for monthly and annual reporting. Aggregates can be created for administrative units (counties, school districts), for W-2 service regions, and for statewide reports.

Most important, relational databases can be distributed over a number of locations and a number of types of data sources on different kinds of computers *without losing any capability for linking data*.

## IMPLEMENTATION

### A Matched Sample of Program Beneficiaries and Tax Returns

Phase 1. The DOR tax record sample lends itself to an expansion that incorporates the beneficiaries of safety-net programs. The last sample drawn contains tax records for 1995. The next sample will be drawn in 1997. Identifiers of all beneficiaries of safety-net programs in November 1996 can be provided to the DOR in time to permit a sampling of the 1996 income tax returns for all beneficiaries.

The 1996 tax return sample will yield less than 100 percent match to beneficiaries in a single month. Some beneficiaries do not file returns. Nonetheless, this procedure has many advantages.

1. The problems entailed in capturing returns by identifiers can be exposed by an activity that will not affect the quality of the cross-section that DOR requires for 1997.

2. Sampling by identifier serves two functions. It is the technique that the DOR must follow, if it is to convert its cross-sections into a taxpayer panel. It is also the technique that must be followed if DOR data are to be linked to data on current or past beneficiaries of the safety net.

3. The yield of the sample will serve as a basis for estimating the cost of a continuing effort to match DOR and beneficiary samples.

4. The match will create a better understanding of the nonfiling population, including the extent of noncompliance and failure to benefit from homestead tax credits and earned income tax credits in that population.

5. 1996 beneficiaries can be matched to tax returns in later years to track persons who are no longer beneficiaries.

Phase 2. When procedures for sampling by identifiers are in place, new beneficiaries from December 1996 to May 1997 can be added to the sample, refreshing the population of past beneficiaries that are being sampled. Completion of phase 2 creates the capability for continuous refreshing of the beneficiary-tax record match.

Phase 2 is also the first step towards a capability for refreshing tax return samples. Each year persons who did not file in the prior year must be identified and sampled. This group will include a high proportion of young people who were dependents on other returns in prior years.

### A Tax Return Panel

Phase 3. The mechanism used to create a matched sample of beneficiaries and tax returns can also be used to follow tax returns, creating a panel from the 1995 cross-sectioned sample that already has been collected. Several options exist: (a) a small sample of 1996 tax returns could be collected in order to develop mechanisms for capturing tax returns on the basis of a list of taxpayer numbers; (b) 1996 return processing can be examined to identify first-time filers, needed to refresh the panel sample; (c) panel data can be collected in 1997 as a supplement to the 1997 cross-section.

In all cases, the panel sample needs to be designed differently from existing cross-sections. Because income varies substantially from year to year, it is not efficient to follow all the persons in the 1995 sample (Czajka and Schirm, 1990). A logical way to begin the panel is to select from the existing cross-section at the lowest sampling rate used in 1995, 1/244. Higher sampling rates can be used for populations whose characteristics in 1995 indicate that a record continuing income experience is of special interest to the DOR. For example, returns filed in 1995 that received homestead credits could be followed. Returns that show Schedule C or F (proprietors) or returns that show Schedule D or E (capital gains and partnership, royalty, or fiduciary income) could be selected for study, to understand year-to-year variability in those income sources, consistency of industry and employer identifications from year to year, or the proportion of returns that no longer show these sources in succeeding years. Each of these subjects is of interest to tax policy.

When the sample of 1996 returns has been compiled, DOR will be able to understand the cost of incorporating a panel element in its tax return sample. DOR will also have quantitative data on which it can base decisions on the proportions of returns that should be picked cross-sectionally in each year, and the proportions that should be generated by following a panel.

Phase 4. Phase 3 establishes a panel of taxpayers beginning in 1995. Phase 2 established the continuing beneficiary panel. Because the mission of DWD and DHFS requires that they demonstrate the “success” of the

services they render, both agencies would have an interest in tracking the earnings of past beneficiaries. Monthly earnings reports can be developed from Unemployment Tax returns, but those data offer no insight into family arrangements. DWD and DHFS have an interest in total income and family arrangements that can be gleaned from tax returns. Since DWD and DHFS wish to trace histories, their interest lies in a panel of tax returns linked to past beneficiary records.

A panel of relatively poor taxpayers complements the cross-section of predominantly higher-income taxpayers that DOR now samples. Were the DOR to collect panel data for higher-income persons, it could revise its samples to include a larger representation for lower-income families, because following those families is mandated by recordkeeping requirements of PRWO and is desirable for managing the services rendered under W-2 for greatest probabilities of success with W-2 clients.

A panel will enhance the reliability of estimates from the tax record sample. At the same time, the increase in scale of the DOR samples and the focus on family, rather than tax, units, will enhance policy uses of DOR statistics for other agencies. The outcome will be a capability that is able to represent counties, as well as regions of the state.

Phase 5. To represent well-being of people in the state, tax returns will need to be aggregated into “families.” It will be necessary to obtain the list of all who are related by marriage or dependency status to the filer. Tax returns for each of these people will need to be located. Locating returns of dependents, separate filers, and other persons who file from the same address is more complex and expensive than sampling from a known list. Relational database technology assures that it is feasible. New developments in matching algorithms reduce the number of records that must be studied because of variation in the spelling of street names and other characteristics used in matching. Again, this activity is best begun during a period in which the outcome does not affect the data from which revenue estimates are to be generated.

The creation of families will establish a basis for measuring the rate at which existing family units spawn new tax returns. It can also give insight into the entry of dependents in prior years into safety-net programs.

## BENEFITS OF ADMINISTRATIVE PANELS

### Success with Safety-Net Programs

The DWD needs a long-term panel of data on the population of the state for three reasons: It requires knowledge that some clients have successfully overcome problems that lead to dependence; DWD requires knowledge that it is reaching a population that is of social concern; and it must have data on all former clients to assure that the lifetime limitations on program use are met. Another way of stating these information requirements is to say that the program needs to know how applicants are generated from the population at large and what impact services have on clients at later times. The benefit of a record of all experiences with W-2 services is necessary to judge the quality of the program and to understand how to make improvements.

DWD needs to reach out to continuing sources of data on the Wisconsin population to achieve those objectives. The DOR file of tax returns is a low-cost source of data, far less than panel surveys. Lacunae in the tax records can be identified, and special-purpose studies can be mounted on nonfiling persons.

#### Charting Tax Policy in the Next Century

The DOR can benefit from panel data. Panel data are required to understand more about the nonfiling population; they are required to discover equity and efficiency of taxation on noncorporate businesses; panel data are useful in judging the value of inflation adjustments, and other devices that link history to current returns. The DOR likely would hone the quality of its tax models if it were able to sample using taxpayer identifiers.

Using DOR data capabilities in connection with safety-net programs integrates information on earned income tax credits with transfer program income that is not collated from statistical records at the present time.

The need to learn about our population calls for an integrated panel of administrative data on individuals and families that can be used for statistical studies of tax and welfare policies.

### References

- Cilke, James and Roy A. Wyscarver. 1990. *The Treasury Individual Income Tax Model*. Washington, D.C.: U.S. Department of the Treasury, Office of Tax Analysis.
- Czajka, John and Allen L. Schirm. 1990. "Overlapping Membership in Annual Samples of Individual Tax Returns." *Proceedings of the Survey Methods Section of the American Statistical Association*, pp. 413–418.
- Czajka, John and Bonnye Walker. 1989. "Combining Panel and Cross-sectional Selection in an Annual Sample of Tax Returns." *Proceedings of the Survey Methods Section of the American Statistical Association*, pp. 463–468.
- David, Martin H., William A. Gates, and Roger F. Miller. 1974. *Linkage and Retrieval of Micro-economic Data*. Lexington, Mass.: D.C. Heath.
- Duncan, Greg. 1984. *Years of Poverty, Years of Plenty*. Ann Arbor, Mich.: Institute for Social Research, University of Michigan.
- Meyer, Daniel R. and Maria Cancian. 1996. "Life after Welfare: The Economic Well-Being of Women and Children Following an Exit from AFDC." Discussion Paper no. 1101-96, Institute for Research on Poverty, University of Wisconsin–Madison.
- Robbin, Alice and Margaret Hedstrom. 1981. *A Report on Data Processing and Machine Readable Records in the Department of Health and Social Services*. (Wisconsin Survey of Machine Readable Public Records). Madison, Wis.: University of Wisconsin and State Historical Society, January.
- Rubin-Bleuer, Susana. 1996. "Gross Flows from Administrative Data with Missing Waves." *Proceedings of the Survey Methods Section of the American Statistical Association*. (Forthcoming).
- Sailer, Peter and Mike Weber. 1996. "Creating Household Data from Individual Tax Returns." *Proceedings of the Social Statistics Section of the American Statistical Association*. (Forthcoming).
- Seigel, Paul M. and John Coder. 1996. "Use of Administrative Records to Produce Post-Censal (1993) Income and Poverty Estimates for All U.S. Counties." *Proceedings of the Social Statistics Section of the American Statistical Association*. (Forthcoming).



## **Evaluating the Impacts of W-2 on Family Structure and Maternal and Child Health: A Brief Background Paper**

Gary Sandefur and Molly Martin

Our charge was to write a short background paper that examined some possibilities for collecting and analyzing data on family structure, maternal health, and child health in the context of evaluating the impacts of W-2. Although there is some overlap between the domains of family structure and health, in most respects they are two quite different areas. Consequently, this paper has two distinct foci. First, the paper explores options for examining the relationship between the shift from AFDC to W-2 and family structure, including marriage, divorce, and out-of-wedlock childbearing. The association between welfare programs and family structure has been a critical element in the national debate over welfare reform and in the public discussion of W-2. The official documents that describe W-2, however, make no claims about its potential impact on family formation and dissolution. Nonetheless, W-2 contains within it several changes in the benefit structure that many have argued will influence family decisions, including the treatment of two-parent and single-parent families in the same way and an implicit “family cap,” i.e., no increase in benefits associated with an increase in the number of children.

The second focus of the paper is on options for examining the relationship between the shift from AFDC to W-2 and maternal and child health. W-2, as planned, makes major changes in the provision of health care for participants relative to Medicaid. At this point, we do not know if the federal government will grant the waivers needed to make these changes. One can, however, plan an evaluation of the effects of W-2 on child and maternal health without knowing for certain what kinds of changes in the provision of health care will finally be made. In our discussion below, we assume that the provision of health care in W-2 will be sufficiently different to merit consideration of its effects relative to Medicaid.

The paper discusses three sets of issues regarding the domains of family structure and health. (1) We discuss the hypothetical impacts of W-2 on family structure and child and maternal health. We take seriously the claims of the designers of the program about what it is intended and not intended to accomplish. We do, however, try to think broadly about some possible unintended consequences. (2) We examine sources of data that might be tapped to look at these impacts, concentrating primarily on administrative data and on data collected through surveys. (3) We discuss some of the issues that must be addressed in designing an evaluation for these questions. More specifically, we discuss the merits of combining administrative data, retrospective survey questions, and prospective survey questions.

## I. THE HYPOTHETICAL IMPACTS OF W-2 ON FAMILY STRUCTURE AND CHILD AND MATERNAL HEALTH

### A. Family Structure

Although most of the attention given W-2 has addressed its attempt to increase the work effort of people who apply for public assistance, an appeal to family values is one of the eight philosophical principles guiding the development of W-2. More specifically, Attachment 1 to the Program Narrative states the following principle: “Families are society’s way of nurturing and protecting children. Both parents, whether or not living with their children, are assumed to be equally responsible for their care. All policies must be judged by how well they strengthen the responsibility of parents to care for their children.”

This invocation of family values is somewhat different from that in the federal Personal Responsibility and Work Opportunity Reconciliation Act of 1996. In the conference report on this legislation, the findings section concludes: “Therefore in light of this demonstration of the crisis in our Nation, it is the sense that out-of-wedlock pregnancy and reduction in out-of-wedlock birth are very important government interests and the policy contained in part A of title IV of the Social Security Act (as amended by section 103(1) of this Act) is intended to address this crisis.” If one takes these differences in tone and content seriously, one would conclude that W-2 is less explicitly concerned with illegitimacy and single parenthood than is the national welfare reform legislation. Rather, W-2 tries to ensure that both parents and the grandparents of children born to minor mothers contribute economically to the care of their children. In later discussion of this principle, the authors state that “W-2 focuses on parents in order to help them succeed in their roles as economic providers.”

Nonetheless, W-2 differs from AFDC in the way that two-parent and one-parent families are treated and in the way that noncustodial parents are treated. First, both two-parent and one-parent families are eligible for assistance. The basic criterion for participation is an economic one based on income and assets. This is a major difference from AFDC, which had more restrictive criteria for participation for two-parent families than for one-parent families. Second, single noncustodial parents with child support orders whose family members meet the W-2 income and asset limit may qualify for limited W-2 services, including self-sufficiency planning, work readiness training, and unpaid community work experience.

W-2 also differs from AFDC in the way that it treats minor mothers. W-2 requires the parents of teenage mothers who participate in W-2 to play a larger role than they currently do in AFDC. In determining the eligibility of a minor with children to qualify for child care, transportation, and health care, the income and assets of the parents as well as the teen are taken into account. In addition, W-2 will ask the parents of the teen



to help provide child care in most cases while the teen completes school. The hope of the architects of W-2 is that more teen parents will live with their parents than under AFDC, and that these grandparents will bear more of the financial burdens of caring for their child and grandchild than they do under AFDC.

Given the emphasis of W-2 on improving the efforts of parents to support their children, a potential impact of W-2 on family structure that should be examined is its effect on the role of noncustodial parents in the lives of the children. That is, an evaluation should assess whether or not W-2 increases the economic involvement of noncustodial parents relative to the existing AFDC program, and it should examine whether other aspects of the involvement of noncustodial parents improve relative to what happens under AFDC. Similarly, an evaluation should study the economic and social support provided by grandparents to minor parents and their children under W-2 relative to the levels of participation in AFDC.

It would be unfair to criticize W-2 if it does not lead to reductions in illegitimacy, divorce, and single parenthood, since the official descriptions of the program do not state these as explicit goals. W-2 does, however, provide an opportunity to assess whether treating two-parent and one-parent families in a similar fashion will reduce illegitimacy and single parenthood, and whether the implicit family cap in W-2 reduces out-of-wedlock childbearing among participants in the program.

## B. Child and Maternal Health

The authors of W-2 view the provision of health care to low-income families as an important component of the new system. At this point, Wisconsin has not received the waivers necessary to implement this feature of W-2. It is not clear when and if this will happen. Our discussion below assumes that at some point the state of Wisconsin will be able to implement a W-2 health care plan that differs in significant ways from Medicaid.

The authors of W-2 proposed a replacement for Medicaid based on the following principles: (1) W-2 will provide benefits to workers who do not have employer coverage and will try to prevent employers from dropping coverage for employees who might be covered by W-2; (2) families will share in the costs of health care coverage, and their contribution will increase with their income; (3) W-2 will endeavor to cover as many people as possible while keeping coverage affordable; and (4) W-2 will rely on managed care.

These principles lead to substantial changes in the health care provided to participants relative to AFDC and Medicaid. An income and asset test will be used to determine if families with dependent children and pregnant women are eligible for the program. The income test allows higher incomes than required for participation in some other parts of W-2; families can continue to participate until their income exceeds 200

percent of the federal poverty line. Coverage will be provided by HMOs with slightly less comprehensive coverage than available through Medicaid. Participants must pay a premium based on a sliding scale structure. A major change is that W-2 participants must accept employer-provided coverage if it is available, whereas current Medicaid recipients could reject employer coverage in favor of Medicaid if they desired.

These changes may have an impact on the access to health care and maternal health, infant health, and the health of children. Given the expansion of the potential participant pool, a greater percentage of the low-income population may receive health care coverage. On the other hand, the required copayments may prove difficult for some potential participants. Consequently, an evaluation should assess the effects of W-2 on access to health care among the low-income population and its effects on the health of participants.

