

# When Non-Compliance Carries the Day: Evaluating the Effectiveness of an Employment Program for the Severely Mentally Ill

Evaluation Review  
2022, Vol. 46(5) 555–577  
© The Author(s) 2021  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/0193841X211049685  
[journals.sagepub.com/home/erx](https://journals.sagepub.com/home/erx)  


Alberto Martini<sup>1</sup>, Enrico Rettore<sup>2</sup>,  
Gian Paolo Barbetta<sup>3</sup> , and Fabio Sandrolini<sup>4</sup>

## Abstract

Mental disease is increasing worldwide and people who suffer from it show lower employment rates and lower earnings. Various approaches have been tried to increase the employment rate of people with mental disease. In the US, empirical studies show that *individual placement and support* (IPS)—a rapid transition to the job market, with some external assistance—is effective in increasing the employment rate of the mentally ill. Europe lacks such evidence. The study assesses the impact of an IPS-like program undertaken in Italy on the employment rate of people with severe mental illness. The analysis is based on a RCT that tests whether offering people with severe mental illness the support of a “job coach” increases their chances of being employed. Moreover, using a battery of tests—one of which is made available by the RCT

<sup>1</sup>Università del Piemonte Orientale, Alessandria, Italy

<sup>2</sup>Università di Padova, Padova, Italy

<sup>3</sup>Università Cattolica del Sacro Cuore, Milano, Italy

<sup>4</sup>ASVAPP, Torino, Italy

## Corresponding Author:

Gian Paolo Barbetta, Dipartimento di economia e finanza, Università Cattolica del Sacro Cuore, Via Necchi 5, 20123, Milano, Italy.

Email: [gianpaolo.barbetta@unicatt.it](mailto:gianpaolo.barbetta@unicatt.it)

itself—we show that the large non-compliance with the protocol showed by the RCT is ignorable. This motivates estimating the impact of carrying out a traineeship during the experimental period on employment using non-experimental methods. The study finds that the availability of a coach boosts the patients' chances of finding a traineeship during the experimental period; moreover, undertaking a traineeship almost doubles the likelihood of being at work one year later.

**JEL CODES:** J78, J48, J38

### Keywords

mental illness, job search, RCT, traineeship

## Introduction

Mental illness (which includes disorders such as anxiety, depression, schizophrenia, and bipolar) is increasing in prevalence worldwide. According to (Layard, 2016a, 2016), in the rich countries surveyed by the (OECD, 2012), mental illness accounts for half of all illness up to the age of 45, making it the most prevalent disease among working-age people. Mental disorders have severe consequences on the lives of individuals and society. One of the most relevant consequences on individuals is the reduction of their capacity to be actively employed. Studies analyzing the association between mental illness and labor market outcomes suggest that people with mental illness experience lower employment rates, lower earnings, higher work absences, and lower productivity (Banerjee et al., 2017; Chatterji et al., 2007, 2011; Frijters et al., 2014; OECD, 2012). The lack of jobs is a source of distress for people suffering from mental illness. When the consequences for society are considered, one should remember that mental disorder is a major burden for the public purse as it accounts for close to half of disability benefits in many countries (Layard 2016b). Moreover, “if mentally ill people received treatment so that they had the same employment rate as the rest of the population, total employment would be 4% higher, adding many billions to national output” (Layard, 2016a).

While most people with psychiatric disorders experience difficulties in finding and retaining a job, those with severe mental illness (SMI) suffer in particular (OECD, 2012). Various approaches have been tried to increase the employment rate of people with SMI, from providing sheltered jobs to mandatory hiring quotas for private employers. In the US, over the last 30 years the *Supported Employment* movement has consistently shown that *individual placement and support* (IPS)—characterized by a rapid transition

to the job market, with some external assistance, as opposed to a long stay in sheltered jobs—represents the most effective way to increase the employment rate of people with SMI. A vast body of evidence on the effectiveness of IPS has accumulated since the 1990s, with multiple randomized controlled trials (RCTs) demonstrating the superiority of supported employment over a variety of other rehabilitation models, including group skills training, sheltered workshops, and psychosocial rehabilitation services. Supported employment is now widely accepted as the evidence-based practice for increasing the employment of people with SMI, and efforts are under way to disseminate it beyond the US (Becker and Bond 2020; Bond et al. 2012).

Individual placement and support is also spreading in Europe, with experiences taking place in several countries. However, in Europe this approach lacks robust evidence of its effectiveness, with very few exceptions. The EQOLISE (Enhancing the Quality of Life of Individuals with Supported Employment) project—which involved six European countries in a multi-site RCT to test the effectiveness of IPS versus the more traditional vocational training approach—showed that IPS was superior to the vocational approach in terms of the number of people entering the job market, the number of days and hours worked, and the amount of money earned (Burns et al. 2007). However, an RCT that took place in the UK a few years later and tested IPS against traditional vocational services showed no differences between the treatment and control groups in obtaining employment in the short term (Howard et al. 2010). Only medium-term results (Heslin et al. 2011) showed statistically significant differences between the two treatments. Given the small number of rigorous trials, the effect of IPS in Europe is far from being fully assessed. Moreover, differences in regulations and labor market conditions do not allow an uncritical acceptance of results obtained in the US.

In this paper, we try to reduce this gap in knowledge by reporting the results of *Lavoro&Psiche* (L&P), an RCT undertaken in Italy aimed at testing the effectiveness of a home-grown version of supported employment. The demonstration involved 311 people and provided half of them with a “job coach” for at least 24 months. Our results show that the provision of a traineeship (favored by the job coach) had a relevant impact on the probability of being at work during the first year after the treatment ended. Our results are quite relevant as they can provide Italian and European policy-makers with new evidence about the effectiveness of the IPS approach.

In analyzing the data from L&P, one must deal with the problem raised by the massive non-compliance with the original RCT protocol: in fact, many individuals ended up with an actual treatment status different from the one randomly determined. We find that in the case of L&P, non-compliance is ignorable to the estimation of the causal effect. We reach this result using an array of tests—including one test made available by the randomized setting of L&P—consistently showing that individuals experiencing a traineeship are

comparable to those not experiencing it, *as if* the two groups of individuals were randomly determined. As a straightforward implication, we compare these two groups of individuals to estimate the average causal effect of carrying out a traineeship on subsequent (short-term) labor market outcomes.

