China and Italy School of Policy 中国－意大利政策培训项目

Florence, 30th October 2013

Programme evaluation: the counterfactual approach

Leonardo Grilli

UNIVERSITÀ DEGLI STUDI FIRENZE

DiSIA

DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

grilli@disia.unifi.it

local.disia.unifi.it/grilli

Evaluation of a programme

- ***Programme***: policy intervention aiming at changing a given condition or behaviour

- ***Question***: has the programme been effective? (impact evaluation)

- ***Problem***: disentangle the effect of the program from other effects (e.g. the natural evolution of the phenomenon)

Three aspects of a programme

For example, a training programme for unemployed people

- ***Target population*** (of persons, firms…)
  - unemployed people residing in Tuscany
- ***Treatment*** (intervention)
  - training programme consisting in 30 hours of lessons + a stage
- ***Outcome*** [often several outcomes are relevant]
  - duration of unemployment
  - wage
  - …

Difficulties in the evaluation

- The effect of the programme is confounded with other effects
  - natural evolution of the phenomenon
  - differences between treated and untreated subjects unrelated to the programme
- Heterogeneity in the response of subjects
  - the programme could be beneficial for young people and ineffective for the elderly
- Heterogeneity in the implementation of the policy
  - courses offered in Florence could be different from those offered in Pisa

Causal effect: definition

- Evaluating the impact of a programme entails to ascertain if the intervention actually is the cause of what we observe
- ***Potential outcomes***: for each unit (person, firm …), before the treatment is assigned there are two possible outcomes:
  - Y(*treated*) = outcome if treated
  - Y(*untreated*) = outcome if not treated ('control')
- ***Causal effect***: for each unit, it is a comparison between the two potential outcomes, usually the difference Y(*treated*)-Y(*untreated*)

Causal effect: example

- Let us consider an intervention aimed at increasing wages of workers
- Suppose Mario Rossi has the following outcomes
  - potential outcome if treated =15 euros/hour
  - potential outcome if not treated =13 euros/hour

    → Mario Rossi's causal effect=15–13 = 2 euros/hour

Causal effects at the individual level cannot be estimated …

- Each unit is either treated or untreated, thus one potential outcome is observed and the other one is missing
- The causal effect for any given unit cannot be estimated since one of the two outcomes is missing by definition, namely it is a ***counterfactual***
  - if Mario Rossi is assigned to the intervention, the outcome under treatment is observed (it is 15), whereas the outcome under no-treatment cannot be observed, it is a counterfactual (no hope to know it) → the casual effect for Mario Rossi is not observable because 15 - ? = ?

… but causal effects at the population level can be estimated

- The average causal effect for a population can be estimated: we need a population where some units are treated (*treatment group*) and other units are untreated (*control group*) → under certain assumptions, the difference in the average outcome in the two groups is a valid estimate of the *average causal effect*
  - we cannot estimate the causal effect for Mario Rossi, but we can estimate the average causal effect for a population of workers (provided we have data on two groups of workers: a group of workers assigned to the intervention, and a group not assigned)

Counterfactual causal inference

- The key of the *counterfactual* (or *potential outcomes*) approach is that we can estimate an average causal effect on a population by constructing appropriate counterfactuals (i.e. the missing potential outcomes)
- The main difficulty is the ***selection bias***: units in the treatment group may be systematically different from units in the control group even before the intervention
  - e.g. unemployed people who choose to enrol in a training course are usually different from those not enrolling in terms of age, education, ability, motivation

Selection bias and identification

- Selection bias implies that the treatment group and the control group are systematically different even before the intervention
- In a formula, $D = E + S$
  - D = observed average difference in the outcome between treated group and control group
  - E = true causal effect (average difference in the outcome due to the treatment)
  - S = selection bias (average difference in the outcome due to selection)
- To estimate the causal effect E, we need a method to eliminate S (so called 'identification strategy')
  - most techniques are devised to yield $S \cong 0$, so that $D \cong E$ (namely, the causal effect is approximately equal to the observed difference)