## II. SOURCES OF DATA

We concentrate on two major sources of data on family structure and child and maternal health: administrative data and surveys. At this point, it is not clear what kind of administrative data system will be designed to accompany W-2 or to meet the reporting requirements of the new federal legislation. We can only speculate on what kinds of data this system might collect. In addition, other forms of administrative data can be used to measure some of the domains of interest in evaluating W-2.

Several national surveys include questions that collect information on family structure and health. None of these surveys has sufficiently large samples in Wisconsin to permit use in evaluating W-2. Further, these surveys are clearly not designed to meet the needs of evaluators of programs such as W-2. They contain, however, questions that could be modified and incorporated into a survey instrument designed for evaluation purposes. In our discussion of examples of questions that might be useful, we focus on the National Longitudinal Survey of Youth (NLSY) and the NLSY Mother/Child Supplement to illustrate some potential uses of existing survey questions. We also discuss some of the questions in Wisconsin's Family Health Survey.

Our discussion in this section deals primarily with collecting information and very little with how one might use this information to assess the effects of W-2.

### A. Involvement of Noncustodial Parents in the Lives of Their Children

Administrative data may permit the examination of the economic support provided to participants by noncustodial parents with child support orders. Noncustodial parents whose children are eligible for W-2 will

have access to certain services, but not paid employment, that might improve their ability to meet their child support obligations. If, however, evaluators wished to examine broader aspects of the involvement of noncustodial parents, they will probably have to collect data on visits and other features of this involvement.

The NLSY Mother-Child Supplement asks questions regarding the financial support and social involvement of noncustodial parents. The information about financial support is limited, and other surveys probably do a more complete job of collecting information on financial support. With respect to the social involvement of the noncustodial parent, some of the information relates to child care for the youngest child: the survey asks who or what type of institution serves as the principal and secondary child care providers for the youngest child. One of the choices the respondent can select is “child’s other parent/stepparent.” The respondent is also asked where each of her children usually lives: in the respondent’s household, with the other parent, with other relatives, etc.

B. Involvement of Grandparents in the Lives of Teen Parents and their Children

The administrative records on participants in W-2 will probably include, or could easily be designed to include, information on living arrangements. Consequently, evaluators will know with whom minor parents are living. Administrative records are unlikely to include much information on child care and other support provided by grandparents, so any analyses of this would require collecting data directly from the teen parents and/or the grandparents.

The NLSY Mother-Child Supplement has a series of questions on the living arrangements of the mothers and whether their parents help with child care for the youngest child. The survey also has specific questions as to whether the mother/step/grandmother of the respondent and whether the father/step/grandfather of the respondent is present in the respondent’s household. This household composition data combined with the age and marital status information for the respondent can be used to determine who in the survey are unmarried teen mothers currently living with their parents.

In addition, the NLSY Mother-Child Supplement asks questions about whether the parents of the mother provide primary or secondary child care for the youngest child. If the grandmother is the principal or secondary child care provider, the survey collects information about the grandmother’s work patterns for the last four weeks. Also, the data record whether anyone outside of the household helps pay for the principal child care services for the youngest child and what relationship to the mother this person has. In this manner, one could determine the nonresidential support grandparents give to young mothers.

A final indicator of grandparents' involvement with the mother and child is whether they provide financial assistance. If another person pays for half of the respondent's living expenses, data are collected about this person's relationship to the respondent and the total income this person gives the respondent. The respondent can identify her mother, father, or parents as the source of this income. Though the respondent is asked if she received any income from a friend or relative in the past year, this assistance can not be attributed to a particular person or relation of the respondent.

C. Marriage and Divorce

Administrative data available through W-2 will probably include basic information on marital status. This information is likely to be fairly minimal, however, since marital status is not a criterion for determining eligibility.

Another potential source of administrative data on marriage and divorce is the vital statistics system in Wisconsin. This system collects marriage and divorce certificates and issues regular reports on marriage and divorce. Appended to this paper are the certificates of marriage and divorce that are recorded by the state. The data that are collected would not, however, be very useful for examining marriage and divorce among participants or potential participants in W-2, since these data include no information about their economic status and do not specify their residence below the city, village, or township level. This suggests that if an evaluation is to monitor the effects of W-2 on marriage and divorce, it will have to collect data on marriage and divorce directly.

The NLSY gathers retrospective marriage and divorce histories through a series of questions. During the first interview, the respondent is questioned about their current marital status and the number of marriages by the time of the first interview. Then, the month and year of the first marriage are recorded, as are the month and year that the marriage ended, if it has terminated. Then, the month and year of the most recent marriage are reported. In the following waves of interviews, the survey obtains data about the first, second, and third changes in marital status since the last interview and the month and year of each change.

D. Out-of-Wedlock Childbearing

The data collected by the state of Wisconsin on births is much more extensive than the data collected on marriages and divorces. The Appendix contains the original certificate of live birth and the birth certificate worksheet that are filed with, and processed by, the Center for Health Statistics of the Wisconsin Department of Health and Social Services. The state of Wisconsin reports yearly on out-of-wedlock childbearing by city,

village, and township. Further, the address collected on the worksheet that includes zip codes make it possible (if the data are of sufficient quality and permission can be obtained) to look at out-of-wedlock childbearing in more detailed geographical areas. In addition, the collection of social security numbers on the worksheet make it possible (again assuming quality and permission) to match birth records with other administrative records. Staff at the Center for Health Statistics have informed us that most birth certificate worksheets do include the complete addresses and social security numbers of the mothers. They state that these data are of good quality, since it is in the interest of the mother to provide the information in order to receive her child's social security card. The appendix contains the forms that are required to request access to these data.

An evaluator could also collect original retrospective data on births using fertility history questions that are available in several major national surveys. Birth and marital histories are recorded in the NLSY. The day, month, and year of the birth of the first through fifth child are catalogued. One could compare these dates with those obtained in the marital history section (described above) to determine whether each child was born in or out of wedlock.

#### E. Maternal Health

The W-2 administrative data system will necessarily collect and monitor information that is used to determine eligibility for the medical care component of W-2. In addition, data on the birth certificate worksheets could be used to create a system for monitoring the health of W-2 and other low-income mothers. The state of Wisconsin collects and reports for each birth the date at which prenatal care began and the number of prenatal visits. This information could be collected and monitored for sub-groups of births in areas with high rates of W-2 participation or for mothers who can be matched via social security numbers with their W-2 records.

One could also collect original data on prenatal care and other aspects of maternal health by adopting questions from existing surveys. A wealth of information concerning the prenatal health and care a woman receives during pregnancy is contained in the NLSY Mother-Child Supplement. Data are gathered as to the number and month of prenatal visits, whether sonograms and amniocenteses were conducted, the results of these tests, the maternal weight gain during pregnancy, the gestation duration, and the use of vitamin supplements, alcohol or cigarettes during the pregnancy. Questions such as these could be adopted to determine the mother's health during pregnancy.

Wisconsin collects information on individuals' health status in its Family Health Survey. The best prenatal care information available in this survey is whether any one in the household is currently participating or has participated during the last twelve months in the Healthy Start program. Healthy Start provides health

care for pregnant women and young children who might have problems paying for health care. If someone is participating or has participated in the program, they are identified from among the household list. There is a possibility that data about prenatal doctor appointments may be contained in the question about the person's reason for the last doctor's visit. The category "other" would apply for prenatal care and the respondent is supposed to specify the reason for the visit.

F. Infant Health

It is unlikely that the administrative data system in W-2 will collect much information on the health of infants. Again, an evaluator could use data from the birth certificate worksheet to monitor some aspects of the health of infants born to mothers participating in W-2. The state of Wisconsin collects and reports for each birth the birth weight of the baby. In addition, the worksheet contains information on the APGAR score, an overall measure of newborn health, and information on abnormal conditions and congenital anomalies of the baby.

One could also adopt existing survey questions to collect information on infant health. Infant health measures are included in the NLSY Mother-Child Supplement. Data relating to the birth of the child include the duration of hospital stay at birth for the infant, the length and weight of the newborn, and whether the child was born early or late and if so, by how many weeks. The survey obtains the infant's health status in the first year of life through questions concerning immunization, wellcare check-ups, and health care for illnesses. For wellcare and illness treatment, data collected includes the months in which these visits occurred, the type of place at which the infant received the care (i.e., HMO, community health center, private clinic, etc.), and with regards to illnesses, what the major symptoms were.

The Wisconsin Family Health Survey acquires the general health status of every person in the household, and therefore all infants. Other health questions include the time since the last general physical examination, the reason for the last doctor visit, and the number of times in the past year the infant has seen a doctor for some health care. A few questions involve hospitalization and emergency care. These include whether the child was hospitalized overnight in the past year, whether the infant has needed emergency room care in the past year, and the nature of the health problem requiring this medical attention. There are no specific questions relating to infant care needs.

G. Child Health

The W-2 administrative data system is also unlikely to contain much information on child health. The major source of administrative data on child health would have to come from HMOs that were providing health care to participants in W-2. In addition, one could adopt existing survey questions on child health.

The Mother-Child Supplement of the NLSY contains a wide array of information regarding child health, including physical, social, psychological, and developmental health. We are only concerned with the physical health of the child. First, NLSY collects data about the extent and nature of any accidents the child has experienced in the past year. Illnesses requiring medical attention experienced in the last year are also duly investigated. Wellcare information for children is obtained by asking the last time the child saw the doctor for a routine health check-up.

The survey gathers extensive information about possible physical limitations the child experiences. Parents are asked if their child has a condition that limits school attendance, school work, and/or usual childhood activities. The parents also give information as to whether these conditions require treatment by a medical professional, medicine or drugs, and/or special equipment. The Mother-Child Supplement classifies the nature of the condition through specific questions from a long list of health limitations. Possible conditions include the following: learning disability, asthma, heart condition, serious hearing difficulty, and orthopedic handicap. The length of time that the child has had this limitation is coded in years, and there is an option to code the response as "all his/her life." The NLSY Mother-Child Supplement offers an abundance of information concerning child health.

The Wisconsin Family Health Survey asks similar questions with respect to child health, but without as much detail. The survey asks about the overall health status of the child, the last time the child has seen the doctor, the reason for the last doctor visit, and the number of times in the last year the child has seen the doctor for some health care. Questions concerning hospitalization and emergency room care include whether the child was in the hospital overnight in the last year, and if so, for what reason, whether the child was treated in an emergency room in the last year, and if so, for what problem during the last emergency room visit. If the child could not attend school or usual activities were limited during the last two weeks owing to illness or other health-related problems, the length of time of this limitation, the nature of the health problem, and whether a doctor was seen for this problem are recorded.

### III. ALTERNATIVE DESIGNS FOR COLLECTING AND ANALYZING DATA

The two major alternatives for collecting data appear to be reliance on administrative records or collecting original data from a sample of families. It is clear that if one is to conduct a comprehensive evaluation of the effects of W-2 on family structure and health, one would have to collect original data from participants and some comparison group.

A more limited evaluation of selected domains with administrative data might, however, be informative. The administrative data that are most promising are probably those that involve out-of-wedlock childbearing, infant health, and maternal health, especially if the more detailed information available on the birth certificate worksheet can be accessed by an evaluator, matched with state tax records to identify low-income individuals, and matched with W-2 records to identify W-2 participants. One could also go back in time and construct similar matches for AFDC recipients and low-income individuals who were not participating in AFDC.

One option would be to construct a history of the rate of out-of-wedlock childbearing, birth weights, and prenatal care for the low income population of Wisconsin for several years back. This history would include all low-income mothers, and would note whether or not they were participating in AFDC at the time their child was born. One could then construct similar records for the years after W-2 goes into effect. One could also record other characteristics of the mothers from their birth certificates, tax records, and program data. Such a data set creates the potential for examining some of the effects of AFDC/Medicaid and W-2 on out-of-wedlock childbearing, maternal health, and infant health.

Perhaps the most appropriate way to think of the possibilities for administrative data is their use to monitor changes or the lack of change in out-of-wedlock childbearing, maternal health, and infant health among subpopulations over time. But it would be difficult to use them for a convincing assessment of the impact of W-2.

This suggests that evaluating W-2 requires moving beyond administrative data to methods that collect high-quality information on family structure, maternal health, infant health, and child health and that allow one to rule out some of the other possible alternative explanations of any changes over time.



## Evaluation of Child Care Services under the Wisconsin Works Program (W-2)

Karen Fox Folk

Adequate provision of child care services will be crucial to the successful transition of welfare participants to employment. Mothers' child care responsibility, particularly the cost of child care, is the single greatest barrier to employment of mothers of young children (Maynard, 1995; Oliker, 1995). Among single Wisconsin mothers receiving AFDC in 1993, two-thirds had at least one preschool child. In addition, 40 percent had estimated market child care costs of more than half their earnings if they worked full-time at \$6 per hour (Cancian and Meyer, 1995.). Economic labor supply models have documented the strong positive effects of providing child care subsidies on labor force participation of single AFDC mothers (Connelly, 1990; Kimmel, 1995). Convenience and reliability of child care arrangements also have been related to AFDC recipients' ability to remain employed and off welfare (Siegel and Loman, 1991; Weber, 1987; Wolf and Sonenstein, 1991).

This paper reviews the proposed system of W-2 child care services and presents hypothesized impacts of these services on families and the child care market. Various evaluation strategies are discussed in the context of a comprehensive evaluation of W-2.

### CHILD CARE SERVICES AS PROPOSED UNDER W-2

The provision of child care services to all who are willing to work is an integral part of the W-2 welfare reform program. Program policies include offering individuals "the necessary support services, including child care, and health care to enable them to move into a work setting" (Department of Workforce Development, 1996, p. 19). Under W-2, child care subsidies are proposed to be available to all families with incomes below 165 percent of the federal poverty line (FPL) who also meet the W-2 assets test.<sup>1</sup> Families must have one or more children younger than 13 years and in need of care so that the parent(s) may be employed. A larger pool of low-income families will thus be eligible for health care and child care assistance than the pool of families eligible to participate in a W-2 employment position. W-2 participants must have family incomes below 115 percent of the FPL.

---

<sup>1</sup>The current income level at which families qualify for child care assistance is 215 percent of the federal poverty line, and there are waiting lists for assistance in many counties.

To increase the supply of providers for mothers mandated to work, a provisionally certified category of providers will be added to the current system of licensed and certified child care. Currently, regulated providers are required to have varying amounts of training and continuing education. The new category of provisionally certified providers will be subject to a criminal background check, and one health and safety site visit. They will not be required to have the 15 hours of child care training currently required of certified providers. They will be reimbursed at 50 percent of the maximum rate for licensed family child care, whereas certified family providers will be reimbursed at 75 percent of the maximum rate. Maximum reimbursement rates are set on the basis of local market surveys of licensed child care facilities.

Families will be free to choose any certified or licensed child care arrangement for subsidized care, including relatives if the relative becomes provisionally certified. A subsidy for care in the child's home by any coresident is not allowed. As proposed in the W-2 legislation, all families will be required to pay some of their child care costs— 7.5 percent for families below 75 percent of the FPL, 10 percent for families from 75 to 95 percent of the FPL, and increasing 1.3 percent for each 1 percent increase in income as a percentage of the FPL above 95 percent of the FPL. The percentage of cost paid by parents rises rapidly for incomes above the poverty line. Those with incomes at 100 percent of the FPL pay 16 percent of child care costs; those at 120 percent of the FPL pay 42 percent of costs. Parents pay 100 percent of child care costs at 165 percent of the FPL.<sup>2</sup> The W-2 child care system was to take effect August 1, 1996; income cutoffs and copayments are now under review because of protests by parents and concern about the effects on child care providers.

### The Administration of Child Care Services

W-2 agencies may be private contractors or county social services departments. W-2 agencies will primarily do case management to guide participants toward increasingly self-sufficient employment. W-2 agencies will determine eligibility for child care services but, in many cases, separate county human/social service departments will administer child care services. County departments will help eligible parents to find child care providers, determine child care copayments, and use a computerized statewide reporting system to document child care provider services. The state computer system will issue biweekly checks to providers based on county reports. In addition, the county will be responsible for conducting criminal background checks, certifying child care providers, and setting maximum reimbursement rates based on local market surveys.

---

<sup>2</sup>These copayments are much larger than those in the current system of child care subsidies. Current copayments are very small, are based on income level and number of children only, and do not vary with the cost of the care used.

