Working Paper Series

Evaluating welfare reform under program heterogeneity and alternative measures of success

Luis Ayala
Magdalena Rodríguez

ECINEQ WP 2010 – 186

ECINEQ 2010-186
October 2010

www.ecineq.org

Evaluating welfare reform under program heterogeneity and alternative measures of success

Luis Ayala*
Universidad Rey Juan Carlos

Magdalena Rodríguez
Instituto de Estudios Fiscales

Abstract

This paper aims at presenting 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)– which comprises heterogeneous subprograms. We test whether work-related subprograms perform better than general activities aimed at improving life skills. We also try to identify which work-related subprogram works best. The availability of a large database of administrative records (over 50,000 spells) matched with a special survey conducted for former welfare recipients makes possible to develop different types of evaluation strategies on the basis of multiple participation states. Our results show that intensive employment activities yield remarkably better results than general work-related schemes or life skills activities. However, increasing work participation does not automatically lift participants out of material hardship.

Keywords: welfare, poverty, program evaluation, multiple treatments.
JEL Classification: I38, I31

* **Addresses of correspondence Luis Ayala (corresponding author):** Facultad de Ciencias Jurídicas y Sociales, Universidad Rey Juan Carlos, Paseo Artilleros s/n, 28032 Madrid, SPAIN, luis.ayala@urjc.es, Fax: +34 915532796. **Magdalena Rodríguez:** Instituto de Estudios Fiscales, Cardenal Herrera Oria, 378, 28035 Madrid, SPAIN, magdalena.rodriguez@ief.minhac.es

INTRODUCTION

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 with the aim of increasing work incentives and reducing costs. As a consequence, there is a wide array of options for recipients to participate in work-related activities. The heavy emphasis on engaging recipients in work activities has rekindled interest in exploring whether these reforms have resulted in higher levels of well-being for participants. There has been much discussion over U.S. policy changes with considerable empirical evidence on the relevant outcomes.[1] 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 providing intensive training and educational opportunities for recipients.[2]

Work-related reforms of welfare policies have also given rise to a considerable European literature. A huge range of experiences has already been assessed including extensive activation programs in the field of Social Security and labor market policies in Nordic countries [Sianesi (2004, 2008), and Carling and Richardson (2004)], specific targeted welfare-to-work initiatives in the Netherlands [Van Oorschot (2002), and 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), Lydon and Walker (2005), Gregg *et al.* (2009)], welfare-to-work programs and job search enforcement in Germany (Huber *et al.*, 2009) or 'insertion contracts' embedded in minimum income programs in France and Spain [Zoyem (2001), Ayala and Rodríguez (2006a), and Terracol (2009)]. In general terms, there is voluminous evidence showing that policy changes appeared to have mattered.

Even though substantial research has been carried out, caution is important when evaluating these policies. Firstly, we have relatively little insight into the different effects the activities grouped under the notion of work-related welfare schemes have brought about.

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

[2] See Cancian *et al.* (1999), Freedman *et al.* (2000), Moffitt (2001), Barnow and Gubits (2002), Bloom *et al.* (2004), and Dyke *et al.* (2006).

This is due to the heterogeneous nature of such activities. In many countries, these programs are not mutually exclusive and welfare recipients can participate in different activities. This fact raises complex methodological issues the standard evaluation literature does not deal with. The standard binary treatment model of only two states could be extended to multiple states, but it needs to be suitably revised [Imbens (1999), Lechner (2001, 2002), Sianesi (2008), Imbens and Wooldridge (2009)].

Secondly, there is growing evidence that suggests that the effects of these programs may differ between outcomes. It can be difficult to see the whole picture of the programs' results if evaluation focuses exclusively on employment. While most of the new programs have been designed to move welfare recipients into the labor market, the ultimate goal of these policies is to improve the economic self-sufficiency of these households. In practice, the assessment of welfare reforms depends crucially on the indicators chosen to measure the programs' outcomes. As reviewed by Cancian and Meyer (2004), little systematic analysis deals with the question of whether different measures of success are capturing the same thing. The programs' success can vary greatly depending on whether independence from Public Assistance or income poverty are used as outcome indicators instead of employment results [Blank (2002), Grogger and Karoly (2005)]. Studies in welfare program evaluation have also paid special attention to the effects on material hardship. While some authors have found that low-wage working mothers experienced higher levels of hardship than welfare recipients [Edin and Lein (1997), Danziger *et al.* (2002)], others concluded that material circumstances of single mother families improved modestly after welfare reform in the U.S. [Winship and Jencks (2004), Meyer and Sullivan (2006)].

Consequently, one of the outstanding challenges in the evaluation of welfare programs is to analyze the effects of different and simultaneous treatments in work-related welfare reforms –program heterogeneity– on different types of outcomes. This paper aims at presenting 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)– which comprises heterogeneous subprograms. In Southern Europe, active labor market policies usually embedded in new designs of welfare programs coexist with a variety of initiatives aimed at promoting life skills. Therefore, recipients can simultaneously participate in very different actions and can be considered as treated, as they should take part in some specific activity.

Many reasons make this program particularly valuable for a comparative evaluation of welfare reforms. Firstly, to the extent that benefit levels fall between the low payments in the U.S. and the generous ones that are made in Northern Europe, this analysis can provide an interesting comparison for other countries. Secondly, the IMI approach to self-sufficiency seems less punitive than the U.S. approach based on strict conditions, sanctions, and time limits. Thirdly, while the context and program details are different than those in the U.S. or the Nordic countries there are remarkable similarities with the experience of other Southern European Countries. Our results could also apply to some extent to other countries with similar conditions.

In order to measure the relative effectiveness of the different activities available for IMI recipients, we address two different questions: first, we test whether work-related subprograms perform better than general activities aimed at improving life skills; second, we try to identify which work-related subprogram works best. We perform propensity score matching in a setting of heterogeneous treatments. Our results show that intensive employment activities –such as subsidized employment or engagement in social enterprises– yield remarkably better results than general work-related schemes or life skills activities in terms of employment and subjective well-being. However, increasing work participation does not automatically lift participants out of material hardship.

