EVALUATING WELFARE REFORM UNDER PROGRAM HETEROGENEITY AND ALTERNATIVE MEASURES OF SUCCESS

Luis Ayala
Instituto de Estudios Fiscales / Universidad Rey Juan Carlos
Magdalena Rodríguez
Instituto de Estudios Fiscales

INTRODUCTION[1]

Over the last years, there has been a lively debate on the success of welfare reforms in terms of achieving better results in labor market participation and economic well-being. Most OECD countries have enacted major welfare reform legislation, being the main aim of the implemented changes increasing work incentives. When it comes to public policy discussions of welfare programs, there is no doubt that fostering transitions from welfare to the labor market has been a subject of increasing concern to policy-makers. As a result, there is, in practice, a wide array of options for recipients to participate in work-related activities, like job preparation skills, job search assistance, intensive training, social enterprises and many other actions. These initiatives should increase labor participation, both by improving employment-related skills and by establishing larger incentives for job search and employment.

The heavy emphasis on engaging recipients in work activities has rekindled interest in exploring whether these reforms have given rise or not to higher levels of well-being for participants. U.S. policy changes have been much discussed, with considerable empirical evidence on the relevant outcomes[2]. A substantial amount of research has been devoted to following persons as they left welfare in order to analyze their behavior and well-being after participation in work-related activities. Differential effects are found when considering the results of both work-first strategies –trying to push recipients into the labor market as rapidly as possible– and long-term programs –focused on human capital developments through intensive training and educational opportunities for recipients.

Work-related reforms of welfare policies have also given rise to a considerable European literature. Despite the guaranty of a minimum level of resources still is a corner-stone of the European social model, efforts at getting welfare recipients into the workforce have also produced substantial changes in these countries. There is a huge range of experiences already assessed, like extensive activation programs in the field of Social Security and labor market policies in Nordic countries [Sianesi (2004), and Carling and Richardson (2004)], specific targeted welfare-to-work initiatives in the Netherlands [Van Oorschot (2002), and

[2] For a synthetic overview see Moffitt and Ver Ploeg (2001), Blank (2002), and Grogger and Karoly (2005).

Van den Berg *et al.* (2004)], new policies focusing on low-income families with children combining Social Assistance reforms with earned income tax credits in the United Kingdom [Blundell and Meghir (2002), and Hills and Waldfogel (2004)], or 'insertion contracts' embedded in minimum income programs in Southern Europe [Zoyem (2001), and Ayala and Rodríguez (2006a)].

Despite substantial economic research literature has accumulated, there still are, however, considerable caveats in the evaluation of these policies. Estimating the effects of the changes described above creates real evaluation challenges. Some of these challenges have been adequately addressed by standard evaluation literature. There is, however, a need for research to provide a more complete picture of the advantages and limits of this new type of welfare developments.

First, we have relatively little insight into the different effects caused by the heterogeneity in the measures grouped under the notion of work-related welfare schemes. In many countries, these programs are not mutually exclusive and welfare recipients can participate in different activities. This fact raises complex methodological issues, much different from the standard evaluation literature, mostly focusing on the U.S. programs. As stressed by Sianesi (2001), the basic framework in which a program is administered at a point fixed in time, and individuals are either treated or nor treated does not always works well regarding other welfare schemes. As rigorously argued by other authors, the standard binary treatment model of only two states could be extended to the case of multiple states, but needs to be suitably revised [Imbens (2000), Lechner (2001, 2002)].

Second, there is growing evidence that labor market indicators are not always the best measures for an adequate understanding of the programs' effectiveness. A rapidly expanding literature has focused on earnings and working hours as essential indicators of welfare evaluation. In practice, however, the assessment of welfare reforms crucially depends on the indicators chosen to measure the programs' outcomes. In many countries, current debates about the success of welfare reforms have been obscured by the use of inconsistent indicators of success. As reviewed by Cancian and Meyer (2004), little systematic analysis tackles the question of whether or not different measures of success are capturing the same thing. In the U.S., different authors have posed the question of a sizeable sensitivity of conclusions to alternative ways of measuring the success of the

reforms. The success of the programs can greatly oscillate as independence from Public Assistance, income poverty or material hardship are used as outcome indicators [Blank (2002), Grogger and Karoly (2005), and Meyer and Sullivan (2006)]. Alternative indicators can only lead to similar conclusions if they measure the same type of processes.

This paper aims to present an assessment of welfare reforms under a framework of program heterogeneity and alternative measures of success. We focus on a specific welfare program –Madrid's *Ingreso Madrileño de Integración* (IMI)– comprising heterogeneous subprograms. Recipients can simultaneously participate in very different actions. IMI is a standard program within the complex set of national and regional schemes existing in South Europe. In some of these countries, like France or Spain, new welfare designs were introduced some years before reforms were implemented in other OECD countries. As pointed out by Sianesi (2001) for the Swedish case, the Southern European experience of welfare programs puts into question the standard evaluation strategy followed by most of the U.S. literature. Every recipient is treated, as she/he should take part in some specific activity attempting both an upgrading of life skills as well as promoting higher levels of labor skills. The first set of measures includes overall actions developed to guarantee the basic pre-conditions of social participation. They consist of a variety of services comprising such different topics as general life skills or different measures aimed to make it easier for some families to sustain their daily routines. The second set of actions focuses on the achievement of higher labor market participation rates of low-income households.

A key point in choosing the IMI program is the availability of a large data set (over 50,000 spells) with very detailed information on the recipients' characteristics and the different kinds of treatments. This data can be matched with a special survey conducted for welfare leavers covering very different dimensions of the households' economic well-being some years after they participated in the program. On the one hand, the richness of the data makes possible to develop different types of evaluation strategies on the basis of multiple participation states. On the other, by matching administrative records and survey data we avoid some endogeneity problems and potential biases caused by the heterogeneity of welfare duration across participants in these treatments. Time spent on welfare could have a negative effect on labor outcomes regardless of participation in work-related activities. This problem, however, can be accounted for if data on the duration of welfare spells from administrative records is merged with updated information about the economic success of

welfare leavers. The comparison of the performance of the different treatments using alternative measures of success may contribute to the development of a more comprehensive body of alternatives to setting up more efficient welfare designs.

The remainder of this paper is organized as follows. Next section summarizes research related to welfare evaluation from the viewpoint of program heterogeneity and alternative outcome variables. We turn then to the particular design features of the IMI program and the available data. In the following section, we review different approaches to deal with the problem of evaluation in a framework of multiple states and alternative outcomes. In section four we test the extent to which the results are sensitive to alternative definitions of success comparing the performance of the different treatments. The paper ends with a brief list of conclusions.

1. BACKGROUND

As outlined above, welfare reforms across the OECD countries have taken very different forms. A variety of strategies has been developed under diverse objectives and circumstances. In some countries, work-first strategies have been at the heart of welfare changes. Other countries have opted for longer-term measures including intensive training and human capital investments into traditional welfare programs. One immediate consequence of this diversity is the huge range of possible assessments. Additionally, the burden of collection of evaluation procedures and methods can be extremely large depending on the specificity of the reforms' objectives.

The parameters for an adequate evaluation can also change according to the different economic agents' viewpoints. For the recipients the key question is how utility increases with the new measures. From the perspective of the programs' designers, the key elements can be varied. They include both the promotion of self-sufficiency as a long-standing way of materializing social rights and solidarity principles as well as imperative cost reductions. There are, therefore, many alternatives in order to draw an adjusted picture of the programs' outcomes.

Both questions become especially complex when welfare recipients can take part simultaneously in heterogeneous sub-programs. As stressed by Frolich (2004), active labor

market policies, usually embedded in new designs of welfare programs, consist of public employment programs, on-the-job training, retraining, job search assistance, wage subsidies, etc. In Southern Europe, these sub-programs coexist with a variety of initiatives aimed at promoting life skills: general life skills, family stabilization, children's schooling, etc. On the one hand, it is necessary to assess both the effects of participation in a given program compared with not participation at all as well as the effects of participation in a program with having participated in a different one. On the other hand, combinations of sub-programs can give rise to very different combinations of outcomes.

1.1. *Fundamentals of welfare reform*

The effectiveness of welfare reforms has been a major focus in the evaluation literature. The gist of these policies has been to move welfare recipients into the labor market. One controversy has been over the relative merits of work-first strategies and human capital and intensive training programs. An outstanding result of the review of the literature is the mixed and sometimes discrepant findings of empirical studies[3]. As Dyke *et al.* (2004) have shown, broad areas of disagreement exist concerning the effects of both kinds of measures. A potential explanation for not finding a well-defined picture of results has to do with the very different economic rationale of the two types of measures. Work-first strategies are mainly targeted to push recipients into the labor market as rapidly as possible while long-term programs focus on human capital developments through intensive training and educational opportunities for recipients.

Human capital developments in a welfare framework make the potential social rate of return on training investments in welfare programs quite high. The improvement of labor skills should reduce welfare dependence while increasing the recipient's living standards. Anyway, these results largely depend on the commitment of recipients to staying long enough in the program-operated training and subsidized employment devices. The opportunity cost of participation could be too high, especially in periods of strong economic growth with higher earnings and employment levels [Moffitt (2002), and Ayala and Rodríguez (2006a)]. As rigorously argued by Grogger (2005), the economic rationale of work-first strategies is also based on human capital theory: work today should raise

---

[3] See Cancian *et al.* (1999), Freedman *et al.* (2000), Moffitt (2001), Barnow and Gubits (2002), Blank (2002), Bloom *et al.* (2004), and Grogger and Karoly (2005).

experience tomorrow, which in turn should raise future wage offers and reduce dependency on welfare.

In general terms, there is voluminous evidence showing that policy changes appeared to have mattered in the mid-1990s in the United States, although a strong labor market unambiguously helped increased work among welfare recipients (Blank, 2002). However, there is no conclusive evidence about substantial differences between the results of the two aforementioned strategies. Using matching methods Hotz *et al.* (2000) found that in the long-term those who receive intensive training had better results than those who were put into work-first programs. Estimates based from censored regression methods also show that the later strategy raised wages modestly (Grogger, 2005). Results found by Freedman *et al.* (2000) put into question, however, the traditional hypothesis that training will make welfare recipients better off in the long run. The gains to experience among welfare leavers who passed trough work-first strategies were larger than the gains to education and training. Anyway, existing findings are largely dependent on the time period considered in the evaluation research. Blank's (2002) examination of the number of years used in each study yields important insights into the effects of each kind of strategy. While evaluations supporting better results for short-term strategies had three-year follow-up surveys at the most, Hotz *et al.* (2000) estimates are based on data covering welfare-to-work participants for up to nine years. Human capital programs seem to work better with longer-term evaluations.

European studies focusing on the assessment of the new welfare policies' results seem consistent with the extensive body of empirical evidence found for the U.S. experience. In general terms, they give general support to the notion that intensive training and other human capital sub-programs yield better results than work-first strategies. As a matter of fact, one of the huge differences between U.S. and European labor markets for low-income households is that they are considerably more restricted in the European case, with considerably lower employment rates in Europe for these individuals. In any case, as rather different models coexist results are significantly different across countries. In the Netherlands, for instance, the results of human capital measures in Social Assistance programs show that these policies are far from being effective (Van Oorschot, 2002). In Spain, however, recent research finds that long-term measures embedded in welfare programs have made low-income households less dependent on government and more

self-sufficient because of the development of intensive training and the setting of a friendly labor environment (Ayala and Rodríguez, 2006a).