## THE IMPACT OF W-2 ON CHILD CARE SERVICES

Adequate and timely provision of child care services is necessary for W-2 programs to achieve their goal of moving families to higher levels of self-sufficiency through employment. The broadening of full-time work requirements to all except mothers with a child less than 12 weeks creates a demand for child care services unprecedented in any previous welfare to work demonstration projects. Most demonstration projects were voluntary for mothers with children under age six, and volunteers with small children may have been those with lower-cost child care options. Existing research documents the need for child care services to promote self-sufficiency, but not the likely pattern of child care choices when most welfare recipients are mandated to full-time work (Bane and Ellwood, 1994).

Provision of child care services under the W-2 program will have short- and long-term impacts on parents, on children, and on the child care market for all employed parents. Possible impacts, based on low-income child care research and what is known about the current child care market, are summarized below.

1. The demand for child care services under W-2 is likely to exceed supply, especially in the short run.

The existing supply of regulated child care is a small proportion of the estimated demand when W-2 work requirements are fully instituted. An additional problem is convenient location of child care. Rural facilities are geographically sparse; urban residents often must use buses to transport children to and from care and also for travel to and from employment. This has been cited as a problem in two studies of employed Wisconsin AFDC recipients (Cochran, 1989; Weber, 1987). The increased demand for certifying new providers and processing subsidies could lead to an administrative bottleneck. W-2 legislation does not include increased funding for child care certification and for the new reports required of county social service departments.

The W-2 plan expects the new category of provisionally certified providers to increase supply, but the economic incentives to provide care in this way are not large. In 1994, certified family care providers (who are the same as provisional providers, except for the 15 hours of training required) had annual net incomes of \$5,132 for a 58-hour work week (Burton et al., 1995). Provisionally certified and certified family care providers may care for a maximum of three children other than their own. They must submit weekly or biweekly attendance reports to obtain state reimbursement and collect the copayment amount from parents.

The full-time work requirements of W-2 will also make use of relative care less feasible than it is for current AFDC recipients. More than half of poor families rely on relatives to provide child care, but that care

is most often provided part time and is less reliable than formal market child care (Kisker et al., 1989; Siegel and Loman, 1991).

Shortages of certain types of child care already exist for all families, not just low-income families. This includes infant care, before and after school care, care for special needs children, and night and weekend shift care (Hofferth and Phillips, 1991; Maynard et al., 1990). Night and weekend shift care will be a particular problem since one-third of poor mothers work weekends and one-half of working poor parents have rotating or changing work schedules (Hofferth, 1995).

2. The proposed schedule of W-2 child care copayments may lower net family incomes for families who must purchase care to meet work requirements.

Because the child care copayments are figured as a percentage of the cost of care, families with more children will face substantially higher copayments than smaller families. Larger families placed in a community service job or W-2 Transitions work activity will make higher child care copayments from a monthly grant of the same size that smaller families receive. All families who increase their wages above the poverty level will face rapidly rising child and health care copayments that will offset increased earnings at high marginal rates. These high marginal rates provide little incentive to participants to increase self-sufficiency by working more hours or advancing to a higher wage.

3. The quality of child care for children of W-2 participants is likely to be lower than under the current system and may have long-term consequences on child development.

Higher copayments are likely to cause many parents to move children from higher-quality, licensed, center or family child care to lower-cost providers with no child care training. Several studies found training of providers related to increased quality of care (Helburn et al., 1995; Maynard et al., 1990). State regulations require more training for licensed family child care and licensed center care than for certified family care.

Many current AFDC recipients rely on informal care by relatives or friends and more are likely to use this form of lower cost care under W-2. Galinsky, et al., (1994) found that care by relatives is of lower quality than that in regulated child care homes and this is especially true for those relatives caring for children out of necessity rather than by choice. Less formal care is also more subject to breakdowns that jeopardize employment (Kisker et al., 1989; Siegel and Loman, 1991).

The economic incentives under W-2 for parents to shift children to lower-quality, possibly substandard child care undermine the original intention of child care subsidies for at-risk children—to foster child

development and improve school readiness. Among the possible long-term consequences when children enter school are increased tax costs to enable schools to provide remedial education and deal with behavioral problems.

4. W-2 child care provisions may decrease the supply of regulated child care for all families.

There is disagreement about the possible effects of W-2 child care on the general child care market (Davidoff, 1996; *Wisconsin State Journal*, 1996). W-2 planners envision greater freedom of choice of care for welfare recipients and an increased supply of both formal and informal child care. Child care advocates expect and are reporting a shift by parents from regulated child care providers to less expensive unregulated care because of the increased copayments. They predict a decline in the number of licensed centers available for all children. Under the current voucher system in place since 1985, parents may choose either regulated centers or regulated forms of family child care with the same low copayment. The majority of parents currently receiving vouchers use licensed center child care.

#### EVALUATING CHILD CARE SERVICES UNDER W-2

Evaluating the child care services component of W-2 using a pre/post or experimental/control design faces similar problems to those inherent in a comprehensive evaluation of the entire W-2 program. In the pre-welfare-reform world, in which few families were affected by work requirements under AFDC/JOBS, the provision of child care services did not have the same importance as it does in the post-reform world, where full-time work is mandated for most W-2 recipients. Establishing a counterfactual, i.e., what the nature of public intervention would have been without the change in place (see the companion paper by Robert Haveman), is again complicated by this changed nature of the need for child care services under a mandatory work requirements program. In addition, following its philosophy of providing support for self-sufficiency through employment, the W-2 program extends child care (and health care) services to a pool of low-income families that is larger than the pool of potential W-2 eligibles with mandated employment. Focusing an evaluation only on participants in W-2 employment positions will not capture the impact of child care subsidies on the larger pool of low-income working families.

Evaluating child care services under W-2 requires an assessment of how well these services meet the two goals inherent in providing subsidized child care to low-income families. The first goal is to enable parents to be employed and to move toward economic self-sufficiency; the second goal is to provide care for

disadvantaged children that will result in school readiness and, in the future, healthy and productive workers. Evaluation designed to assess the extent to which W-2 child care services facilitates those goals would examine the effects of the program on parents' employment and children's well-being. An intermediate evaluation goal would be to assess the process of providing child care services, i.e., does the system provide the needed help without administrative delays or bottlenecks, and does it do so equally well in various communities across the state with differing child care markets and employment opportunities?

Important evaluation questions include:

1. Do W-2 participants receive the help they need when they need it to obtain child care (process evaluation)?
2. How do the characteristics of child care and subsidies (i.e., cost, convenience, reliability) affect the ability of W-2 participants to obtain and retain employment (process and outcome evaluation)?
3. How do child care arrangements affect children's well-being; are arrangements stable and of adequate quality (outcome and impact analysis)?
4. How do W-2 child care subsidy provisions affect the ability of the child care market to provide quality child care for low-income children (impact analysis)?

The first two questions are of the greatest importance in evaluating the success of W-2 in meeting its goal of increasing self-sufficiency through employment. The third question is a broader research question assessing both the role of child care and the possible effects on children of W-2 work requirements, reduced time with their own parents, and increased time in substitute care. The answer would have implications for the long-term effects on the youngest generation of potential welfare recipients and for future welfare policy reforms. The fourth question attempts to assess the possible unintended consequences of W-2 and the changes in the current child care subsidy system on the entire child care market; this is a question related to the broader goal of appropriate child development and improved school readiness among disadvantaged children.

Answering some of the above questions could be integrated into a process evaluation of the administrative efficiency of W-2, or an evaluation of the movement of W-2 participants up the job self-sufficiency ladder. This paper considers two possibilities: (1) what could be learned from existing administrative and other child care system data, and (2) what a longitudinal panel study of a sample of low-income households would add to a comprehensive evaluation.

#### Using Administrative and Other Existing Child Care Data

##### *Administrative Data.*

The only reliable pre-W-2 administrative data on low-income subsidized child care simply track the number of children served, number of families, and total dollars expended within each category of child care funding by counties and the state. This data provides no information on the relationship between child care subsidies and parent's employment, the focus of the first two evaluation questions, and the most important. Under W-2, a reporting system which creates a longitudinal record for each individual to monitor time limits within each employment tier and overall will be necessary. The current state administrative system for tracking W-2 participants, monitoring time limits, applying sanctions, etc., does not include information on child care arrangements, cost, or subsidies. However, a computerized system for issuing child care vouchers and recording individual usage data from each county is scheduled to be operating in early 1997. Data on individual households receiving subsidies could theoretically be linked to W-2 employment tracking data, but both systems are new, and initial debugging may muddy data collected in the early stages of operation.

Federal reporting requirements of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 may provide useful evaluation data to track both child care subsidies and use of child care.<sup>3</sup> These federal reporting requirements place a much larger burden on the state than the present system can report and it is as yet unclear how strictly they will be enforced. In addition, the monthly collection requirement is difficult, since many families in unsubsidized employment, receiving only child care and/or health care benefits, would not meet regularly with the W-2 agency staff. Thus there would be no data on the family and employment characteristics of those low-income families in unsubsidized employment to match with data on child care use from the child care subsidy reporting system.

Using administrative data collected by the state would be the least expensive evaluation strategy for child care services if data from the two newly created computerized tracking systems for W-2 and for child care can be merged. This would require the full cooperation of state agencies and full compliance with federal reporting requirements. If those requirements are met, and data are structured as longitudinal records for

---

<sup>3</sup>States must collect information monthly and submit reports quarterly to the U.S. Department of Health and Human Services (DHHS) on those receiving child care subsidies including: "family income; county of residence; the gender and age of children receiving assistance; whether the family includes only one parent; sources of family income, including employment, IV-A cash assistance, housing assistance, food stamp assistance, and other assistance programs; the number of months the family has received benefits; the type of child care in which the child is enrolled; whether the child care provider is a relative; the cost of child care for families; and the average hours per week of care." Information on child care subsidies must also be reported to HHS every six months. These reports must include: "the number of child care providers that received funding; the monthly cost of child care services and the subsidy cost portion; the number of payments made to providers through vouchers, contracts, cash, and disregards under public benefit programs, listed by the type of child care services provided; the manner in which consumer education information was provided to parents, and the total number, without duplication, of children and families served." (APWA, NGA, NCSL Analysis, 1996, p. 23.)

individuals, changes in the type of care used, its cost, and other features could be evaluated. The reliability of data that are not vital for issuing payments or tracking time limits could also be questionable if county-level workers inputting the data have little motivation to maintain accuracy; this has been true with previous administrative reporting systems (see the companion paper by Piliavin and Courtney).

The missing element in an evaluation using administrative data would be the link between child care arrangements and employment changes. It would not be possible to determine if child care difficulties caused employment loss or changes in employment caused child care changes. In addition, the stability of child care arrangements could be tracked, but no evaluation of their quality would be possible using these data, and without quality measures, only the crudest analyses of the potential effects of child care on child well-being could be done. Another limitation is that administrative data would contain information only on active W-2 participants, not on those who left the program because of child care or employment difficulties. But one strength of using these data is that they would supply information for all counties of the state, allowing comparisons of local variation in the effectiveness of W-2 agencies and answering the question of whether participants receive the child care help they need when they need it.

#### *Child Care Resource and Referral Data*

The child care resource and referral system (CCR&R) has data on the supply of licensed and certified care and on parent requests for help in finding child care. These data would be most useful in examining the fourth question, that of possible adverse effects on the child care market arising from W-2 subsidies and copayment schedules. Currently 17 CCR&Rs serve all 72 Wisconsin counties. To track historical and current data, funding would be needed to download and merge data from the 17 CCR&R datasets into a statewide dataset. Data are available for all 17 CCR&Rs beginning in 1995; these would provide one to two years of pre-W-2 baseline data for the state.

CCR&R data on the child care market include information on enrollments and open slots, new facilities and closures for regulated care in licensed centers, licensed family centers, certified family child care homes, and licensed school-age programs. A weakness is that CCR&R data are limited to the regulated child care market, providing no information on changes in use of child care by nonsubsidized relatives or other informal care, and no data on the supply of unregulated providers.

CCR&R data will continue to be collected by the 17 individual CCR&Rs and can provide comparative data on changes in the regulated child care market under W-2. The new category of provisionally certified providers will be included as regulated child care facilities and tracked by CCR&Rs.



Data on individual parent requests could also provide some indication of changes in demand for particular types of care. CCR&R data on parent requests include the number of requests for each type of child care, and special considerations such as need for evening care or special needs children. For each request, the characteristics of the family are also recorded: age of the child for whom care is requested, ethnic background, family income level, and whether family receives public assistance. Although not all AFDC recipients receive CCR&R assistance, information on AFDC and W-2 receipt could allow some analysis of pre- and post-W-2 child care demand.

#### Longitudinal Panel Study before and during W-2

To adequately answer the four questions posed above, a longitudinal panel study of W-2 recipients, tracking their child care and employment history, is needed. The analysis of administrative data is unlikely to provide sufficient detail to adequately evaluate the role of child care and child care subsidies in W-2 participants' employment history. A repeated cross-sectional survey of low-income families as W-2 progresses could provide the greater detail needed on connections between child care cost and availability and employment choices. Such surveys could gather retrospective histories for the prior six months or a year from respondents. However, with the new focus of welfare reform on changing long-term behaviors of the welfare-dependent (see the companion paper by Thomas Corbett), only a multiyear longitudinal design can assess the complex factors influencing individual life course events. Such a design would also have the potential for longer term assessment of child outcomes (Cancian and Wolfe, 1996).

Ideally, data collection should start before full W-2 program begins, to provide information on the effect of W-2 requirements on changing employment and child care patterns of recipients. The sample should be large enough to include low-income families with children of varying ages. It should not be restricted to W-2 participants since child care subsidies are available to all families who meet income and assets requirements.

Sampling from the population of all families with incomes less than 165 percent of the FPL is important, for several reasons. It is important to track the effects of W-2 for current AFDC recipients who exit the system because of the work mandates or their inability to comply with work requirements. Including low-income families also captures young mothers who choose to work part-time, an option not allowed under W-2 work participation requirements. Finally, low-income families who do not receive subsidies provide a comparison group for assessing the positive effects of subsidies on employment and quality of care. A number of studies have shown improvements in quality among those receiving child care subsidies under the existing system (Berger and Black, 1992; Hofferth and Wissoker, 1992).

The proposed study would gather information on receipt of child care subsidies, any child care problems that affect employment, and the cost and other characteristics of care used. A longitudinal history of employment, income, and child care changes, and direct questions about the reasons for these changes would allow analyses of the link between child care and ability to improve economic self-sufficiency through employment. Rough measures of the quality of care could be gathered from parents' reports of staff/child ratios and warmth of caregiver's relationship to the child. These are measures that have been correlated with more objective direct observations of child care quality (Galinsky et al., 1994).

The strengths of this evaluation approach are that the study could be designed to adequately answer major evaluation questions related to W-2 child care services. The necessary data could also be gathered in conjunction with other evaluations of the W-2 program. The most obvious limitation is the cost of gathering such data from a hard-to-reach and highly mobile population. Cost would also prohibit sampling all counties in the state; instead, representative urban and rural samples would be needed.

There are several reasons that investing in a comprehensive evaluation of W-2 child care services is important. First, the proposed W-2 full-time work requirements for mothers of young children are historic and unprecedented. W-2 work requirements exceed federal block grant requirements and will place extreme demands on the child care market. Under these conditions, the impact of the cost and availability of child care on mothers' ability to become self-sufficient through employment should become much clearer, but only with adequate evaluation. Second, the addition of subsidies for minimally regulated providers reverses the previous philosophy that publicly funded child care should be regulated to provide a minimum of good-quality care for low-income children. Including measures of the impact of W-2 on the quality of care is necessary both to assess the short-term effects on children and to estimate the long-term effects on school readiness and other child outcomes. An adequate evaluation of the impact of W-2 child care services will provide crucial information for improving Wisconsin welfare reform and for the design of welfare reform by policymakers in other states.