The structure of this paper is as follows. The opening section summarizes the particular design features of the IMI program and the available data. The second section reviews different approaches to deal with the problem of evaluation in a framework of multiple states and alternative outcomes. The third section tests the extent to which the results are sensitive to alternative definitions of success by comparing the performance of the different treatments. Concluding remarks are presented in the final section.

1. THE IMI PROGRAM

1.1. *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. The Spanish system of means-tested benefits has

drastically changed over the last decades. A characteristic trait of the system has been the gradual coverage of different demographic groups by means of specific welfare sub-systems. This expansion gave rise to a certain softening of social conflict, but created additional social segmentation by adding differentiation in the type of income guarantee to differences already existing in earnings and income mobility. The final result is a close-woven welfare network. Despite attempts at rationalization a complex mixture of overlapping benefits survives and significant gaps in the system's overall coverage persist. Unemployed individuals who have previously worked have access to two types of benefits: the contributory benefit and the assistance subsidy. In order to gain access to the first level, workers must have previously paid social contributions to cover this risk. The unemployment subsidy is for people whose contributory benefit has expired and who have income of less than 75 percent of the National Minimum Wage.

Welfare schemes are designed for those individuals who have exhausted their rights to unemployment benefits and who are not working. General risk of poverty is covered by regional schemes and Social Assistance in Spain is completely decentralized. The Madrid Program can be considered an 'average' program within the complex set of regional schemes existing in Spain and Southern Europe. As in other European schemes, all households are entitled to IMI access if they have used up entitlement to other income maintenance programs.

Previous studies have pointed out that employability and belonging to an ethnic minority 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 ability to enter the labor market. Consequently, they need to be dealt with differently. Previous research has also provided information on the IMI's recidivism determinants (Ayala and Rodríguez, 2010). These studies have found that activities to maximize the duration of the off-welfare spells should focus on implementing reforms that improve recipients' chances of leaving the program to enter into more stable forms of employment and allocate a greater amount of resources to promote the 'insertion' of specific groups.[3] There is also evidence from

[3] Insertion is the general term used in these countries to summarize the different types of activities aimed at improving life and labor skills of welfare participants. In general terms, it means higher levels of social participation.

standard binary treatment evaluation that participation in work-related activities reduces recidivism rates among IMI recipients (Ayala and Rodríguez, 2006a).

Among the different institutional features of the program, the 'insertion activities' represent the most prominent trait in a comparative framework. Once benefits are approved by the program's managers, recipients must sign an 'insertion contract' with the welfare agencies. Participation in these contracts is mandatory while recipients receive benefits. Initially, they are intended to improve the recipients' self-sufficiency through an individualized design of 'insertion' activities adjusted both to individual and households' characteristics. The idea of co-responsibility is at the heart of the program. Individual assessment is conducted when recipients enter the program and social services support is provided to help these households to address specific challenges. The contents of the contracts are negotiated by both sides resulting in a final plan of specific public intervention for each household.

Therefore, the program's outcomes could be conditioned by the efficacy of caseworkers in allocating individuals to government programs. Recent studies have stressed the key role of caseworkers in the assignment of welfare recipients to different work-related programs [Bloom *et al.* (2003), Behncke *et al.* (2009), Lechner and Smith (2007), Huber *at al.* (2009)]. Most of the evidence shows that caseworkers matter in allocation decisions. Available data from our survey somewhat confirm that caseworkers also play a relevant role in assigning IMI recipients to the different treatments. Nevertheless, most respondents state that, while the final decision was taken by the caseworker, it was the result of a negotiation process.[4]

Every recipient, therefore, has to join a specific program, bringing about 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 different activities. A broad classification of the activities can be made by breaking down the existing activities into two categories. The first set of activities includes overall actions developed to guarantee the basic preconditions of social participation. They consist of a variety of services comprising such different topics as general life skills, family mediation, children's schooling and

[4] Households were asked on the assignment process. Results do not show big differences in the observed frequencies between the different treatments. Almost one third of the assignments were the result of an agreement under caseworkers' predominance and approximately 30 percent were made on the basis of a fifty-fifty agreement between caseworkers and recipients.

activities aimed at making it easier for some families to continue their daily routines or helping recipients recognize their strengths.

A second set of activities specifically aim at improving recipients' employment opportunities (labour skills). There is also certain heterogeneity among these sub-programs. There are, first, various general services designed to improve the recipients' labor market opportunities; second, there are specific actions trying to push recipients into the labor market as soon as possible, including social enterprises and subsidized employment. 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. Usually, they are conducted by government agencies and non-profit organizations. These entities work with a variety of targeted populations, including long-term unemployed.

1.2. *Data*

In this study, we match the program's administrative records –covering the whole history of the program– with a specific survey conducted in 2001. This survey covers very different dimensions of the households' economic well-being some years after their participation in the program. The merging of the two sources may prove successful in creating a comprehensive dataset for evaluation purposes. On the one hand, by examining administrative records, we have very detailed information on the recipients' characteristics at the moment of entering the program. On the other hand, the IMI survey allows us to assess these households' economic well-being some time after participating in the different treatments. Furthermore, by matching administrative records with outcome variables, such as employment, income, and living conditions, we can correct some potential biases in evaluation related to omitted information on the characteristics of previous welfare participation.

The examination of the program's administrative records allows us to study a very diverse set of socioeconomic characteristics of IMI recipients. We have information on over 50,000 spells in the program corresponding to 39,200 households. 8,500 of them had left the program at some stage and then re-entered it at least once. Recipients' characteristics

include some of the variables highlighted as ideal for analyzing welfare populations, such as the existence of structural problems (social isolation, alcohol abuse and drug addiction) or the development of behavior associated with marginal situations (prostitution or begging).