A challenging question deserving further attention is the evaluation of programs with mixed activities. We have relatively little insight into which are the effects of these new welfare policies in contexts of multiple treatments. As outlined above, the basic framework used in most of the U.S. evaluation literature in which a program is administered at a point fixed in time, and individuals are either treated or nor treated does not always works well when welfare schemes allow recipients to participate in different activities. In practice, differential effects are found when considering the results of both strategies within the framework of the same program (Hotz *et al.*, 2000). Bloom and Michalopoulos (2001) found that a possible combination of a work-first strategy and education should improve employment and income prospects for welfare recipients.

Empirical evidence is also scarce in the case of European welfare schemes. This could be partially explained for the prevalence of binary treatments in many national welfare programs. However, there is growing evidence that the array of simultaneous sub-programs have considerably increased in many of these countries. Some research has been done regarding active labor market policies for unemployed individuals. Using a multiple states evaluation framework for the differential performance of the Swedish measures, Sianesi (2001) found that both relative to one another and compared to more intense job search in open unemployment the more similar a program was to a regular job, the higher the program's benefits to its participants. By defining a causal evaluation model for the Swiss region of Zurich handling the issues of individual heterogeneity and treatment heterogeneity, Lechner (2002) showed that the multiple treatment approach can lead to valuable insights. More precisely, the best outcomes corresponded to further training and temporary wage subsidy. There is a considerable need, however, to do additional work toward developing a more accurate picture of the results of welfare schemes with mixed activities promoting both life as well as labor skills.

1.2. *Alternative indicators of economic success*

A crucial issue in the evaluation of welfare reform is the outcome considered in order to assess the programs' effectiveness. Despite most of the new programs have been designed

to move welfare recipients into the labor market, the final goal of these policies is improving the economic self-sufficiency of these households. Employment and earnings could be considered as an intermediate goal for achieving higher levels of economic well-being. In other words, income poverty or material hardship could be so important indicators as working hours, earnings or employment rates. In practice, attempts to evaluate both work-first and human capital strategies have almost exclusively focused on examining the changes in the later kind of outcomes.

A key question is the sensitivity of evaluation results to different outcome indicators. As stressed by Cancian and Meyer (2004), measurement issues are fundamental for an adequate assessment of the reforms' strengths and limits. These authors find found a sizeable sensitivity of conclusions to alternative ways of measuring the success of TANF reforms through independence from Public Assistance, income poverty and material hardship indicators. Substantial differences can even be found when considering some of these specific indicators. This is the case of the net income-increasing or poverty-reducing impacts surveyed by Blank (2002). While most studies calculating poverty among welfare leavers find very high rates, the magnitude of the estimated effects considerably diverges.

In addition to labor market indicators, reviewed above, income poverty measures have been frequently used for welfare evaluation purposes. There is abundant empirical evidence for the post-reform results in the U.S. Interest in testing whether welfare recipients are financially better off working or remaining in welfare has heightened recently. Renewed concern has been fed by concerns about the idea that it does not pay to move from welfare to work. Increasing work participation does not automatically lift welfare leavers out of poverty, as family income may not increase appreciably. In this sense, the reduction of income poverty has been a subject of increasing concern to policy-makers and poverty indicators have been gradually used as outcomes in the assessment of work-related welfare policies.

Somewhat contrasting evidence comes from an extensive body of empirical results[4]. While Ellwood (2000), Meyer and Rosenbaum (2000), Schoeni and Blank (2000), and Acs and Loprest (2001) found that after leaving welfare financial benefits increased –only with modest improvements in the later study–, Bavier (2002) and Cancian *et al.* (2003) obtained

---

[4] See discussion of this issue by Blank (2002) and Danziger *et al.* (2002).

very different results[5]. As stressed by Danziger *et al.* (2002), this unsatisfactory result could be due to the inability of most studies to include all income sources. In the first study based on panel data that includes detailed information on both welfare leavers and welfare stayers (*Women's Employment Study*) they found that despite almost half of the work-reliant were poor and about half experienced at least two material hardships, after participation in welfare single mothers were financially better off working that remaining as non-working recipients. Moffitt and Winder (2004) put into question this final conclusion using data from the *Three-City Studio* (Boston, Chicago and San Antonio). The main reason of this difference is that the women in this study obtain less additional income from other family members[6].

If the success of welfare reform is going to be measured trough the improvements in the economic self-sufficiency of recipients many would claim that its effects should be evaluated in terms or their material hardship-reducing impacts. Material hardship takes many forms that are not necessarily captured by any other measure. A relative large literature has traditionally shown a very limited correlation between income poverty and measures of hardship. Since the seminal contribution of Townsend (1979) a long line of research has attempted to identify a strong and significant relationship between income and different conditions of multidimensional deprivation. Most of this literature has focused on the United Kingdom [Desai (1986), Hutton (1991), and Berthoud *et al.* (2004)]. More recently, the setting up of the European Community Household Panel (ECHP) supposed an excellent opportunity to test the equivalence between the two measures in a larger set of European countries. The general result is the lack of correlation between income measures and material hardship [Callan *et al.* (1993), Nolan and Whelan (1996), Layte *et al.* (2001), Tsakloglou (2006), and Watson and Maître (2006)].

It is not surprising, then, that the effects on material hardship had been a major focus in the recent literature on the evaluation of welfare programs. Following up on previous contributions testing the correlation between the official poverty line and measures of material hardship [Mayer and Jencks (1989), Rector *et al.* (2001)], some studies have begun

---

[5] Between half and three-quarters of families entering TANF in New Jersey, Washington, and Wisconsin had a family income below the poverty line two years later (Cancian *et al.*, 2003).

[6] As stated by Blank (2002), leaver studies typically do not ask about income or earnings among other family members, which several studies find to be a major reason why total incomes among single mothers are rising in the U.S.

to consider whether various measures of success are identical, including income poverty and material hardship[7]. The evidence is again mixed. Edin and Lein (1997) found that low-wage working mothers experienced higher levels of hardship than welfare recipients. Using ten types of material hardship, Danziger *et al.* (2002) also found that wage-reliant mothers were worse than when they were on welfare. On the contrary, Winship and Jencks (2004) found that food-related problems declined among welfare recipients with the social policy changes of the 1990s. This trend parallels changes in the official poverty rate during the same years. Consumption data also suggests that the material circumstances of single mother families improved modestly after welfare reform (Meyer and Sullivan, 2006).

Among the huge range of options to evaluate the success of work-related welfare reforms, other indicators that suit well the final goals of the programs are those specifically reflecting the notion of 'welfare independence'. The usual economic interpretation of welfare dependence refers to an analysis of the duration of welfare spells. Wider definitions should force one to take into account whether the reforms will ultimately attenuate the intergenerational transmission of welfare by promoting alternative work values. From a more pragmatic perspective, some studies define dependence as the use of any government benefit available only to those with low income. Cancian and Meyer (2004) distinguish work-related welfare payments from cash assistance not directly tied to a work activity, finding that only about one-quarter of recipients were independent according to that measure.

Recidivism patterns have also been used for evaluation of the reforms' effectiveness. As a high percentage of recipients return to the programs in the near term, there is a growing awareness of the importance of designing public intervention more in keeping with these new forms of dependence. Carrington *et al.* (2002) found that leavers in the later half of the nineties were much less likely to return to welfare than leavers in the early part of that decade. This could be interpreted as a positive effect of welfare reform in terms of higher levels of self-sufficiency. There is also evidence for Southern Europe and sub-programs promoting life and labor skills. Using matching methods and considering recidivism rates and the duration of off-welfare spells as outcome variables, Ayala and Rodríguez (2006a) suggest potentially successful interventions.

---

[7] In contrast to the extensive literature in Europe on several dimensions of poverty, the work in the U.S. is scarce (Short, 2005).

A last group of indicators also used for the assessment of welfare reforms refers to a varied set of social difficulties. According to the weight assigned to actions developed to guarantee the basic pre-conditions of social participation, social difficulties could be included among the key outcomes. There are some limits, however, to clearly differentiate social problems and indicators of material hardship. In the European literature, if children do not receive needed medical care this should be considered as a very relevant social problem. This item, however, is frequently included into conditions of material deprivation in the U.S. Several reviews suggest that personal and family challenges might impede welfare recipients' ability to find jobs [Kalil *et al.* (1998), Danziger *et al.* (2000)]. These problems might be related to physical disabilities, health limitations, substance abuse, or mental health problems of parents or children. They also might include family breakdown or instability, involvement with the child welfare system, housing instability, child care and transportation problems, limited English proficiency, and prior felony convictions. There is even some empirical evidence about the effects of welfare reform on domestic violence. Rigger and Staggs (2004) showed that inclusion of generous work supports in welfare legislation may help reduce domestic violence.

The real question is whether the different measures also suggest different levels of success for welfare reforms. As stated above, Cancian and Meyer (2004) found substantial diversity in the patterns of success of welfare reform in the U.S. as different outcome variables were used. This lack of correlation is undoubtedly associated with a potentially high sensitivity to the outcome chosen as reference for evaluation purposes. Complexity increases as more treatments are taken into account. The scant literature on multiple welfare treatments – mostly focused on employment– does not address the problem of alternative outcomes. Therefore, the development of a multiple treatments/multiple outcomes matrix could be a crucial improvement for the evaluation of heterogeneous welfare programs.

2. THE IMI PROGRAM

2.1. *Institutional features of the program*

The program analyzed in this study is the Madrid Regional Government's welfare program (IMI), which was set up in 1990. Social Assistance in Spain is completely decentralized and the Madrid's program can be considered an 'average' program within the complex set of

regional schemes existing in Spain and Southern Europe. Some conclusions could then be extrapolated to other similar programs. Potential claimants can apply for benefits only if they have used up entitlement to other income maintenance programs. Like other European systems, the main difference from U.S. programs is that IMI access is not only allowed to female lone-parent households, but also to couples without children, single individuals or male-headed families.

A distinctive characteristic among other European countries is that benefits are considerably lower. Many households receive income support insufficient to lift them over the poverty threshold[8]. Nominal benefits are far below the minimum wage. Most welfare programs in Spain tax 100% of other social benefits as well as earned incomes[9]. However, the IMI introduced some exceptions to encourage labor market participation, such as the compatibility of earnings and benefits during some months, or the decision not to consider specific means-tested benefits for elderly household members in determining household benefits. Benefits are granted for one year, automatically renewable.

Previous studies have pointed out that belonging to an ethnic minority and employability are the main determining factors leading to lengthened spells in the program, with visible signs of a certain degree of duration dependence (Ayala and Rodríguez, 2006b). These results show that there are different kinds of recipients depending on their possibilities for entering the labor market, needing to be dealt with differently. Previous research has also provided information on the IMI's recidivism determinants (Ayala and Rodríguez, 2004). According to their results, measures to maximize the duration of the off-welfare spells should focus on implementing reforms improving recipients' chances of leaving the program to enter into more stable forms of employment and allocating a greater amount of resources to promote the 'insertion' of specific groups.