### References

- APWA, NCSL, NGA (American Public Welfare Association, National Governors' Association, and the National Conference of State Legislatures). 1996. "The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (Conference agreement for H.R. 3734) Analysis." Available on the World Wide Web at <http://www.apwa.org/reform/analysis.htm>, August 9, 1996.
- Bane, Mary Jo and David T. Ellwood. 1994. Welfare Realities: From Rhetoric to Reform. Cambridge, Mass.: Harvard University Press.
- Berger, Mark C. and Dan A. Black. 1992. "Child Care Subsidies, Quality of Care, and the Labor Supply of Low-income, Single Mothers." Review of Economics and Statistics, 74: 635–642.
- Burton, A., M. Whitebook, and L. Sakai. 1995. Valuable Work; Minimal Rewards. A Report on the Wisconsin Child Care Work Force. Washington, D.C.: The National Center for the Early Childhood Workforce and the Wisconsin Early Childhood Association (WECA).
- Cancian, Maria and Daniel R. Meyer. 1995. A Profile of the AFDC Caseload in Wisconsin: Implications for a Work-Based Welfare Reform Strategy. IRP Special Report 67, Madison, Wis., September 1995.
- Cochran, Susan. 1989. Post-AFDC Child Care Assistance Report, June, 1989. Madison, Wis.: Wisconsin Department of Health and Social Services, Office of Policy and Budget, Evaluation Section.
- Connelly, Rachel. 1990. "The Effect of Child Care Costs on the Labor Force Participation and AFDC Reciprocity of Single Mothers." Institute for Research on Poverty Discussion Paper no. 920-90, Madison, Wis.
- Davidoff, Judith (1996). Will W-2 Hurt Your Children? Isthmus, April 5–11, 1996 (Madison, Wis.).
- Department of Workforce Development (DWD), State of Wisconsin. 1996. Draft Request for Proposals for Wisconsin Works, Memo from Richard Wegner, Acting Secretary, June 18, 1996.
- Galinsky, Ellen, Carollee Howes, Susan Kontos, and Marybeth Shinn. 1994. The Study of Children in Family Child Care and Relative Care. New York: Families and Work Institute.
- Helburn, Suzanne, M. L. Culkin, J. Morris, and N. Mocan. 1995. "Cost, Quality, and Child Outcomes in Child Care Centers." Executive Summary. University of Colorado at Denver, Economics Department.
- Hofferth, Sandra L. 1995. "Caring for Children at the Poverty Line." Children and Youth Services Review, 17(1/2): 61–90.
- Hofferth, Sandra L. and Deborah A. Phillips. 1991. "Child Care Policy Research." Journal of Social Issues, 47(2): 1–13.
- Hofferth, Sandra L. and Douglas A. Wissoker. 1992. "Price, Quality and Income in Child Care Choice." Journal of Human Resources, 27(1): 70–111.

- Kimmel, Jean. 1995. "The Effectiveness of Child-care Subsidies in Encouraging the Welfare-to-work Transition of Low-income Single Mothers." American Economic Review, 85(2): 271–275.
- Kisker, Ellen E., Rebecca Maynard, Anne Gordon, and Margaret Strain. 1989. The Child Care Challenge: What Parents Need and What Is Available in Three Metropolitan Areas. Report to Department of Health and Human Services. Princeton, NJ: Mathematica Policy Research.
- Maynard, Rebecca A. 1995. "Subsidized Employment and Non-Labor-Market Alternatives for Welfare Recipients." In Demetra Smith Nightingale and Robert H. Haveman, eds., The Work Alternative: Welfare Reform and the Realities of the Job Market. Washington, D.C.: Urban Institute Press.
- Maynard, Rebecca, Ellen E. Kisker, and Stuart Kerachsky. 1990. Child Care Challenges for Low-Income Families: The Minority Female Single Parent Demonstration. N.Y.: Rockefeller Foundation.
- Oliker, Stacey J. 1995. "Work Commitment and Constraint among Mothers on Workfare." Journal of Contemporary Ethnography, 24: 165–194.
- Siegel, Gary L. and L. Anthony Loman. 1991. Child Care and AFDC Recipients in Illinois: Patterns, Problems and Needs. Report prepared for Division of Family Support Services, Illinois Department of Public Aid. St. Louis: Institute of Applied Research.
- Weber, Joan. 1987. Child Care in Wisconsin: Provision of Care, Supervision, and Education for All Children Including Those from AFDC Families. Madison, Wis.: Wisconsin Department of Health and Social Services, Bureau of Planning, Division of Policy and Budget, March, 1987.
- Wisconsin Department of Health and Social Services. 1979. Child Care in Wisconsin, WRWP#9, February, 1979. Madison, Wis.: Wisconsin Department of Health and Social Services, Welfare Reform Study Staff.
- Wisconsin State Journal (1996). "Wisconsin Day Care Bind Could Pinch Welfare Reform Plans." February 19, 1996.
- Wolf, Douglas A. and Freya L. Sonenstein. 1991. "Child-Care Use among Welfare Mothers: A Dynamic Analysis." Journal of Family Issues, 12(4): 519–539.

## Potential Effects of Wisconsin Welfare Reform (W-2) on Child Support

Stuart Kipnis

Wisconsin Works (W-2) is likely to have a significant impact on child support enforcement, including the establishment of paternity and of child support orders and the collection of support from noncustodial parents. Several aspects of W-2 may increase the motivation of custodial parents to cooperate with child support enforcement, particularly the provision that the full child support payments by the noncustodial parent will go directly to the family. With a time limit on W-2 benefits of five years, child support payments will become a critical source of support for custodial parents, both those who are working and those who are not able to find employment once time limits are reached. I briefly discuss the current child support system in Wisconsin, changes in the system made under W-2, and any deviations from the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA). The implications for child support order establishment and compliance are highlighted, as are ways to measure the effect of W-2 on child support enforcement. I also discuss possible research designs for measuring the impact of W-2 on child support.

### THE CURRENT CHILD SUPPORT ENFORCEMENT SYSTEM

In the case of nonmarital births, the initial step in obtaining a child support order is the establishment of paternity. Paternity is sought in all cases in which the mother is receiving public assistance, except where good cause is given (for example, a history of violence by the father). Research indicates that paternity establishment is the primary obstacle in getting support orders for children born out of wedlock (Danziger & Nichols-Casebolt, 1990). A dramatic increase has occurred in paternity establishments since 1985, both nationwide and in Wisconsin (Committee on Ways and Means, 1994). In Wisconsin, the majority of cases in which paternity has been established, as well as the vast majority of divorce cases involving children, have legal child support orders (Meyer et al., 1996).

Once an order has been entered by the court, several methods of enforcement are designed to increase compliance: routine withholding of the child support obligation from wages, intercept of tax refunds, parent locator services, and law enforcement efforts, including jailing the noncustodial parent and taking away his or her driver's license. Custodial parents who received AFDC have been provided a financial incentive to cooperate with the child support enforcement agency. They received the first \$50 of payments collected, the remainder going to the state to offset assistance payments. Compliance with child support orders has increased

over time in Wisconsin with these enforcement efforts (Meyer et al., 1994). In addition, noncustodial parents in several Wisconsin counties who are delinquent in child support payments have been able to participate in the Children First demonstration project, allowing them to take part in job training activities ordinarily reserved for custodial parents.

#### CHILD SUPPORT ENFORCEMENT UNDER W-2

Under W-2, there will be many similarities with the current child support enforcement system. Those custodial parents applying for W-2 services must cooperate fully with paternity and child support order establishment efforts, unless good cause is shown for not doing so. The current enforcement system, including the use of immediate income withholding, tax intercepts, periodic review and modification of orders, and law enforcement efforts, will continue.

W-2 brings, however, a number of significant changes in the child support enforcement system. Custodial parents who are receiving a W-2 benefit will be entitled to the full child support payment by the noncustodial parent, rather than only the first \$50 collected. Assume a custodial parent was initially not receiving child support and received a W-2 grant of \$555/month for a community service job. That parent's monthly income would then be \$555/month. If the custodial parent began receiving child support payments of \$200/month, the full payment would go to the family, and the W-2 grant will remain at \$555/month. That parent's monthly income would rise to \$755/month. These changes in the payment of child support to parents on public assistance are designed to emulate real-world work conditions. The increase in monthly income due to child support could, however, be partially offset by an increase in copayments for child care and health care (see the paper by Karen Folk for a discussion of child care copayments).

This change in the "pass through" (the amount given to the family) should significantly increase the motivation of many custodial parents to cooperate with the child support enforcement agency in establishing paternity and establishing and enforcing child support orders. Also, it has been widely acknowledged that, in AFDC cases, the limited amount of child support payments going to the custodial parents and the children has served as a disincentive in the paying of child support by noncustodial parents (Edin, 1995). Therefore, it can be reasoned that the receipt of the full amount of the payment by the custodial parent should serve to increase the noncustodial parent's incentive to pay support.

Of course, this benefit of W-2 is only realized in cases where a child support order is both established and is being paid by the noncustodial parent. In 1993, less than 50 percent of Wisconsin families on AFDC

received any child support payments (Committee on Ways and Means, 1994). Therefore, unless both the establishment and the enforcement of child support orders increase dramatically under W-2, many families will be left without this important means of income.

It should be noted that the federal legislation allows states to pass through some or all of the child support payments to parents receiving welfare assistance **only if the states use their own funds**. Therefore, Wisconsin is being significantly more generous in its approach to child support for families receiving W-2 assistance than either the current system or the approach that many states will likely take under PRWORA.

Another change that may affect child support enforcement efforts is the nature of W-2 as a time-limited public assistance program. Facing a 5-year lifetime limit on the receipt of W-2, custodial parents will likely view child support as an essential element for their children's well-being. In fact, it can be argued that child support payments will be the primary safety net for parents who have been cut off from assistance, either because they did not comply with the W-2 requirements or they reached the time limits. In addition, charging premiums and copayments for health care and child care may have an effect on child support enforcement efforts, as an increase in child support payments can be used to offset these potentially substantial health care and child care costs.

#### POSSIBLE EFFECTS OF W-2 ON CHILD SUPPORT

Several factors that were mentioned above point to an increase in paternity and child support order establishment and an improvement in enforcement of child support orders under W-2. Custodial parents will have an incentive to receive child support, as payments go in full to the family without a reduction in W-2 assistance and will serve as a valuable source of income should the custodial parent be sanctioned for noncompliance or be cut off of assistance by reaching a time limit. Further, noncustodial parents may be encouraged to pay child support, as the full amount of the support will go to the custodial parents and the children.

Establishment of child support orders and collection of payments may, however, decrease under W-2. Currently, the cost of the child support enforcement system is offset by the collection of payments when the custodial parent is on AFDC. Innovations in child support enforcement, such as immediate income withholding and periodic review and modification of orders, are considered beneficial not only because they help increase collections, but also because they have a benefit/cost ratio to the state that is significantly greater than 1. Under W-2, the primary beneficiary of the changes in the child support enforcement system will be custodial parents

and their children; the fiscal benefit to the state no longer exists. It is not known what affect this will have on child support enforcement efforts. It can be argued, on the other hand, that the state will be enlightened by the enormous importance of maintaining a strong child support enforcement system, particularly in the benefit of payments to children and the heightened importance that child support payments will have as a safety net under W-2. Further, child support receives such broad support from the public and from both major political parties that it would be quite difficult to decrease efforts to establish and enforce child support orders.

#### MEASURING THE EFFECTS OF W-2 ON CHILD SUPPORT

Several measures can be used to assess the effect of W-2 on child support enforcement. The initial step for many W-2 cases will be to establish paternity. Measuring paternity rates involves calculating both the number of paternities established and the number of nonmarital births during a given period. It is important that both the numerator (paternities established) and the denominator (nonmarital births) reflect the same time period. For example, assume there were 10,000 nonmarital births in 1996. To accurately compute the rate of paternity establishment for these births requires determining which of these children had paternity established, either in 1996 (for example, 5,000) or in later years (perhaps 2,500). The paternity rate for 1996 would be  $7,500/10,000$  or 75 percent. Therefore, when paternity is established, it is essential to collect the child's birth date.

Calculations of the percentage of orders established and compliance with orders is much more straightforward. To determine the rate of order establishment requires simply dividing the number of orders established by the universe of cases where an order could be established. For example, we may want to know what percentage of all W-2 custodial parents (for example, 50,000) have child support orders (perhaps 30,000). This is calculated by dividing the number of W-2 custodial parents with child support orders by the total number of W-2 custodial parents; 60 percent of W-2 custodial parents have child support orders in the above example. Alternatively, we may be concerned with percentage of orders established for subgroups of the W-2 population, for example teenage parents or new W-2 applicants, as well as parents who are not receiving W-2.

Compliance is also generally straightforward, the numerator being actual dollars collected (for example, \$2 million) and the denominator being dollars owed (for example, \$4 million), for a compliance rate of 50 percent. The primary difficulty in determining rates of compliance occurs if the support orders are not all fixed-dollar orders, for example \$200/month. In several states, including Wisconsin, orders are sometimes



entered either totally or partially as percentage-expressed awards (for example, 20 percent of income or \$100/month, whichever is higher). With payment amounts varying based on the income of the noncustodial parent, it becomes more difficult to measure compliance. For more information on measuring compliance for these orders, see Bartfeld and Garfinkel (1996).

While paternity establishment and child support order establishment and compliance can be measured using administrative data, other effects of W-2 on child support can only be determined with other data. Suppose we want to determine whether an increase in child support payments by noncustodial parents under W-2 is associated with an improvement in the relationships between the parents and between the noncustodial parents and their children. Clearly, administrative data cannot provide much help with assessing the strength of relationships. Instead, it would be necessary to collect data through intensive interviewing and/or surveys of the parents or observations of the interactions between the parents and children. This sort of primary data collection can also be used to obtain information on informal child support arrangements, such as child care provided by the noncustodial parent or members of his or her family; gifts, including birthday presents, clothes, sports equipment, and diapers; and money paid to the custodial parent outside of the formal child support enforcement system.

## RESEARCH DESIGNS

Several research designs can be used to evaluate the effect of W-2 on child support. The strengths and weaknesses of each approach are discussed in the companion papers, including those by Glen Cain, Robert Haveman, and Irving Piliavin and Mark Courtney. Perhaps unique among the potential outcomes of W-2, we have excellent longitudinal data on child support orders and collections in Wisconsin, as well as in many other states (see Meyer et al., 1994, for a discussion of Wisconsin child support data). It is, therefore, quite feasible to compare before and after W-2 the rates of order establishment for cases with paternity already established, level of child support orders, and compliance with child support orders by means of administrative data. These comparisons can extend from the AFDC period through the W-2 period, including the transition from AFDC to W-2.

Rather than looking specifically at the effect of W-2 on child support enforcement, it may be valuable to see whether the receipt of child support plays a role in contributing to successful outcomes for the W-2 recipient. First, it is necessary to control for likely differences in the demographic composition of families receiving child support from those not receiving any payments from the noncustodial parent. Such differences

may include education, work history, and age of the custodial parent. We can then compare W-2 families receiving child support payments with similar W-2 families not receiving child support. Do these families differ in terms of success with the W-2 work requirements, ability to move up the W-2 tiers (for example, from W-2 Transitions to community service jobs and so on), access to child care and medical care, and avoidance of out-of-home placements in the child welfare system?

Another potentially valuable analysis involves comparing Wisconsin, which passes through the entire child support payment to custodial parents who are on public assistance, with other states where any payments of child support are either kept fully or partially by the state to offset public assistance payments. Does the motivation for cooperating with the child support enforcement agency differ for both custodial and noncustodial parents in Wisconsin compared to these other states? How have rates of child support order establishment and compliance changed over time in Wisconsin in relation to states without the full pass through of child support to the families? Are there differences in attitudes about child support enforcement among child support workers, the public, and political officials in Wisconsin and the other states? This last question is important, because in Wisconsin, W-2 payments will not be reduced by child support payments. In many other states, public assistance payments will continue to be fully or partially offset by child support payments.