The remainder of this paper is structured as follows. The “Employment Situation of the Mentally Ill” section discusses the employment situation of people with SMI and the interventions that have tried to improve the employment rate of these people, the “Lavoro & Psiche Random Assignment Demonstration” section describes the L&P treatment and the design of the RCT, the “Results of the Demonstration” section discusses the results of the RCT, while the final section draws some conclusions.

## Employment Situation of the Mentally Ill

The extremely low employment rate of people with mental illness represents a serious challenge to public policy in every country (Dean et al. 2015, 2018). Here, we focus on the most severe forms of mental illness, like schizophrenia and bipolar disorder, leaving aside the less serious ones such as clinical depression and chronic anxiety.

For people with SMI, a salary is only one of the many benefits of employment: work can also provide a sense of purpose, improve self-esteem, and even reduce psychiatric symptoms (Llena-Nozal et al. 2004; McGurk et al. 2005). Rowland and Perkins (1988) identified four benefits of work, namely as a restorative psychological process, improving self-concept, offering a protective effect, and a social dimension. Positive and meaningful employment experiences have been linked to improved self-concept and self-efficacy, higher subjective well-being, regained self-esteem, and improved engagement in work activity with associated symptom reduction. Most importantly, work offers hope, which is vital to recover from mental illness (Boardman et al. 2003). Work exclusion erodes self-confidence and creates a sense of isolation and marginalization.

Nonetheless, there is a lack of consensus on the appropriate mix of remedies that should be applied to address the low level of employment among people with SMI. Opinions widely vary from those who still believe that the mentally ill should be confined to protected or sheltered environments to those who advocate immediate placement into the job market with some external support for job search and retention.

In addition to the sheltered/un-sheltered jobs controversy, other issues—that we only mention—are on the table. The first concerns the most suitable way to eliminate the work disincentives built into the disability insurance system, an issue relevant for most countries but mainly debated in the US and UK. The second is the usefulness of a special type of affirmative action, comprising imposing hiring mandates on private employers, as is done in most

of continental Europe (for example, in France, Germany, Spain, and Italy), as a potentially counter-productive way to overcome the barriers faced by people with mental illness.

Attitudinal as well as structural barriers prevent people with mental illness from becoming active participants in the job market, even when compared to persons with physical impairments. Some barriers are related to the symptoms of the disease themselves, such as cognitive impairments that can be found in attention deficit, psychomotor speed, verbal, learning and executive function (McGurk et al. 2005). Other barriers originate in the workplace itself. Stigma associated with mental illness creates a reluctance to hire the mentally ill and fosters low expectations of mentally ill workers in general (Hill et al. 2016). Several different approaches have been tested over time to overcome these barriers.

Since the late-1980s, a vast movement—called *Supported Employment*—has developed in the US, pushing for a whole new approach to tackle the employment situation of people with SMI. This movement claims that a rapid transition to the job market—as opposed to a long stay in sheltered jobs—is the most effective way to deal with the barriers faced by people with SMI. Supported employment is distinguished from other approaches such as vocational rehabilitation by its emphasis on rapid job search (rather than extensive vocational assessment or training), placement in real jobs (rather than set-aside jobs for persons with disabilities), integration of vocational and clinical services, attention to individual preferences with respect to job types, disclosure of one's psychiatric condition to employers, and follow-along support to facilitate job retention (McGurk et al. 2005).

A body of evidence on the effectiveness of this approach has accumulated since the 1990s, with multiple RCTs—mostly undertaken in the US—demonstrating the superiority of supported employment over a variety of other rehabilitation models, including group skills training, sheltered workshops, and psychosocial rehabilitation services (Kinoshita et al. 2013). Supported employment is now accepted as the evidence-based practice for SMI, and efforts are underway to disseminate it beyond the US (Bond et al. 2012). *Individual placement and support* is the term most often used by these attempts at dissemination.

The IPS approach is now spreading in Europe, with countries such as Belgium (De Greef 2020), England (Melleney and Kendall 2020), Italy (Rizza and Fioritti 2020), Norway (Sveinsdottir et al. 2020), Spain (Hilarion et al. 2020), Switzerland (Hoffman and Richter 2020) and the Netherlands (van Weeghel et al. 2020) discussing—or directly experiencing—this new practice.

Its diffusion notwithstanding, European countries lag behind the US when it comes to testing the effectiveness of this approach. To our knowledge, only two RCTs are in fact available in Europe. The EQOLISE project—a European multi-site/multi-country RCT—tested the effectiveness of IPS versus the

more traditional vocational training approach. In the six European centers involved in the RCT, 312 patients with SMI were randomly assigned to receive IPS ( $n = 156$ ) or vocational services ( $n = 156$ ). Patients were supported for a maximum of 18 months. Despite the extensive differences in labor market regulations, organization, and culture of mental health services among Europe and the US, the EQOLISE RCT showed that IPS was superior to the vocational approach in terms of the number of people entering the competitive market, the number of days and hours worked, and the amount of money earned. The primary outcome was the difference between the proportions of people entering un-sheltered employment in the treatment and control groups (55% vs. 28%) (Burns et al. 2007), representing intention-to-treat estimates. To our knowledge, this is the only trial testing the IPS approach and involving more than one European country.

A second RCT (the SWAN (Supported Work and Need) study) took place in the UK, where 219 individuals with SMI were randomly assigned to IPS or traditional vocational services and assessed one and 2 years later. The study showed “no evidence that IPS was of significant benefit in achieving competitive employment for individuals in South London at 1-year follow-up” (Howard et al. 2010), while at 2-year follow-up the treatment arm showed a higher employment rate (22% vs. 11%,  $p = 0.041$ ), but the employment level was quite low regardless (Heslin et al. 2011). Different timing of the two RCTs and different job market conditions of the cities in which they were undertaken (several European cities vs two boroughs of South London), could explain the striking differences in overall outcome levels between the EQOLISE and the SWAN studies. Given the small number of rigorous trials testing the IPS approach and the differences between their findings, the results of the L&P demonstration are particularly interesting.