Among the different institutional features of the program, the 'insertion measures' constitute its most prominent trait in a comparative framework. Once benefits have been approved by the program's managers, recipients must sign an 'insertion contract' with the welfare agencies. Participation in 'insertion contracts' necessarily occurs while recipients are receiving IMI benefits. Initially, these contracts are intended to improve the recipients' self-sufficiency through an individualized design of 'insertion' measures adjusted both to

---

[8] Adequacy rates, defined as the ratio of benefits over poverty thresholds, are 57.8%, 37.6% and 35.1% for people living alone and couples with one and three children, respectively.

[9] A similar problem is found in the French *Revenu Minimum d'Insertion* (Gurgand and Margolis, 2005).

individual and households' characteristics. The primary foundation is the idea of co-responsibility. Individual assessment is conducted when recipients enter the program and social services support is provided to help these households address specific and family challenges. The contents of the contracts are negotiated by both sides, fixing a final plan of specific public intervention for each household.

Every recipient, therefore, should be engaged in a specific program, giving place to a very different scenario than that depicted by the standard theory of program evaluation. A very relevant issue is that recipients can simultaneously participate in very different activities. Among them, a broad classification can be made breaking down the existing measures into two categories (Table 1). The first set of measures includes overall actions developed to guarantee the basic pre-conditions of social participation. They consist of a variety of services comprising such different topics as general life skills, family stabilization, children's schooling, measures aimed to make it easier for some families to sustain their daily routines or helping recipients recognize their strengths. They try to achieve a balance between easing and accommodating barriers to employment and make the majority of ICs contents.

[TABLE 1]

A second set of measures specifically aim to improve the employment opportunities of recipients. Among these measures there is also certain heterogeneity. There are various general services targeted at the improvement of the labor market opportunities of recipients, specific actions trying to push recipients into the labor market as soon as possible and intensive training sub-programs. 'Insertion projects' comprise widely targeted labor services, intensive training, and social enterprises. The common purpose of these actions is the achievement of basic labor skills and the establishment of a friendly work environment as necessary first steps in the transition to competitive employment. Social enterprises are relatively similar to some of the experiences embedded in the U.S. paid work experience programs. Some of these measures stand out as the most important public attempt at including human capital components in this welfare program. Usually, they are conducted by government agencies and non-profit associations. These entities work with a variety of targeted populations, including long-term unemployed.

2.2. *Data*

The evaluation of the success of the different 'insertion' sub-programs included in the IMI can be made using different datasets. In this study, we match the program's administrative records –covering the whole history of the program– with a specific survey of IMI leavers conducted in 2001. This survey covers very different dimensions of the households' economic well-being some years after they participated in the program. The merging of the two sources may prove successful for creating a comprehensive dataset with evaluation purposes. On the one hand, by exploiting administrative records we have very detailed information on the recipients' characteristics at the moment of welfare participation. On the other hand, the survey of IMI leavers allows us to assess these households' economic well-being some time after leaving welfare. Furthermore, by matching administrative records with outcome variables like employment, income, and living conditions, we can correct some potential biases in evaluation exercises related to omitted information on the characteristics of previous welfare participation.

The cleaning of the program's administrative records allows us to examine a very varied set of socioeconomic characteristics of IMI recipients. We have information on over 50,000 spells in the program, corresponding to slightly more than 39,200 households. Of these, 8,500 have left the program at some stage and then re-entered it at least once. Recipients' characteristics include some of the variables various studies have highlighted as ideal for analyzing welfare populations (Mainieri and Danziger (2001), Goerge and Joo Lee (2001)), such as the existence of structural problems (social isolation, alcohol abuse and drug addiction) or the development of behavior associated with marginal situations like prostitution or begging.

A descriptive analysis of the IMI data allows us to give a preliminary assessment of the characteristics of recipients. Table 2 differentiates between the households that completed a spell in the program at some time between 1990 and 2001 and the households that were receiving benefits at the moment of collecting the data. The data on age show a larger presence of middle-aged individuals among household heads. Concerning the differences between completed and ongoing spells, the lower proportion of young people and the greater presence of individuals over 55 in the former stand out[10]. Frequencies of recipients'

---

[10] This is because of the transfer of recipients to the national non-contributory pension scheme.

gender suggest that the program has been increasingly used by women. Regarding household size and type, small households stand out in general. People living alone make up a third of total households and have gained in relative weight over time. The presence of single-parent households is also striking. As expected, educational levels are low as shown by the huge percentage of recipients whose highest attainment is primary education. However, no straight inferences should be made regarding the possibilities for finding a job. Employability frequencies reveal that a non-negligible segment of recipients could access employment now[11].

[TABLE 2]

A set of variables provides information on different social problems that accompany the lack of income. Five types of social problems stand out among IMI recipients[12]. The first is related to health problems, be they general health problems or those derived from the consumption of drugs and alcohol, as well as from mental illnesses. Another group constitutes social pathologies arising from insolvency in situations of debt, including non-payment for dwellings. A third problem involves belonging to an ethnic minority[13]. There also are some recipients suffering from severe mental health problems that limit their chances of becoming economically self-sufficient. A final problem is the development of behavior associated with social alienation, such as begging or prostitution, although these groups are not really relevant in quantitative terms.

As stated above, there is also available information on the performance of IMI recipients after leaving welfare. A specific survey was conducted by the Madrid government in 2001 including very detailed information on both participation in sub-programs during their time on IMI as well as different dimensions of their current economic situation. The sample size of the survey is around 2,300 households, obtained by stratified random sampling. The population of welfare recipients was divided into four strata and a simple random sample was selected from each stratum. The variables used to define the strata were the date of entry, type of exit, duration of IMI participation and municipality size.

---

[11] Employability is a variable defined by social workers the first time future clients apply for benefits. It takes the lowest level if there are no possibilities of working because of physical deficiencies and a maximum level if recipients could be already in the labour market.

[12] The variables on social problems are reported by the social workers from their observation of the recipients

[13] Belonging to an ethnic minority is not in itself a social problem. It is regarded as such in so far as belonging to an ethnic minority limits a person's possibilities of social integration. Most individuals classified into this group are Gypsies.

[TABLE 3]

The database is fairly informative because it contains very detailed information on the participation in the different 'insertion' activities included in the IMI program. The different subprograms included in the survey are general information, general counseling, continuous individual support, psychological support, legal support, children intervention, family mediation, group activities, assistance for getting to other benefits, access to specific employment offers, general job search assistance, training, subsidized employment and social enterprises. There is also information on different dimensions of current economic well-being, like employment, subjective economic well-being, material hardship and social difficulties. Some data on socioeconomic characteristics are also collected in the dataset, like age, gender, household type, marital status, educational attainment or labor status. Table 3 summarizes some of the most relevant characteristics of the leavers' sample. All this data allow us to define an evaluation framework in which multiple treatments give rise to multiple outcomes.

3. METHODOLOGY

The lack of knowledge about the main results of the Southern European model of social assistance makes it necessary to evaluate the extent to which alternative packages of 'insertion' measures give rise or not to substantial improvements in the participants' economic well-being. The analysis of the IMI program can provide new insights into the effects of relevant reforms that have not still motivated an extensive amount of research. Since the IMI features are very similar to those of other welfare schemes developed in Spain, France, Portugal and Italy, some of the results could be extrapolated to these programs. From a methodological point of view, the IMI program meets some characteristics offering interesting challenges for welfare evaluation. As abovementioned, program heterogeneity and multiple outcomes have been questions insufficiently addressed by the standard empirical strategies mostly followed by U.S. researchers.

When the aim is to evaluate the effect of a specific sub-program on alternative outcome variables, an initial question concerns the comparison state. The IMI leavers' survey provides very detailed information on fourteen different treatments, mixing measures

aimed at upgrading life skills and work-related initiatives. For the purpose of this study, the different treatments were aggregated into four different and mutually exclusive groups: non-participation in specific work-related sub programs, participation in general labor-oriented measures, participation in labor-intensive measures and participation both in general labor-oriented measures as well as in labor-intensive measures. The first group (treatment) comprises overall actions developed only to guarantee the basic pre-conditions of social participation (general information, general counseling, continuous individual support, psychological support, legal support, children intervention, family mediation, assistance related to other social benefits, and group activities). The second group (treatment) includes general labor services for recipients (access to specific employment offers, general job search assistance, and training). The third group (treatment) involves those actions more specifically targeted to foster transitions from welfare to work (subsidized employment and social enterprises). The fourth group (treatment) refers to the possibility of taking part both in the second and third groups.

To the extent that every recipient participates in one of the four defined treatments we must focus on the relative effectiveness of each treatment. Among the relevant options for policy-makers, three specific questions should be addressed. First, we can evaluate the effect of participation in some work-related scheme as compared to participate only in general measures promoting life skills. Second, we can assess the effects of participation in each one of the specific work-related schemes (general, intensive and mixed) compared to participation in general life skills activities. Third, we can also examine the relative effectiveness of each specific labor-oriented treatment. Participation in general labor measures can be compared to taking part in more intensive activities. Both treatments can be also compared to mixed strategies combining both general labor services as well as more intensive activities. Table 4 summarizes the pair-wise comparisons.

[TABLE 4]

The question of which are the major outcomes of each one of the possible treatments as compared to a different treatment leads us to choose a particular method of evaluation. A well-known problem of causal inference is how to estimate treatment effects in observational studies in situations where some individuals are exposed to a treatment, but with no methods of experimental design to get a control group. Taking simply the

difference between the outcomes of the respective treatments would lead to a selection bias. A literature based on direct comparisons of experimental and non-experimental findings has shown the strengths and limits of non-experimental causal studies[14]. In general terms, matching methods have been highlighted as producing valid estimates of program impacts. The fundamental basis of matching evaluation is to re-establish experimental conditions when no such data are available. It is possible to build up a sample counterpart by pairing each participant with non-participant recipients. A necessary assumption is conditional independence between non-treated outcomes and program participation (Rubin, 1977).

The limitation for matching is that it relies on a sufficiently rich comparison group. As the number of observable covariates increases, there are growing problems for finding exact matches for each of the treated units. In a seminal study, Rosenbaum and Rubin (1983) suggested the use of the probability of receiving treatment conditional on covariates (propensity score) to reduce the dimensionality of the matching problem. As stressed by Becker and Ichino (2002), if this balancing hypothesis is satisfied, observations with the same propensity score must have the same distribution of observable characteristics independently of treatment status. This means a random exposure to treatment and control, and treated units should be on average observationally identical. In practice, matching on the propensity score is essentially a weighting scheme (Heckman *et al.*, 1998).

Most of the evaluation literature of welfare reforms using matching estimators rests, however, on a basic framework in which a program is administered at a point fixed in time, and individuals are either treated or nor treated. For an adequate evaluation of the IMI program, it is necessary to extend the standard binary treatment model of only two states to the case of multiple states. We follow here the approaches proposed by Lechner (2001, 2002) and Sianesi (2001), introducing some variation in the estimates of the propensity score.