### References

- Bartfeld, Judith & Garfinkel, Irwin. 1996. "The Impact of Percentage-Expressed Child Support Orders on Payments." Journal of Human Resources, 31, 794–815.
- Committee on Ways and Means, U.S. House of Representatives. 1994. 1994 Green Book. Washington, D.C.: U.S. Government Printing Office.
- Danziger, Sandra & Nichols-Casebolt, Anne. 1990. "Child Support in Paternity Cases." Social Service Review, 64, 458–474.
- Edin, Kathryn. 1995. "Single Mothers and Child Support: The Possibilities and Limits of Child Support Policy." Children and Youth Services Review, 17, 203–230.
- Meyer, Daniel, Bartfeld, Judith, Garfinkel, Irwin, & Brown, Patricia. 1996. "Child Support Reform: Lessons from Wisconsin." Family Relations, 45, 11–18.
- Meyer, Daniel, Garfinkel, Irwin, Bartfeld, Judith, & Brown, Patricia. 1994. "An Evolving Child Support System." Focus, 16, 1–14.



## **Potential Effects of Wisconsin Welfare Reform (W-2) on Families Who Have Very Young Children with Disabilities and Special Health Care Needs**

George Jesien

The overall purpose of this paper is to explore how upcoming changes in Wisconsin's welfare reform may affect families who have very young children with disabilities. This paper provides an overview of Wisconsin's Birth to Three program, describes how the program is typically implemented at the local county level and then describes a series of possible indicators that may serve to elucidate the impact of W-2 on this special population.

Having a newborn challenges most families to adapt to the demands of the infant by shifting priorities, changing schedules, and garnering the necessary resources to clothe, feed, and take care of the growing infant (Begun, 1996). Having an infant with a physical or mental disability or a special health care need further stresses the resources of any family in terms of the additional needs that the infant poses, the time needed to interact with medical staff and developmental therapists, and the emotional strain of wanting to do everything possible to optimize his or her chances for developing normally (Turnbull & Turnbull, 1990). Raising and nurturing an infant or toddler with disabilities or special health care needs in an economically well-off and intact family poses complications and associated stresses sufficient to have national policy makers realize that federal assistance is needed to identify, assess and provide services for the needs of infants and toddlers with disabilities and their families (Smith & McKenna, 1994). Addressing these needs in situations where families are living in poverty or are headed by single or teenage parents poses significant additional obstacles and difficulties to overcome (Erwin, 1996). Families in poverty face the synergistic and additive effects of all three sources of stress: adapting to the new infant, learning about and addressing the needs posed by the disability, and at the same time continuing the struggle for securing food, shelter, and the other necessities of life (Zigler, 1995).

### **THE BIRTH TO THREE EARLY INTERVENTION PROGRAM**

National policy makers recognized the seriousness of the challenge facing a family with an infant or toddler with a disability by passing Public Law 99-457 in 1986, now referred to as Part H of the Individuals with Disabilities Education Act (IDEA). The legislation changed the "landscape" of services available for very young children with disabilities and has had far-reaching effects on systems of services throughout the country

(Jesien, 1996). One of the legislation's main purposes was "to enhance the capacity of families to meet the special needs of infants and toddlers with disabilities" (IDEA, 1991). Policy makers accepted that the needs posed by having a child with a disability were a challenge to all families regardless of socioeconomic or educational levels (Smith & McKenna, 1994).

This legislation gave states the option to develop, over a span of five years (two annual extensions were subsequently provided), a comprehensive and coordinated state system for the provision of family-centered early intervention services (Trohanis, 1994). In 1991, Wisconsin passed its own early intervention legislation, HHS 90. The Wisconsin law is patterned after the federal legislation. It designated the Department of Health and Family Services (DHFS) as the lead agency and, in effect, established a mandate for early intervention services for all eligible children and families.

The eligibility criteria in the federal law targeted children with a developmental delay, children having a physical or medical condition that has a high probability of resulting in developmental delay, and—at the discretion of the state—children who are at risk of having substantial developmental delays if early intervention services are not provided (IDEA, 1991). The definition of developmental delay was to be defined by each state. Wisconsin defined a developmental delay as a 25 percent delay in one or more areas of development and did not choose to serve children at risk but did include a category of children who have "atypical development" (HHS 90).

The law establishes a set of "core services" which are to be provided to all families at no cost. These services include:

a. Evaluation—an interdisciplinary process to determine a child's initial and continuing eligibility for early intervention services.

b. Development of an Individualized Family Service Plan (IFSP)—developed in conjunction with the family, detailing the expected outcomes, providing information on the developmental status of the child, and with the parents' permission including information on parent resources, priorities, and concerns. The IFSP contains the type, frequency, and intensity of services that will be provided and also contains transition plans for children as they "graduate" out of the program at three. The IFSP is to be redeveloped annually and reviewed every six months.

c. Service Coordination—provided by an assigned individual for each child and family to monitor and assist in the implementation of the IFSP. It is this person's responsibility to link the family with community resources, to help parents identify their own and their child's needs, and to serve as a monitor to assure that the IFSP is implemented.

d. Transition Services—preparation of the child for moving into the next environment whether it is a special education program in the public schools, Head Start, preschool, child care, or home environment.

Other early intervention services are provided based on the IFSP. These services include special instruction, developmental therapies, audiology, nursing care necessary for the child to benefit from early intervention, and parent education and counseling. If parents permit, third party insurance can be billed for therapy services such as occupational, physical, and speech therapies. Medical assistance has also been billed for therapy services in many counties. Developmental or educational services which are typically not billable to third party insurers or Medical Assistance are provided at no cost to families by Birth to Three providers. It should be noted that Wisconsin is in the process of developing a fee for service provision within the program where families will undergo a “determination of ability to pay.” If deemed able to pay a fee, they will be charged for early intervention services but not core services. This provision is expected to go into effect sometime toward the end of 1996 or beginning of 1997.

Currently, approximately 105 providers of Birth to Three services serve the 72 counties of Wisconsin. The counties are designated as the administrative agent for the program administered by DHFS. Counties have flexibility in how they implement the program but are expected to meet state guidelines. Some counties provide direct services to families using county employees, while others provide only service coordination through the county and contract with other agencies for remaining services. Still other counties contract out for all services, including evaluation, service coordination, and early intervention services. The number of children served by individual counties varies greatly, ranging from Milwaukee, where more than 2000 children were served last year to 10 counties that each served fewer than 10 children in 1995. Dane County served the second highest number of children, with more than 400 children receiving services in 1995.

Children receiving services in the Birth to Three program vary greatly in regard to the degree and type of disability. Children range from those who have a 25 percent delay in speech and language with minimal other compromises in other developmental areas to those with severe disabilities who may be technologically dependent for food and breathing capability. Some children with multiple and complex needs may be involved with as many as 15 provider agencies and personnel addressing their medical, health and developmental needs. Many of the children in the program have had extensive hospital stays prior to their referral to the Birth to Three program.

Services are also provided in a range of locations: the home, child developmental center, family or child care setting, hospital or clinical offices. Based on data submitted to the federal government in December of 1995, approximately 45 percent of children in Wisconsin receive most of their services in the home, 35 percent

receive them in a center or classroom for children with disabilities, 15 percent in an outpatient clinics, and 2.5 percent receive the majority of their services in a family or child care setting (Hoffman, 1996). Settings are to be chosen based on the preferences of the family and the needs of the child. In actuality, some program settings for services are determined more on the basis of what is available or where openings for children exist rather than preferences of the family.

The number of children served in the state has slowly increased over the last five years. In 1990, approximately 2233 children were in the program on a December 1 count basis (required by the federal government), whereas 3919 children were served on December 1, 1995. The total number of children served during the year is considerably larger because children enter and leave the program during the year. For example, the total number of children served in 1995 is 7205.<sup>1</sup>

**Number of Children Served on December 1 and Total for the Year (1995)**

<b>County</b>	<b>December 1 Count</b>	<b>Total Number Served for Year</b>
Milwaukee	855	2053
Dane	213	415
Waukesha	184	394
Brown	154	292
Rock	127	256

#### DESCRIPTION OF LOCAL SERVICES AND INFORMATION COLLECTED

Each child who is referred to the Birth to Three program is evaluated to determine whether he or she is eligible for early intervention services. An interdisciplinary team of at least two individuals (many evaluations are conducted by three or more disciplines) is convened. The team obtains a developmental history and determines functioning level and needs in the developmental areas of communication, physical, social, cognitive, and adaptive development. Data collection methods and evaluation instruments vary from county to county and provider to provider.

---

<sup>1</sup>Data on the number of children served by year and county was provided by State Birth to Three Program staff in DHFS, the Division of Supported Living.



Once the child is determined eligible, the early intervention team, made up of least two disciplines, the service coordinator (who can also represent one of the disciplines), and the parent develops the IFSP. The IFSP contains the parents' statements on their expectations, goals, and outcomes for intervention. Additionally, with parent consent, information on the family makeup, their resources and priorities are obtained, as well as their statement of other needs that they may have in addressing the needs of their child.

Although the evaluation instruments and format of the IFSP vary greatly, the State does require some consistent data elements to be collected across the Birth to Three programs and reported on an annual basis.

The data elements include:

- child's name
- client characteristics, e.g. type of disability
- social security number
- parents' name
- county of residence
- birth date
- gender
- ethnicity
- start & end dates of service
- services provided
- location of service-home, center, hospital, etc
- discharge reason

Optional elements include the child's address and telephone number and other program codes. Individual programs have additional information that is kept in different formats but would be easily accessible. This information includes reason for eligibility, name of service coordinator, type of health insurance, whether the child has MA or SSI, and native language. Each county is charged to maintain an individual record on each child served by the Birth to Three program and maintain the files for seven years after the child leaves the program. Many counties delegate this responsibility to the provider agencies with whom they contract.

Additional information could be obtained from the provider programs through the service coordinator who works with the family, if appropriate permission can be obtained from the parents and program administrators. Data, such as whether family members are employed, child care needs, who provides child care, where services are delivered, levels of parent participation, dropout rates, etc. could be obtained from individual

programs. Also, when the state fee schedule is initiated and ability to pay determinations are made, programs will have income data on the families served.

Numerous programs could be contacted about collecting information for a potential future study. This author has contacted a number of program managers who have expressed interest in potentially participating in a study of the effects of W-2 on families with young children with disabilities. The effects of W-2 on families and the services that Birth to Three programs provide have been an expressed interest of numerous Birth to Three program staff. They would serve as motivated agents for data collection. The information they obtained could also be expected to be reliable because of their close relationships with families and the high likelihood that a trusting relationship exists.

Two potential counties to participate in a study are Milwaukee and Dane. These two are selected because they serve a large number of families with young children with disabilities, have a fairly large population of ethnically diverse families that will be affected by upcoming welfare reform developments, and provide a range of Birth to Three provider structures and organizations. If rural counties are also of interest, a number of them would have to be selected to gain sufficient numbers of families receiving Birth to Three services. For example, Adams County serves six families at any point in time. Clearly, additional counties could be selected to get better representativeness of the rural areas of the state. Below is some general information on the two selected counties:

**Milwaukee County**, with a population of 951,884 (estimated in 1992), serves approximately 850 children on a point in time basis and approximately 2000 children annually. Estimates are that more than 80 percent of the families are MA eligible and receiving AFDC while the overall percentage of families below the poverty income level in the county is 12.6 percent. The Birth to Three program is administered by the Milwaukee Human Services Office and contracts with approximately 17 providers serving the county's families. The impact of W-2 may be expected to be strongest in Milwaukee because of the large number of families on AFDC that participate in the Birth to Three program. The majority of services in Milwaukee are conducted in center-based programs that also provide child care services as children are attending their Birth to Three program.

**Dane County**, with a population of 380,306 (1992 estimate) and 4.8 percent of families below the poverty line (1995 data, Kaplan & Brown, 1996) serves approximately 280 families at any point in time and more than 400 annually. The county has two Birth to Three providers, one for metropolitan Madison, Bridges for Families, which serves approximately 150 families at any one point in time and 250 during the year. The other program, Connections, serves families residing in the rest of the county. This program serves slightly

fewer families. Of the Madison families served by the Bridges program, approximately 50 percent are on AFDC.

#### POTENTIAL EFFECTS OF W-2

Given the additional demands on families' financial, time and emotional resources that having an infant or toddler with a disability places, it is imperative that the impact of W-2 on this specific and especially vulnerable population be carefully evaluated. The Wisconsin Council on Developmental Disabilities estimates that of the approximately 7000 families who have a very young child with a diagnosed disability and are participating in the Birth to Three Program, 2250 will be required to participate in the work requirements of W-2. The largest percentage of families (approximately 1500) will be in Milwaukee.

Families with a child with a disability typically have additional costs for special foods, clothing, and equipment. Health care is typically a major concern for many families with a child with disability. Many children have conditions that require frequent medical attention, such as children who may be gavage fed, have a tracheostomy for breathing, or may be at home on either oxygen support or heart and breathing monitors. Often parents learn specific procedures that are necessary for the care of the child. Adequate and affordable health insurance is an ongoing and persistent challenge for many families. Child care, in many cases, poses a special challenge to find willing and qualified providers who can meet the needs of the child with a disability. Last, parents of children with disabilities need greater flexibility in work situations, because they often have to take their children to medical and therapeutic appointments for assessment, care, or intervention. Children with disabilities may be more susceptible to bacterial and viral infections so that parents may need more excused time from work to take care of a sick child or to meet with health, social service, or early intervention staff.

The manner in which the W-2 program is implemented and the degree of flexibility that is available to parents of children with disabilities will in large part determine the degree to which parents are able to balance the demands of specialized care for their child and the imperative to work. Significant impacts are likely in four areas; a) ability of family to remain intact, b) availability of parents in attending to the special needs of their child in working with Birth to Three program personnel, c) the physical and developmental well being of children, and d) increased demands on community child care systems and individual providers to include children with disabilities. Each of these potential impacts and possible measurable indicators are described below.

Ability of Family to Remain Intact

Parents who are needed at home because of the illness or incapacity of another member of the family will be placed at the W-2 Transitional Level. The amount provided at this level (\$628/month plus food stamps) may not be sufficient to care for the child with a disability or special health care need and other members of the family, especially in larger families. With the additional costs of raising an infant with a disability, parents may feel forced to choose to place their child outside the home (foster care, adoption, child care institution) in order to be able to meet the needs of the other children in the family. And new mothers may be much less willing to bring children home after lengthy hospitalizations who require intensive health care maintenance if they have to return to work 12 weeks after the birth of their child.

**Potential Indicators:**

- Frequency of or requests for out-of-home placements such as foster care, child care institutions, or adoptions
- Increased number of babies abandoned at hospitals, especially in neonatal intensive care units

Parent Availability to Participate in Services/Programs for the Child with a Disability

Because of work schedules parents may have less time to spend with their children and may have less time available to address their specific needs associated with a disability. Parents may be unable to make the appointments with early intervention staff or medical or therapeutic staff. Parents may also become less in touch with the developmental progress of their child and be less aware that a developmental delay even exists.

**Potential Indicators:**

- Changes in delivery of services models with trend toward center-based and away from home-based services
- Decreased rates of attendance at IFSP meetings, transition meetings, parent education opportunities such as support groups and informational meetings
- Decreased rates of follow-through on clinic appointments, well baby checks, therapy sessions and reevaluation
- Decreased implementation of developmental intervention strategies proposed by Birth to Three staff
- Lower rates of referrals from parents and possibly even public health nurses and pediatric clinics and increased rates of referrals for child care providers
- Increased rates of children dropping out of the Birth to Three program
- Increased rates of eligible families rejecting Birth to Three services
- Decline in rates at which IFSP goals are achieved based on annual reviews

### Maintaining the Health and Well-Being of a Child with Disabilities

Group child care can be a significant benefit for many children, but not for all. Placing medically fragile children in child care situations may expose a significant number of infants to increased health risks. Quality infant care is difficult to find under the best of conditions. Quality infant care for infants and toddlers with disabilities and special health care needs is rare in most of Wisconsin's communities. The lack of quality care and potential decreased attention of parents may put the health of many children at risk. Changes in health care for W-2 participants may limit the amount and extent of health services that children and families receive. Additionally, adding stressors to already stretched and struggling families without the provision of supports may result in increased child abuse incidents.

