Alcune considerazioni sulla impact evaluation in campo sanitario

(Some considerations on impact evaluation in healthcare system)

Marta Giachello

marta.giachello@economia.unige.it

Università degli Studi di Genova, Dipartimento di Economia

AIES_XXIII National Conference, Napoli, 27-28 settembre 2018

Rationale

- The last decades have seen an increasing interest in matter of policy evaluation, consequently also impact evaluation studies are increased.
- Impact evaluation studies have been mostly employed for analysing polices concerning unemployment, education, environment and public welfare.
- However the impact evaluation has not been systematically employed to evaluate the effect of health policies.

Outline

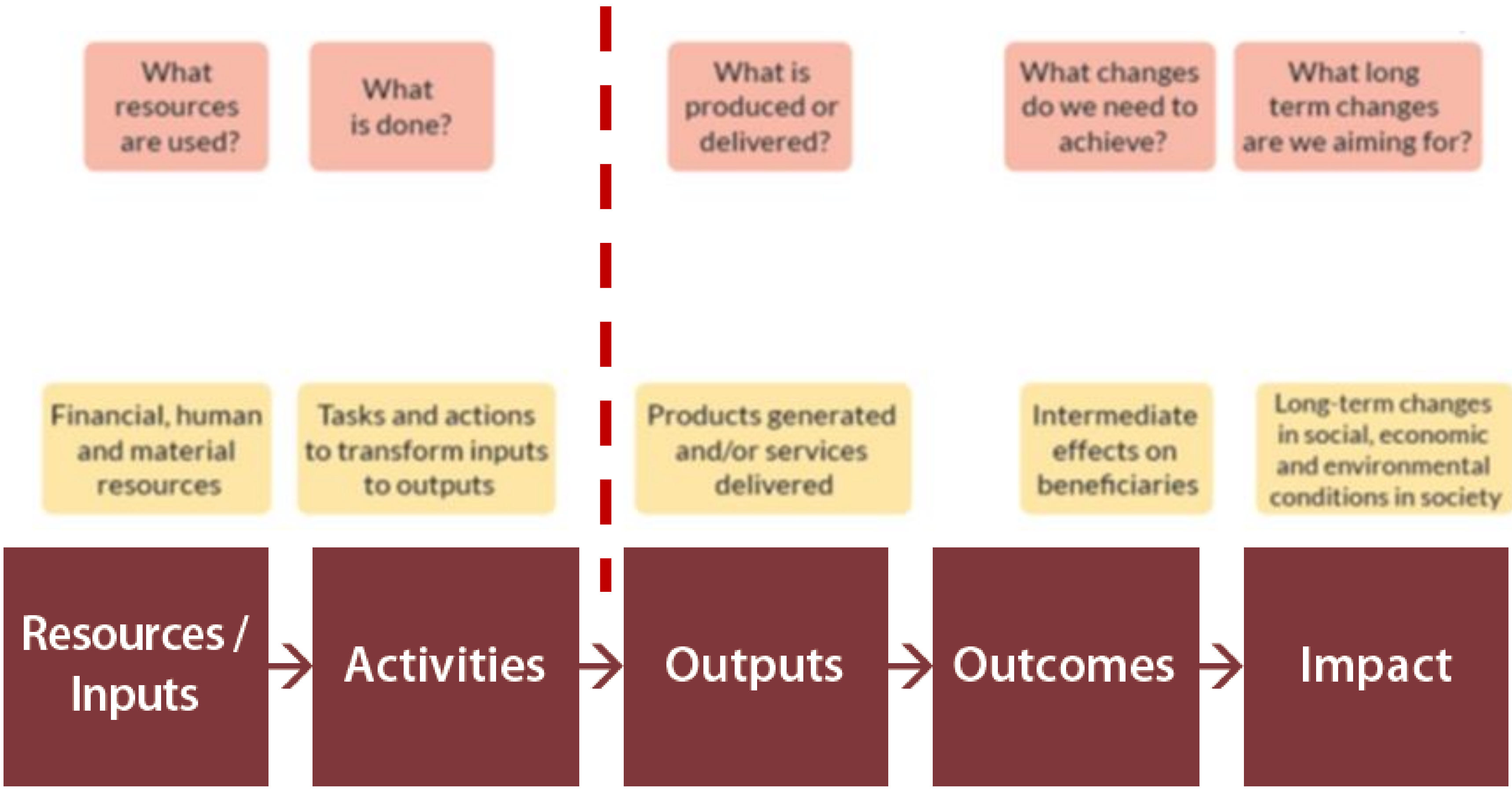
1. What is impact (policy) evaluation?
2. Main *ex-post* and *quantitative* methods
3. Some examples
4. Future case study Liguria

1.

What is impact evaluation?

- The impact evaluation quantifies the impact of the policy or program, with the aim of analysing the **causal relation** between the intervention and the medium-long term effects produced on individuals, families and society in a broader meaning.
    - *What* is the object of evaluation?
    - *When* is it possible to evaluate?
    - *How* is it possible to evaluate?

What?



(Sources: UN-Habitat 2012; ICRC, 2008)

When?

- **Ex ante**: before the intervention. To verify the coherence of the program in order to prefigure scenario.
- **On going**: evaluation *in itinere*. Use to check the development and progress of the policy.
- **Ex post**: to verify the achievement of the objectives and the impact on single individuals and on the whole society.

How?