## Lavoro & Psiche Random Assignment Demonstration

### Treatment

In 2008, partly stimulated by the first-ever involvement of an Italian mental health service in a multi-site multi-country social experiment such as EQOLISE, the Cariplo Foundation—the largest philanthropic foundation in Italy—launched the L&P demonstration. This demonstration was aimed at producing evidence on the effectiveness of a new strategy for increasing employment among severely mentally ill patients.

The placement model is a flexible version of the orthodox IPS, offering people with SMI intensive counseling and coaching by caseworkers for a period varying between 24 and 36 months. The main aim of the caseworkers was to prepare their patients for entering the job market through a series of preliminary activities (such as assessing their skills and preparing their

resumé) completed in a short time, and a traineeship as a first tool to enter the job market. Traineeships have recently become the most common path to a permanent job for young people, as well for people with disabilities, in Italy. According to Italian legislation, a traineeship for people with disability is a period (not longer than 24 months) of paid professional training in a specific company. While the company has some incentives (such as disbursing a smaller salary and being exempt from paying some social security contributions), it must prepare and manage a specific training plan for each trainee.

The focus on traineeships represents the most relevant difference compared to the orthodox IPS approach. Offering traineeships seemed like a wise decision given that the deepening of the great recession starting in 2008 made the “rapid job search” at the basis of the IPS appear more like wishful thinking than sound policy: in fact, while the project was running, the general unemployment rate in Italy shot up from 6.7% in 2008 to 12.7% in 2014, and the chances of finding permanent jobs for people with SMI (as well as other weak segments of the workforce) were strongly reduced.

Beside traineeships, the L&P protocol was very similar to the traditional IPS approach: thanks to their relatively limited workload, caseworkers could concentrate on each individual patient and pay attention to their specific preferences, caseworkers interacted closely with clinical services to deal with specific needs, they fully disclosed the psychiatric condition of their patients to perspective employers, and they followed up with both their patients and employers to facilitate job retention.

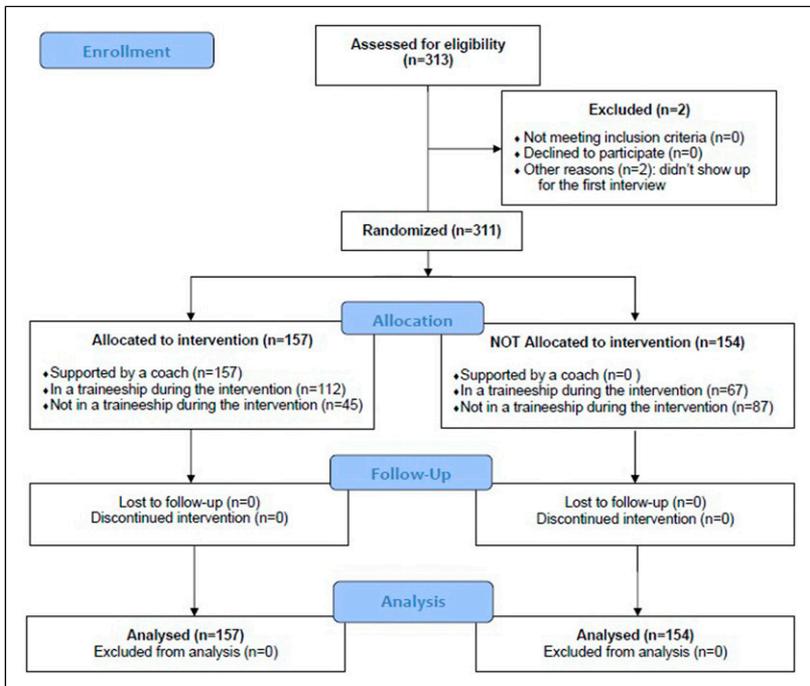
### *Method (Design of the Demonstration)*

The L&P demonstration was conducted between 2009 and 2012 and involved the 29 public mental health centers located in four provinces of Lombardy (a region of almost 10 million people in Northern Italy): Como, Lecco, Milano, and Sondrio. The recruitment of patients started in late 2009 and continued throughout the whole of 2010. Upon referral to the project by the mental health center staff, the organization in charge of project evaluation administered a face-to-face baseline interview with all patients, obtained their informed consent, and randomly assigned patients into the experimental and control groups. Patients were immediately notified by the center staff.

For budgetary reasons, a crucial design choice was made in favor of randomizing patients within sites (the 29 local mental health centers) rather than randomizing at the site level. This choice had the advantage of minimizing the number of patients to be treated. To reach the same power, randomizing sites would have required more treated cases, thus making the study more expensive than the foundation had anticipated. Of course, randomizing across individuals within sites increased the possibility of spillovers.

By the end of 2010, 313 patients with SMI had been referred by the staff of the public mental health centers. While two patients did not show up, 311 of them went through the baseline interview: 157 were randomly assigned to the treatment group, and 154 to the control group. For an average period of about two-and-a-half years, the treated group members received help in their job search by trained, qualified, and committed caseworkers (the so-called job coaches), carrying a light workload of 12-13 patients, while the control group had access to the job placement services normally available to all mental patients in public centers. All randomized patients went through the follow-up interview, and for all of them we were able to measure the outcome variables described in the next paragraph; therefore, attrition was not an issue for the RCT. [Figure 1](#) reports the flow diagram of the RCT.

[Table 1](#) contains descriptive statistics for the treatment and control groups, whereby all information was obtained from the baseline interviews. The randomization was successful, and the two groups are well balanced.