#### **Potential Indicators:**

- Increased hospitalizations for preventable illnesses
- Increased emergency hospitalizations
- Lower immunization and well-baby visit rates for young children
- Increased reports of unsubstantiated and substantiated child abuse and neglect
- Increased requests for out-of-home placements

### Changing Demands on the Child Care System

This area may prove to be the most difficult for families with children with disabilities and special health care needs. Even though laws such as the Americans with Disabilities Act (ADA) prohibit child care providers from discriminating against children with disabilities, many centers and family care homes are reluctant to take children with disabilities either because of perceived or actual additional time that these children may take or because they feel they do not have the training or knowledge to effectively care for children with special needs. Milwaukee personnel estimate that more than 1000 additional child care slots will be needed for children with disabilities once W-2 is in full operation.

#### **Potential Indicators:**

- Increased numbers of children with disabilities included in typical child care settings
- Increased referral calls to established information and referral resources for appropriate placements
- Longer waiting lists at established providers
- Increased violations of licensure regulations for child-staff ratios and overcrowding at existing centers

- Increased demand for training and support by child care providers for meeting the needs of children with disabilities

## POSSIBLE METHODS FOR STUDYING IMPACTS OF W-2

Three basic approaches are proposed to obtain information on the impact of W-2 welfare reforms on families and young children with disabilities. These include (a) a case study approach, (b) a panel study of a cohort of families experiencing the changes, and (c) collection and analysis of existing data from selected program sites. The following section will briefly discuss each option.

### Case Study

A case study approach can be used to understand any number of aspects of the impact of welfare reform on the families' routines and resource utilization, and impact on the care of the child with a disability. A case study approach would provide a depth of information from which a greater understanding of the complicated factors at play could be obtained, and variables for potential future studies could be identified. Cross-sectional descriptive case studies where families are selected based on blocking variables such as type and severity of disability, size of family, and need for services could be used to obtain both qualitative and quantitative information on the differential effects of various elements of W-2.

A variation of Yin's (1994) case study methodology could be used to obtain both process and outcome data associated with the care of children with disabilities, work requirements of W-2, and changes in health care availability. According to Yin, case studies are the "preferred strategy . . . when the focus is on a contemporary phenomenon within some real-life context" (p. 1, emphasis added). A multiple-case-study design could be used, with families selected based on the principal variables associated with potential impacts of W-2. Four of the main blocking variables could include severity of the disability measured by level of delay or level of need for services, size of family, ethnicity, and length of time on previous AFDC program. A common case study protocol format could be used for the conduct of the case studies across program sites.

### Panel Study

A second option is to develop a panel study of families in selected program sites. Families could be surveyed either through structured personal or telephone interviews to gather data on factors such as level of care for the child with a disability, use of specialized services, child care arrangements, child health status, family well-being, participation in the early intervention program, amount and intensity of therapeutic services,

typical daily routines, and transition process into programming after early intervention. These same factors could be sampled in six to eight month intervals over the coming two or three years to establish short- and longer-term trends of change within families and the manner in which the child with disability's development is nurtured. Impacts of various elements of the W-2 program as parents go through training and the various employment options can be monitored along with the progress of the child and the services that he or she receives.

Personnel to conduct the panel study could be selected either from outside or within the early intervention programs. Service coordinators are a possible choice. They could be trained and reliability rates could be established. Personnel hired from outside of the program may be seen as more objective, yet may have greater difficulty in obtaining information from families participating in early intervention who may be reluctant to speak openly to unknown interviewers. Service coordinators typically have well-established relationships and rapport with families in the program.

#### Review of Records

A third, less comprehensive, approach would be to conduct a review of records with existing information. Programs that have sizable populations that will be affected by welfare reform could be selected. Past records could be reviewed to determine baseline data on rates of a series of program participation and family indicators. Possible data elements include rates of referral, program dropout, refusals of services, "no shows" for therapy and home visits, suspected child abuse reports, referral for out-of-home placements, attendance at IFSP, suspension of services because of hospitalizations, parent meeting participation, and intervention outcome goals achieved. Once baseline information on data elements has been established, these same factors could be tracked to identify systematic changes over time. The advantage of this approach is that it is much less intrusive and costly, but it suffers from a lack of depth, since all data elements are secondary indicators of what may actually be occurring within families. Another strategy to fill in the details of the impact of W-2 would be to additionally interview service coordinators and other providers using critical incidence methods to obtain more information on what they observed happening in families and their utilization of services as they were going through changes brought about by W-2

#### Summary

Each of the approaches has merits and disadvantages in the scope and depth of data that would be obtained. The amount of human and financial resources available will clearly present parameters for the

breadth of study possible. Given the interest that W-2 has generated among both parents and providers of Birth to Three Early Intervention services, there would likely be a great deal of motivation by program staff to participate in any evaluation of the impact of W-2 on families and the service system. Birth to Three staff could serve as a valuable resource to obtain information from parents and their communities.

The children and families enrolled in Birth to Three programs are a well-defined population who have regular and sustained contact with services and providers in their communities. Families who have eligible children and are in poverty are in an especially vulnerable state, facing multiple needs and challenges simultaneously. The study of the impact of W-2 on these families is critical to help assure that public policy designed to optimize the chances for financial independence does not have the exact opposite effect by mitigating the impact of community developmental, medical, and social support efforts and endangering the continued development of very young children with special needs.



### References

- Begun, A. L. 1996. "Family Systems and Family-Centered Care." In P. Rosin, A. D. Whitehead, L. I. Tuchman, G. S. Jesien, A. L. Begun, and L. Irwin, eds., *Partnerships in Family-centered Care: a Guide to Collaborative Early Intervention*. Baltimore: Paul H. Brookes Publishing Co.
- Erwin, E. 1996. *Putting Children First: Visions for a Brighter Future for Young Children and Their Families*. Baltimore: Paul H. Brookes Publishing Co.
- Hoffman, C. 1996. "Data on Young Children with Special Needs." Unpublished ms. Wisconsin Council on Developmental Disabilities, Madison, WI.
- HSS 90: Early Intervention Services for Children from Birth to Age 3 with Developmental Needs. Wisconsin Administrative Code. Register, June, 1993, No. 450.
- Individuals with Disabilities Education Act (IDEA) of 1990, PL 102-119. (October 7, 1991). Title 20 U. S. C. 1400 et seq; *U. S. Statutes at Large*, 105, 587-608.
- Jesien, G. 1996. "Challenges for the Future of Early Intervention." In P. Rosin, A. D. Whitehead, L. I. Tuchman, G. S. Jesien, A. L. Begun, and L. Irwin, eds., *Partnerships in Family-centered Care: A Guide to Collaborative Early Intervention*. Baltimore: Paul H. Brookes Publishing Co.
- Kaplan, T. & Brown, P. 1996. *Wiskids Count Data Book: A Portrait of Child Education in Wisconsin by School District*. Wisconsin Council on Children and Families, Madison, Wis.
- Smith, B. J. , & McKenna, P. 1994. "Early Intervention Public Policy: past Present and Future." In L. J. Johnson, R. J. Gallagher, M. J. La Montagne, eds., *Meeting Early Intervention Challenges: Issues from Birth to Three* (2nd ed. , pp. 251-264). Baltimore: Paul H. Brookes Publishing Co.
- Trohanis, P. 1994. "Continuing Positive Changes Through Implementation of IDEA." In L. J. Johnson, R. J. Gallagher, M. J. La Montagne, eds., *Meeting Early Intervention Challenges: Issues from Birth to Three* (2nd ed. , pp. 167-182). Baltimore: Paul H. Brookes Publishing Co.
- Turnbull, A. P. & Turnbull, H. R. 1990. *Families, Professional and Exceptionality: A Special Partnership*. New York: Merrill Publishing.
- Yin, R. K. 1994. *Case Study Research: Design and Methods* (2nd ed. ). Thousand Oaks. Calif.: Sage Publishing Co.
- Zigler, F. E. 1995. "Meeting the Needs of Children in Poverty." *American Journal of Orthopsychiatry* 65(1): 6-9.



## Welfare Reform and Education

John F. Witte

### I. INTRODUCTION

This short paper outlines the potential effects and issues linking welfare reform and the education of poor children. Because the policy regime we are trying to anticipate is not clear, and because many of the educational effects of welfare reform will be indirect, the paper discusses a wide set of potential research issues. The issues and questions will vary from state to state as reform efforts vary. One way to organize and understand these complex and imbedded issues is to consider issues which affect the student and family, then those affecting schools and districts, and finally those affecting communities and state educational systems.

#### Relevant Comparison Groups

To what, and to whom, should the new welfare reform regime be compared? The eligibility of families for welfare will change, and that changing population could be tracked. The low-income population (families below the poverty line) could also be used as a reference group. Some of those families would have been eligible for AFDC under the old regime and some not. Finally, comparisons could be made to matched sample units (schools, districts, etc.) in a state or locality not undergoing reform.

For the purposes of this proposal, however, the major comparison is to “pre-eligibles,” or families who were eligible for AFDC before the reforms. Thus the primary purpose of this study is to understand how the circumstances for these families change. Even though reforms are assumed to tighten eligibility, some new families will be eligible for some services under some reforms. These newly eligible families could also be tracked, but I suggest that they form a secondary population group (and I do not believe it will be a large group in most states). The impact of reforms on the entire low-income population could also be studied, but, for reasons outlined below, that seems very difficult. Given the research problems, if we could describe the effects on pre-eligibles we would make a major contribution to understanding the effects of welfare reform on the education of affected children.

#### Definitions

The following terms will be used in this paper:

**Welfare Reform:** Reforms of the current AFDC system are assumed, at a minimum to include:

1. A mandatory work requirement;

2. A monthly limit on benefits, both over a continuous period (e.g., 24 months) and over a lifetime (e.g., 5 years);
3. Increased support in terms of employment, and health and child care benefits.

**Pre-eligible:** Families and children who were eligible for AFDC, SSI, etc., prior to the enactment of welfare reform.

**Educational achievement:** The knowledge, skills, and educational capabilities of students as measured by standardized achievement tests, grades, authentic assessments, and other indicators of learning.

**Educational attainment:** Number of years of education, graduation rates.

**Family Resources.** Standard indicators of socioeconomic status (income, occupation, and education) plus time and family capacity to devote to student education.

## II. EDUCATIONAL EFFECTS ON CHILDREN AND FAMILIES

The most consistent set of variables predicting educational achievement and attainment are those related to the socioeconomic status (SES) of families. Although nationally representative educational data bases are often not disaggregated to study SES effects within low-income populations, studies of inner-city school districts confirm the general effect (Witte and Walsh, 1990). Thus the anticipated welfare reform effects on family income, family stability, and parental employment are likely to affect the educational achievement and attainment of children of pre-eligible families and teenage pre-eligibles. The most obvious hypothesis would be that lower levels of family resources (including time, as defined above), would adversely affect educational achievement and attainment. However, as suggested by point 8 below, the opposite must also be considered.

This general hypothesis suggests a number of more refined issues:

1. Will educational outcomes be affected at all? Because education will be indirectly affected by welfare reform, and because education outcomes may already be quite low for this population, it may be that education is not affected by welfare reform.
2. What are the changes in resources devoted to education and what explains the change? For example, do some families, when faced with shrinking resources, shift resources to human capital investment in an effort to retard educational decline? Or (more likely?) are resources devoted to education, especially parental time, reduced?
3. Assuming we can measure the adjustment of family resources devoted to education:
  - a. What is the magnitude of the reduction?
  - b. What factors explain various strategies?
4. If educational achievement and attainment decline, what are the explanations?

- a. Are parent expectations changed?
  - b. Is parental involvement affected?
  - c. Is the importance of education in the family diminished relative to other factors?
  - d. Are children under less parental supervision?
5. Will welfare reform affect pre-school attendance and specifically Title 1 programs for disadvantaged children? How will this effect vary with work requirements for welfare recipients?
  6. How will welfare reform affect after-school situations for both elementary children and teenagers? Will working parents compound the “latch-key” problem? Will reforms affect after-school “activities” for teenagers?
  7. Will welfare reform affect household mobility? Does that affect student mobility between schools? Does student mobility affect achievement and attainment?
  8. Is it possible that reduced resources and required work enhance the importance and value of education? The argument is that when the assured subsidy is no longer available, parents will reassess the importance and value of education for themselves and their children as the path to stable incomes and future prosperity.

### III. EFFECTS ON SCHOOLS AND SCHOOL DISTRICTS

The effect of welfare reform on schools and school districts will also be indirect. The most obvious question is: If resources decline for families in poverty-concentrated school districts, will there be negative consequences for these schools and districts? This question assumes that as resources for families in the community decline, the resources available to school districts and schools will also be diminished. It further assumes that resources directly affect student outcomes—an assumption contested by some researchers (Hanushek, 1994).

This general issue can be broken down into more precise research questions:

1. Will Chapter 1 enrollment increase?
2. Will there be an impact on the number of students designated with disabilities?
3. Will there be more remedial education in schools? If yes, what will be the impact on general education funding and staff hiring, morale, and turnover?
4. Will the SES composition of poverty-concentrated school districts be further eroded through middle-class flight? Will there be an increase in the segregation of schools and districts in terms of race and class?
5. Will schools experience greater student mobility? What will be the consequences for curriculum, school autonomy, and teacher morale?

6. How will the changes impact the financial base of public schools? Which school districts will be winners? Which will be losers?

The impacts of welfare reform will obviously vary greatly between school districts (and therefore schools). More affluent districts will undoubtedly be much less affected. However, school districts with larger concentrations of poor families may also be affected differently. In many states those concentrations occur in either inner-city or rural districts. Some of the issues above, such as white flight (4) and increased mobility (5) may not be relevant for rural districts. Others may be more important, such as (1). Thus any evaluation strategy needs to separate districts on the basis of underlying demographic and geographic differences.

A generally more positive scenario is also possible. Over time, as welfare reform reduces dependency and adds to the employment base, inner-city districts could become economically more viable, creating a more solid financial base and a reduction in the dysfunctional educational outcomes outlined above. The time frame for this possibility is unclear, and thus the more immediate impacts are likely to be the negative results hypothesized above.

#### IV. EFFECTS ON COMMUNITIES AND STATES

Consideration of the educational effects of welfare reform on communities and states becomes even more speculative, because the causal links involve more layers and more complexity. In addition, the time frame in which effects might show up is longer, thus stretching out the evaluation and making causal inferences more difficult. However, these effects cannot be ignored. For example, in nearly all states, either property wealth or income is a major factor in state education aid formulas. As property values and income shift, so does the state commitment and the distribution of aid between districts. Because equalization of expenditures on education between districts remains a critical policy and legal issue, the effects of welfare reform could be dramatic.

Additionally, one of the longest standing policy issues in education is the integration of schools. Although there appears to be almost a surrender on that front in terms of public policy (Orfield and Eaton, 1996), welfare reform and other policies affecting urban areas may nonetheless have an effect on the increasing segregation of residential neighborhoods and schools.

Finally, education has long tentacles in communities. There is very solid evidence that education achievement and (more so) attainment affects economic viability, property values, crime, gang activity, etc. Welfare reform may directly affect these conditions, but there may also be a residual, and lag effect coming through education.

These potential outcomes suggest several concise research questions:

1. Will state financial support of schools be affected? Will special aids be necessary for heavily affected districts?
2. Will there be an effect on the residential segregation of the community?
3. Does the change in educational achievement and attainment have an effect on the economic viability of the community? On crime rates? On gang activity?
4. Will these financial changes lead to more centralized, state-centered control of education?

#### IV. RESEARCH AND DATA PROBLEMS

Establishing causality between changes in welfare programs and changes in education systems, achievement, and attainment will obviously be difficult. As welfare reforms are implemented (and this process is not simple to monitor), policy and institutional changes that have nothing to do with welfare reform will be occurring in the educational arena. Further, the linkages between any of these changes and decisions and behaviors within families, schools, and communities will be difficult to establish. Thus the evaluative tasks outlined above will not be simple to analyze.

However, if efforts are not made to track these changes, then the political null hypothesis—no effect—may be accepted as reality. Thus an evaluation of the effects of welfare reform on education is necessary. Two data acquisition strategies are needed, and each has its limitations.