- **Qualitative method**: it is informative to evaluate the economic context, to understand the people behaviour and the perception of the policy. Usually surveys and interviews are employed. Important for process and formative evaluation.

- **Quantitative method**: based on data collection and statistical and econometric methods. They can answer research questions both on process and impact evaluation

- **Mixed method approach**: it combines the two methods above.

2.

Main *ex-post* and *quantitative* methods

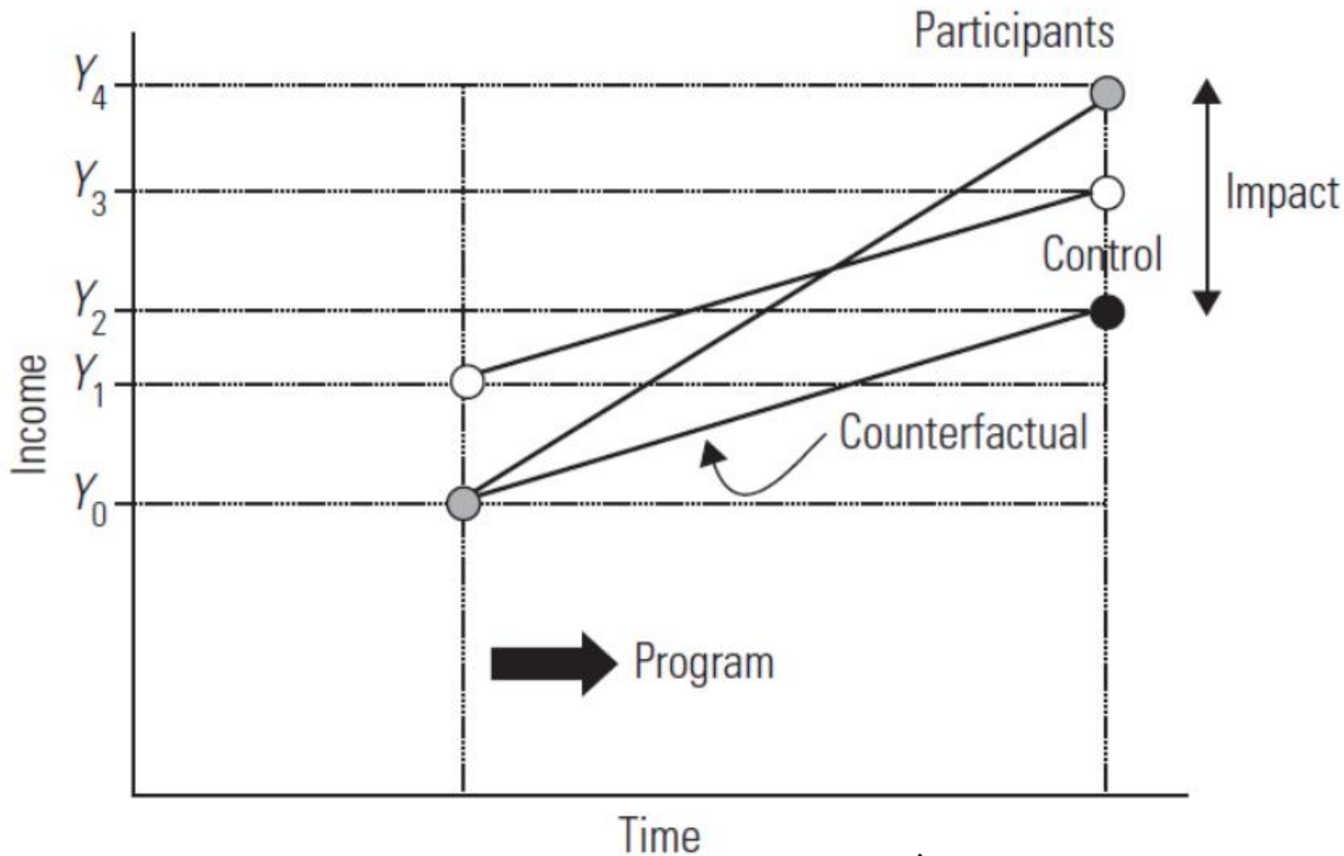
Counterfactual

- The problem of the **unobservability** of the **counterfactual** which represents *what would have happened in the absence of the policy.*

- Also referred in literature as "*the fundamental problem of causal inference*" (Holland, 1986).

- The challenge: to find a good **comparison** (control) **group** which should be the most similar to the **treated group** except for the treatment status.

Impact estimation: with and without comparison group



(Source: Khandker et al,. 2009)

Selection bias

The "*Do hospitals make people healthier?*" (Angrist e Pischke, 2008)

- pay attention at the *naive comparison*

$$potential\ outcome = \begin{cases} Y_{1i} & se\ D_i = 1 \\ Y_{0i} & se\ D_i = 0 \end{cases}$$

$$Ob.Diff. = E[Y_{1i}\,|D_i = 1] - E[Y_{0i}\,|D_i = 0]$$

$$Ob.Diff. = E[Y_{1i}\,|D_i = 1] - E[Y_{0i}\,|D_i = 0] + E[Y_{0i}\,|D_i = 1] - E[Y_{0i}\,|D_i = 1]$$

$$Ob.Diff. = \underbrace{E[Y_{1i}\,|D_i = 1] - E[Y_{0i}\,|D_i = 1]}_{ATE} + \underbrace{E[Y_{0i}\,|D_i = 1] - E[Y_{0i}\,|D_i = 0]}_{Selection\ Bias}$$