Identification strategies

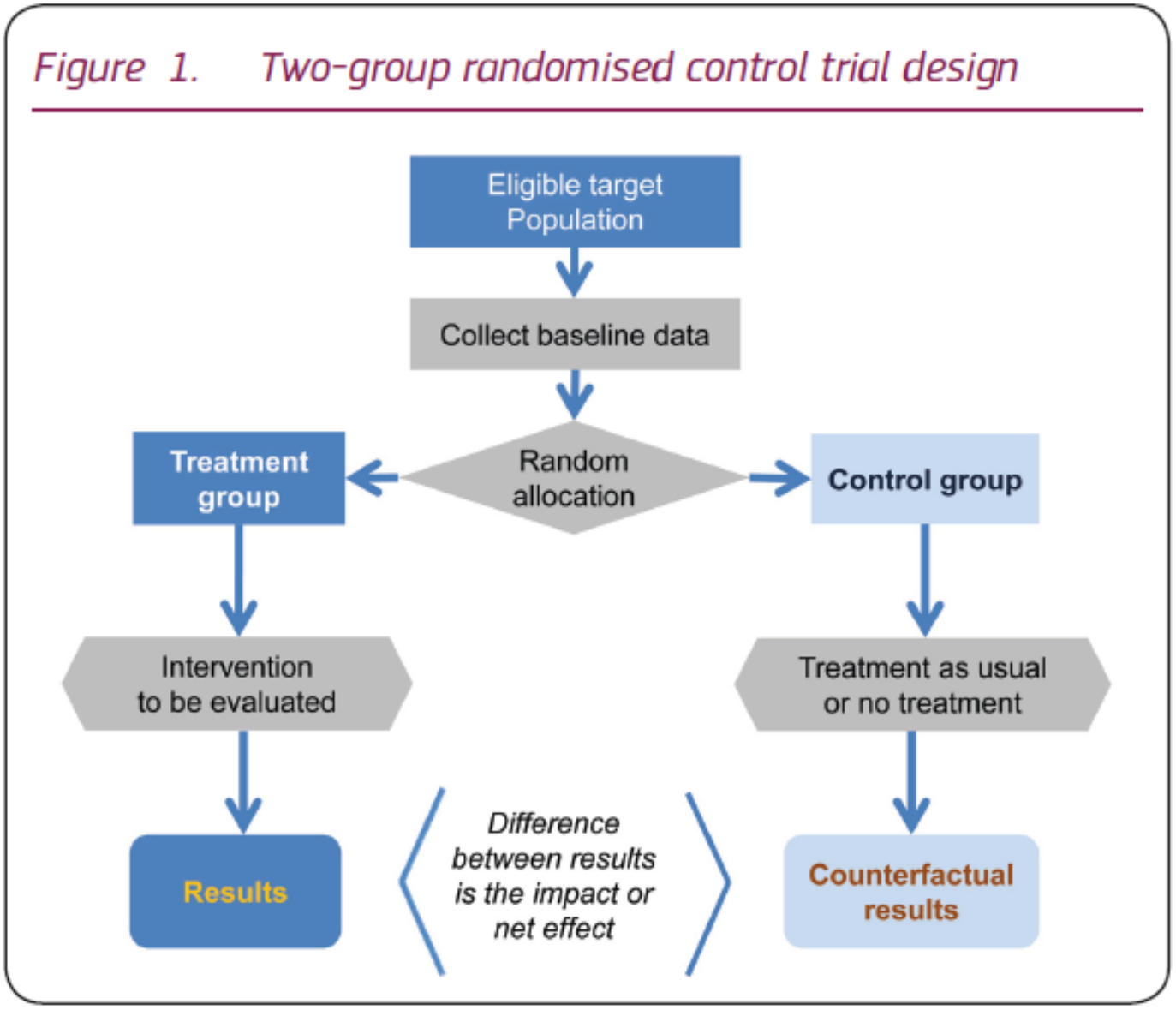
Selection bias can be eliminated or reduced by

- Design (how data are collected)
  - Randomised experiments
- Methods of analysis (how data are analysed)
  - Matching, e.g. propensity score matching
  - DiD (Difference-in-Differences)
  - RD (Regression Discontinuity)
  - Instrumental variables