The survey of IMI recipients was conducted by the Madrid Government in 2001 including very detailed information on both participation in subprograms during their time in IMI and different dimensions of the current economic situation. The sample size of the survey is about 2,300 households, obtained by stratified random sampling from the program's administrative records. The population of ex-welfare recipients was divided into four strata and a simple random sample was selected from each stratum.[5] The variables used to define the strata were: date of entry, exit type, duration of IMI participation and town size.

The survey contains detailed information on participation in the different 'insertion' activities included in the IMI program. The different subprograms considered in the survey are general information, general counseling, continuous individual support, psychological support, legal support, children intervention, family mediation, group activities, assistance to obtain 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 economic well-being, such as employment, subjective economic well-being, material hardship and social difficulties. We will use these dimensions to evaluate the outcomes of the program.[6] Some data on socioeconomic characteristics such as age, gender, household type, marital status, educational attainment and labor status are also collected in the dataset.

2. METHODOLOGY

The purpose of this evaluation is to identify which alternative packages of 'insertion' activities bring about substantial improvements in the participants' economic well-being.
The IMI survey provides very detailed information on fourteen different treatments, mixing activities aimed at upgrading life skills and work-related initiatives. For the purpose

[5] The survey was conducted using as initial universe those households that had been in the program at a given moment in the previous decade. Therefore the sample may include both households who are still in the programs and welfare leavers.
[6] We will use the survey data to identify treatments and outcomes. Socioeconomic information from administrative records of these households will be used to estimate the probability of taking part in a given treatment. Therefore, the sample we use in our identification strategy is 2,300 households.

of this study, the different treatments were aggregated into four different and mutually exclusive groups: non-participation in specific work-related subprograms (life-skills only) (n=811), participation in general labor-oriented activities (n=594), participation in labor-intensive activities (n=113) and participation both in general labor-oriented activities and in labor-intensive activities (n=331). The first group (treatment) comprises overall actions developed only to guarantee the basic preconditions 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).[7] The second group (treatment) includes general labor services for recipients (access to employment offers, general job search assistance and training). The third group (treatment) involves more intensive actions to foster transitions from welfare to work (subsidized employment and social enterprises). The fourth group (treatment) refers to individuals taking part simultaneously in the second and third groups.

To the extent that every recipient participates in one of the four defined groups, we focus on the relative effectiveness of each treatment. Among the relevant options for policy-makers, three specific questions can be addressed. First, we evaluate the effects of participation in some work-related scheme as compared to participation in general activities promoting life skills. Second, we assess the effects of participation in each one of the specific work-related schemes –general, intensive and mixed– as compared to participation in general life skills activities. Third, we also examine the relative effectiveness of each specific labor-oriented treatment. Table 1 summarizes the pair-wise comparisons.

[TABLE 1]

A key question in the analysis of the different effects is the selection problem arising when treatment assignment is not random. Observable traits differ for the treatment groups (see Table 1 in the annex). Therefore, it would be misleading to just compare individuals between treatment groups. 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. In this paper we use propensity score matching estimators to build up a sample counterpart by

[7] Individuals in groups 2, 3 and 4 can simultaneously participate in any of the life-skills activities included in the first group.

pairing each participant with non-participant recipients with similar characteristics. As it is widely known, a necessary assumption is conditional independence between non-treated outcomes and program participation (Rubin, 1977).

The main limitations of matching are this assumption and that it relies on a sufficiently rich comparison group. As the number of observable covariates increases, it becomes more difficult to find 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.

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 fixed point 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 [Imbens (1999), Lechner (2001, 2002), Sianesi (2008), Imbens and Wooldridge (2009)]. Given a framework of (M+1) mutually exclusive subprograms (treatments), every individual will have one observable outcome $\{Y_0, Y_1,\ldots,Y_M\}$. Participation in one of the predefined mutually exclusive subprograms is indicated by $S \in \{0,1,\ldots, M\}$. We are interested in the effects of participation in one subprogram (a) compared to participation in other subprogram (b) for individuals who took part in a:

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

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

Under the assumption that conditional independence holds also in the multiple-states framework, evaluation requires observing all the characteristics (X) of the program's recipients affecting both the probability of participation in the respective subprograms and

the outcome variables. All participants in subprogram *a* need to have a counterpart in group *b* for each *X* (Imbens, 1999). [8] We can select from the participants in *b* a control group whose distribution of observed variables is as similar as possible to the distribution in the group of participants in subprogram *a*. This requires:

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

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

For an adequate comparison of the outcomes of the different subprograms, we need a balancing score function ($g(X)$) of the recipients' characteristics. Since we are interested in the pair-wise comparisons of the different subprograms, we need to find a balancing score ensuring the balancing of the *X*'s in the two subpopulations of interest for each comparison:

$$E[Pr(S=a|X,\ S\in\{a,b\})|g(X)] = Pr(S=a|X,\ S\in\{a,b\}) \equiv P^{a/ab} \tag{3}$$

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

$$E(Y^b|\mathrm{S}=a) = E_{P^{a/ab}}[E(Y^b|S=b,\ P^{a/ab}(X))|S=a] \tag{4}$$

As discussed by Lechner (2002), two different approaches can be used to estimate 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 entails 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. The reduced-form approach is not prohibitive when used with a relatively low number of comparisons (seven in our case). We also avoid the problem of the structural approach when one choice equation is misspecified biasing all conditional probabilities.

[8] 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 individuals receiving either treatment *a* or treatment *b*.

A relevant question is the selection of X's for balancing the different subsamples in each pair-wise comparison. The resulting quality of the matched samples has informed our choice. We have used the administrative records to select the characteristics necessary to estimate the propensity score. The covariates considered where the number of social problems, single-parenthood, educational level, unemployment rate at entry, household size, number of children, single persons, gender –they all were measured at the moment of entering the program– and some variables related to the administrative process including the duration of the treatment.