Given a framework of (M+1) mutually exclusive sub-programs (treatments), every welfare leaver will have one observable outcome $\{Y_0, Y_1,…,Y_M\}$. Participation in one of the

---

[14] The seminal contribution of LaLonde (1986) gave rise to an abundant literature comparing the effects on trainee earnings of an employment program run as a field experiment with the estimates that econometric methods without experimental data might have produced. Dehejia and Wahba (1999, 2002), and Smith and Todd (2004) use the same data from the National Supported Work Demonstration to test propensity score matching estimators.

predefined mutually exclusive sub-programs is indicated by $S \in \{0,1,\ldots, M\}$. The standard definition of average treatment effects should be extended to address the question of several pair-wise comparisons of treatments. The seven comparisons previously enunciated can be developed considering a very straightforward procedure. We are interested in the effects of participation in one sub-program (*a*) compared to participation in other sub-program (*b*) for a former welfare recipient who took part in program *a*:

$$\tau^{a,b}_0 = E(Y^a - Y^b | S=a) = E(Y^a | S=a) - E(Y^b | S=a) \tag{1}$$

where $\tau^{a,b}$ represents the expected effect for a former welfare recipient randomly drawn from the group participating in sub-program *a*. As in the case of the standard binary treatment, we need a counterfactual to estimate $E(Y^b | S=a)$.

Assuming that the assumption of conditional independence holds also in the multiple-states framework –the effects of participation in a sub-program are independent of the assignment mechanism for any given value of a vector of characteristics–, evaluation requires to observe all the characteristics (*X*) of the program's recipients affecting both the probability of participation in the respective sub-programs as well as the outcome variables[15]. All participants in sub-program *a* need to have a counterpart in group *b* for each *X*. We can select from the participants in *b* a control group in which the distribution of observed variables is as similar as possible to the distribution in the group of participants in sub-program *a*. This requires:

$$0 < Pr\,(S=a|\ X=x) < 1 \quad \text{for } x \in \widetilde{X} \tag{2}$$

and guarantees that all treated recipients have a counterpart in the other group.

As stated above for the standard binary treatment model of only two states, Rosenbaum and Rubin (1983) illustrated the use of the probability of receiving treatment conditional on covariates (propensity score) to reduce the dimensionality of the matching problem. Imbens (1999) showed that the properties of the propensity score for only two states hold

---

[15] As we are only interested in the pair-wise comparisons of the programs defined, the assumption of conditional independence can be relaxed by requiring to hold only for the groups of welfare recipients receiving either treatment *a* or treatment *b*. See Sianesi (2001).

also in models with multiple states. The propensity score is the conditional probability of participating in a sub-program given X:

$$Pr^a(X) = Pr(S=a|X) \qquad (3)$$

For an adequate comparison of the outcomes of the different sub-programs we need a balancing score function ($g(X)$) of the recipients' characteristics. As stressed by Sianesi (2001), since we are interested in the pair-wise comparisons of the different sub-programs, we need to find a balancing score ensuring the balancing of the X's in the two subpopulations of interest for each comparison. We need a balancing function such that:

$$E[Pr(S=a|X, S \in \{a,b\})|g(X)] = Pr(S=a|X, S \in \{a,b\}) \equiv P^{a/ab} \qquad (4)$$

Equation (1) can then be calculated, once the counterfactual is estimated as:

$$E(Y^b|S=a) = E_{P^{a/ab}}[E(Y^b|S=b, P^{a/ab}(X))|S=a] \qquad (5)$$

If the aforementioned assumptions hold in the multiple states framework we only need, then, the propensity score to evaluate the different sub-programs. As discussed by Lechner (2002), two different approaches can be used to modeling the respective propensity scores for matching. One approach consists of specifying and estimating a multiple discrete-choice model, such as multinomial logit or probit model (structural approach). A second approach is estimating all conditional probabilities between possible pairs of choices directly (reduced-form approach). This second approach closely mirrors the usual propensity score approach for binary treatments. As we have defined seven pair-wise comparisons the reduced-form approach is not so prohibitive than when using more disaggregated analysis. Additionally, the problem of using the structural approach is that if one choice equation is misspecified all conditional probabilities could be misspecified.

A relevant question is the selection of X's for balancing the different sub-samples in each pair-wise comparison. Following Sianesi (2001), the resulting quality of the matched samples has guided our choice, for each pair-wise comparison, of the specification for $g(X)$. We have used both the administrative records –including the duration of welfare

participation– as well as the leavers' survey to select the characteristics necessary to estimate the propensity score.

To compare sub-program *a* and sub-program *b* for participants in sub-program *b*, each participant in the former group is matched to one or more participants in group *b* based on the balancing score. Different weighting procedures were selected for associating the sets of participants. The results shown above have been obtained with kernel matching estimators.

Estimates of the effects of 'insertion' activities on the recipients' economic well-being by using propensity score estimators are only reliable if the matching produces credible control groups. Figure 1 plots the diagrams of the estimated propensity scores for the seven pair-wise comparisons. The horizontal axis displays the cumulative units from lowest to highest propensity scores and the vertical axis shows the propensity scores of the treated and control units. The solid and dashed lines largely coincide. The matching is especially high in those units with the highest propensity score. Despite there are slight differences between the seven comparisons, the fit is acceptable in general terms.

[FIGURE 1]

A final methodological issue is the definition of outcome variables. As stated above, results can be highly sensitive to the dimensions chosen for the assessment of the former recipients' economic well-being. Despite most of the insertion activities aim to foster transitions from welfare to work, labor indicators do not always account for changes in the possibilities of social participation for these households. Subjective well-being, material hardship or social difficulties could be so relevant for the self-sufficiency of welfare leavers than earnings or working hours' indicators.

We have chosen four different dimensions to evaluate the relative effectiveness of the different sub-programs under study. Despite there is not always available information for a complete assessment of the outcomes in each dimension, the leavers' survey allows us to draw a very comprehensive picture of results. The first dimension focuses on employment outcomes and two indicators are used: whether or not the household head is currently employed and if there is a legal contract and the payment of employer payroll taxes. The

second dimension refers to subjective economic well-being and poverty. We have defined four indicators for this dimension: subjective poverty, changes in living standards compared to ten years ago (self-assessed), and two indicators of inter-generational well-being deducted from two specific questions asking for comparisons in the leavers' economic situation: first, compared to the one enjoyed by their parents at present, and, second, compared to that of their parents when they were the same age.

A third dimension comprises different indicators of material hardship. Most of them are related to housing conditions[16]. Given the different nature of the available indicators, it seems clear that all the items considered carry a different weight in the households' economic well-being. Arithmetic addition implicitly imposes a severe value judgement because it does not differentiate the weighting of each material condition or necessity. We need to decide, then, how to summarize a wide array of conditions into a material hardship index. In the literature on multidimensional poverty different indicators have been applied to derive a synthetic measure of multiple deprivation. As stated by Brandolini (2000), the different conditions can be summarized into an index of material hardship:

$$Z_i = \Sigma_j \, w_j \, z(d_{ij}) \tag{6}$$

or an index of living standard:

$$L_i = \Sigma_j \, w_j \, l(d_{ij}) \tag{7}$$

where $z(\cdot)$ and $l(\cdot)$ are non-increasing and non-decreasing functions, respectively, of the amount d_{ij} possessed by the ith household ($i=1,...,n$) of the jth attribute ($j=1,...,J$), and w_j is the corresponding weight. While some authors give an equal weight to each item, the core of this line of research interprets deprivation as a relative situation. One of the most common strategies is to apply weighting systems which give more importance to the lack of goods considered necessary by larger groups of the population (Halleröd (1994, 1995)) or, alternatively, to the goods which are most widely owned in a society (Desai and Shah (1988)). For the material hardship index we use a normalised weight calculated as:

---

[16] The conditions included in the index of material hardship are homelessness, running water, hot running water, having electricity, having gas, inside toilet with running water, bath or shower, wash basin, kitchen, oven, refrigerator, washing machine, dishwasher, heating, telephone, mobile phone, car, and van.

$$w_j = \frac{v_j}{\sum_j v_j}$$

where w_j is the proportion of people not lacking item j. Thus, the weights attached to each item are functions of the spread of the good or activity among the whole population, compared to the spread of the other goods or activities considered. The commodities enjoyed by most of society are given more weight. We also define an index of material well-being using the complementary options.

A final dimension focuses on social and economic difficulties. We have defined five indicators of social difficulties. They include the number of problems and specific indicators for housing, health and individual or family problems[17]. All the conditions have also been summarized in a single indicator of social problems, using the abovementioned weighting schemes. We have also created an indicator of economic problems as the number of economic difficulties.

4. RESULTS

A key question in the evaluation of welfare reforms putting into action heterogeneous sub-programs is whether or not it provides credible evidence about the specific effects of each treatment. While the drawbacks of matching estimators are widely known[18], we can expect them to be very helpful for drawing some conclusions about the long-term effects of the programs under study. In order to measure the relative effectiveness of the different sub-programs we estimate average effects for each pair-wise comparison and the four types of the abovementioned indicators. We address two different questions in each case: first, we test if work-related sub-programs perform better than general measures aimed at improving life skills; second, we try to identify which work-related sub-program works best.

---

[17] We have classified the different types of social difficulties into three groups. Housing difficulties include inadequate housing conditions, overcrowding, excessive housing spending, and non-payment of dwelling. Health problems comprise general health problems, alcohol abuse, drug abuse, mental health problems, and access to medical care. Personal and family problems include school attendance, familiar conflicts, eviction, social isolation, domestic violence, and gambling addiction.

[18] See Blundell (2000), Smith and Todd (2004), and Imbens (2004).

One of the main thrusts of the IMI's development has been the provision of skills to welfare recipients to keep an attachment to the labor market. Considering the achievement of higher employment rates as a relevant goal, we might expect a better performance of work-related measures as compared to like-skills measures. The primary result of Table 5 is that employment effects are substantially higher for those sub-programs aimed at improving labor opportunities. Although the approaches to foster transitions from welfare to work differ among the different sub-programs evaluated, they all give rise to substantially higher employment rates than general life-skills measures. The direction of these effects is not unexpected since work-related sub-programs spend a substantial amount of resources to promote higher levels of labor participation.

[TABLE 5]

While the basic information that participating in work-related measures is associated to better employment outcomes is confirmed, it is of interest that the three kinds of treatments –general measures, intensive measures and mixed strategies– present a very different relative effectiveness. Among the three different options, intensive labor policies targeted to welfare recipients stand out as those with the highest capacity to increase employment opportunities. It seems that it is even better to assign most of the available resources to direct measures promoting employment than combining these actions with very general job assistance measures. It could be the case that placing welfare recipients – especially those hardest to employ– into very different daily labor routines could reduce the program's efficiency.

If attention is focused on more specific indicators of economic success the evidence is somewhat contrasting. Our second binary indicator shows not only the possibility of being employed but having a legal contract and being covered by employer's contributions. The most relevant result is that very general labor measures are not enough to guarantee a stable position in the labor market. They produce very modest positive effects as compared to general life-skills measures and their results also seem very limited compared to those stimulated by intensive training or mixed strategies. The latter seem, in this case, the most suitable strategy for achieving the best results.