Randomised experiments

- The ideal strategy: if the treatment is assigned at random the treated and untreated groups do not have any systematic difference → the selection bias is $S \cong 0$
- Limitations:
  - Unfeasible in some cases (e.g. incentives to depressed areas)
  - Ethical issues
  - Careful planning, expensive implementation
  - The experiment may change the behaviour of the units (so called 'randomisation bias')
  - Non-compliance, missing responses

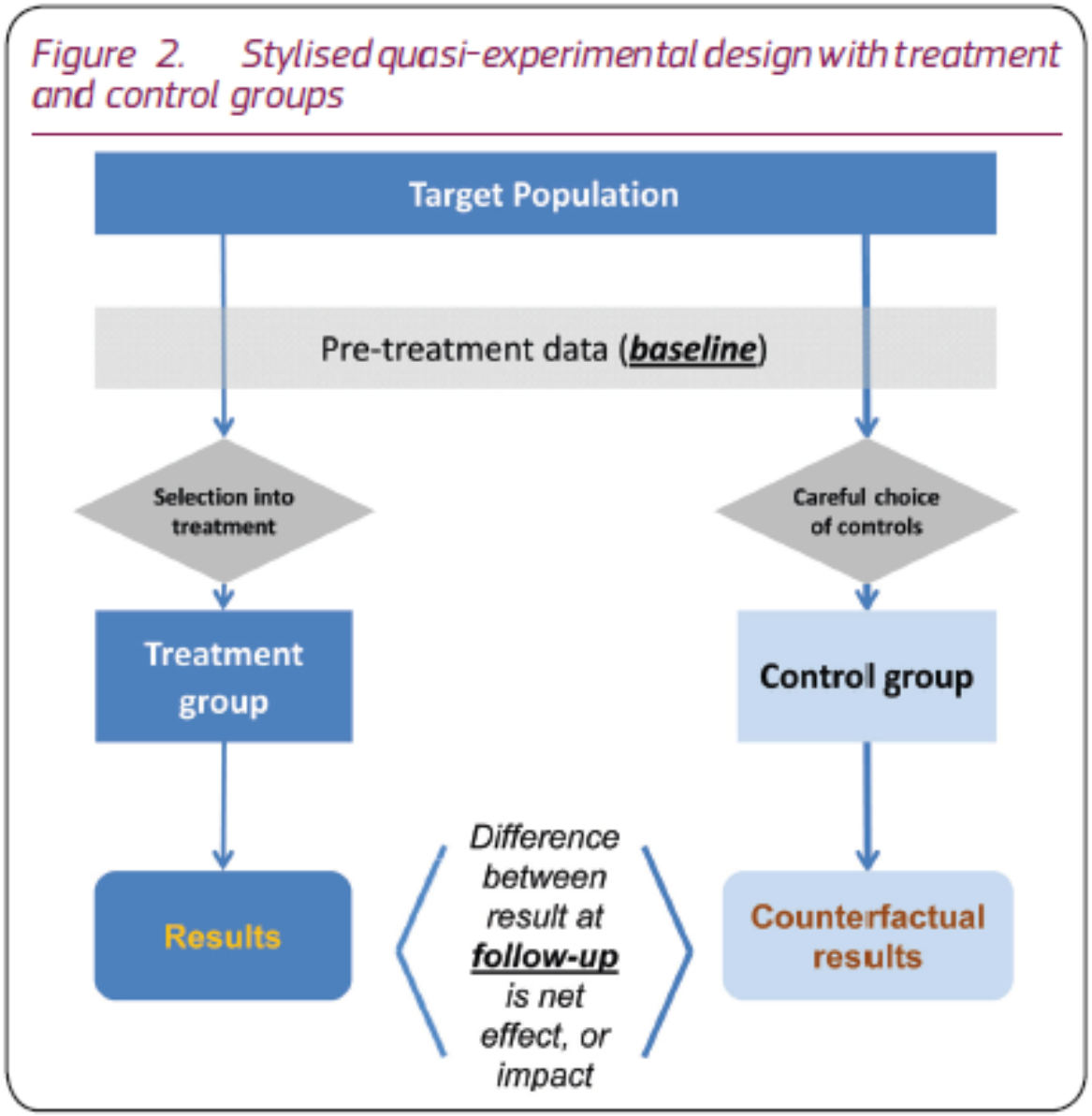
Figure 1. Two-group randomised control trial design