**Figure 1.** Flow diagram of the RCT.

**Table I.** Pre-Treatment Characteristics of Randomized Patients.

	Patients in the treatment group	Patients in the control group	p-value
<i>N</i>	157	154	
<b>Gender</b>			
Female	38%	38%	0.89
<b>Age</b>			
Mean	36.1	35.7	0.66
<b>Age group</b>			
Less than 24 years old	8%	8%	0.80
25–34	38%	33%	0.41
35–44	41%	47%	0.24
More than 45 years old	14%	11%	0.43
<b>Education</b>			
Primary school	1%	1%	0.98
Middle school	39%	35%	0.42
Vocational school	17%	19%	0.71
High school diploma	29%	33%	0.47
University degree	5%	5%	0.97
Missing	8%	6%	0.69
<b>Diagnosis</b>			
Bipolar disorder (F31 of the ICD-10 classification)	15%	14%	0.93
Borderline personality disorder (F60.3)	13%	12%	0.65
Paranoid personality disorder (F60.0)	10%	8%	0.46
Schizoid personality disorder (F60.1)	7%	5%	0.51
Schizophrenia and related disorders (F20)	55%	61%	0.26
Lost work capacity <sup>a</sup>			
More than 45%	61%	57%	0.55
Average lost work capacity (%)	77.11	78.84	0.42
<i>N</i>	95	88	
<b>Previous work experience</b>			
Had a job in 2008	38%	33%	0.35
Had a job in 2009	28%	27%	0.88
In traineeship in 2008	8%	11%	0.41
In traineeship in 2009	10%	13%	0.44

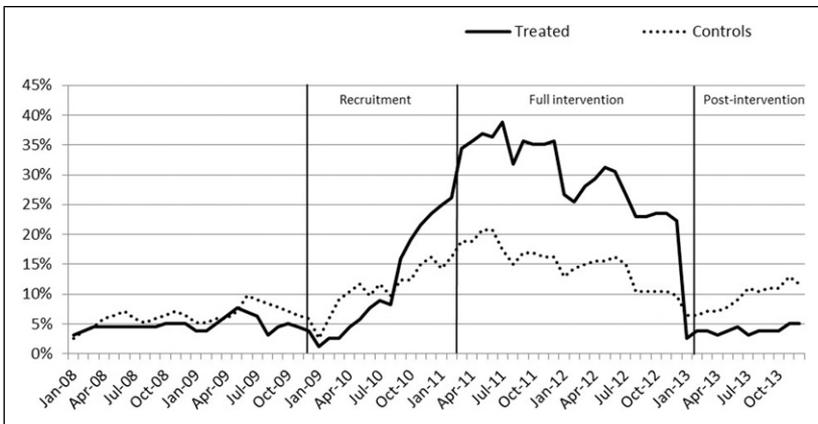
<sup>a</sup>This is a measure (certified by a public body) of the work productivity lost by patients with a mental impairment. The exact lost work capacity is only available for people with a certified disability and with a lost work capacity higher than 45%.

## Outcome Variables

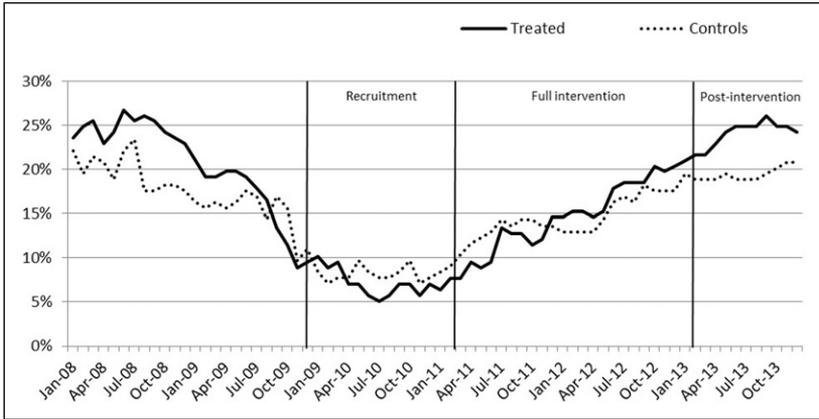
The outcomes presented in this paper are limited to employment, even though other aspects of the life of people with SMI might improve with the intervention, mainly their health condition. The reason for this choice is two-fold: on the one hand, employment remained the specific focus for the whole demonstration; on the other hand, data availability admittedly carried some weight. Around 2008, Italy made a huge leap in the quality of its employment data, introducing a legal obligation for every employer to communicate to the local Public Employment Service Office information on all job contracts started or stopped (this is called the COB archive, which stands for “Comunicazioni OBbligatorie” or “mandatory communication”). Thanks to the COB data, we were able to collect monthly individual work and training data, thus reconstructing the individual histories of all patients from 2008 through to the end of 2013. [Figure 2](#) shows the monthly figures for the fraction of the treated and control patients who had a traineeship and [Figure 3](#) the fraction who had paid employment (note that a traineeship does not count as employment).

## Results of the Demonstration

In this section, we move from the standard estimate of the *effect of making a coach available*—the so-called intention to treat (ITT)—to discover that it is statistically equal to zero. So is the estimate of the *effect of the actual experience of a traineeship*, an estimate we obtain using the random assignment



**Figure 2.** Percentage of patients in a traineeship.



**Figure 3.** Percentage of patients with a job.

to the treatment group as an instrumental variable (IV) for the actual experience of a traineeship. The bottom line is that due to non-compliance with the L&P protocol there is insufficient statistical precision in the original L&P protocol to provide a reliable estimate of the causal effects.

Nonetheless, we provide the results of an array of tests—including one test made available by the randomized setting of L&P—consistently showing that individuals experiencing a traineeship are comparable to those not experiencing it, regardless of their L&P treatment status, *as if* the two groups of individuals were randomly determined. As a straightforward implication, we compare these two groups of individuals to estimate the average causal effect of carrying out a traineeship on subsequent (short-term) labor market outcomes.

### Effect of Making a Coach Available

First, we estimate the effect of making the support of the coach *available* to patients, the so-called *ITT* effect. This is undertaken in equation (1)

$$ITT = P(L = 1|Z = 1) - P(L = 1|Z = 0) \tag{1}$$

where **L** is the working condition (1 = having a paid job for at least 1 day during the year that followed the treatment administration; 0 = being out of work for the whole year) and **Z** is the randomly determined assignment of patients to a coach (1 = being assigned to a coach; 0 = not being assigned to a coach).

**Table 2.** Specification Tests and Estimates of the Employment Effect of L&P.

	Estimates of the employment effect of L&P			Specification tests	
	(1) OLS (outcome on randomly assigned status)	(2) IV	(3) OLS (outcome on actual status)	(4) OLS for those sometime in a traineeship (T = 1)	(5) OLS for those never in a traineeship (T = 0)
Assigned to a coach ( <b>Z</b> )	0.053 (0.051)	—	—	0.014 (0.074)	-0.017 (0.069)
Took a traineeship ( <b>T</b> )	—	0.189 (0.177)	0.185 (0.050) (0.063) <sup>a</sup>	—	—
Constant	0.251 (0.035)	0.165 (0.105)	0.166 (0.038)	0.343 (0.059)	0.172 (0.042)
N	311	311	311	179	132

Column 1 reports the ITT estimate; column 2 reports the LATE estimate; column 3 reports the ATT estimate, if there is no selection bias; columns 4 and 5 report the [Black et al. \(2017\)](#) test. The binary outcome is 1 for patients at work at least 1 day during the post-experimental year. Standard errors in parentheses.