##### Monitoring Trends in Basic Aggregate Education Data

Education trends will have to be monitored at the school and district level. In most states, data will have to be acquired directly from school districts. Outcome data will include achievement test data, drop out rates, grade points, postsecondary enrollments, etc. Data will also be necessary for a number of “independent” variables: race, gender, free-lunch qualification, welfare recipient, disability status, “at-risk” status, ESL status, and perhaps something on family structure. Nearly all school districts will have those data, broken down by school. The problem will be causally linking welfare reform to changes in educational outcome measures. School districts are in constant flux, and establishing tight links to welfare reform will be difficult.

##### Family-Based, Longitudinal Studies

A longitudinal study monitoring the effects of welfare reform on education, as well as other aspects of the lives of pre-eligible families will allow us to begin to establish some of the causal linkages unavailable through the aggregate trend data. However, monitoring family decisions, resource allocations affecting education, and other relevant family conditions is not easy. A simple interview or phone survey panel study may not suffice. More intensive ethnographic or tracking studies may be needed.

## V. RESEARCH DESIGN SUGGESTIONS

Welfare reform provides some advantages in terms of research design and data collection. First, we know, more or less where poverty families live and are concentrated. Thus the impact of welfare reform on education is likely to be local, because of the concentration of pre-eligibles in a few locations in most states. In Wisconsin, for example, half the current case load is in Milwaukee and about two-thirds of the children in the Milwaukee Public School District are in pre-eligible families. Second, prereform data on welfare, employment, and education are likely to be available so that comparisons can be made over time.

This concentration of welfare reform effects suggests a geographic evaluation design focusing on a small number of sites. There will obviously be a trade-off in terms of generalizations, but without in-depth analysis we will never overcome the causality problems and may not even be able to describe adequately the primary effects of welfare reforms on student education.

Site selection should vary by the size of the community and the concentration of poverty. At a minimum, two sites should be selected:

1. A medium-sized city (500,000–750,000). Larger cities make problems of causal inference more difficult and exasperate data collection. In Wisconsin, Milwaukee would be the city.
2. A rural-poverty site (or sites) in the same state. I would anticipate quite different effects in the two sites.

If resources permit, middle-sized cities and towns and additional rural sites could be added. Regardless of the number, my suggestion would be to focus the research on a few sites to be studied in depth rather than many sites to be simply “monitored.”

Within these sites, a number of research strategies should be used. As indicated above, district and school-level data would be critical to assessing the overall effects on students, schools, and districts. Also essential would be family-level data, trying to capture the changing choices and behavior affecting family resources and approaches to education. Longer-term results on communities and the state would require data



on the relevant variables at those levels. In the short term I would focus data collection and research on students and families and on schools and school districts.

Student and family data would probably require a survey- or interview-based panel study, coordinated with individual student data acquired from schools. If the surveys were to begin before the reforms were put in place, we could record prereform behavior (of students and parents), resource allocation (of both children and parents), parental expectations, and student achievement and aspirations. The panel would track these into the post-reform period. School outcome measures for the panel would follow a similar pattern using the school/district-level data described in the last section.

The types of data available through districts and schools are outlined above. School districts provide good outcome measures, which vary by school level, but relatively poor “independent” variables on students and families. For that reason, most national studies (High School and Beyond and the National Education Longitudinal Study) have to link district and school data with family and student data. In these studies, data are acquired by selecting a school sample and a student sample and then interviewing parents, teachers, and administrators and acquiring district and school-level data to match the student. I doubt the resources would be available to make these links on anything other than a site-specific basis. That is another reason for the design strategy indicated above. However, freedom of information restraints and other legal considerations may make the data linkages between families and school districts problematic even with that form of targeted design.

Finally, there is consideration of an appropriate timeline for this research. The natural inclination is to suggest a long timeline, so that impacts can be recorded and long-term outcomes analyzed. This may be particularly important for young children, because quantitative measures of achievement and attainment are generally minimal before middle or high school. However, given the necessity for relatively short-term feedback, limited resources, and limited political patience, a study ranging from five to seven years may be the outer limit. If that is the case, the studies should be designed to analyze grade modules (e.g. pre-kindergarten through grade 4; grade 5 to grade 8; and grade 8 to grade 12). This will allow one to understand effects at the relevant levels. The trade-off is that one would not be able to analyze a full developmental model.

**References**

- Hanushek, Eric A. 1994. Making Schools Work: Improving Performance and Controlling Costs. Washington, D.C. : The Brookings Institution.
- Orfield, Gary and Susan E. Eaton. 1996. Dismantling Desegregation: The Quiet Reversal of *Brown v. Board of Education*. New York: New Press.
- Witte, John F. and Daniel Walsh,. 1990. "A Systematic Test of the Effective Schools Model." Education Evaluation and Policy Analysis 12 (Summer): 188–212.

## **Welfare Reform and Child Welfare Services: Issues of Concern and Potential Evaluation Strategies**

Mark E. Courtney

The possible impact of welfare reform, including Wisconsin's W-2 program, on the child welfare services system is a cause of concern. There has been considerable speculation about the potential of welfare reform to result in an increase in child maltreatment and the movement of large numbers of children into out-of-home care (OOHC) settings (e.g., foster homes, kinship foster homes, group homes and children's institutions). Little empirical evidence exists, however, on which to base firm predictions. The poor understanding of how the child welfare system and its clients will fare in the new era raises a number of questions. In what ways might welfare reform affect the demand for child welfare services? What existing data sources might be used to try assess such effects? What measures beyond the use of existing data might need to be taken to ensure an adequate evaluation of welfare reform's impact on the child welfare system? In this paper I try to answer these questions using Wisconsin's W-2 proposal as an example of welfare reform, and in the process, to inform the discussion of the evaluation of state-level welfare reform in general.

### **WHAT ARE THE POTENTIAL EFFECTS OF W-2 ON CHILDREN AND FAMILIES AND HOW MIGHT THEY BE EXPECTED TO HAVE AN IMPACT ON THE CHILD WELFARE SYSTEM?**

W-2 is likely to significantly change the circumstances of the children of poor families in Wisconsin in ways that are difficult to predict at this time. Some children may benefit if W-2 contributes to improving the financial condition of families by moving parents into the workforce. In contrast, some children may suffer if W-2 leads to the cutoff of benefits to families who fare poorly under the program's work requirements or if families are unable to acquire adequate child care. Several issues stand out as important to consider in developing a method to assess the consequences of W-2 for Wisconsin's children.

#### Trends in Out-of-Home Care in Wisconsin and Their Relationship to the AFDC Program

Since 1988 the statewide OOHC caseload has grown by 77 percent, from 4,891 in 1988 to 8,649 at the end of 1995.<sup>1</sup> About four-fifths of this growth has taken place in Milwaukee County, which now accounts

---

<sup>1</sup>Mark E. Courtney and Linda Park, *Out-of-Home Care in Wisconsin: 1995* (Madison, Wis.: School of Social Work, University of Wisconsin, 1996).

for about half of all placements in Wisconsin. These numbers do not include unlicensed kinship foster care, which best estimates suggest accounts for about 1,500 placements in Milwaukee County. Over half of children in OOHC in Wisconsin come from families that were eligible for AFDC at the time the child was placed in such care. If Wisconsin is similar to other states, the vast majority of children entering OOHC whose families are not eligible for AFDC come from families with near-poverty incomes. In other words, children in OOHC come almost exclusively from poor or near-poor families. This is not to say that all, or even a large proportion, of poor families maltreat their children. In fact, in Wisconsin fewer than one in twenty families who receive AFDC have a child placed in OOHC.

The rapid growth in OOHC has taken place during the same time that restrictions have been placed on public assistance programs in Wisconsin and that AFDC caseloads have been declining. It is unclear what role, if any, AFDC program changes and benefit reductions have played in the growth of OOHC during this period. It is clear, however, that caseload reductions in public assistance programs in Wisconsin have not led to a commensurate decrease in the number of children needing to be placed out of their homes.

Although only a small proportion of families receiving public assistance abuse or neglect their children, there is a clear association between material deprivation of families, child maltreatment, and the placement of children in OOHC. The incidence of abuse and neglect is approximately 22 times higher among families with incomes below \$15,000 per year than among families with incomes over \$30,000 per year.<sup>2</sup> Moreover, a time-series analysis of the number of referrals to child protective services in Los Angeles County that controlled for the overall trend and seasonal variation in reports found that a 2.7 percent cut in benefit levels in 1991 was associated with an increase of about 12 percent in the monthly number of protective services referrals, while a 5.8 percent cut in AFDC benefits implemented in two phases during late 1992 was associated with an approximate 20 percent increase.<sup>3</sup> The relationship between poverty and child maltreatment is most pronounced in the area of child neglect, which is the most common form of child maltreatment and the primary reason children are placed out of home. Thus, if W-2 or any other welfare reform program results in a significant increase in poverty, it is likely to contribute to increased demand for child welfare services. Conversely, lifting families out of poverty would be expected to reduce the demand for such services.

---

<sup>2</sup>The 1993 National Incidence Study of Child Abuse and Neglect provides data on the relationship between family income and the incidence of child maltreatment. See press release from the U.S. Department of Health and Human Services, "Survey Shows Dramatic Increase in Child Abuse and Neglect," September, 18, 1996.

<sup>3</sup>Robert E. Sherman, unpublished report to Norm Zimlich, of the Institute for Human Services Management (Seattle, Wash.: Institute for Human Services Management, April 5, 1995).

### Changes in the Economic Well-Being of Families

The impact of W-2 on the economic well-being of families is likely to vary depending on the circumstances of families. In some cases W-2 will result in a net improvement in family resources due to enhanced earnings combined with government support for child care and health care. These benefits will most likely accrue to families with relatively well-functioning wage earners and relatively fewer children. In other cases, W-2 will result in a decrease in family resources. Even full participation in W-2 subsidized jobs, community service jobs, and W-2 Transitions activities will result in a net decrease in income for large families. More important, some parents will be unable or unwilling to participate in required W-2 activities. These families will be “sanctioned,” resulting in a loss of income. The time-limit provisions of W-2, if fully implemented, will also eventually result in some number of families being denied assistance.

### Impact on Children of Work Requirements of W-2

Many parents who would not work at all or would work part time under current public assistance arrangements will choose to work full time because of the incentives/sanctions structure of W-2. The impact on children of this change in parental involvement in the work force is difficult to predict but will almost certainly vary depending on family circumstances.

Some parents will adapt well to work. This could lead to an improvement in parental self-esteem and parenting behavior. It might also serve as a source of improved self-esteem for the children in these families. In some cases, children placed in child care when their parents are working may receive better care than they did in their own homes. Families with strong support systems to help with child care responsibilities (e.g., extended family and friends) are likely to cope best with the added responsibility of work.

Other parents will have difficulty coping with the combination of parenting responsibilities and work. Some parents are marginal caregivers when providing full-time care of their children. Many of them may neglect or abuse their children when faced with the combined stress of parenting and work. In some cases a general lack of parental supervision may lead children to experience decreased school performance, be subjected to extrafamilial abuse, or engage in delinquent behavior. In addition, some parents may have difficulty finding safe and affordable child care for their children. Isolated parents with poor or nonexistent support systems will probably fare worst in balancing the demands of work and parenting. (Of course, as mentioned above, some parents will simply fail to work and will be denied benefits under W-2.)

### Consequences for Children of the Apparent Contradiction between the Goals of W-2 and Those of the Child Welfare Services System

W-2 and the recently enacted federal welfare reform legislation are primarily concerned with encouraging parents to work instead of relying on public assistance. Families are rewarded for work both financially and through government support for child care and health care. When parents refuse to work or do not appropriately participate in work activities (e.g., are late for work), W-2 sanctions families through denial of benefits. The purpose of the child welfare system is to protect children from maltreatment while preserving the integrity of families whenever possible, based on the assumption that a child's family is generally the best place for a child to be raised. Though material deprivation often contributes to the circumstances bringing families in contact with the child welfare system, poverty alone is not grounds for placement of children in OOHC. Nevertheless, in the process of trying to preserve families, the child welfare system often provides support to families in the form of targeted financial assistance (e.g., help with paying utilities or rent), in-kind benefits (e.g., child care) and various social services. Whatever its limitations, the cash-based public assistance system complemented the child welfare services system: it provided minimal financial support to poor families regardless of whether or not they chose or were able to work.

In contrast, provision of assistance to poor families by the child welfare system may be seen to undercut the work-related sanctions of W-2, even though the present child welfare system does not have the resources to give ongoing financial support to poor families. Moreover, employees of W-2 agencies which are working under performance contracts may see the child welfare system as a vehicle for ridding themselves of responsibility for parents who are poor prospects for work.

It is impossible to anticipate at this time how the child welfare system may evolve in response to rapid changes in the social safety net for families, or how the managers of W-2 programs will attempt to make use of child protective services. If W-2 results in a noticeable increase in child poverty and neglect, political pressure may result in increased intervention to protect children through placement in OOHC. Alternatively, in the absence of new resources, the system may be forced to narrow its grounds for intervention and respond only to extremely desperate situations. Given Wisconsin's decentralized administration of child welfare programs, the response to change is likely to vary widely from county to county.

### The Uncertain Role of Kinship Care under W-2

To date, Wisconsin has made little concerted effort to develop policy in the area of kinship foster care, despite the fact that licensed and unlicensed kinship foster homes account for between 25 and 35 percent of all

OOHC placements in Wisconsin. Currently kinship care appears to account for over half of OOHC in Milwaukee County.

W-2 calls for the development of a new category of kinship care. Presumably this is at least partly a response to the perceived need for out-of-home placement resources for children whose parents are not able to care for them when confronted by the requirements of W-2. Although not specified in the W-2 legislation, current plans call for a kinship care reimbursement rate that is slightly lower than existing payments to non-legally-responsible relatives (NLRRs) under the AFDC program. Proposed procedural requirements for approving kinship homes and reviewing the care of children in these homes are much less stringent than existing permanency planning requirements for children in OOHC.

Changes in kinship care arrangements proposed under W-2 raise a number of questions. For example, what effect will changes in regulation and reimbursement of kinship care have on the supply of kinship care providers in Wisconsin? Will the altered context of kinship care contribute to changes in the number of children entering OOHC? How will children fare in kinship care settings?

There is no way to know in advance if, or how, W-2 will affect the rate of child maltreatment and need for child welfare services. Nevertheless, the issues raised above call for close attention to the child welfare system in evaluating of the consequences of W-2.

#### WHAT DATA SOURCES MIGHT BE USED TO EVALUATE THE IMPACT OF W-2 ON THE CHILD WELFARE SYSTEM?

Existing management information systems and data systems to be implemented under W-2 could provide for an assessment of changes over time in child maltreatment and the transition of children from poor families receiving public support to out-of-home care. Before summarizing the potential use of these systems for evaluating W-2, a description of the systems and their strengths and weaknesses is in order.

##### Data on Child Maltreatment

Wisconsin collects data on reported child maltreatment via the completion of the Division of Community Services #40 form (DCS-40) by county child welfare workers (a copy of the DCS-40 form and a summary of data field codes is found in Appendix A). Unique child identifiers are not recorded on the DCS-40, meaning that it is impossible to link data on child maltreatment reports to data from public assistance, child support, or OOHC data systems. Therefore, about the only possible use of the DCS-40 data for evaluation

purposes would be to measure aggregate change over time in the number of maltreatment reports, and even that possibility is limited by other problems with the DCS-40, including intercounty variation in completion of the form. In particular, the standard for reporting maltreatment as “substantiated” (i.e., meeting the criteria listed in statute) appears to vary from county to county. Moreover, there is some reason to believe that reporting numbers can be affected by resource constraints. For example, preliminary data from the 1993 National Incidence Study of Child Abuse and Neglect indicate that the proportion of maltreated children whose situations were investigated by child protective services authorities declined from 44 percent in 1986 to 28 percent in 1993.<sup>4</sup> In a number of cases media attention to sensational child maltreatment cases has been shown to lead to rapid, though generally transitory, increases in child maltreatment reports. All of these problems call into question the meaning of changes in the number of reports over time.