As stated above, employment outcomes can be considered just an intermediate output of welfare reform. Despite most of the sub-programs analyzed aim to move welfare recipients into work, the final goal of these policies is achieving more direct effects on the households' economic well-being. In this sense, it is interesting to test whether or not participating in the different sub-programs helps to improve the economic situation of these households by means of job tenure. Among other dimensions, income poverty has been at the heart of the mainstream approach to identifying economic well-being regarding welfare reform. Unfortunately, the IMI leavers' survey does not provide information on households' incomes. Nevertheless, the survey comprises a set of questions allowing subjective-welfare measurement.

Subjective poverty indicators are given in Table 6 for each pair-wise comparison. Several points are worth mentioning. First, it is interesting to note that there are substantial differences between the treatments under study. In general terms, work-related measures not only contribute to higher employment rates but also help to reduce income poverty measured on a subjective base. Second, intensive employment programs appear to be the most efficient policy, especially if they are not combined with more general labor-related activities. Results for the latter show that welfare-to-work sub-programs not resting on specific forms of subsidized employment or social enterprises do not produce substantially better results than measures aimed to make it easier for some families to sustain their daily routines.

[TABLE 6]

The leavers' survey also allows us to make some intertemporal analysis. More precisely, households are asked to report changes in their economic situation over the last decade. We have defined a dummy variable representing upward variation in self-assessed income. It might be expected that households moving from welfare to employment would report improvements in their economic situation. A virtuous circle could take place of getting employed, higher earnings and increasing disposable income. Furthermore, previous studies have found a significant relationship between unemployment and subjective income insecurity (Ayala *et al.*, 1999).

Somewhat contrasting evidence comes however from the estimated effects for the different treatments. In general terms, the variation range of the estimated average effect is narrow. Work-related activities therefore do not seem to have substantial effects on the intertemporal changes in the economic well-being of welfare participants. Anyway, we should be especially cautious on possible outcomes extracted from self-reported income. As stated by Ravallion and Lokshin (2002), the welfare inferences drawn from answers to subjective survey questions are clouded by concerns about measurement errors and how latent psychological factors influence observed respondent characteristics.

Two additional indicators also refer to subjective economic well-being but focusing on intergenerational comparisons. We can assess the effects of participation in specific sub-programs on intergenerational gains both comparing ex-IMI recipients to their parents at present and comparing them to their parents at their age. Columns three and four of Table 6 show the respective average effects. Apparently, work-related measures might also generate positive effects from this perspective. This relationship however does not hold across the different indicators and treatments. In any case, to draw again possible policy inferences we would need detailed longitudinal information. It must also be noted that rigorous evaluation requires comparing the outcomes of these households with respect to their parents in a hypothetical situation of welfare participation. Otherwise, we would be putting an excessive pressure on programs with a naturally limited scope.

A third dimension for the assessment of the different sub-programs' performance is material well-being. As discussed above, we have combined indicators of material hardship to create a composite measure. We assign weights considering the proportion of households lacking (or not lacking) the respective item. The idea is that the higher the proportion of households with a particular item, the greater the extent to which the item may be deemed to be a necessity. Table 7 presents results for both indicators. Comparing across treatments shows that participating in work-related activities does not sharply alter the households' levels of material well-being. Average effects are rather small in most cases suggesting that participating in work-related schemes has not led to reduced material hardship.

[TABLE 7]

One possible explanation for the little impact of targeted welfare-to-work programs on material well-being is that increasing work participation does not automatically lift welfare leavers out of material hardship. Family income may not increase appreciably and structural forms of deprivation, as housing, remain unresolved. This conclusion however should be treated cautiously. On the one hand, there might be relevant methodological issues affecting evaluation. As discussed by Winship and Jencks (2004), welfare leavers who have experienced serious problems of material hardship are hard to find and the representativeness of hardship variables in the survey could be limited. Additionally, it must be noted that composite or summary measures provide additional information on the concurrence of various hardships, but are at risk of obscuring detail in the individual components. On the other hand, the effects on material well-being could be interpreted as a kind of Pareto improvement. The average effects suggest that participation in work-related measures do not materially harm welfare recipients and, furthermore, these activities help to increase their employment levels

Results also seem to support the idea that participation in work-related activities would lead to larger reductions in poverty (self-assessed) than in material hardship. This conclusion is in keeping with well-known previous empirical evidence for Spain. Past research has shown only a very moderate association between poverty and hardship measures, both considered in static as well as in dynamic perspective [Martínez and Ruiz-Huerta (2000), Pérez-Mayo(2005), and Navarro and Ayala (2006)].

While the general result that participating in work-related measures is associated to reduced effects on material well-being is confirmed, it is important to note again that the three labor treatments give rise to different outcomes. Participation in intensive work-related activities turns out to be more effective than taking part in general labor-oriented measures. The average effect is anyway small. More surprisingly, mixed activities show worse results. Engaging recipients in multiple activities may impose certain inefficiencies, especially if they are linked to bureaucratic and prolonged processes.

While income poverty and other monetary measures of well-being usually capture transitory deprivation, most social difficulties are likely to be more affected by structural processes. On one side, they require long-term and individually tailored interventions. On the other side, a reasonable assumption can be made that specific life-skill measures can

work better than some kinds of work-related programs. Table 8 shows the average effects for the seven pair-wise comparisons and the five indicators defined in the previous section. When we compare the estimated impacts of the different sub-programs, we see that the average effects are again very small. Results give general support to the notion that work-related programs have a very low incidence on social problems as compared to more general life-skill measures. This conclusion holds considering both the numerical raw indicator of social problems as well as the summary measure. Anyway, as in the case of material hardship outcomes, our findings provide no general evidence that participation in work-related activities aggravates social problems among welfare recipients.

[TABLE 8]

There are however substantial differences across types of social problems and treatments. As it might be expected given the previous results for material hardship indicators –mainly based on housing conditions–, housing problems do not improve, in general terms, when households take part in labor oriented sub-programs instead of doing it in more general activities. Health problems, on the contrary, are clearly reduced in that case. To the extent that households rank this dimension as the most relevant one in the assessment of well-being, work-related programs turn out to be more effective from this perspective than life-skill measures. A very different picture emerges however from data on family and individual problems. Labor-related measures have little to do with social difficulties like familiar conflicts, eviction, social isolation, domestic violence, or gambling addiction. In most cases, the number of economic problems do not decrease with participation in specific sub-programs.

Regarding the average effects for each pair-wise comparison, we find very similar results to those previously highlighted for material well-being. Among work-related measures, only those focused on intensive employment actions seem to be efficient in reducing the incidence of social difficulties. Participating in mixed work-related schemes yields again very negative results in terms of alleviating these problems.

These findings can be very helpful for better understanding of the program evaluated. In practice, program heterogeneity can cause a variety of results depending on the variable chosen as outcome. In this sense, a final matrix of treatments and outcomes can be created

in order to assessing the final effects of a huge range of options. This matrix might contribute to the development of a more comprehensive concept of alternatives to setting up new designs of the program. It can provide useful feedback on whether the different-sub-programs are generating impacts consistent with long-term expectations and policy-makers can choose different combinations of inputs depending on political priorities. If the main target of the program is improving employment for welfare recipients there is no doubt that some alternatives do, indeed, perform better than others. If the priority is to minimize the number of social problems, the matrix allows us to identify which combination of sub-programs yields better results. Collecting and interpreting the empirical findings from this matrix may therefore help to provide new insights to welfare reform.

[TABLE 9]

Table 9 shows how the effects of the different sub-programs largely depend on the outcome variable. Nonetheless, we find some evidence that could help to clarify the available array of options to policy-makers. One might expect employment and material well-being or the former and social problems to be positively correlated, so that engaging welfare participants in work-related activities would improve other dimensions of these households' economic well-being. However, our results put into question the traditional view that transitions from welfare to work produce gains in the very different dimensions of economic well-being. What this means is that there is not an universal solution for the different problems posed by welfare populations. Policy-makers must frequently make hard decisions subject to very complex restrictions.

Anyway, the multiple states/multiple outcomes matrix can be an useful tool for handling different options. The corresponding analyses of files and columns can be powerful avenues for the consideration of various options that often confront the programs' managers in deciding the way ahead. If they had to decide between fostering participation in work-related sub-programs or general measures promoting life skills the matrix informs us that, in general terms, the former activities have a positive effect on employment and poverty without harming substantially neither material well-being nor social difficulties. Therefore, if the success of the programs is going to be measured trough employment indicators there is no doubt that enrollment in work-related activities should be

encouraged. If the main goals of 'insertion' measures are promoting social participation and improving living standards, life-skills sub-programs should be the chosen option.

Regarding employment measures, there is an additional binary choice. Once that more emphasis in work-related measures becomes the policies' guideline, a decision must be made on the best way to promote employment and economic well-being. The overarching finding from the paper is that intensive employment measures, like subsidized employment or engagement in social enterprises, yield remarkably better results than more general work-related schemes. They cause higher levels of employment and subjective well-being, reduce poverty and improve different kinds of social difficulties. However, the comparison of the different treatments and outcomes places into question the possibility of combining both strategies. There seems to be a certain kind of administrative threshold limiting the relative effectiveness of mixed programs.

5. CONCLUSIONS

Major policy changes have increased interest in outcomes for participants in welfare reforms. In most countries, the major goal of the enacted reforms has been to reduce the dependence of low-income households on government support by improving employment opportunities while continuing to maintain a social safety net for qualified families. This study has assessed the effects of Madrid's *Ingreso Madrileño de Integración* (IMI). This program –comprising heterogeneous sub-programs– is standard among the existing schemes in South Europe. Regarding other welfare models, the development of heterogeneous sub-programs providing different 'insertion' services was its central and most prominent change. This model opens challenging questions for program evaluation. Instead of the basic framework used in most welfare evaluation literature based on the standard binary treatment model of only two states, program heterogeneity gives place to multiple states. The merging of two different datasets –administrative records and a leavers' survey– has allowed us to estimate the average effects of the different sub-programs using pair-wise comparisons and a huge set of outcome variables.

The picture portrayed by the different data we use in the paper is generally consistent and our findings can be useful for better understanding of the program evaluated. For policy-makers, work-related measures appears at first sight attractive, since they could allow to

concentrate resources to reduce welfare participation and improving employment. The results of this paper, however, lead us to caution against drawing oversimplified conclusions. Empirical evaluation of a variety of sub-programs suggest that the effects of the different 'insertion' activities are rather varied depending on the outcome variables. Work-related measures seem successful in moving low-income families away from dependence on welfare policies by promoting higher employment levels. The long-run impact of these changes on economic well-being is likely to be positive, as evidence indicates that time in the labor market improves future employability.

However, despite the higher employment rates of former welfare recipients who have participated in work-related sub-programs, labor participation does not seem enough to allow them to achieve better results in terms of material well-being or social problems. If the overriding goal of social policy is to reduce material deprivation and social difficulties, there is no doubt that work-related policies are not completely suitable. Anyway, strict assessments on the validity of this measures could be misleading. The estimated average effects suggest that participation in work-related measures do not materially harm welfare recipients while increasing employment levels.

In order to contribute to the development of a more comprehensive concept of alternatives we have created a multiple treatments/multiple outcome matrix. This matrix provides information on the expected outcomes for different policy choices. Among the different alternatives for defining work-related sub-programs, intensive employment measures like subsidized employment or engagement in social enterprises yield remarkably better results than general work-related schemes. They cause higher levels of employment and subjective well-being, reduce poverty and improve different kinds of social difficulties.