Ex post and quantitative methods

- **Non experimental:** they don't employ a comparison group, they just compare before and after intervention the same group
- **Experimental**: random assignment at treated and control group (under eligibility criteria)
- **Quasi-experimental**: they employ the comparison group but there is not the random assignment.
  - Matching methods
  - Double difference
  - Instrumental variable
  - Regression discontinuity
  - Synthetic control group

Random assignment

- Randomized Controlled Trial (RCT) is considered as the *gold standard* of clinical research
- Different type of randomization (symple, block, stratified) and different level
- Solve the problem of Selection bias

$$Ob.Diff. = \underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{ATE} + \cancel{\underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{Selection\ Bias}}$$

- Criticality: ethical issue; not so practice, there are organizzative difficulties in employing it for evaluation of broadly health policy

Matching (1/2)

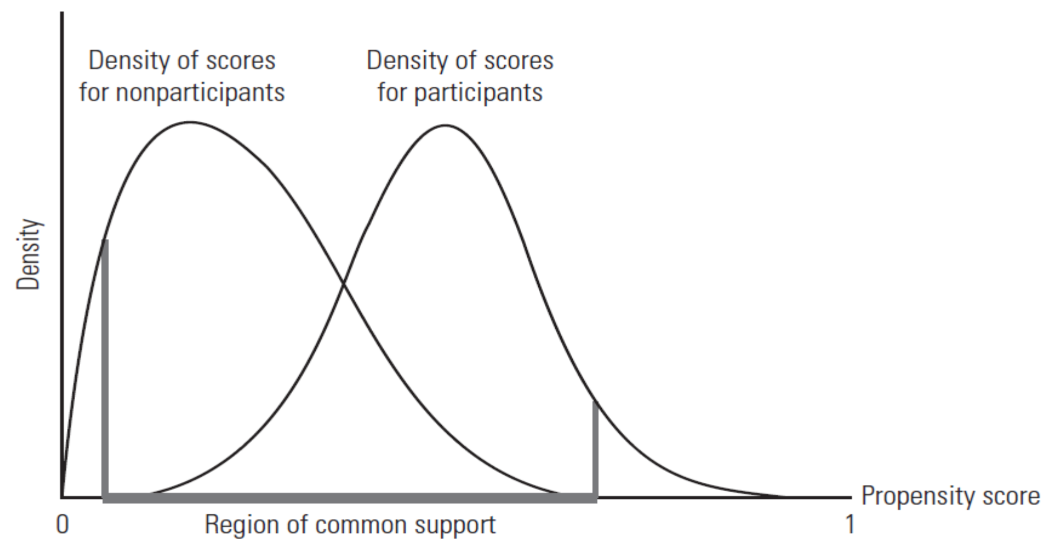
- Second best after randomization: it tries to mimic randomization
- There are many proposals: the most used is Propensity Score Matching (PSM)
- The PSM creates a comparison group based on the participation probability at the treatment given the observable characteristics
- The ATE is measured based on the mean difference in outcomes across comparison and treatment groups.

Matching (2/2)

Validity conditions:

- **Conditional independence** participation at the treatment is entirely based on observable characteristics, also referred to as "*unconfoundedness*" (Rubin, 1990)

- **Common support** at least **overlap condition** ensures that treatment units have a comparison observation in the nearby propensity score distribution, there is overlap between the treated and untreated subsamples



(source: Khandker et al., 2009)

Difference in difference (DD)

- Especially used for panel data
- One of the most employed methods: "*Designing Difference in Difference Studies: Best Practices for Public Health Policy Research*", (Wing et al., 2018)
- Condition: units should be observed at least in two time periods
- It allows **heterogeneity** unobservable in participants but assumes that it is **time invariant**
- It is possible to estimate the average impact of the treatment using DD as follows, a sort of before-after comparison.

$$Average\ Impact_{DD} = E(Y_{1T} - Y_{0T} | T_1 = 1) - E(Y_{1C} - Y_{0C} | T_1 = 0)$$

Instrumental Variable (IV)

- Very powerful but it's a hard challenge to find a good instrument,
- The IV must be highly correlated with the program and its localisation but it must have no correlation with unobservable variables which influence the outcome and there is no omitted variables.
- The aim is to delete the correlation between the regressor (T) and the error term (ε). In order to do that the IV (Z) should satisfy:
  - $Cov\ (Z, T) \neq 0$, a correlation between Z and T
  - $Cov\ (Z, \varepsilon) = 0$, no correlation between Z and ε

Regression Discontinuity (RD) (1/2)

- This method is focused on the discontinuity and delays of the implementation of the program