County-level data could conceivably be used to evaluate in selected counties the impact of welfare reform on the incidence of child maltreatment and the provision of child welfare services to families. Counties sometimes maintain their own child welfare data systems, and some of these systems may have more reliable data than that obtained from statewide data systems. In addition, child maltreatment data in some counties could include child identifiers, facilitating the linkage of data on child maltreatment to data on OOHC and/or family participation in public assistance programs. Last, some county data systems collect information on purchased services that are provided to families during their involvement with child protective services. Unfortunately, there do not appear to be any counties in Wisconsin with data systems that improve in any significant way on available state child maltreatment data systems.<sup>5</sup>

#### Data on Out-of-Home Care

The state collects data on out-of-home care placements via the completion of the Human Services Reporting System (HSRS) Substitute Care module by county child welfare workers and clerical staff (a copy of the HSRS substitute care module and data field codes is found in Appendix B). All out-of-home care placements, with the exceptions of placements in secure detention facilities and unlicensed kinship care, are recorded in the HSRS data, but licensed kinship care placements are not distinguished within HSRS from foster

---

<sup>4</sup>U.S. Department of Health and Human Services, “Survey Shows Dramatic Increase.”

<sup>5</sup>At a Madison, Wisconsin, meeting on September 20, 1996, for Wisconsin county officials interested in evaluating the impact of W-2 on children and families, none of the officials (including several from larger counties in the state such as Dane, Milwaukee, and Rock) believed that any county had a data system capable of child-specific tracking of child maltreatment reports, out-of-home placement histories, or service provision to families and children.



family home placements. (It may be possible to obtain some historical data on AFDC-supported kinship care through the AFDC reporting system). The HSRS system does not generate payments to providers, and there is fairly minimal monitoring of county compliance. As a result, while it is believed by those operating the system that nearly all children who are placed out of home for at least a few days show up in the HSRS system, cases are sometimes not closed in a timely manner, and data fields that are not subject to system edits (e.g., family income and some child characteristics) are not routinely filled out by workers. HSRS could be linked to historical AFDC data. Moreover, HSRS will continue to operate under W-2 and might therefore be used to monitor movement of children from families participating in W-2 into out-of-home care.

As in the case of child maltreatment reports, it is conceivable that county data could be used to fill in gaps in HSRS data (e.g., cost data on services provided to children and families and inclusion of data on kinship foster care placements). Unfortunately, no county has a data system that significantly improves on the limitations of HSRS that are noted above.

## POTENTIAL EVALUATION STRATEGIES

Discussion of the potential uses of existing administrative data to evaluate the impact of W-2 on the child welfare system and its clients illustrates the strengths and weaknesses of this approach. There are two advantages to using administrative data. First, it is a relatively inexpensive and timely source of information. Second, in some cases it makes possible the comparison of data from before the implementation of W-2 with data generated after implementation. In particular, data on child maltreatment reports and OOHC—with all of their problems—are available from the late 1980s onward and will continue to be available in the foreseeable future.

There are several disadvantages to using administrative data to evaluate the impact of W-2 on child welfare services. The statewide data are of somewhat questionable reliability. The OOHC data do not cover unlicensed kinship care and do not distinguish between licensed kinship homes and foster family homes. Considering the prominent role of kinship care under W-2, this is a serious limitation. Most important, a simple interrupted time-series of child maltreatment reports or OOHC entries—the most likely use of administrative data—will be hard to interpret meaningfully. As has already been mentioned, many factors can significantly affect maltreatment reports, and the same can be said for OOHC caseloads. A particular problem in the current context is the recent enactment of significant changes in the Wisconsin child protection statutes that define child maltreatment. Similarly, W-2 is being implemented at the same time that the Wisconsin Department of Health

and Family Services is taking over the operation of the child welfare services system in Milwaukee County. This is not to say that there is no use in monitoring changes in reports or caseloads over time, only that alternative explanations of any observed changes will abound.

Although administrative data can generally be used to monitor certain events of interest in a person's life, they are usually not very helpful in assessing the *reasons* for these events. Herein lies the central limitation of administrative data: learning that abuse reports are up or that OOHC entries are down does not tell us why these things have occurred.

One alternative to the use of administrative data for evaluating the impact of W-2 on child welfare services would be the collection of data over time on the circumstances of families participating in W-2. This could take the form of surveys or interviews. A cohort or panel design would be preferable, so that changes in family and child status could be measured over time. This approach would have the advantage of measuring outcomes in more detail than is possible using administrative data. Furthermore, at least in theory it offers the opportunity to examine the intervening variables that contribute to particular outcomes, a crucial missing link in administrative data. For example, interviews with program participants might show that a substantial proportion of W-2 families have children placed in OOHC, that this occurs almost exclusively among families where parents have been sanctioned due to program time limits, and that the removal happens shortly after sanctions are imposed. Clearly this would be a more interesting, and more policy-relevant, finding than simply noticing that OOHC placements had gone up over time.

A panel study is no panacea. Collection of survey or interview data would be much more logistically difficult and costly than analyzing administrative data. Moreover, conducting a panel study would most likely require adequate administrative data to select and track an appropriate sample. It is very unlikely that a panel study would be undertaken solely to monitor the impact of W-2 on child welfare services. As a result, it is quite possible that any panel created would not be large enough to generate adequate statistical power when it comes to outcomes such as child maltreatment and OOHC. For example, if 5 percent of AFDC families have a child placed in OOHC (over five times the placement rate for the overall population), even a doubling of the OOHC placement rate for W-2 families would still mean that only 10 percent of children in W-2 families are removed from the care of their parents. Thus, it would take a very large sample of families to measure major effects on the child welfare system, let alone medium-size effects. Perhaps most important, a panel study in Wisconsin does not offer the opportunity of comparing outcomes before and after W-2, since W-2 is already being implemented and will be well under way before a panel could be created.

It is important to note that Wisconsin is not necessarily a good model for assessing the impact of welfare reform on the child welfare system, given its relatively poor child welfare data capacity. Some states already have the ability to track children's paths through the child welfare system from child abuse report through a variety of transitions within the system to eventual exit. In some cases these states can also link data from their public assistance systems to child welfare services data. Moreover, federal financial support for the development of statewide automated child welfare information systems (SACWIS)<sup>6</sup> currently exists under the Omnibus Budget Reconciliation Act of 1993 and has been extended for one year under the Work Opportunity and Personal Responsibility Reconciliation Act of 1996. This support is being used by many states to develop much more sophisticated data management systems than those currently in existence, many of which integrate child welfare, public assistance, child support, and juvenile court data. Needless to say, such integrated data systems will increase the potential use of administrative data in assessing the impact of welfare reform on children in general, and on the child welfare system in particular. Unfortunately, Wisconsin has not chosen to develop a SACWIS at this time.

#### A "HAPPY MEDIUM?"

Perhaps the use of a combination of administrative data and selective analysis of additional data on families who receive child welfare services could realize the benefits of administrative data while filling in some gaps. Analysis of entries to OOHC can serve as an example. HSRS data can now be linked to AFDC data to identify children who are removed from families who receive AFDC. Similarly, it is likely that it will be possible to link data on W-2 participation of families to data on children being placed in OOHC from these families. A comparison of the number of OOHC entries among AFDC families to OOHC entries among W-2 families would be more interesting for evaluation purposes than a simple comparison of total OOHC entries before and after the implementation of W-2.

This analysis of administrative data could be augmented by examination of child welfare case records and/or interviews with parents of children placed prior to the implementation of W-2, and, similarly, for children placed after W-2 is under way. Case records, particularly required court reports, contain information about the circumstances under which children are placed in OOHC. Interviews with parents would provide

---

<sup>6</sup>Mark E. Courtney and Raymond C. Collins, "New Challenges and Opportunities in Child Welfare Outcomes and Information Technologies," *Child Welfare*, 73(5): 359-378.

more detailed information, but would be more costly and present some logistical difficulties (e.g., human subjects protections). This approach would facilitate more rigorous testing of hypotheses concerning whether W-2 per se contributes to the movement of children into OOHC than would be possible using administrative data alone. It would be more expensive than the exclusive use of administrative data, but would be much less costly than a panel study, and would be targeted at the population of interest (children placed in OOHC), thus making for easier generation of an adequate sample.

Given the limitations of statewide child maltreatment reporting data, it will not be possible to selectively augment administrative data analysis in order to assess the impact of W-2 on child maltreatment for Wisconsin as a whole. It may be possible, however, to conduct in several counties the sort of analysis discussed above, to the extent that the counties have the capacity to uniquely identify maltreated children in a manner that allows for linking child maltreatment data to AFDC and W-2 participation.

## CONCLUSION

Much concern over the implementation of welfare reform has focused on the possibly negative impact on children. Little attention has been paid to how policy makers and the public will know if children have been helped, or harmed, by the rapid reconstruction of the social safety net. In Wisconsin, at a minimum, administrative data should be used to examine changes over time in the movement of children from families that receive public assistance into OOHC. Depending on resources and political will, it may also be possible to delve more deeply into the circumstances that lead poor children to be removed from their homes, and in the process come to better understand how W-2 might contribute to these transitions.

Appendices A and B of “Welfare Reform and Child Welfare Services: Issues of Concern and Potential Evaluation Strategies”, by Mark E. Courtney, are not available online. These appendices consist of the following two double-sided forms.

### **Appendix A**

Wisconsin DHSS  
Division of Community Services  
DCS-40 (Rev. 1/86)

CHILD ABUSE & NEGLECT  
Investigation Report

Front: form  
Back: codes for use in various fields

### **Appendix B**

Department of Health & Social Services  
Division of Community Services  
DCS-27 (Rev. 1/93)

HSRS Children in Substitute Care Report

Front: form  
Back: codes for use in various fields

## Assessing the Impact of W-2 on Homelessness

Irving Piliavin

There are no data on recent trends in the number of homeless people in the United States or, so far as has been ascertained to date, in the various states across the nation. However, homeless agency officials in Milwaukee and Madison, the two largest cities in Wisconsin, believe that homelessness has risen greatly in the past 8 to 10 months and that the increase is largely due to the elements of welfare reform as these have already been implemented in Wisconsin. Beginning in April of 1996, a period in which demand for shelter care is normally lessened, the use of emergency shelters dramatically increased. Agencies are operating at capacity and daily turn down requests for lodging. The overload results from two conditions: a large increase in the number of requests for shelter and a large decrease among shelter users' ability to find other housing.

The increase in requests for shelter care has come from both families and single individuals. Agencies in Milwaukee experienced increased requests from both groups, and agencies in Madison experienced increased requests primarily from single individuals. The increase in family requests is believed to be due at least in part to the Pay for Performance reform introduced by the state early in the spring of 1996. In Madison, the increase in requests from single individuals is believed to be due largely to the reduction in Dane County general assistance grants, also implemented in 1996. There are, however, no available systematic data to substantiate these views.

The following comments take up three approaches for obtaining data that would permit the study of the relationship between homelessness and experiences with welfare reform efforts. The discussion notes some advantages and weakness in each. Although the comments are not intended to be relevant to the study of welfare reform generally, the specific examples of data availability will be based on information provided by social service personnel in Wisconsin.

The most detailed approach for determining this relationship is through the use of *panel studies* of families who have made use or applied for benefits from these programs. The studies would permit examination, over the period of research, as to how various experiences with programs (e.g., receipt, disqualification, rejection) alter the likelihood of becoming homeless. The panel study design would permit the observation of events typically thought of as implying homelessness, but which are not ascertainable through most alternative designs. These events include the doubling up with friends or families, residence in shanties, and "living on the streets." A possibly major shortcoming of panel studies, however, derives from the likelihood that homelessness will be a relatively rare outcome among applicants or participants in welfare or welfare reform efforts. Thus, reliable estimates of its occurrence will require quite large samples and high costs.

A less expensive, more broadly based but less sensitive approach to obtaining data relevant to the study of the impact of reform efforts on the likelihood of becoming homeless relies on the use of welfare reform and homeless *agency records* to compare the use of shelters across time and jurisdictions among those who apply to or use welfare and welfare reform efforts. Basically this requires access to identifying information (e.g., names, social security numbers) to match recipients of welfare programs and users of homeless shelter programs. There are two major shortcomings of this approach: (1) its inherent inability to assess non-shelter forms of residential instability (e.g., doubling up) and (2) the absence of agencies that collect statewide data on shelter use. Indeed, information currently available indicates that it may be possible to obtain data on shelter use for only some agencies in some communities. On occasion, the agencies may have information pertaining to all shelter users in a given city. More often, it will probably be the case that the information will be limited to certain categories of shelter users (e.g. families) and only for selected agencies.

The latter is the case in Madison and Milwaukee. There are agencies in both cities, the Salvation Army in Madison and the Health Care of the Homeless in Milwaukee, that maintain data on the identity of families receiving shelter care from most agencies in those communities. There is presumably an agency in Green Bay which does the same. Other agencies doing the same may exist elsewhere but in the absence of a central listing, their identity will take a substantial amount of time.

Before describing the information available on the homeless people from the two agencies cited above, it should be noted that both organizations may require that access to client information be conditioned on clients' permission. This may not be uniformly forthcoming.

Milwaukee. The data maintained by Health Care for the Homeless are supplied by six agencies serving the city's homeless families. This information, to be filled out by an adult member of the family unit requesting shelter, includes names, social security numbers, and some background information. The latter includes current receipt of social insurance and welfare benefits, employment status, and a checkoff question presumed to identify the event precipitating the current homeless episode. However, the categories of response to the last question do not make clear whether the loss of eligibility for welfare or W-2 is a precipitating factor. And clearly the response categories do not make clear why there might have been a loss of eligibility. The Health Care for the Homeless does not collect information on those who apply for shelter care but are denied. This failure is important because estimates as to the number and conditions precipitating the crises of homeless families will possibly be in substantial error. There is some possibility that Health Care for the Homeless will be willing to increase the scope of information that it collects for shelter users and that it will also begin to collect information on applicants denied service. However, a necessary condition for these changes would be

that the agency be provided funds to carry out these activities. A copy of the information form maintained by the agency is attached to this report.

Madison. The Salvation Army differs substantially from Health Care for the Homeless in the scope of information it collects on homeless families. It is the gatekeeping agency for homeless families in Madison. Any family seeking agency shelter or vouchers for housing must pass through this agency. The Salvation Army collects much more detailed information on shelter users than does Health Care for the Homeless. In addition to the data collected by the Milwaukee agency the Salvation Army intake report provides data on the number of family members seeking shelter, including children, previous shelter use, medical information, reasons for seeking shelter, current income, and sources of income.

Perhaps more important, the Salvation Army maintains data on applicants for shelter care, whether or not they are accepted. The data recorded include detailed information on family members (e.g., education, race, social security number) and public assistance status. The meaning of some categories of response are at times ambiguous (e.g. reasons for being sanctioned by public assistance agencies are not given). A copy of the information form maintained by the Salvation Army is also attached to this report. It should be noted that the Salvation Army has retained completed copies of these forms for several years, thus perhaps providing opportunity to study in this community the numbers and circumstances of families requesting shelter use over a period that includes a number of important welfare policy changes in Wisconsin. Although the agency has a policy that use of this information requires permission from shelter users, there may well be ways to avoid this requirement for data that are not current.

Finally, if either of the approaches discussed above is not possible, the only data available to researchers on the number and character of the homeless in Wisconsin will be those which simply indicate the *numbers of people receiving and requesting shelter services*. These data will obviously be open to various conflicting and unresolvable interpretations. As such they will be of little value in trying to understand whether and how welfare reform may affect the problem of homelessness.



Example forms attached “Assessing the Impact of W-2 on Homelessness”, by Irving Piliavin, are not available online, but are described below.

**Form 1** (1 page)

Health Care for the Homeless of Milwaukee, Inc.

HCHP Form 1: Client Initiation & Client Change in Status  
HCHP-F1; REV4 (5-82)

**Form 2** (1 page)

Health Care for the Homeless of Milwaukee, Inc.

HCHP Form 1: Client Initiation & Client Change in Status  
HCHP-F1; rev1

**Form 3** (2 pages)

Dane County Emergency Shelter  
Network Intake Form  
305-456 (2/92)  
05/01/1991