In terms of public intervention, however, this finding should be considered cautiously. Given the high levels of heterogeneity among welfare populations, possible inferences should be restricted to certain households. In practice, work-first strategies can only be a solution for a segment of the recipients' population. For people who are totally unfit for employment, an upgrading of life skills through specific non-labour related interventions should result more efficient.

In any case, our findings provide new evidence to address some of the central questions posed on the current debate on welfare reform. As other countries are discussing similar

reforms, our results could contribute to watch the welfare experiments in Southern Europe with greater interest. New evidence on approaches considering simultaneous sub-programs combining life and labor skills and different types of outcomes might inform and partially shape the future public policy agenda in the welfare reform debate.

References

Acs, G., and Loprest, P., and T. Roberts, 2001. Final synthesis report on the findings from ASPE's "leavers" grants. Report to the U.S. Department of Health and Human Services. Washington, DC: Urban Institute.

Anderson K.H., Burkhauser, R.V. and J.E. Raymond, 1993. The Effect of Creaming on Placement Rates Under the Job Training Partnership Act. Industrial and Labor Relations Review 46, 613–624.

Angrist, J.D., Imbens, G.W. and D.B. Rubin, 1996. Two-Stage Least Square Estimation of Average Causal Effects in Models with Variable Treatment Intensity. Journal of the American Statistical Association 90, 468–472.

Ayala, L. and M. Rodríguez, 2004. Multiple Occurrence of Welfare Recipiency: Determinants and Policy Implications. Instituto de Estudios Fiscales, Working Paper 13/2004.

Ayala, L. and M. Rodríguez, 2006ª. The Latin Model of Welfare: Do 'Insertion Contracts' Reduce Long-Term Dependence?", *Labour Economics* (forthcoming).

Ayala, L. and M. Rodríguez, 2006b. Barriers to employment and welfare dynamics: evidence from Spain. Journal of Policy Modeling (forthcoming).

Ayala, L., Ruiz-Huerta, M., R. Martínez, 1999. Poverty and Social Policy in Spain in a Period of Political Transition. In Hutton, S. and Redmond, G. (eds.): Poverty in Transition Economies. London: Routledge.

Barnow, B. and D.B. Gubits, 2002. Review of Recent Pilot, Demonstration, Research and Evaluation Initiatives to Assist in the Implementation of Programs under the Workforce Investment Act. U.S. Department of Labor.

Bavier, R., 2002. The Impact of Welfare Reform on Families in Data from the Survey of Income and Program Participation. Washington, D.C.: Office of Management and Budget.

Becker, S.O. and A. Ichino, 2002. Estimation of Average Treatment Effects Based on Propensity Scores. The Stata Journal 2, 358–377.

Blundell, R., 2000. Evaluation Methods for Non-Experimental Data. Fiscal Studies 21, 427–468.

Blank, R.M., 2002. Evaluating Welfare Reform in the United States. Journal of Economic Literature 40, 1105–1166.

Bloom, H.S., Michalopoulos, C., 2001. How Welfare and Work Policies Affect Employment and Income: a Synthesis of Researc. New York: Manpower Demonstration Res. Corp.

Bloom, H.S., Michalopoulos, C., Hill, C.J., and Y. Lei, 2004. Can Nonexperimental Comparison Group Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs? The Review of Economics and Statistics 86, 156–179.

Blundell, R., 2000. Evaluation Methods for Non-Experimental Data. Fiscal Studies 21, 427–468.

Blundell, R. and C. Meghir, 2002. Active Labour Market Policy *vs.* Employment Tax Credits: Lessons from Recent UK Reforms. Working Paper 2002:1, Institute for Labour Market Policy Evaluation.

Brandolini, A. And G.D'Alessio, 2000. Measuring Well-Being in the Functioning Space. 26th General Conference of The International Association for Research in Income and Wealth, Cracow, Poland, 2000.

Callan, T., Nolan, B., and C.T.Whelan, 1993. Resources, Deprivation and the Measurement of Poverty". Journal of Social Policy 22, 141-172.

Cancian, M., Haveman, R., Kaplan, T., Meyer, D.R., and B. Wolfe 1999. Work, Earnings and Well-Being After Welfare. In: Danziger, S. (Ed.), Economic Conditions and Welfare Reform. MI: W.E. Upjohn Institute Employment Research.

Cancian, M. and D.R. Meyer 2004. Alternative Measures of Economic Success among TANF Participants: Avoiding Poverty, Hardship, and Dependence on Public Assistance. Journal of Policy Analysis and Management 23, 531–548.

Carling, K. and K. Richardson, 2004. The Relative Efficiency of Labor Market Programs: Swedish Experience from the 1990s. Labour Economics 11, 335–354.

Carrington, W.J., Mueser, R. and K.R. Troske, 2002. The Impact of Welfare Reform on Leaver Characteristics, Employment and Recidivism. IZA Discussion Paper nº 561, Bonn.

Danziger, S., Heflin, C., Corcoran, M., Oltmans, E., and H.Wang, 2002. Does It Pay to Move from Welfare to Work?. Journal of Policy Analysis and Management 21, 671-692.

Danziger, S.K., Corcoran, M., Danziger, S., Heflin, C., and A. Kalil, 2000. Barriers to the Employment of Welfare Recipients. In R.Cherry and W. Rodgers (eds.). Prosperity for All? The Economic Boom and African Americans. New York: Russell Sage Foundation.

Dehejia, R.H. and S. Wahba, 1999. Causal Effects in Nonexperimental Studies: Reeevaluating the Evaluation of Training Programs. Journal of the American Statistical Association 94, 1053–1062.

Dehejia, R.H., and S. Wahba, 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. The Review of Economics and Statistics 84, 151–161.

Desai, M. and Shah, A., 1988. An Econometric Approach to the Measurement of Poverty, Oxford Economic Papers 40, 505-522.

Dyke, A., Heinrich, C.J., Mueser, P.R., and K.R. Troske, 2004. The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes. 2004 Meeting of the Society of Labor Economists.

Edin, K., and L.Lein, 1997. Making ends meet: how single mothers survive welfare and low-wage work. New York: Russell Sage Foundation.

Ellwood, D.T., 2000. The Impact of the Earned Income Tax Credit and Social Policy Reforms on Work, Marriage, and Living Arrangements. National Tax Journal 53, 1063-1105.

Freedman, S., Friedlander, D., Hamilton, G., Rock, J., Mitchell, M., Nudelman, R., J., Schweder, A. and L. Storto, 2000. National Evaluation of Welfare-to-Work Strategies. Manpower Demonstration Research Corporation, New York.

Frolich, M., 2004. Finite-sample Properties of Propensity-Score Matching and Weighting Estimators. The Review of Economics and Statistics 86, 77–90.

Goerge, R.M. and B. Joo Lee, 2001. Matching and Cleaning Administrative Data. In: Ver Ploeg, M., Moffitt, R.A. and C. Citro (Eds.), Studies of Welfare Population: Data Collection and Research Issues. National Academy Press, Washington.

Grogger, J. (2005). Welfare Refom, Returns to Experience, and Wages: Using Reservation Wages to Account for Sample Selection Bias. NBER Working Paper nº 11621.

Grogger, J., and Karoly, L.A., 2005. Welfare Reform: Effects of a Decade of Change. Cambridge, MA: Harvard University Press.

Gurgand, M., and D.N. Margolis, 2005. Does Work Pay in France? Monetary Incentives and the Guaranteed Minimum Income. IZA   iscusión Paper nº 1467, Bonn.

Halleröd, B., 1994. Poverty in Sweden: a New Approach to the Direct Measurement of Consensual Poverty. Umea Studies in Sociology, nº 106.

Halleröd, B., 1995. The truly poor: diret and indirect consensual measurement of poverty in Sweden. Journal of European Social Policy 5, 111-129.

Heckman, J.J., Heinrich, C., and J. Smith, 2002: The Performance of Performance Standards. Journal of Human Resources 37, 778-811.

Heckman, J.J., Ichimura, I., and P. Todd, 1998. Matching as an Econometric Evaluation Estimator. Review of Economic Studies 64, 605-654.

Hills, J., and J. Waldfogel, 2004. A "Third Way" in Welfare Reform? Evidence from the United Kingdom? Journal of Policy Analysis and Management 23, 765–788.

Hotz, V.J., Imbens, G.W., and J.A. Klerman, 2000. The Long-Term Gains from Gain: a Re-Analysis of the Impacts of the California GAIN Program. NBER Working Paper 8007, Cambridge, MA.

Imbens, G.W., 2000. The Role of Propensity Score in Estimating Dose-Response Functions. Biometrika 87, 706–710.

Imbens, G.W., 2004. Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. The Review of Economics and Statistics 86, 4–29.

Kalil, M. Corcoran, S. K. Danziger, R. Tolman,K. Seefeldt, *et al.*, 1998. Getting Jobs, Keeping Jobs, and Earning a Living Wage: Can Welfare Reform Work?. IRP Discussion Paper no.1170-98, University of Wisconsin–Madison.

LaLonde, R.J., 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. American Economic Review 76, 604–620.

Layte, R., Nolan, B., and C.T.Whelan, 2001. Reassessing Income and Deprivation Approaches of Poverty in Ireland. The Economic and Social Review 32, 239-261.

Lechner, M., 2001. Some Practical Issues in the Evaluation of Heterogeneous Labor Market Programs by Matching Methods. In Lechner, M., and F. Pfeiffer (eds.), Econometric Evaluations of Active Labor Market Policies in Europe. Heidelberg: Physica/Springer.

Lechner, M., 2002. Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies. The Review of Economics and Statistics 84, 205–220.

Lichter, D.T., and R. Jayakody, 2002. Welfare Reform: How Do We Measure Success? Annual Review of Sociology 28, 117–141.

Mainieri, T. and S. Danziger, 2001. Designing Surveys of Welfare Populations, Report from the Workshop on Designing Surveys of Welfare Recipients. Ann Arbor, Michigan.

Martínez, R. and J. Ruiz-Huerta, 2000. Income, Multiple Deprivation and Poverty: An Empirical Analysis Using Spanish Data. 26th General Conference of The International Association for Research in Income and Wealth, Cracow, Poland.

Meyer, B., and D. Rosenbaum, 2000. Making Single Mothers Work: Recent Tax and Welfare Policy and Its Effects. National Tax Journal 53, 1027-61.

Meyer, B.D., and Sullivan, J. (2006). Consumption, income, and material well-being after welfare reform. NBER Working Paper Series nº11976.

Moffitt, R., 2001. From Welfare to Work: What the Evidence Shows? The Brookings Institution, Policy Brief 13/2002.

Moffitt, R., 2002. Welfare Programs and Labor Supply. In: Auerbach, J.M. and M. Feldstein (Eds.), Handbook of Public Economics, vol.4. North-Holland: Amsterdam.

Moffit, R., and M. Ver Ploeg (Eds.), 2001. Evaluating Welfare Reform in an Era of Transition. National Academy Press, Washington, DC.