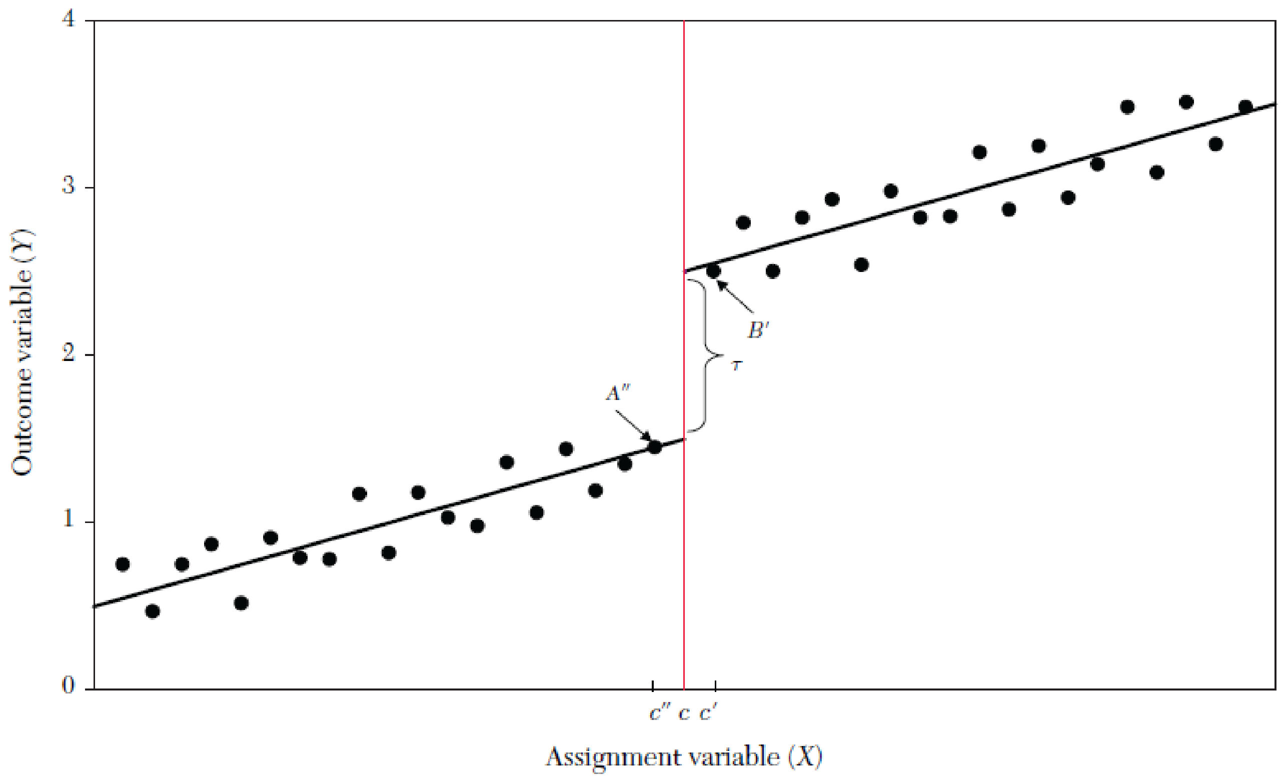
- The assignment at the treatment is determined if the "*running variable*“ (X_i) overcomes a cutoff ($\bar{X}$)

- *Regression discontinuity designs are underutilized in medicine, epidemiology, and public health* (Moscoe et al., 2015).

- This method works good if people are not able to precisely manipulate their assignment to the treatment, this implicate that the variation in the treatment status in the neighbourhood of the cutoff results as a randomization

Regression Discontinuity (RD) (2/2)

- **Example of Linear Regression Discontinuity Desing**



(source: Lee and Lemieux, 2010)

Synthetic Control Method

- Considered an extension of difference in difference (DD)

- Useful when in the context:

  - the difference between the treated and control group aren't time invariant, then it is really difficult to find a suitable control group

  - small number of units cross sectional but they have a long longitudinal dimension

- A "***synthetic***" comparison group is created by using a weighted average of the units chosen from the donor pool. (See Abadie et al., 2015).

3.

Some examples

Randomized Controlled Trial (RCT)

Baicker K., Taubman S. L., Allen H. L., Bernstein M., Gruber J. H., Newhouse J. P…. and Finkelstein A. N. (2013). The Oregon experiment—effects of Medicaid on clinical outcomes. *New England Journal of Medicine*, *368* (18), 1713-1722.

- **The effects on clinical outcomes of Medicaid expansion in Oregon 2008 based on lottery drawings (under eligibility criteria)**

Oregon Health Plan Standard closed to new enrolment in 2004, then in 2008 it is opened a new waiting list and 8 random lottery drawings are conducted.
The use RTC approach. People eligible but not randomly selected construct the control group.
Medicaid generated no significant improvement in measured physical health outcomes (in the first 2 year) but it did increase use of health care services, lower rates of depression and reduce out of pocket medical expenditures and raise rates of diabetes detection and management. (Baicker et al., 2013)

Propensity Score Matching (PSM)

Johar M. (2009). The impact of the Indonesian health card program: a matching estimator approach. Journal of health economics, 28(1), 35-53.

➢ **The Impact of the Indonesian Health Card Program (HCP)**

The HCP of 1994 was an effort to improve the nation health conditions. The program targets poor households and provides full price subsidy to medical expenses at public health facilities for all members of the household. Health Cards are distributed by the village heads based on households welfare criteria.
Since there is the presence of longitudinal data, matching technique can be combined with DD estimator, that generates a powerful estimator of ATT. Propensity scores are estimated using a logit model.
The presence of HCP may allow younger household members to receive curative treatments and females in the households to enroll in a contraception program, but it has no significant positive effects on the other household members. Program redesign or redirection of resources may yield a larger impact. (Johar, 2009).

Propensity Score Matching (PSM)

Wagstaff A. and Yub S. (2005). Do Health Sector Reforms Have Their Intended Impacts? The World Bank's Health VIII Project in Gansu Province, China.