To compare subprogram *a* and subprogram *b*, each participant in the former group is matched to one participant in group *b* based on the balancing score. Different procedures were selected for associating the sets of participants. The results we present below have been obtained with nearest neighbor matching estimators without replacement. As stressed by Smith and Todd (2005), replacement reduces bias but in turn increases the variance of the estimator. The problem of matching without replacement is that estimates depend on the order in which observations get matched (Caliendo and Kopeinig, 2008). In our estimates ordering is randomly done. We carried out different sensitivity analyses with other estimators finding similar results.[9]

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. There is a common support requirement for all pair-wise comparisons. Figure 1 plots the different density distributions of the propensity score for each comparison. Overlap in compared propensity scores regions seems to ensure common support across treatment groups. Even though there are slight differences between the seven comparisons –the matching is especially high in evaluations 1,4, and 6– the fit is reasonable. The crucial issue of the overlap condition seems to hold in our estimates.[10]

[FIGURE 1]

[9] Results are available upon request.

[10] Other indicators of matching quality were also estimated. The reduction in the standardised bias suggested by Rosenbaum and Rubin (1985) was estimated for the different variables used to define treated and matched control subsamples. Only gender and single-parenthood in evaluation 3 and unemployment, household size and number of children in evaluation 4 showed differences that were not significant.

A second general question is the definition of outcomes. As stated above, results can be highly sensitive to the dimensions chosen for the assessment of economic well-being. Although most of the 'insertion' activities aim to foster transitions from welfare to work, employment activities do not always account for changes in the opportunities for self-sufficiency. Subjective well-being, poverty or material hardship could be as relevant as earnings or working hours' indicators.

We have chosen three different dimensions to evaluate the relative effectiveness of the different subprograms under study. In spite of the fact that there are some limitations for a complete assessment of the outcomes in each dimension, the survey allows us to draw a very comprehensive picture of results. The first dimension focuses on employment outcomes and two indicators are used: 1) whether the household head is currently employed and 2) whether there is a legal contract and payment of employer contributions. The IMI survey does not provide information on households' incomes. The survey, however, comprises a set of questions allowing the measurement of subjective well-being. We have defined two dummy indicators for this dimension: subjective poverty and changes in living standards as compared to ten years ago (self-assessed).

A third dimension comprises different indicators of material hardship. Most of them are related to housing conditions.[11] 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 judgment because it does not differentiate the weighting of each material condition or necessity. Different indicators have been applied to derive a synthetic measure of multiple deprivations in the literature on multidimensional poverty. The different conditions can be summarized into an index of material hardship:

$$Z_i = \Sigma_j w_j z(d_{ij}) \tag{5}$$

or an index of living standards:

$$L_i = \Sigma_j w_j l(d_{ij}) \tag{6}$$

[11] The conditions included in the index of material hardship are 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.

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 j_{th} attribute (j=1,...,J), and w_j is the corresponding weight. While some authors give an equal weight to each item, the interpretation of deprivation as a relative situation is at the core of this line of research. One of the most common strategies is to apply weighting systems which give more importance to the goods most widely owned in a society. For the material hardship index we use a normalised weight calculated as:

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

where v_j is the number of households not lacking item j in the survey. Thus, the weights attached to each item are functions of the spread of the good among the whole population, compared to the spread of the other goods or activities considered. The commodities most people in society enjoy are given more weight. We also define an index of material well-being using the complementary options. A final measure provides information on the number of housing problems. [12]

3. RESULTS

In order to measure the relative effectiveness of the different subprograms we estimate average effects for each pair-wise comparison and for the three types of indicators we have defined. We address two different questions in each case: first, we test if work-related subprograms perform better than general activities aimed at improving life skills; then, we try to identify which work-related subprogram works best. In each case, we estimate an unconditional estimator which is 100 times the treatment B group mean minus the treatment A group mean divided by the treatment A group mean.

One of the main thrusts of the IMI's development has been the provision of skills to welfare recipients so that they are closer to the labor market. Considering the achievement of higher employment rates as a central goal, we might expect a better performance of

[12] Housing difficulties include inadequate housing conditions, overcrowding, excessive spending, and non-payment of dwelling.

work-related activities as compared to life-skills activities. The main result of Table 2 is that employment effects are substantially higher for those subprograms aimed at improving labor opportunities. Although the approaches to foster transitions from welfare to work differ among the different subprograms evaluated, they all bring about substantially higher employment rates than general life-skills activities.

[TABLE 2]

While results show that participation in work-related activities is associated to better employment outcomes, it is of interest that the three kinds of treatments –general activities, direct employment activities and mixed strategies– present different effectiveness levels. Among the three different options, intensive labor policies intended for 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 activities promoting employment than to combine these actions with very general job assistance activities. 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 employment success, the evidence is somewhat mixed. Our second binary indicator shows the possibility of being employed and having a legal contract covered by employer's contributions. The most relevant result is that very general labor activities are not enough to guarantee a stable position in the labor market. They produce very modest positive effects as compared to general life-skills activities. Furthermore, their results seem very limited compared to those produced by training or mixed strategies. The second group of activities appears to be the most suitable strategy in this case.

As stated above, employment results can be considered as intermediate outcomes. While most of the subprograms analyzed aim to move welfare recipients into work, the final objective of these policies is to achieve more direct effects on the households' economic well-being. For this reason, it is important to test whether participation in the different subprograms helps to improve the economic situation of these households. Among other dimensions, income poverty has been at the heart of the mainstream approach to identifying economic well-being in welfare reform.

Subjective poverty indicators are given in Table 3 for each pair-wise comparison. Several points are worth mentioning. Firstly, it is interesting to note that there are substantial differences between the treatments under study. As a rule, work-related activities not only contribute to higher employment rates, but also seem to reduce income poverty measured on a subjective base. However, results are not significant. Secondly, 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 subprograms not resting on specific forms of subsidized employment or social enterprises do not produce substantially better results than activities aimed to make it easier for some families to continue with their daily routines.