<sup>a</sup>This estimate of the standard error accounts for the selection of the estimator. See the main text for details.

Using equation (1), our estimates show a *statistically non-significant five percentage point difference* in the employment of treatment and control group members in the first full post-program year (Table 2, col. 1).

### *Effect of Experiencing a Traineeship*

Subsequently, we move to the causal effect of carrying out a traineeship. The key point that we make here is that the availability of a coach made a major difference for the probability of securing a traineeship. When comparing Figure 2—which shows the percentage of treated and control individuals who were in a traineeship—with Figure 3—which shows the percentage of the same individuals who had a job—a sharp difference in the time profiles for the two types of positions in the labor market is immediately evident. While the monthly figures for the traineeships closely follow the various implementation phases of the demonstration, the monthly figures for employment completely ignore them. The employment level does not show any major variation, not even in correspondence with the sharp phase-out of the demonstration: all job coaches were laid off in December 2012, a fact known to all involved since the spring of that year. On the contrary, the sharp decline by over 20 percentage points in the level of traineeship between December 2012 and January 2013 simply means that the support of the coach is deemed necessary for the very existence of a traineeship.

From our perspective, the best way of explaining the massive increase in traineeships is that during the implementation of the demonstration, locating firms willing to offer traineeships to mentally ill patients became the major activity of the coaches, thus making it the essential active component of treatment, and a necessary preliminary step for finding a permanent job position on the market. Therefore, given that a traineeship is probably the most significant component of the service provided by the coach, we focus on the causal effect of experiencing a traineeship on the probability of subsequent employment.

### *Non-Compliance Entering Into the Picture*

Based on the scrutiny of Figures 2 and 3, one could say that while the members of the treatment group might have received less than what was stated in the L&P protocol, the members of the control group might have received more than envisioned in the design. In fact, during the recruitment phase, the percentage of patients in a traineeship sharply increases not only for the treatment group (a result expected because of the activities of the coaches) but also the control group. In our opinion, this is due to a complex set of circumstances.

First, as already mentioned, randomizing patients within sites increased the possibility of *contamination*: while the members of the control group might have received services formally different from those of the treatment group, from a practical perspective they might have been sufficiently similar to dilute away any difference in the treatment received by the treated and control patients. This could be the possible consequence of the stress created for the local health staff, who: (i) saw their patients being assigned to either the treatment or control conditions without being consulted (which is the essence of random assignment, an idea that most of those operating in the social sector loath, often with a vengeance), (ii) had to explain to the patients the meaning of the informed consent, and finally (iii) had the unpleasant duty of communicating to each patient the outcome of randomization, namely their status with respect to receiving the support of the coach versus reconfirming their access to the existing services. Given that most of the local health staff in charge of job placement (along with many other duties) had built a network of firms willing to offer traineeships to mentally ill patients over the years, the temptation was present to use such resources to compensate for the “unjust” situation created by the randomization: for many, such a temptation was too difficult to resist, and they caved in very easily. This explains the unusually high level of traineeship activated for the members of the control group. The possible compensation of patients assigned to the control group was acknowledged by one of the implementation reports of the project, which stated that “for several patients, the assignment to the control group increased the chance of receiving support for job placement” by the operators of the health centers (Dusio et al., 2011, 33, our translation). Nonetheless, both implementers and supervisors underestimated the possible relevance of the problem, which appears quite significant when the final data are considered.

Moreover, we might add that the L&P program was met with some hostility, because the existing mental health staff had much more experience than the newly hired job coaches but much less time available to devote to job placement. To summarize, the insertion of L&P in the structured mental health system of the Lombardy region created tensions along some predictable fault lines.

Consequently, a credible interpretation of the demonstration results should carefully consider the issue of *non-compliance*. In this situation, the trouble in reaching credible causal conclusions is that differences in behavior taking place after randomization are *de facto* trying to undo what randomization achieves. In our case, “classical” cross-overs are not possible because the coaches were not allowed to make substitutions or add patients to the list that they received for coaching. Nonetheless, a less classical form of cross-over for the controls—securing a traineeship—is quite visible. As Figure 2 shows, back in 2008-09, before the demonstration started, the monthly rate of utilization of traineeships was about 5% among both groups of patients later to

become the treated and control patients. In 2011-12, we obtain proportions of 71% and 43% of treated and control patients in a traineeship, respectively. As far as non-compliance goes, this situation is difficult to beat (Heckman et al., 2000).

### Formal Approach

As before, let  $\mathbf{Z}$  be the randomly determined assignment of patients to a job coach. Let  $\mathbf{T}$  be equal to one for those carrying out a traineeship during the experimental period and equal to zero for those who did not. This status is *not* randomly assigned as according to the L&P protocol, only the job coach is randomly assigned. However, being assigned to a job coach makes a major difference for the probability of taking part in a traineeship (see Figure 2).

In this situation, two-sided non-compliance is possible, since there are both no-shows—that is, patients assigned to a job coach but not taking part in any traineeship ( $\mathbf{Z} = 1, \mathbf{T} = 0$ )—and cross-overs, that is, patients not assigned to a job coach and nonetheless taking part in a similar traineeship ( $\mathbf{Z} = 0, \mathbf{T} = 1$ ). It should be noted that all traineeships are born equal in the sense that they are all regulated by the same national law. Moreover, given that the L&P is a very minor affair compared to the size of the labor market, holding constant the characteristics of the trainee, the traineeship experience was very much the same across the two groups.