- **Impacts of the Health Sector Reforms (China, only Gansu province)**

The Health VIII was a project supported by the World Bank. It began in October 1998. It was a system wide project. Its goals were to raise the quality, affordability and utilization of the health services (especially poor) and thereby to improve health outcomes and promote financial protection by reducing the direct and indirect costs associated with illness.
They estimated the impact of Health VIII on individual, household, health facility and village outcomes by combining DD and PSM.
Project achieved the intended dampening effect on out-of pocket expenses and reduced incidence of catastrophic health spending. The project had a little impact in the use services. (Wagstaff and Yub, 2005).

Difference-in-difference (DD)

Filippini M., Ortiz L. G. and Masiero, G. (2013). Assessing the impact of national antibiotic campaigns in Europe. The European journal of health economics, 14(4), 587-599.

- **Assessing the impact of public policies on antibiotic use in Europe**

The overuse of antibiotics is the main force driving the increase of bacterial resistance (major threat to public health). The aim of the study was to assess the impact of public education campaigns on antibiotic use in Europe.
To measure the effect of public campaigns on antibiotic consumption they used a DD approach.
The results provide some evidence that public campaigns represents an effective strategy to reduce the use of outpatient antibiotics. Countries that adopt public campaigns succeed in terms of reducing their levels of antibiotic use over time. (Filippini et al., 2013).

Difference-in-difference (DD)

Friebel R., Hauck K. and Aylin P. (2018). Centralisation of acute stroke services in London: Impact evaluation using two treatment groups. Health economics, 27(4), 722-732.

- **Impact evaluation of centralisation of acute stroke services in London**

In 2010 London and Greater Manchester performed a reorganisation of stroke care aimed at streamlining stroke services through service centralisation. 8 Trusts were converted into Hyper Acute Stroke Units (HASUs) in London with the objective of improving stroke process and outcomes (and savings costs in the long run).
They applied a DD analysis using 2 treatment group: HASU and London non-HASUs.
Policy objective was partly achieved by increasing rates of thrombolysis treatment for patient admitted to HASU. 15% of patients still receive acute stroke care in London non-HASUs. Their results could indicate differences in quality of care received at London HASU and Non-HASU Trusts. Non-HASU patients were less likely to receive specific processes, 24hr brain scan and thrombolysis treatment, and they were more likely to die within 7 and 30 days. (Friebel et al., 2018).

Instrumental Variable (IV)

Malkin J. D., Broder M. S. and Keeler E. (2000). Do longer postpartum stays reduce newborn readmissions? Analysis using instrumental variables. Health services research, 35(5 Pt 2), 1071.

- **Do longer postpartum stays reduce newborn readmissions?**

Concerns about potential adverse effect (newborn readmission) of early discharge of newborn have led the federal government and most state legislatures to pass laws mandating minimum 48-hours or 96-hours hospital stays following respectively natural deliveries or cesarean sections. The aim was to evaluate the effects of length of stay on newborn readmissions.
The study employs hour of birth and method of delivery as IV (both are strong predictors of length of stay) to account for unobserved heterogeneity . Hour affects on an eventual extra night, while delivery affects mother's time of recovery.
The results showed that a 12-hours increase in length of stay is associated with a reduction in the newborn readmission rate of 0,6% point, therefore an increase in the length of postpartum hospital stays may result in a decline in newborn readmission and this may yield meaningful benefit (social and economic). (Malkin et al., 2000).

Regression Discontinuity (RD)

Bernal N., Carpio M. A. and Klein T. J. (2017). The effects of access to health insurance: evidence from a regression discontinuity design in Peru. Journal of Public Economics, 154, 122-136.

- **The effects of access to health insurance in Perù**

In developing countries a large number of individuals in not covered by health insurance. It may be a cause of concern: health insurance doesn't only protect against catastrophically high health expenditures, it also encourages people to see doctor and to be more aware of illness. In 2001 was created the Peruvian Social Health Insurance called "Seguro Integral de Salud" (SIS) for individuals outside the formal labour market. They are eligible if a welfare index called Household Targeting Index (IFH) is below a specific threshold. The aim was to estimate the impact of SIS coverage on health care utilization and out-of pocket expenditures.
It is used ad Sharp RD, treatment assignment is a deterministic function of the running variable (IFH). The effects of health insurance coverage are positive for forms of care that are of a more general nature and can be provided by MINSA health care centers at relatively low cost. The coverage has positive effects on the level and the variability of out of pocket spending and that is partly driven by supply limitations. It leads to increase awareness about health problems and may generate form of supplier-induced demand. (Bernal et al., 2017).

Regression Discontinuity (RD)

Palmer M., Mitra S., Mont D. and Groce, N. (2015). The impact of health insurance for children under age 6 in Vietnam: a regression discontinuity approach. Social Science & Medicine, 145, 217-226.

➢ **The impact of health insurance for children under age 6 in Vietnam**

Accessing health services at an early age is important to future health and life outcomes. A policy in 2005 in Vietnam provided health insurance coverage to children under age of six.
The study uses a Fuzzy RD (the treatment assignment is a probabilistic function of the running variable). The basic idea is that children on either side of the cutoff age are similar and almost randomly assigned to insurance coverage.
The policy has been successfully in improving access to outpatient and inpatient care. The study showed no evidence of a substitution from private to public care under insurance and no significant impact on health expenditures and thus on providing financial protection. Results suggest that adopting public health insurance programs fro children under 6 may be an important instrument to improve service utilization in low and middle income country context. (Palmer et al., 2015).