Source: European Commission, Design and commissioning of counterfactual impact evaluations. doi: 10.2767/94454

Methods for non-experimental data

- Randomised experiments are rare in the social sciences
- Therefore, we need methods for non-experimental data (also called 'quasi-experimental data', or 'observational data') that essentially seek to mimic randomisation
- In non-experimental settings a unit is treated or untreated depending on unknown factors out of the control of the analyst, usually including a free decision of the unit itself
- We say that a unit is 'selected into treatment' when such a unit has entered the treatment, regardless of the reasons



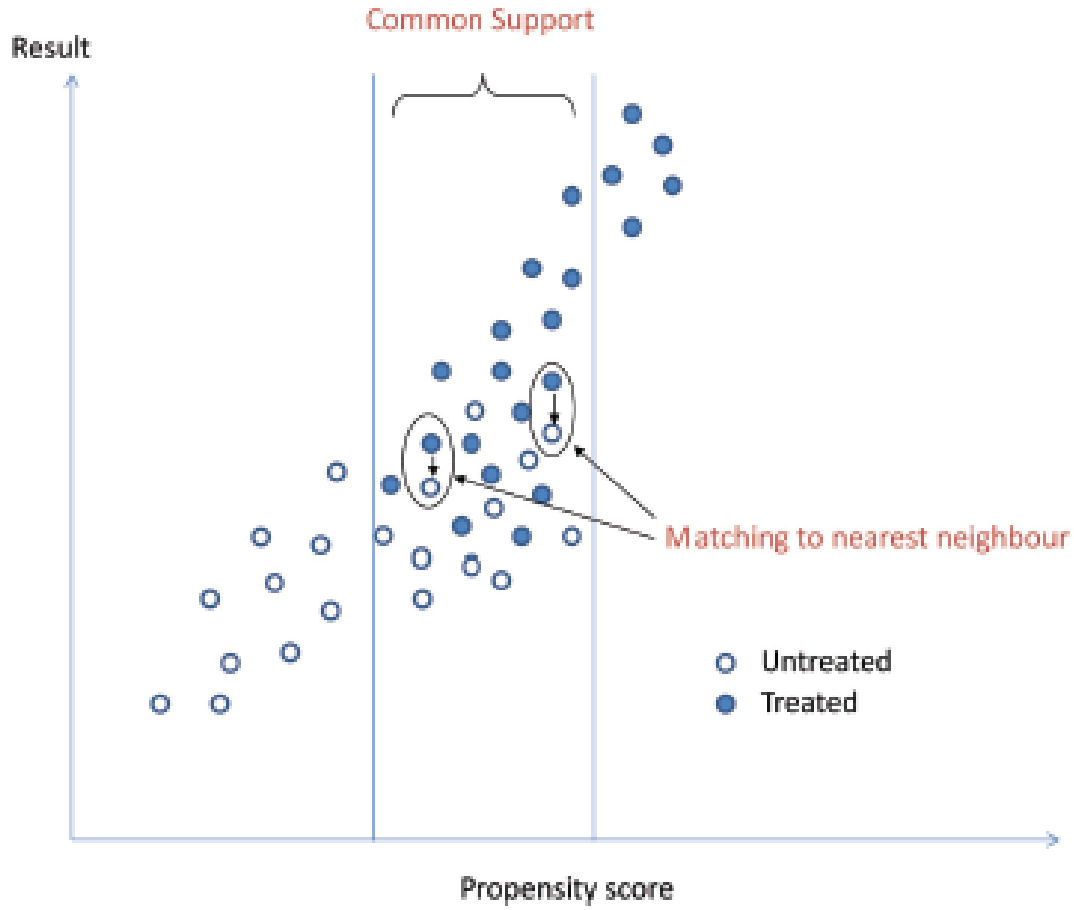
Source: European Commission, Design and commissioning of counterfactual impact evaluations. doi: 10.2767/94454