Following the classical analysis of Angrist, Imbens and Rubin (1996), who elaborate on previous results by (Imbens and Angrist 1991), we use  $\mathbf{Z}$  as an instrumental variable for  $\mathbf{T}$  to estimate the so-called local average treatment effect (LATE). The key condition for the validity of this strategy is that being assigned a job coach must not have any effect on the outcome other than through its effect on the *probability of experiencing a traineeship spell*. This is the so-called *exclusion restriction* on which any IV strategy relies. This setup might seem to be a post-hoc rationalization driven by data availability. Considering as the treatment only something that we observe—namely the traineeship status  $\mathbf{T}$ —in principle might mean ignoring the existence of other, possibly beneficial services provided as components of the treatment. The anecdotal evidence that we gathered suggested otherwise, and we are convinced of its plausibility. Reassuringly, below we show that *this exclusion restriction bears testable implications*, which in fact are not rejected by our data (see the “Testing for the Ignorability of Non-Compliance” section).

The IV strategy also requires an additional condition named *monotonicity*: it must not be the case that some patients taking part in a traineeship if not assigned a job coach would not take part in the traineeship in case they were assigned a job coach. Following our argument above, if the driver of non-compliance is the behavior of the local health staff, monotonicity seems entirely plausible: the bias of the staff is toward providing a traineeship to their

patients regardless of their randomly determined status  $Z$ , that is, moving patients from  $T = 0$  to  $T = 1$  as far as possible, not the other way around. This is sufficient to secure monotonicity.

Finally, the IV strategy identifies the average causal effect only for a specific subset of patients, the so-called *compliers*, namely those who secure a traineeship *if and only if* assigned a case worker. When both the treatment status  $T$  and the instrument  $Z$  are binary variables, the IV estimator is the so-called Wald ratio, in which the causal effect on the outcome of being assigned a job coach is divided by the compliance rate, that is, the difference in the take-up rate between those assigned and not assigned a job coach, respectively, as shown in equation (2)

$$\text{LATE} = \frac{P(L = 1|Z = 1) - P(L = 1|Z = 0)}{P(T = 1|Z = 1) - P(T = 1|Z = 0)} \quad (2)$$

As applied to the case of L&P, the LATE is as large as 18.9 points (Table 2, col. 2), more than three times the 5 points that we obtained from estimating the causal effect on  $L$  of being *assigned* a job coach.

There are two problems with this estimate. First, unsurprisingly, it is fairly imprecise (see the standard errors in col. 2 of Table 2), which is typically the case in most IV applications (Murray 2006). Second, Angrist, Imbens, and Rubin (1996) show that the denominator of the LATE is an estimate of the size of the sub-population of *compliers*. In this instance, it is as large as 28% of the whole sample. Strictly speaking, the estimated average causal effect refers only to this specific sub-population.

As the bottom line so far, both the ITT—the experimental estimate of the causal effect of making a coach available—and the LATE—the IV estimate of the causal effect of a traineeship spell—are positive but imprecisely estimated.

### Testing for the Ignorability of Non-Compliance

Here, we show that patients experiencing a traineeship spell ( $T = 1$ ) during the experimental period are comparable to those not experiencing it ( $T = 0$ ) as if the two groups were randomly determined.

The first test that we propose to support our hypothesis is based on Black et al. (2017) (but, the key insight driving this result dates back to Imbens and Rubin (1997)). They propose a simple test for the *presence of selection bias* in an IV setup. Failing to reject the null hypothesis—that is, failing to find compelling evidence for the presence of selection bias—one can use the (typically much) more precise OLS to estimate the causal effect of  $T$  on  $L$  instead of using the IV. Moreover, one obtains evidence consistent with the case in which the LATE coincides with the average causal effect over the

whole group of patients recruited for L&P, that is, not only an average effect for compliers.

The test amounts to running the OLS regression of the outcome  $L$  on the instrumental variable  $Z$  separately for the two sub-groups indexed by  $T$ . Consider the case of  $T=1$ . All of these patients carried out a traineeship. Those with  $Z=0$  secured it despite having been randomly denied the job coach. In the Angrist, Imbens and Rubin (1996) terminology, they are the *always-takers*. Those with  $Z=1$  secured it either because they were assigned a job coach—the *compliers*—or because they are always-takers. Finding that the outcome  $L$  does not depend (on average) on  $Z$  means that always-takers and compliers are equivalent (on average) with respect to  $L$ . In the case of the sub-group  $T=0$ , the story is the same but the always-takers are replaced by the *never-takers*, namely those who do not carry out a traineeship whether assigned a job coach or not. Of course, this is a *partial* test for the presence of selection bias, in the sense that there is no way to compare compliers to always-takers in case they *did not* carry out a traineeship, nor to compare compliers to never-takers in case they carried out a traineeship.

Relevant to our problem, Black et al. (2017) also note that this test has power against the violation of the exclusion restriction, because if the treatment status  $Z$  had an impact on the outcome via any channel other than the traineeship status  $T$ , then conditioning on  $T$  the outcome would depend on  $Z$ .

Columns 4 and 5 in Table 2 show that the null hypothesis is not rejected for both sub-groups (by the way, also note that both point estimates are very small). Summing up, *we find no evidence of selection bias nor violation of the exclusion restriction*.

The second test that we propose to complement the previous one is a standard in this literature: comparing individuals exposed/not exposed to a traineeship spell with respect to characteristics known to be unaffected by the intervention. In a baseline survey, *prior* to randomization we measured many individual characteristics, including labor market experience in the 2 years prior to the experimental period. In Table 3, we compare patients carrying out a traineeship during the experimental period to those not carrying it out with respect to this set of pre-treatment characteristics. Out of the fourteen characteristics that we consider, only in one single case is the difference between the two groups (marginally) statistically significant ( $0.05 < p\text{-value} < 0.1$ ). Incidentally, most of the differences are also negligible in absolute value. We also test the null hypothesis jointly on the fourteen variables by running a regression of the traineeship status  $T$  on those variables. Again, the F-test—reported in the bottom line of Table 3—does not reject the null hypothesis.

The bottom line of these tests is that all available evidence consistently points to the ignorability of non-compliance, that is, the two groups of patients—those who experienced a traineeship during the experimental period

**Table 3.** Comparison of Patients That Carried Out a Traineeship (T=1) to Those Who Did Not (T = 0) with Respect to Pre-Randomization Characteristics.