Synthetic Control Group

Lépine A., Lagarde M. and Le Nestour A. (2015). Free primary care in Zambia: an impact evaluation using a pooled synthetic control method. Available at SSRN 2520345.

- **Impact of free primary care in Zambia**

In Zambia in 2006 user charges were removed in only 54 of the 72 districts in the country.
The aim was to evaluate the impact of removing user fee on health seeking behaviours and out of pocket expenditure
They construct a synthetic control for each districts by taking a weighted average of the available control units (a higher weight is given to control units that are more similar to the treated unit).
They showed no evidence that user fees removal policy in rural Zambia changed health seeking behaviours, even among the poorest . The policy mostly benefited the richer groups.(Lépine et al., 2015).

Synthetic Control Group

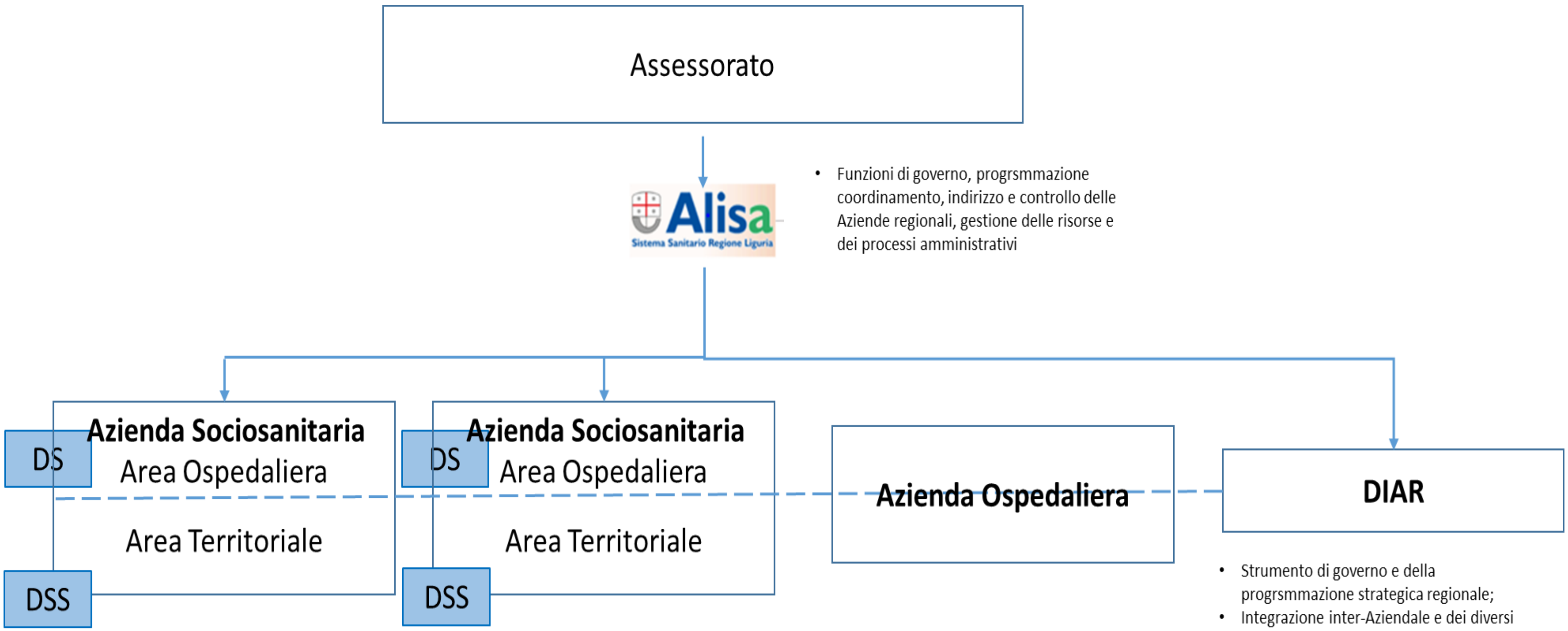
Bruhn C. A., Hetterich S., Schuck-Paim C., Kürüm E., Taylor R. J., Lustig R., … and Weinberger D. M. (2017). Estimating the population-level impact of vaccines using synthetic controls. Proceedings of the National Academy of Sciences, 114(7), 1524-1529.

- **Impact of the introduction of pneumococcal conjugate vaccines (PCVs) on hospitalizations for all-cause pneumonia**

It is critical to measure the impact of the introduction of a new public health policy (such as vaccine). The Pneumococcal Conjugate Vaccines (PCVs) is targeted on a subset of 3 serotypes of pneumococcus. They sought to estimate changes in hospitalizations for all-cause pneumonia associated with the introduction of PCVs.
They used data from 5 countries (Brazil, Chile, Ecuador, Mexico and United States).
They employ a Synthetic controls approach which can be adapted to detect and adjust for unmeasured bias and confounding in the evaluation. For all five of the countries, the introduction of PCVs was associated with substantial and significant declines in hospitalizations for all-cause pneumonia among infants; whereas estimates for young and middle ages adults did not report a decline in hospitalizations for pneumonia after vaccine introduction in children in any of the five countries. (Bruhn et al., 2017).