Matching

- Treated and untreated units may be different due to
  - **Selection on observables**, i.e. the selection into treatment depends on observed characteristics (e.g. age, education)
  - **Selection on unobservables**, i.e. the selection into treatment depends on unobserved characteristics (e.g. ability, motivation)
- Matching methods aim at building a control group that is similar to the treatment group for all the observed characteristics → the selection bias due to observables vanishes
- Usually, the matching is based on the propensity score

  propensity score = probability that any given unit is selected into treatment, expressed as a function of observed characteristics
- Key assumption: selection into treatment is driven solely by observable characteristics (selection on observables)

Figure 3. Illustration of the prospensity score approach



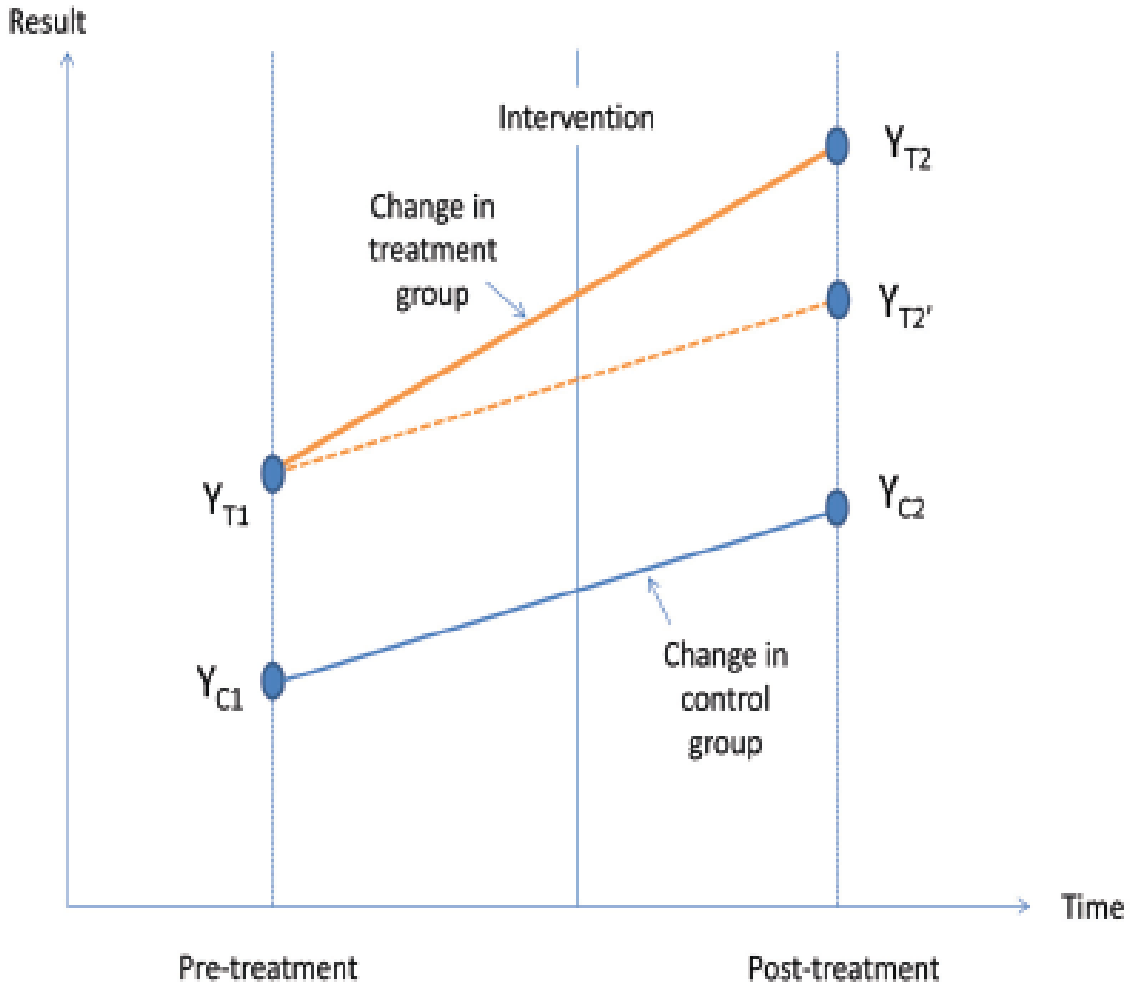
Source: European Commission, Design and commissioning of counterfactual impact evaluations. doi: 10.2767/94454

The process is essentially repeated until every treated case is matched to an untreated case within the region of common support.
The 'nearest neighbour' to any member of the treatment group is the control group observation with the closest propensity score. Once two groups have been formed, mean results can be compared in order to obtain an estimate of impact.