	(1) T = 0	(2) Difference between T = 1 and T = 0	(3) <i>p</i> -value for the difference
Age	36.80	-0.70	0.41
Years of schooling	10.62	0.17	0.61
Female	0.37	0.01	0.80
Lost work capacity (%)	79.45	-2.52	0.25
Lost work capacity>45%	0.55	0.06	0.28
Bipolar disorder	0.14	0.00	0.97
Borderline personality disorder	0.11	0.02	0.59
Paranoid personality disorder	0.08	0.01	0.72
Schizoid personality disorder	0.04	0.03	0.32
Schizophrenia and related disorders	0.61	-0.06	0.29
Had a job in 2008	0.36	0.00	0.98
Had a job in 2009	0.27	0.01	0.90
In traineeship in 2008	0.07	0.05	0.15
In traineeship in 2009	0.08	0.07*	0.06
F test	0.90		0.55

\*0.05<*p*-value<0.10. The F-test is for the null hypothesis of *joint* irrelevance of the whole set of characteristics in the regression of the treatment status T.

and those who did not—are comparable. Of course, being based on a specific—even if rich—set of observable characteristics, these tests provide only partial evidence in favor of the hypothesis of no selection bias. Based on this evidence, we take the post-program observable difference in work experience between the two groups as an estimate of the average causal effect for those recruited in L&P: as a result of a traineeship, the probability of having a paid job 1 year later rises from 0.17 to 0.35 (this is the OLS estimate in [Table 2](#), col. 3). The estimated impact is highly significant ( $t$ -stat=3.7).<sup>3</sup>

There is a potential issue with the estimated standard error of this estimate of the average causal effect as it does not consider that the selected estimator results from a formal test designed to choose between two competing estimators: the OLS and IV (see the discussion in [Guggenberger and Gitanjali 2012](#)). Under the null hypothesis of the ignorability of non-compliance, in hypothetical repeated sampling the null would be rejected with a probability as large as the level of the test. To account for this, we inflate the sampling variance of the OLS by evaluating the weighted average of the sampling

variances of the IV and OLS (columns 2 and 3, respectively, in [Table 2](#)) with weights 0.95 and 0.05, respectively, i.e. the probabilities of accepting/rejecting the null when it holds true in a test with a level of 0.05. The resulting standard error is marginally larger than the conventional OLS one, but it does not affect our main conclusion (see the second standard error in [Table 2](#), col. 3, last but one row).

## Conclusions

Based on the results of our set of specification tests, we conclude that comparing the *actual* treatment group to the *actual* control group (the OLS estimate in [Table 2](#) column 3) provides an estimate of a meaningful causal effect of the L&P intervention, notwithstanding the strong degree of non-compliance with the original protocol. Paradoxically, the main result that we obtain from the L&P demonstration—originally designed to be an RCT—exploits the random assignment to the two treatment arms only to validate a non-experimental estimator. Even if in principle one cannot exclude that the two actual treatment arms differ with respect to something unobservable to us, the whole evidence that we have points to the equivalence of the two groups, as if they had been randomly selected.

The validity of the OLS estimator provides two major advantages over using the randomly assigned treatment status as an IV for the actual status. First, there is a large gain of precision switching from the IV to the OLS estimator (compare columns 2 and 3 in [Table 2](#)): using the OLS, we safely conclude that experiencing a traineeship spell during the experimental period approximately doubles the probability of having a paid job 1 year later, raising it from 0.17 to 0.35. Second, the IV strategy estimates the average causal effect for a very specific sub-population (in this case, it amounts to approximately 28% of the pool of L&P participants), whereas the OLS provides a valid estimate for the whole pool of L&P participants.

As a result of these estimates, we can state that a traineeship should be considered as a valid tool to increase the chances of people with SMI finding a job. In this respect, our result represents an original contribution to the debate about the role of IPS in fostering the employment of people with SMI. We show that when made easier and more fluid thanks to the role played by a coach or a social worker, a traineeship—a form of a light and cheap contractual relationship between workers and employers—represents an effective way to introduce people with SMI into the job market. The light contractual relationship represented by traineeship can be very helpful in countries characterized by strong public regulation of the job market. In fact, the lower cost (compared to standard contracts) and the ease of interrupting the contract can overcome the reluctance of entrepreneurs to hire people with SMI.

The estimate of the cost-benefit of this intervention is well beyond the scope of this paper. However, the data produced during the RCT show that the average cost of patients belonging to the treatment group for a treatment lasting about 24 months is in the range of €7,700. This cost should compare to the benefit represented by the different earnings of the patients attending a traineeship compared to the patients not attending a traineeship. Most unfortunately, these data are not available, whereby more research in this area is necessary to assess the cost-effectiveness of this IPS interventions.

## Acknowledgments

The Lavoro & Psiche demonstration was entirely financed by the Cariplo Foundation, whose generous support is acknowledged. We thank five anonymous reviewers for helpful suggestions.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Gian Paolo Barbetta  <https://orcid.org/0000-0001-5865-3766>

## Notes

1. See Heckman et al. (2000) for a situation that may rival our case.
2. They elaborate on previous results by Imbens and Angrist (1991).
3. Our results do not change when we control for a set of pre-randomization individual covariates. Moreover, using the number of days at work during the post-experimental year as an outcome, we obtained the same results. Results available on request.

## References

- Angrist, J., Imbens, G., & Rubin, D. B. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434), 444-455.
- Banerjee, S., Chatterji, P., & Lahiri, K. 2017. "Effects of Psychiatric Disorders on Labor Market Outcomes: A Latent Variable Approach Using Multiple Clinical Indicators." *Health Economics* 26, 184-205.