Future case study
Liguria

The reform (1/3)



- Regista del rafforzamento dell'attività assistenziale territoriale e delle risposte ai bisogni sociosanitari e di protezione sociale.

- Strumento di governo e della progrsmmazione strategica regionale;
- Integrazione inter-Aziendale e dei diversi livelli di assistenza delle attività sanitarie e socio sanitarie;
- Centro di coordinamento tecnico e professionale

The reform (2/3)

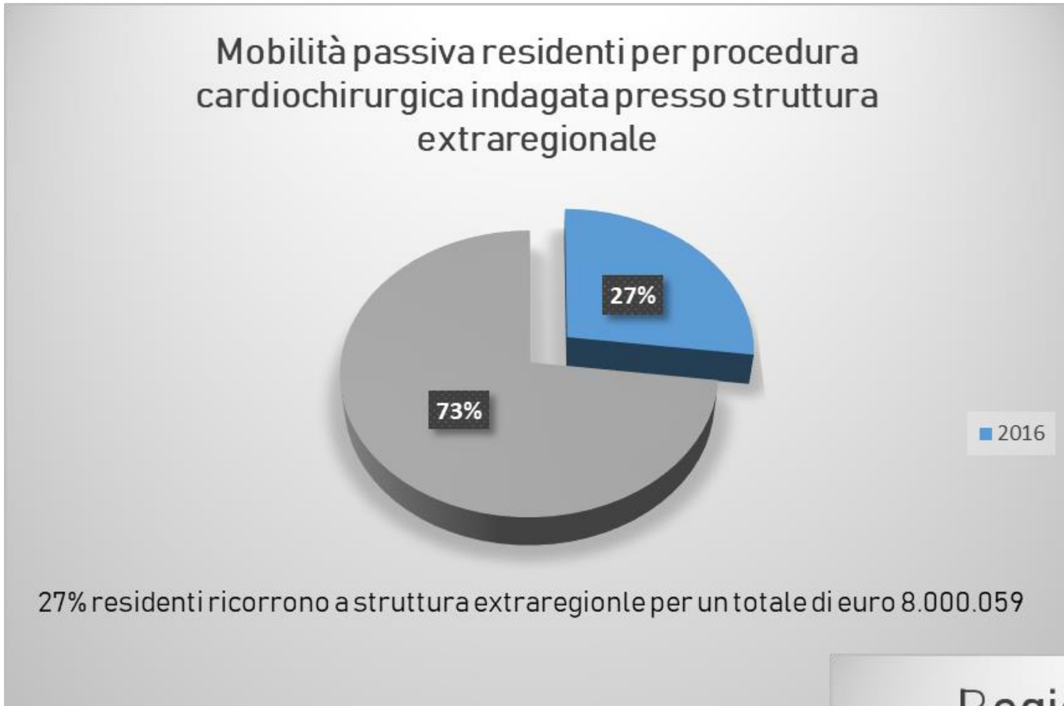
- Evolution towards a network model
- Institution of :
  - Alisa- function of governance
  - Dipartimenti Interaziendali Regionali (DIAR)- coordination and integration finality
- Currently 5 DIARs are activated:
  - Accident & emergency
  - Cardiovascular
  - Neuroscience
  - Onco hematological
  - Surgical
- The awareness of stakeholders: formative sessions
- General and specific obiectives of DIAR

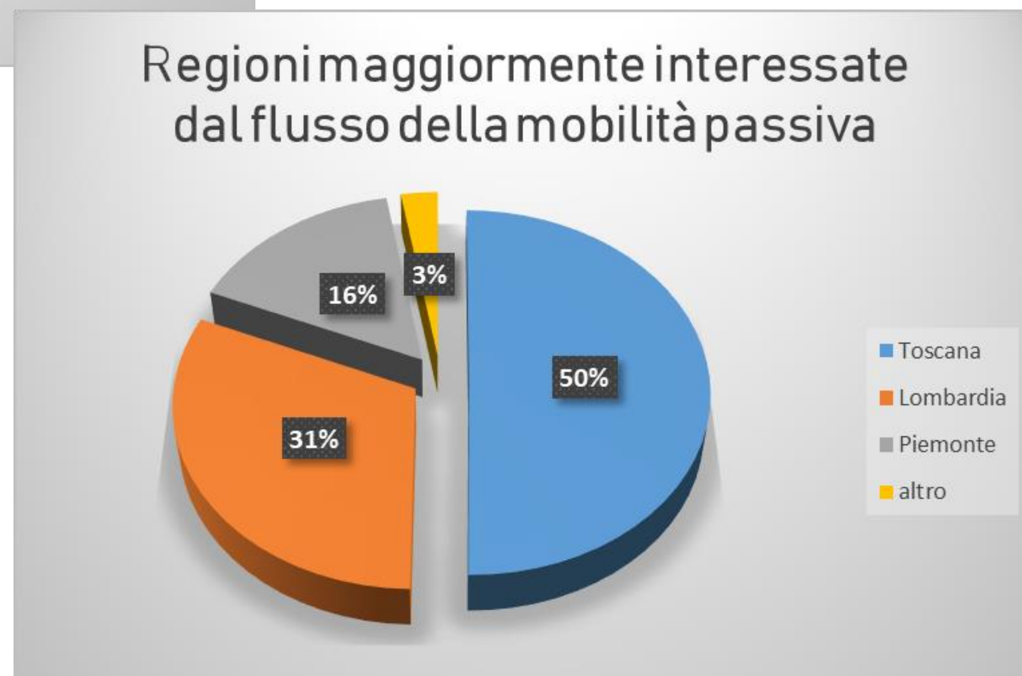
The reform (3/3)