[TABLE 3]

The survey also enables us to carry out 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 of getting employed, higher earnings and increasing disposable income could take hold. Furthermore, previous studies have found a significant relationship between unemployment and subjective income insecurity in Spain (Ayala *et al.*, 1999).

Somewhat mixed evidence comes from the estimated effects for the different treatments. As a rule, the range of variation of the estimated average effects is narrow and are not significant. Work-related activities do not seem to have substantial effects on the intertemporal changes of economic well-being among welfare participants. In any case, we should be especially cautious about possible outcomes resulting from self-reported income. As stated by Ravallion and Lokshin (2002), concerns about measurement errors and the influence of latent psychological factors on observed respondent characteristics limit the validity of the well-being inferences drawn from answers to subjective survey questions.

A third dimension for the assessment of the different subprograms' performance is material well-being. As discussed above, we have combined indicators of material hardship

to create composite measures of material deprivation and living standards. We assign weights according to 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 4 presents results for both indicators. A comparison across treatments shows that participation 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 participation in work-related schemes has not reduced material hardship. A similar result is also found for the measure summarizing the number of housing problems. Work-related programs seem to produce worse results than life-skills activities.

[TABLE 4]

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 out of material hardship. Family income may not increase significantly and structural forms of deprivation, like housing, remain unresolved. However, this conclusion should be treated cautiously. On the one hand, relevant methodological issues affecting evaluation may be present. 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 also be interpreted as a kind of Pareto improvement. While work-related sub-programs do not materially harm welfare recipients, 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 prior empirical evidence for Spain. Past research has shown only a very moderate association between poverty and hardship measures, both considered in static and dynamic perspective [Pérez-Mayo (2005)].

While the results show that participation in work-related activities is associated to limited effects on material well-being, 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 for reducing material hardship than involvement in general labor-oriented activities. The average effects are small anyway and in most cases are not significant.

These findings can be very helpful for a 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 help to assess the overall effects of a huge range of options. It can provide useful feedback on whether the different-subprograms are generating impacts consistent with long-term expectations. Policy-makers can also choose different input combinations depending on political priorities. If the main goal of the program is to improve employment for welfare recipients, there is no doubt that some alternatives do indeed work better than others. If the priority is to minimize material hardship, the matrix allows us to identify which combination of subprograms yields better results.

[TABLE 5]

Table 5 shows how the effects of the different subprograms 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 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 bring about improvements in the different dimensions of economic well-being. This implies that there is not a universal solution to the different problems posed by welfare populations. Policy-makers must frequently make hard decisions subject to very complex restrictions.

In any case, the multiple states/multiple outcomes matrix can be a useful tool for handling different options. The corresponding analyses of files and columns can be powerful tools for evaluating the various options that often confront programs' managers in deciding the way ahead. If they have to decide between fostering participation in work-related

subprograms or general activities promoting life skills, the matrix shows that, for the most part, the former activities have a positive effect on employment and poverty without harming substantially material well-being. Therefore, if employment indicators are used to measure the programs' success, participation in work-related activities should be encouraged. If the main objective of 'insertion' activities is improving living standards, the options are not so well defined.

Regarding employment activities, there is an additional binary choice. Once the determination to implement work-related activities 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 activities, such as subsidized employment or engagement in social enterprises, yield remarkably better results than more general work-related schemes or the combination of both strategies.

4. CONCLUSION

Major policy changes have increased interest in outcomes for participants in welfare reforms. In most countries, the main 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 paper has assessed the effects of Madrid's *Ingreso Madrileño de Integración* (IMI). This program is standard among the existing schemes in Southern Europe. Compared to other welfare models, the development of heterogeneous subprograms providing different 'insertion' services is the main difference it presents. The merging of two different datasets – administrative records and a survey– together with the use of pair-wise comparisons and a huge set of outcome variables allow us to estimate the average effects of the different subprograms.

The picture emerging from the different evaluations is generally consistent. For policy-makers, work-related activities appear attractive at first sight, as they imply a concentration of resources to reduce welfare participation and improve employment. The results of this paper, however, lead us to caution against drawing oversimplified conclusions. Empirical evaluation suggests that the effects of the different 'insertion' activities are rather varied depending on the outcome variables. By increasing employment levels, work-related

activities seem successful in moving low-income families away from dependence on welfare policies. The long-run impact of these changes on economic well-being could be positive. Work today should raise experience tomorrow and raise future employability. However, higher labor participation does not seem enough to enable low-income families to achieve better results in terms of material well-being. If the overriding goal of social policy is to reduce material deprivation and social hardship, there is no doubt that work-related policies are not entirely suitable. In any case, strict assessments on the validity of these activities could be misleading.

Among the different work-related subprogram options, intensive employment activities, such as subsidized employment or engagement in social enterprises, yield remarkably better results than general work-related schemes. They result in higher levels of employment and subjective well-being. In terms of public intervention, however, this finding should be taken with caution. 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 those unemployable, an upgrading of life skills through specific non-labor related interventions should result more efficient.

Our findings provide new evidence to address some of the central questions the current welfare reform debate has raised. As other countries are discussing similar reforms, our results could contribute to examine the welfare experiments in Southern Europe with greater interest. New evidence on approaches that consider heterogeneous subprograms and different types of outcomes might inform and partially shape the future public policy agenda in the welfare reform debate.

References

Ayala, L. and M. Rodríguez, 2006a. The Latin Model of Welfare: Do 'Insertion Contracts' Reduce Long-Term Dependence?, *Labour Economics* 13, 799-822.

Ayala, L. and M. Rodríguez, 2006b. Barriers to Employment and Welfare Dynamics: Evidence from Spain. Journal of Policy Modeling 29, 237-257.

Ayala, L. and M. Rodríguez, 2010. "Explaining welfare recidivism: what role do unemployment and initial spells have?. *Journal of Population Economics* 23, 373–392.