- Becker, D. R., & Bond, G. R. 2020. "Commentary on Special Issue on Individual Placement and Support (IPS) International." *Psychiatric Rehabilitation Journal* 43(1), 79-82.
- Black, D., Joo, J., LaLonde, R., Smith, J. A., & Taylor, E. J. 2017. "Simple Tests for Selection: Learning More from Instrumental Variables." *CESifo Working Paper Series* N. 6392.
- Bond, G., Drake, R. E., & Becker, D. R. 2012. Generalizability of the Individual Placement and Support (IPS) Model of Supported Employment Outside the US. *World Psychiatry* 11, 32-39.
- Boardman, J., Grove, B., Perkins, L., & Shepherd, G. 2003. Work and Employment for People with Psychiatric Disabilities. *British Journal of Psychiatry* 182, 467-468.
- Burns, T., Catty, J., Becker, T., Drake, R. E., Fioritti, A., Knapp, M., Lauber, C., Rossler, W., Tomov, T., van Busschbauch, J., Whita, S., & Wiersma, D., The EQOLISE Group. 2007. "The Effectiveness of Supported Employment for People With Severe Mental Illness: A Randomized Controlled Trial in Six European Countries." *The Lancet* 370, 1146-1152.
- Chatterji, P., Alegria, M., Lu, M., & Takeuchi, D. 2007. "Psychiatric Disorders and Labor Market Outcomes: Evidence from the National Latino and Asian American Study." *Health Economics* 16, 1069-1090.
- Chatterji, P., Alegria, M., & Takeuchi, D. 2011. "Psychiatric Disorders and Labor Market Outcomes: Evidence from the National Comorbidity Survey-Replication." *Journal of Health Economics* 30, 858-868.
- Dean, D., Pepper, J. V., Schmidt, R., & Stern, S. 2015. "The Effects of Vocational Rehabilitation for People with Cognitive Impairments." *International Economic Review* 56(2), 399-426.
- Dean, D., Pepper, J. V., Schmidt, R., & Stern, S. 2018. "The Effects of Vocational Rehabilitation Services for People with Mental Illness." *Journal of Human Resources* 12(1), 1-37.
- De Greef, V. 2020. "Analysis of Barriers and Facilitators of Existing Programs in Belgium for the Purpose of Implementing Individual Placement and Support (IPS)." *Psychiatric Rehabilitation Journal* 43(1), 18-23.
- Dusio, F., Martini, A., & Nava, L. (2011). *Lavoro&Psiche. Analisi critica di implementazione*. Mimeo.
- Frijters, P., Johnston, D. W., & Shields, M. A. 2014. "The Effect of Mental Health on Employment: Evidence from Australian Panel Data." *Health Economics* 23, 1058-1071.
- Guggenberger, P., & Gitanjali, K. 2012. "On the Size Distortion of Tests After an Overidentifying Restrictions Pretest." *Journal of Applied Econometrics* 27, 1138-1160.
- Heckman, J., Hohmann, N., Smith, J., & Khoo, M. 2000. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment." *The Quarterly Journal of Economics* 115(2), 651-694.

- Heslin, M., Howard, L. M., Leese, M., McCrone, P., Rice, C., Jarrett, M., Spokes, T., Huxley, P., & Thornicroft, G. 2011. "Randomized Controlled Trial of Supported Employment in England: 2 Year Follow-Up of the Supported Work and Needs (SWAN) Study." *World Psychiatry* 10, 132-137.
- Hilarion, P., Bonet, P., Cid, J., Otin, J. M., Espallargues, M., Molina, C., & Suñol, R. 2020. "Implementation of the Individual Placement and Support Pilot Program in Spain." *Psychiatric Rehabilitation Journal* 43(1), 65-71.
- Hill, M. J., Maestas, N., & Mullen, K. J. 2016. "Employer Accommodation and Labor Supply of Disabled Workers." *Labour Economics* 41, 291-303.
- Hoffmann, H., & Richter, D. 2020. "Supported Employment in Switzerland—Are We on Track?" *Psychiatric Rehabilitation Journal* 43(1), 72-75.
- Howard, L. M., Heslin, M., Leese, M., McCrone, P., Rice, C., Jarrett, M., Spokes, T., Huxley, P., & Thornicroft, G. 2010. "Supported Employment: Randomised Controlled Trial." *British Journal of Psychiatry* 196(5), 404-411.
- Imbens, G. W., & Angrist, J. D. 1991. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2), 467-475.
- Imbens, G. W., & Rubin, D. B. 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *The Review of Economic Studies* 64(4), 555-574.
- Kinoshita, Y., Furukawa, T.A., Kinoshita, K., Honyashiki, M., Omori, I. M., Marshall, M., Bond, G. R., Huxley, P., Amano, N., & Kingdon, D. 2013. "Supported Employment for Adults with Severe Mental Illness." *Cochrane Database of Systematic Reviews* 9, CD008297.
- Layard, R. (2016a). *The Economics of Mental Health* (p. 321). Bonn, Germany: IZA World of Labor.
- Layard, R. 2016b. "Why We Should Spend More on Mental Health." *American Journal of Medical Research* 3(1), 188-206.
- Llena-Nozal, A., Lindeboom, M., & Portrait, F. 2004. "The Effect of Work on Mental Health: Does Occupation Matter?" *Health Economics* 13, 1045-1062.
- McGurk, S.R., Mueser, K.T., & Pascaris, A. 2005. "Cognitive Training and Supported Employment for Persons with Severe Mental Illness." *Schizophrenia Bulletin* 31(4), 898-909.
- Mellency, L., & Kendall, T. 2020. "Individual Placement and Support (IPS) in England." *Psychiatric Rehabilitation Journal* 43(1), 76-78.
- Murray, M. P. 2006. "Avoiding Invalid Instruments and Coping with Weak Instruments." *Journal of Economic Perspectives* 20(4), 111-132.
- OECD. (2012). *Sick on the Job? Myths and Realities About Mental Health at work*. Paris, France: Mental Health and Work, OECD Publishing.
- Rizza, R., & Fioritti, A. 2020. "Is Individual Placement and Support an "Active" Labor Market Policy?" *Psychiatric Rehabilitation Journal* 43(1), 60-64.
- Rowland, L. A., & Perkins, R. E. 1988. "You Can't Eat, Drink or Make Love Eight Hours a Day. The Value of Work in Psychiatry." *Health Trends* 20, 75-79.

- Sveinsdottir, V., Bull, H. C., Evensen, S., Reme, S.E., Knutzen, T., & Lystad, J. U. 2020. "A Short History of Individual Placement and Support in Norway." *Psychiatric Rehabilitation Journal* 43(1), 9-17.
- van Weeghel, J., Bergmans, C., Couwenbergh, C., Michon, H., & de Winter, L. 2020. "Individual Placement and Support in the Netherlands: Past, present, and Future Directions." *Psychiatric Rehabilitation Journal* 43(1), 24-31.