Difference-in-Differences (DiD)

- The DiD method can be used when the outcome variable has been measured before and after the intervention for both the treatment group and the control group
  - Before intervention (time 1): $D(1) = S(1)$
  - After intervention (time 2): $D(2) = E + S(2)$
  - DiD: $D(2) - D(1) = E + [S(2) - S(1)]$
  - If the difference due to selection bias is constant, i.e. $S(2) = S(1)$, then $D(2) - D(1) = E$
- The DiD method accounts for selection on observables (by using matching or regression) and it also accounts for selection on unobservables, as long as their effect is time-constant

Figure 4. Illustration of difference-in-differences approach



Source: European Commission, Design and commissioning of counterfactual impact evaluations. doi: 10.2767/94454

Y_{C1}= outcome of control group before treatment

Y_{T1}= outcome of treatment group before treatment

Y_{C2}= outcome of control group after treatment

Y_{T2}= outcome of treatment group after treatment

DiD = $(Y_{T2} - Y_{C2}) - (Y_{T1} - Y_{C1})$

Research

At the *Department of Statistics, Informatics, Applications* of the University of Florence there is a research group (F. Mealli, A. Mattei, C. Rampichini, L. Grilli) about statistical issues related to causal inference

- **Mealli F.**, **Rampichini C.** (2012) Evaluating the effects of university grants using regression discontinuity designs, Journal of the Royal Statistical Society A, 175, 775-798.
- Frumento P., **Mealli F.**, Pacini B., and Rubin D.B. (2012) Evaluating the Effect of Training on Wages in the Presence of Noncompliance, Nonemployment, and Missing Outcome Data, Journal of the American Statistical Association, 107, 450-466.
- **Mattei A.**, **Mealli F.** (2011) Augmented designs to assess principal strata direct effects. Journal of the Royal Statistical Society B, 73, p. 729-752.
- Bartolucci F. , **Grilli L.** (2011) Modelling partial compliance through copulas in a principal stratification framework. J. American Statistical Association, 106, pp. 469-479.
- Schwartz S., Li F., **Mealli F.** (2011) A Bayesian Semiparametric Approach to Intermediate Variables in Causal Inference, J. American Statistical Association, 106, 1331-13444.
- Arpino B., **Mealli F.**(2011) The specification of the propensity score in multilevel observational studies, Computational Statistics and Data Analysis, 55, 1770-1780
- Grilli L., **Mealli F.** (2008) Nonparametric Bounds on the Causal Effect of University Studies on Job Opportunities Using Principal Stratification. Journal of Educational and Behavioral Statistics, 33, 1, 111-130.
- **Mattei A.**, **Mealli F.** (2007) Application of the Principal Stratification Approach to the Faenza Randomized Experiment on Breast Self-Examination. Biometrics, 63, p. 437-446

References

- Guides from public institutions (European Commission, World Bank)
  - Design and commissioning of counterfactual impact evaluations (2012) *European Commission* http://ec.europa.eu/social/BlobServlet?docId=10805&langId=en
  - Evalsed Sourcebook: Method and techniques (2013, draft) *European Commission* http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf
  - Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2010). *Impact evaluation in practice*. Washington DC: The International Bank for Reconstruction and Development. *World Bank* http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf
- An excellent review on impact evaluation in Italian:
  - Ugo Trivellato: La valutazione degli effetti di politiche pubbliche: paradigma controfattuale e buone pratiche. *Politica Economica* XXVI, n. 1, 2010
- An excellent review about technical issues:
  - G. W. Imbens and J.M. Wooldridge: Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature* 2009, 47:1, 5–86