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.

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

Behncke, S, Frölich, M., and M. Lechner (2009). Targeting Labour Market Programmes. Results from a Randomized Experiment, Swiss Journal of Economics and Statistics 145, 221-268.

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

Bloom, H.S., Hill, C.J., and J.A. Riccio, 2003. Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments. Journal of Policy Analysis and Management 22, 551-575.

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. and C. Meghir, 2002. Active Labour Market Policy *vs.* Employment Tax Credits: Lessons from Recent UK Reforms. Swedish Economic Policy Review 8, 13-37.

Caliendo, M. and S. Kopeinig, 2008. Some Practical Guidance For The Implementation Of Propensity Score Matching. Journal of Economic Surveys 22, 31-72.

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.

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.

Dyke, A., Heinrich, C.J., Mueser, P.R., and K.R. Troske, 2006. The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes. Journal of Labor Economics 24, 567-608.

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

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.

Gregg, P., Harkness, S. and S. Smith, 2009. Welfare Reform and Lone Parents in the UK. Economic Journal 119, 38-65.

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

Huber, M., Lechner, M., Wunsch, C. y Walter, T. 2009. Do German Welfare-to-Work Programmes Reduce Welfare and Increase Work?. IZA Discussion Paper nº 4090.

Imbens, G.W., 1999. 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.

Imbens, G. and J.M. Wooldridge, 2009. Recent Developments in the Econometrics of Program Evaluation. Journal of Economic Literature 47, 5-86.

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.

Lechner, M., and J. Smith (2007). What Is the Value Added by Caseworkers?. Labour Economics 14, 135-151.

Lydon, R. and I. Walker, 2005. Welfare to work, wages and wage growth. Fiscal Studies 26, 335-370.

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.

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

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.

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.

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.

Sianesi, B., 2008. Differential effects of active labour market programs for the unemployed. Labour Economics 15, 370-399.

Smith, J., and P. Todd, 2005. Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators? Journal of Econometrics 125, 305–353.

Terracol, A., 2009: "Guaranteed Minimum Income and Unemployment Duration in France," Labour Economics 16, 171-182

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.

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

Evaluation	*Treatment A*	*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 intensive work-related schemes	Participation in a general work-related scheme
6	Participation in mixed work-related schemes	Participation in a general work-related scheme
7	Participation in mixed work-related schemes	Participation in an intensive work-related schemes

Table 2
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.
***Significant at 99%, **Significant at 95%, *Significant at 90%.

Table 3
Effects on Poverty and Subjective Well-Being
(PS matching estimates)

TREATMENT	OUTCOME VARIABLES	
	Subjective poverty	Economic situation compared to 10 years ago
Non-participation in a work-related scheme	0.366	0.237
	(0.482)[1]	(0.425)
Participation in a work-related scheme	0.339	0.251
	(0.474)	(0.434)
Average effect	-7.4	5.9
Non-participation in a work-related scheme	0.375	0.226
	(0.484)	(0.418)
Participation in general work-related schemes	0.393	0.256
	(0.489)	(0.437)
Average effect	4.8	13.3
Non-participation in general work-related schemes	0.312	0.243
	(0.464)	(0.429)
Participation in intensive work-related schemes	0.282	0.230
	(0.453)	(0.423)
Average effect	-9.6	-5.3
Non-participation in a work-related scheme	0.375	0.226
	(0.485)	(0.418)
Participation in mixed work-related schemes	0.300	0.225
	(0.459)	(0.419)
Average effect	-20.0**	-0.4
Participation in general work-related schemes	0.393	0.251
	(0.489)	(0.434)
Participation in intensive work-related schemes	0.277	0.260
	(0.450)	(0.441)
Average effect	-29.5**	3.6
Participation in general work-related schemes	0.370	0.268
	(0.483)	(0.443)
Participation in mixed work-related schemes	0.299	0.229
	(0.458)	(0.421)
Average effect	-19.2**	-14.6
Participation in intensive work-related schemes	0.326	0.226
	(0.471)	(0.420)
Participation in mixed work-related schemes	0.301	0.225
	(0.460)	(0.418)
Average effect	-7.7	-0.4

[1] Standard deviation in brackets.
**Significant at 95%

Table 4
Effects on Material Hardship
(PS matching estimates)

TREATMENT	OUTCOME VARIABLES		
	Material well-being	Material hardship	Housing problems
Non-participation in a work-related scheme	9.991	1.711	0.585
	(2.737)[1]	(1.592)	(0.493)
Participation in a work-related scheme	9.885	1.777	0.639
	(2.778)	(1.581)	(0.480)
Average effect	-1.1	3.9	9.2*
Non-participation in a work-related scheme	9.845	1.755	0.596
	(2.886)	(1.590)	(0.491)
Participation in general work-related schemes	9.889	1.694	0.649
	(2.810)	(1.368)	(0.478)
Average effect	0.4	-3.5	8.9*
Non-participation in general work-related schemes	10.153	1.684	0.552
	(2.545)	(1.603)	(0.498)
Participation in intensive work-related schemes	10.186	1.733	0.546
	(2.354)	(1.697)	(0.501)
Average effect	0.3	2.9	-1.1
Non-participation in a work-related scheme	9.845	1.755	0.596
	(2.886)	(1.590)	(0.491)
Participation in mixed work-related schemes	9.433	2.073	0.645
	(3.123)	(1.877)	(0.480)
Average effect	-4.2	18.1	8.2
Participation in general work-related schemes	9.845	1.695	0.665
	(2.882)	(1.415)	(0.472)
Participation in intensive work-related schemes	10.234	1.691	0.600
	(2.311)	(1.552)	(0.492)
Average effect	4.0	-0.2	-9.8
Participation in general work-related schemes	9.969	1.655	0.640
	(2.756)	(1.377)	(0.480)
Participation in mixed work-related schemes	9.594	2.040	0.648
	(2.959)	(1.884)	(0.478)
Average effect	-3.8	23.3	1.3
Participation in intensive work-related schemes	10.264	1.747	0.630
	(2.149)	(1.657)	(0.485)
Participation in mixed work-related schemes	9.424	2.070	0.639
	(3.136)	(1.872)	(0.481)
Average effect	-8.2	18.5	1.4

[1] Standard deviation in brackets.
*Significant at 90%.