Moffitt, R., and K. Winder, 2005. Does It Pay to Move from Welfare to Work? A Comment on Danziger, Heflin, Corcoran, Oltmans, and Wang. Journal of Policy Analysis and Management 24, 399-409.

Navarro, C., and L. Ayala, 2006. Multidimensional Housing Deprivation Indices with Application to Spain. Applied Economics (forthcoming).

Nolan, B., and T. Whelan, 1996. Measuring Poverty using Income and Deprivation Indicators: Alternative Approaches. Journal of European Social Policy 6, 225-240.

Pérez Mayo, J., 2005. Identifying deprivation profiles in Spain: a new approach. Applied Economics 37, 943-955.

Ravallion, M., and M. Lokshin, 2002. Identifying Welfare Effects from Subjective Questions. Economica 68, 1453-1473.

Rigger, S., and S.L. Staggs. Welfare Reform, Domestic Violence, and Employment. Violence Against Women 10, 961-990.

Rosenbaum, P.R., and D.B., Rubin, 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70, 41–55.

Rubin, D., 1977. Assignment to a Treatment Group on the Basis of a Covariate. Journal of Educational Statistics 2, 1–26.

Schoeni, R.F., and R. Blank, 2000. What has Welfare Reform Accomplished? Impact on Welfare Participation, Employment, Income, Poverty, and Family Structure. NBER Working Paper nº 7627.

Sianesi, B., 2001. Differential Effects of Swedish Active Labour Market Programs for Unemployed Adults in the 1990s. IFS Working Papers W01/25.

Sianesi, B., 2004. An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. The Review of Economics and Statistics 86, 133–155.

Smith, J., 2000. A Critical Survey of Empirical Methods for Evaluating Active Labour Market Policies. Swiss Journal of Economics 136, 1–22.

Smith, J., and P. Todd, 2004. Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators? Journal of Econometrics 125, 305–353.

Tsakloglou, P., 2006. Social Exclusion, Poverty and Cash Social Transfer in the European Union in a Longitudinal Perspective. EPUNet Conference 2006.

Van den Berg, G.J., Van der Klaauw, B., and J.C. Van Ours, 2004. Punitive Sanctions and the Transition Rate from Welfare to Work. Journal of Labor Economics 22, 211-241.

Van Oorschot, W., 2002. Miracle or Nightmare? A Critical Review of Dutch Activation Policies and their Outcomes. Journal of Social Policy 31, 339–420.

Watson, D. and B, Maître, 2006. Poverty, Deprivation and Social Class in the EU. EPUNet Conference 2006.

Winship, S., and Jencks, C. (2004). How Did the Social Policy Changes of the 1990s Affect Material Hardship among Single Mothers? Evidence from the CPS Food Security Supplement. John F. Kennedy School of Government nº04-027.

Zoyem, J.P., 2001. Contrats d'Insertion et Sortie du RMI. Économie et Statistique 346-347, 75–102.

Table 1
'Insertion' activities in the IMI program[a]

	Percent
Life skills	*64.7*
Specific childcare	7.4
General social skills	7.1
Adults' schooling	5.3
Children's schooling	9.1
Specific housing actions	5.1
Specific medical assistance	6.5
Daily routines	2.5
Supportive counseling	16.5
Others	5.2
Labor skills	*35.3*
Basic training	17.8
Job assistance	7.9
Insertion projects	9.6
TOTAL	100.0

[a] Households can take part in more than one activity.

Table 2
Socio-Economic Characteristics of IMI Recipients (administrative records)
(*frequency distribution*)

	Completed spells, 1990-2001	Ongoing spells, 2001
AGE		
<26	6.7	11.4
26–35	30.9	29.5
36–45	28.7	26.5
46–55	18.0	19.6
56–65	15.7	12.9
GENDER		
Male	40.3	34.2
Female	59.7	65.6
HOUSEHOLD SIZE		
1 person	25.8	33.4
2 people	20.6	21.1
3 people	20.2	18.6
4 people	15.5	12.1
5 people	8.9	7.6
6 people	4.7	3.9
7 people	2.2	1.9
8 or more people	2.0	1.3
HOUSEHOLD TYPE		
Single person	25.8	33.4
Lone-parent household	31.6	37.6
Other households with children	20.1	12.0
Other households without children	22.5	17.0
EDUCATION		
Does not read or write	10.3	13.6
No academic qualifications (only reads and writes)	20.6	21.6
Primary Education	36.7	35.5
Middle School Education	18.1	15.8
Secondary Education	6.6	6.6
Level 1 Vocational Training	2.9	2.3
Level 2 Vocational Training	1.7	1.4
University Degree	1.3	1.3
Post-Graduate Degree	1.5	1.8
LABOR FORCE STATUS		
Employed	18.0	13.5
Unemployed	59.1	69.0
Inactive	22.9	17.5
EMPLOYABILITY		
Totally unfit for normal work	9.6	8.0
Needs process of social / health recuperation	23.8	37.3
Unemployed needing training / education	21.1	25.4
Could access employment now	32.4	21.3
Does work on hidden economy or equivalent activity	8.3	7.0
Does normal work or equivalent activity	4.8	1.1

Table 2 (cont.)

SOCIAL PROBLEMS[1]		
Drug abuse	5.0	6.0
Alcohol abuse	4.8	4.7
Other mental health problems	8.8	10.9
Other serious health problems	14.9	18.1
Non-payment of dwelling	6.3	7.0
Debt accumulation	9.7	9.4
Beggary	0.8	1.2
Prostitution	0.4	0.7
Social isolation	10.8	15.9
Ethnic minority	11.7	23.2
Number of observations	(41,996)	(7,568)

[1]The categories appearing in social problems are non-excluding dummy variables. A household can therefore suffer from more than one problem. The figures show percentages of recipients affected by each problem.

Table 3
Socio-Economic Characteristics of Ex-IMI Recipients (Leavers' survey)
(*frequency distribution*)

	Percent
AGE	
<26	3.2
26–35	17.6
36–45	30.6
46–55	21.7
56–65	17.7
>65	9.2
GENDER	
Male	37.1
Female	62.9
HOUSEHOLD TYPE	
Couple, no children	6.0
Couple with children	35.2
Lone-parent household	32.4
Other households	1.4
Unknown	25.0
MARITAL STATUS	
Never married	28.7
Married	34.9
Widowed	10.9
Separated	16.4
Divorced	9.1
EDUCATION	
Does not read or write	7.0
No academic qualifications (only reads and writes)	17.9
Primary Education	30.5
Middle School Education	27.0
Secondary Education	1.0
Level 1 Vocational Training	4.1
Level 2 Vocational Training	8.3
University Degree	3.8
Post-Graduate Degree	0.3
LABOR FORCE STATUS	
Employed	29.4
Unemployed	49.1
Inactive	21.5

Table 3 (cont.)

SOCIAL PROBLEMS[1]	
Inadequate housing conditions	25.1
Overcrowding	11.2
Excessive housing spending	40.7
Non-payment of dwelling	19.6
General health problems	27.9
Alcohol abuse	3.4
Drug abuse	3.1
Mental health problems	10.1
Access to medical care	8.8
School attendance	4.3
Family conflicts	5.9
Eviction	3.7
Social isolation	13.7
Domestic violence	0.7
Gambling addiction	1.8
Number of observations	(2,299)

[1]The categories appearing in social problems are non-excluding dummy variables. A household can therefore suffer from more than one problem. The figures show percentages of recipients affected by each problem

Table 4

Treatment A	*Comparison treatment B*
1. Participation in a work-related scheme	Non-participation in a work-related scheme
2. Participation in general work-related schemes	Non-participation in a work-related scheme
3. Participation in intensive work-related schemes	Non-participation in a work-related scheme
4. Participation in mixed work-related schemes	Non-participation in a work-related scheme
5. Participation in general work-related scheme	Participation in intensive work-related schemes
6. Participation in general work-related scheme	Participation in mixed work-related schemes
7. Participation in intensive work-related schemes	Participation in mixed work-related schemes

Table 5
Employment effects
(PS matching estimates)

TREATMENT	OUTCOME VARIABLES	
	Employment (currently)	Legal contract and payroll taxes
Non-participation in a work-related scheme	0.262	0.382
	(0.440)[1]	(0.486)
Participation in a work-related scheme	0.292	0.468
	(0.455)	(0.499)
Average effect	11.5	22.5
Non-participation in a work-related scheme	0.244	0.382
	(0.430)	(0.486)
Participation in general work-related schemes	0.256	0.393
	(0.437)	(0.489)
Average effect	4.9	2.9
Non-participation in general work-related schemes	0.269	0.407
	(0.444)	(0.492)
Participation in intensive work-related schemes	0.402	0.563
	(0.493)	(0.499)
Average effect	49.4	38.3
Non-participation in a work-related scheme	0.244	0.382
	(0.430)	(0.486)
Participation in mixed work-related schemes	0.296	0.580
	(0.457)	(0.495)
Average effect	21.3	51.8
Participation in general work-related schemes	0.240	0.400
	(0.427)	(0.490)
Participation in intensive work-related schemes	0.385	0.500
	(0.489)	(0.503)
Average effect	60.4	25.0
Participation in general work-related schemes	0.265	0.412
	(0.442)	(0.493)
Participation in mixed work-related schemes	0.313	0.558
	(0.464)	(0.497)
Average effect	18.1	35.4
Participation in intensive work-related schemes	0.376	0.473
	(0.487)	(0.502)
Participation in mixed work-related schemes	0.296	0.577
	(0.457)	(0.495)
Average effect	-21.3	22.0

[1] Standard deviation in brackets

Table 6
Effects on poverty and subjective well-being
(PS matching estimates)

TREATMENT	OUTCOME VARIABLES			
	Subjective poverty	Economic situation compared to 10 years ago	Economic situation compared to parents' (today)	Economic situation compared to parents' (same age)
Non-participation in a work-related scheme	0.366	0.237	0.182	0.325
	(0.482)[1]	(0.425)	(0.387)	(0.469)
Participation in a work-related scheme	0.339	0.251	0.192	0.329
	(0.474)	(0.434)	(0.394)	(0.470)
Average effect	-8.0	5.6	5.2	1.2
Non-participation in a work-related scheme	0.375	0.226	0.168	0.315
	(0.484)	(0.418)	(0.374)	(0.465)
Participation in general work-related schemes	0.393	0.256	0.183	0.314
	(0.489)	(0.437)	(0.387)	(0.465)
Average effect	4.6	11.7	8.2	-0.3
Non-participation in general work-related schemes	0.312	0.243	0.184	0.333
	(0.464)	(0.429)	(0.388)	(0.472)
Participation in intensive work-related schemes				
	0.282	0.230	0.149	0.303
	(0.453)	(0.423)	(0.359)	(0.462)
Average effect	-10.6	-5.7	-23.5	-9.9
Non-participation in a work-related scheme	0.375	0.226	0.168	0.315
	(0.485)	(0.418)	(0.374)	(0.465)
Participation in mixed work-related schemes	0.300	0.225	0.215	0.362
	(0.459)	(0.419)	(0.412)	(0.482)
Average effect	-25.0	-0.4	21.9	13.0
Participation in general work-related schemes	0.393	0.251	0.177	0.319
	(0.489)	(0.434)	(0.382)	(0.467)
Participation in intensive work-related schemes	0.277	0.260	0.195	0.318
	(0.450)	(0.441)	(0.399)	(0.468)
Average effect	-41.9	3.5	9.2	-0.3
Participation in general work-related schemes	0.370	0.268	0.182	0.322
	(0.483)	(0.443)	(0.386)	(0.468)
Participation in mixed work-related schemes	0.299	0.229	0.213	0.344
	(0.458)	(0.421)	(0.411)	(0.476)
Average effect	-23.7	-17.0	14.6	6.4
Participation in intensive work-related schemes	0.326	0.226	0.164	0.294
	(0.471)	(0.420)	(0.373)	(0.458)
Participation in mixed work-related schemes	0.301	0.225	0.213	0.364
	(0.460)	(0.418)	(0.410)	(0.482)
Average effect	-8.3	-0.4	23.0	19.2