Table 5

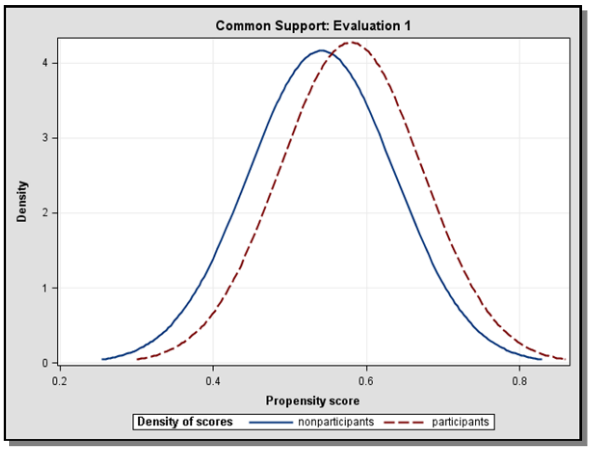
Treatments-Outcomes Matrix

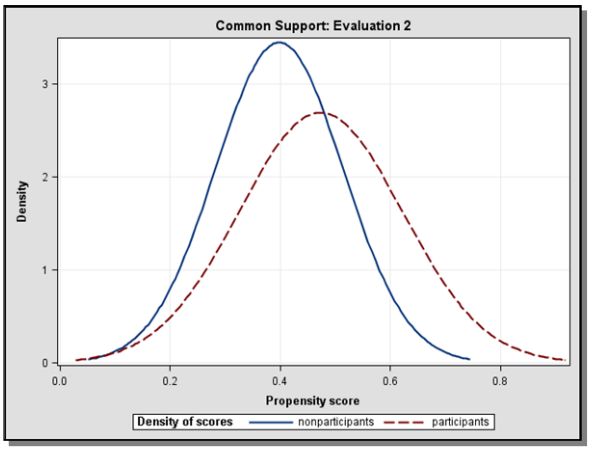
	Comparison 1	*Comparison 2*	*Comparison 3*	*Comparison 4*	*Comparison 5*	*Comparison 6*	*Comparison 7*
Employment (currently)	++*	+	+++**	++*	+++**	++	---
Legal contract and payroll taxes	+++**	≈	+++**	+++***	+++*	+++***	++*
Subjective poverty	-	≈	--	--**	---**	--**	-
Economic situation compared to 10 years ago	+	+	-	≈	≈	--	≈
Material well-being	≈	≈	≈	≈	≈	≈	-
Material hardship	≈	≈	≈	++	≈	+++	++
Housing problems	+*	+*	≈	+	-	≈	≈

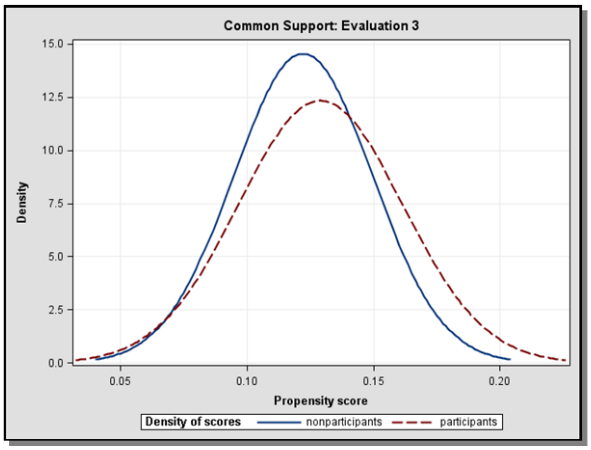
(≈): <5%; (-/+): 5-10%; (--/++): 10-20%; (+++/---): >20%

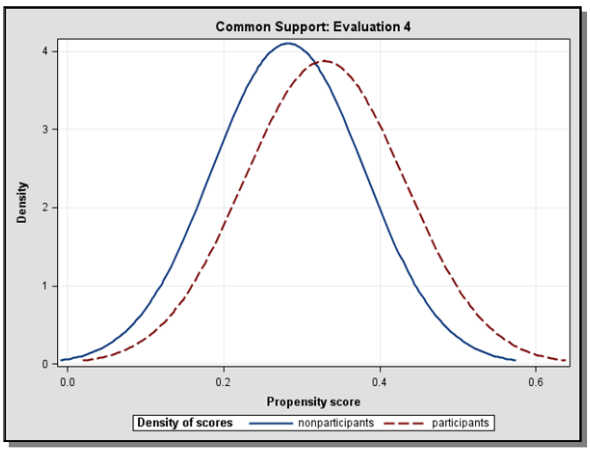
***Significant at 99%, **Significant at 95%, *Significant at 90%.