[1] Standard deviation in brackets

Table 7
Effects on material hardship
(PS matching estimates)

TREATMENT	OUTCOME VARIABLES	
	Material well-being	Material hardship
Non-participation in a work-related scheme	9.991	1.711
	(2.737)[1]	(1.592)
Participation in a work-related scheme	9.885	1.777
	(2.778)	(1.581)
Average effect	-1.1	3.9
Non-participation in a work-related scheme	9.845	1.755
	(2.886)	(1.590)
Participation in general work-related schemes	9.889	1.694
	(2.810)	(1.368)
Average effect	0.4	-3.5
Non-participation in general work-related schemes	10.153	1.684
	(2.545)	(1.603)
Participation in intensive work-related schemes	10.186	1.733
	(2.354)	(1.697)
Average effect	0.3	2.9
Non-participation in a work-related scheme	9.845	1.755
	(2.886)	(1.590)
Participation in mixed work-related schemes	9.433	2.073
	(3.123)	(1.877)
Average effect	-4.2	18.1
Participation in general work-related schemes	9.845	1.695
	(2.882)	(1.415)
Participation in intensive work-related schemes	10.234	1.691
	(2.311)	(1.552)
Average effect	4.0	-0.2
Participation in general work-related schemes	9.969	1.655
	(2.756)	(1.377)
Participation in mixed work-related schemes	9.594	2.040
	(2.959)	(1.884)
Average effect	-3.8	23.3
Participation in intensive work-related schemes	10.264	1.747
	(2.149)	(1.657)
Participation in mixed work-related schemes	9.424	2.070
	(3.136)	(1.872)
Average effect	-8.2	18.5

[1] Standard deviation in brackets

Table 8
Effects on social problems (PS matching estimates)

TREATMENT	OUTCOME VARIABLES					
	Number of social problems	Index of social problems	Housing problems	Health problems	Family problems	Number of economic problems
Non-participation in a work-related scheme	1.779	1.299	0.585	0.415	0.231	5.397
	(1.665)[1]	(1.307)	(0.493)	(0.493)	(0.422)	(3.593)
Participation in a work-related scheme	1.853	1.331	0.639	0.361	0.252	6.217
	(1.619)	(1.258)	(0.480)	(0.480)	(0.434)	(3.576)
Average effect	4.2	2.5	9.2	-13.0	9.1	15.2
Non-participation in a work-related scheme	1.821	1.322	0.596	0.425	0.245	5.407
	(1.675)	(1.310)	(0.491)	(0.495)	(0.430)	(3.622)
Participation in general work-related schemes	1.926	1.364	0.649	0.357	0.262	6.117
	(1.726)	(1.326)	(0.478)	(0.480)	(0.440)	(3.646)
Average effect	5.8	3.2	8.9	-16.0	6.9	13.1
Non-participation in general work-related schemes	1.656	1.226	0.552	0.397	0.228	5.109
	(1.623)	(1.285)	(0.498)	(0.490)	(0.420)	(3.568)
Participation in intensive work-related schemes	1.586	1.184	0.546	0.356	0.241	5.414
	(1.435)	(1.157)	(0.501)	(0.482)	(0.430)	(3.255)
Average effect	-4.2	-3.4	-1.1	-10.3	5.7	6.0
Non-participation in a work-related scheme	1.821	1.322	0.596	0.425	0.245	5.407
	(1.675)	(1.310)	(0.491)	(0.495)	(0.430)	(3.623)
Participation in mixed work-related schemes	1.864	1.356	0.645	0.403	0.246	6.460
	(1.463)	(1.148)	(0.480)	(0.491)	(0.432)	(3.465)
Average effect	2.4	2.6	8.2	-5.2	0.4	19.5
Participation in general work-related schemes	1.935	1.369	0.665	0.352	0.268	6.377
	(1.700)	(1.308)	(0.472)	(0.478)	(0.443)	(3.632)
Participation in intensive work-related schemes	1.677	1.260	0.600	0.333	0.229	5.927
	(1.403)	(1.107)	(0.492)	(0.474)	(0.422)	(3.325)
Average effect	-13.3	-8.0	-9.8	-5.4	-14.6	-7.1
Participation in general work-related schemes	1.850	1.312	0.640	0.343	0.255	6.117
	(1.681)	(1.291)	(0.480)	(0.475)	(0.436)	(3.691)
Participation in mixed work-related schemes	1.926	1.402	0.648	0.406	0.249	6.560
	(1.575)	(1.246)	(0.478)	(0.492)	(0.433)	(3.426)
Average effect	4.1	6.9	1.3	18.4	-2.4	7.2
Participation in intensive work-related schemes	1.645	1.161	0.630	0.269	0.226	5.839
	(1.427)	(1.116)	(0.485)	(0.446)	(0.420)	(3.360)
Participation in mixed work-related schemes	1.865	1.358	0.639	0.405	0.250	6.494
	(1.471)	(1.155)	(0.481)	(0.492)	(0.434)	(3.479)
Average effect	13.4	17.0	1.4	50.6	10.6	11.2

[1] Standard deviation in brackets

Table 9
Multiple States / Multiple Outcomes Matrix

	Comparison 1	*Comparison 2*	*Comparison 3*	*Comparison 4*	*Comparison 5*	*Comparison 6*	*Comparison 7*
Employment (currently)	++	+	+++	++	+++	++	---
Legal contract and payroll taxes	+++	≈	+++	+++	+++	+++	++
Subjective poverty	-	≈	--	--	---	--	-
Economic situation compared to 10 years ago	+	+	-	≈	≈	--	≈
Economic situation compared to parents' (today)	+	+	--	++	+	++	++
Economic situation compared to parents' (same age)	≈	≈	-	++	≈	+	++
Material well-being	≈	≈	≈	≈	≈	≈	-
Material hardship	≈	≈	≈	++	≈	++	++
Number of social problems	≈	+	≈	≈	--	≈	++
Index of social problems	≈	≈	≈	≈	-	+	++
Housing problems	+	+	≈	+	-	≈	≈
Health problems	--	--	--	-	-	++	+++
Family problems	+	+	+	≈	--	≈	+
Number of economic problems	++	++	+	++	-	+	++

(≈): 5%; (-/+): 5-10%; (--/++): 10-25%; (+++/---): >25%

Figure 1. Propensity Score for Treated and Matched Comparison Units

Comparison 1

Comparison 2

Comparison 3

Comparison 4

Comparison 5

Comparison 6

Comparison 7
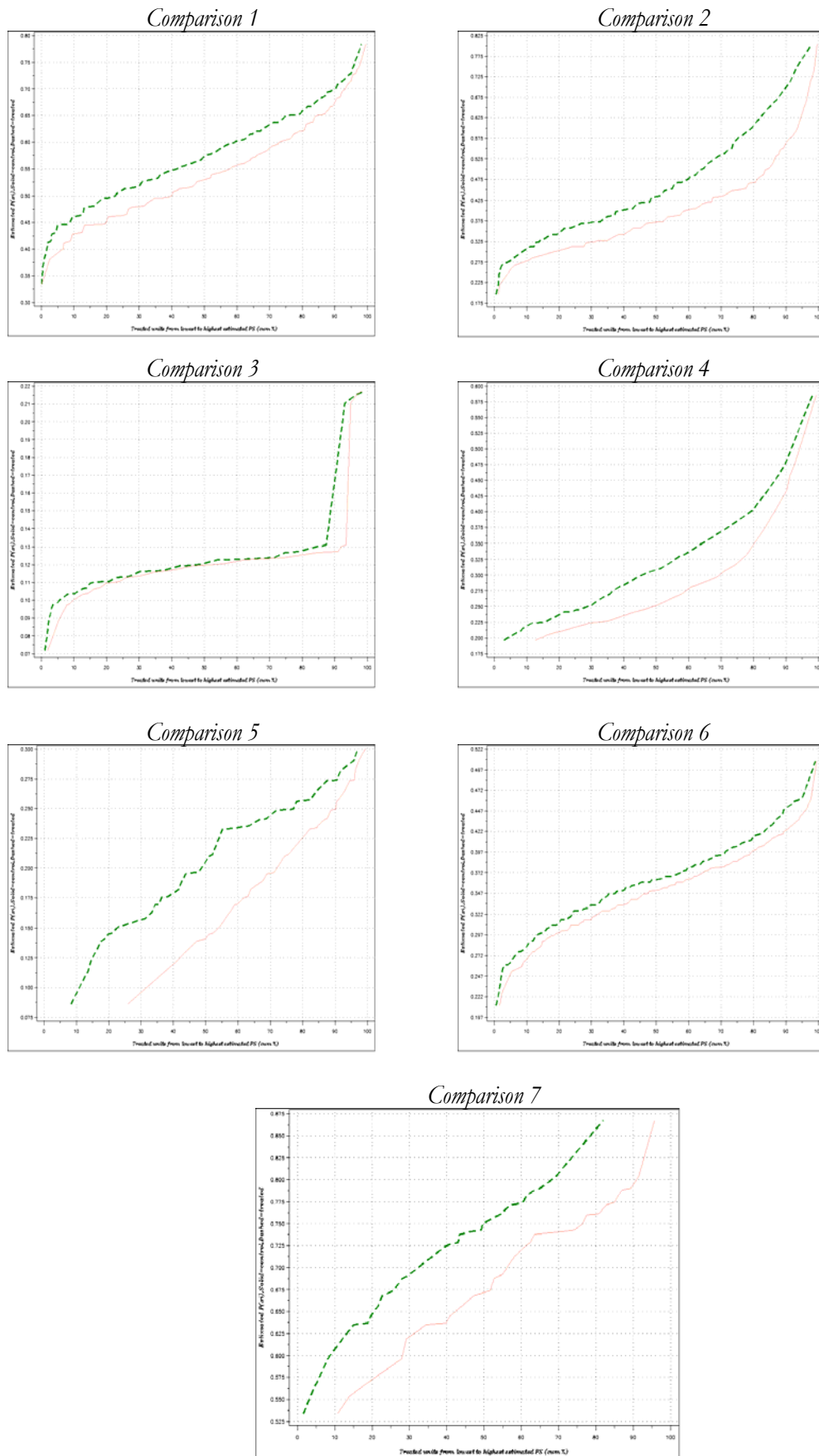
Horizontal axis: cumulative units from lowest to highest propensity scores
Vertical axis: propensity score for treated and control units

49