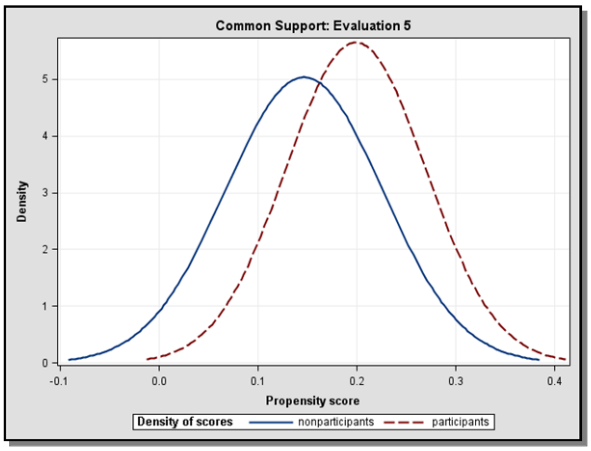
Figure 1. Common Support

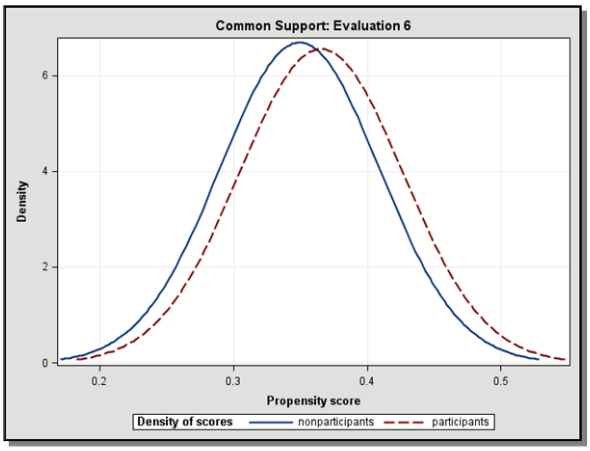


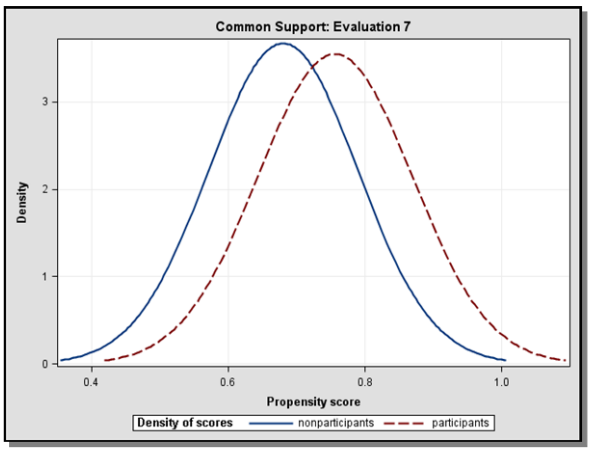












Annex. A.1. Selected descriptive statistics

Evaluation 1	Participants in treatment a		Participants in treatment b	
	Mean	Variance	Mean	Variance
Number of social problems	0.94	0.81	0.98	0.82
Single parent	0.30	0.21	0.35	0.23
Educational level[1]	2.83	1.04	2.94	1.00
Unemployment rate	15.42	14.21	15.60	13.97
Household size[2]	1.92	0.60	1.93	0.53
Number of children[3]	0.78	0.78	0.89	0.79
Single person	0.25	0.19	0.23	0.18
Female	0.65	0.23	0.68	0.22
Evaluation 2	Participants in treatment a		Participants in treatment b	
	Mean	Variance	Mean	Variance
Number of social problems	0.94	0.81	1.04	0.82
Single parent	0.30	0.21	0.37	0.23
Educational level	2.83	1.04	2.85	1.01
Unemployment rate	15.42	14.21	15.47	14.24
Household size	1.92	0.60	1.95	0.51
Number of children	0.78	0.78	0.93	0.79
Single person	0.25	0.19	0.21	0.17
Female	0.65	0.23	0.70	0.21
Evaluation 3	Participants in treatment a		Participants in treatment b	
	Mean	Variance	Mean	Variance
Number of social problems	0.94	0.81	0.84	0.85
Single parent	0.30	0.21	0.28	0.20
Educational level	2.83	1.04	3.04	0.98
Unemployment rate	15.42	14.21	15.49	13.87
Household size	1.92	0.60	1.99	0.55
Number of children	0.78	0.78	0.87	0.79
Single person	0.25	0.19	0.20	0.16
Female	0.65	0.23	0.66	0.23
Evaluation 4	Participants in treatment a		Participants in treatment b	
	Mean	Variance	Mean	Variance
Number of social problems	0.94	0.81	0.94	0.80
Single parent	0.30	0.21	0.33	0.22
Educational level	2.83	1.04	3.06	0.95
Unemployment rate	15.42	14.21	15.87	13.48
Household size	1.92	0.60	1.87	0.55
Number of children	0.78	0.78	0.83	0.79
Single person	0.25	0.19	0.27	0.20
Female	0.65	0.23	0.64	0.23

A.1. (cont.)

Evaluation 5	Participants in treatment a		Participants in treatment b	
	Mean	Variance	Mean	Variance
Number of social problems	1.04	0.82	0.84	0.85
Single parent	0.37	0.23	0.28	0.20
Educational level	2.85	1.01	3.04	0.98
Unemployment rate	15.47	14.24	15.49	13.87
Household size	1.95	0.51	1.99	0.55
Number of children	0.93	0.79	0.87	0.79
Single person	0.21	0.17	0.20	0.16
Female	0.70	0.21	0.66	0.23
Evaluation 6	Participants in treatment a		Participants in treatment b	
	Mean	Variance	Mean	Variance
Number of social problems	1.04	0.82	0.94	0.80
Single parent	0.37	0.23	0.33	0.22
Educational level	2.85	1.01	3.06	0.95
Unemployment rate	15.47	14.24	15.87	13.48
Household size	1.95	0.51	1.87	0.55
Number of children	0.93	0.79	0.83	0.79
Single person	0.21	0.17	0.27	0.20
Female	0.70	0.21	0.64	0.23
Evaluation 7	Participants in treatment a		Participants in treatment b	
	Mean	Variance	Mean	Variance
Number of social problems	0.84	0.85	0.94	0.80
Single parent	0.28	0.20	0.33	0.22
Educational level	3.04	0.98	3.06	0.95
Unemployment rate	15.49	13.87	15.87	13.48
Household size	1.99	0.55	1.87	0.55
Number of children	0.87	0.79	0.83	0.79
Single person	0.20	0.16	0.27	0.20
Female	0.66	0.23	0.64	0.23

[1] five groups, [2] four groups, [3] four groups