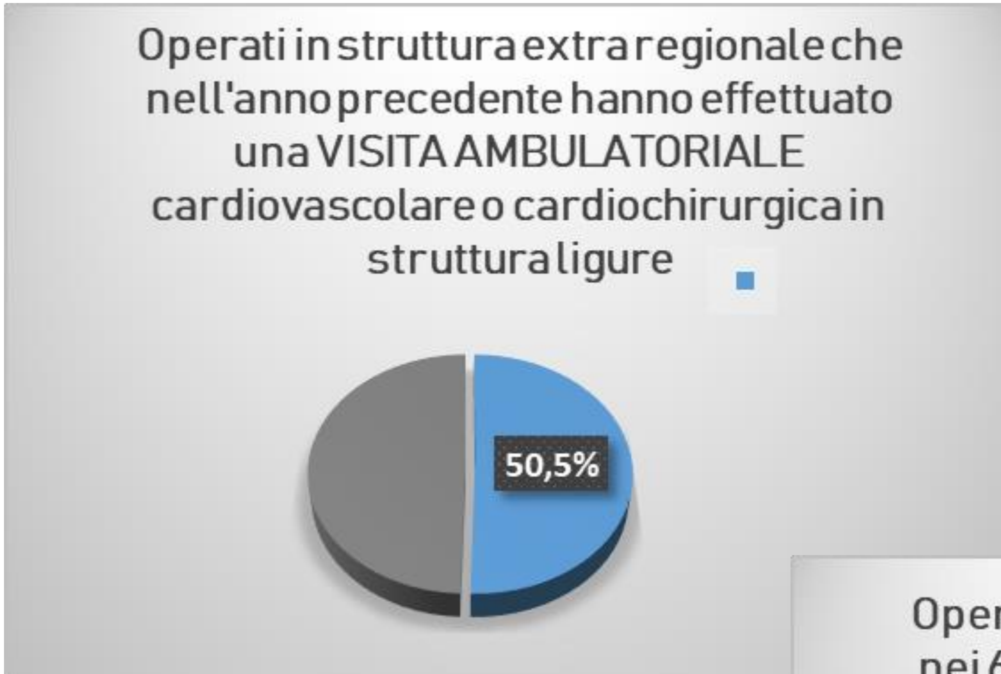
- General purposes which are established for the 5 active DIARs:
  - management of the passive mobility
  - appropriateness of performances and supply setting
  - optimizations of clinical and care pathways
  - clinical and organizational outcomes measures

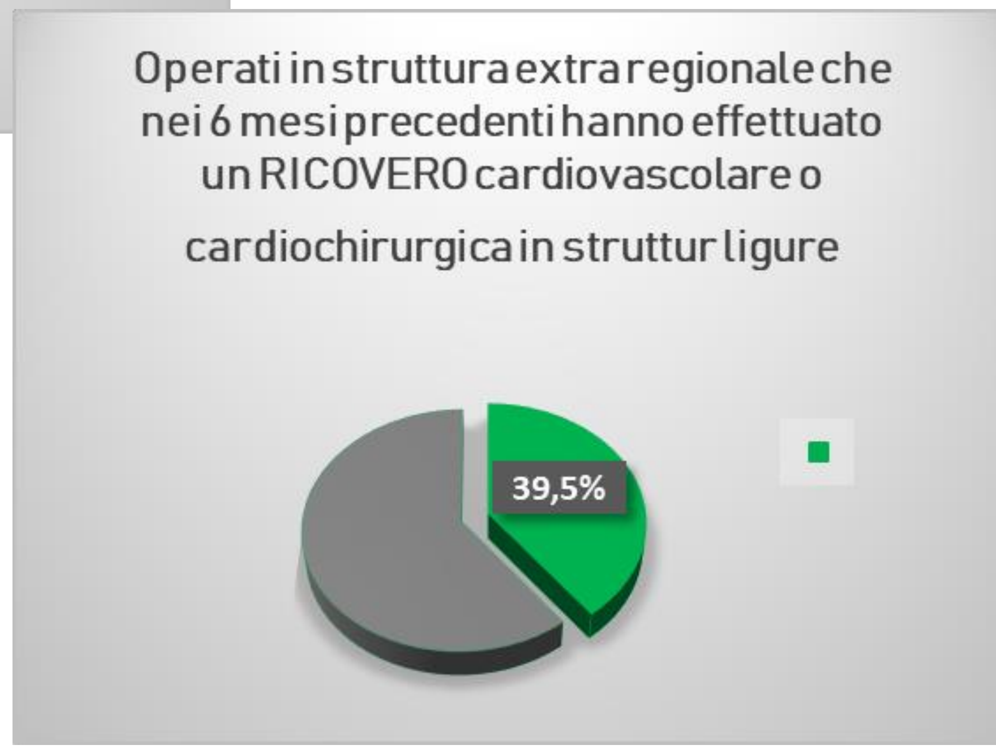
[Del. A.Li.Sa 6/2018 del 15.01.2018 “Indirizzi operativi per le attività sanitarie e sociosanitarie per l’anno 2018”]

- Focus: decrease passive mobility in cardiovascular DIAR
  - cardiovascular surgery
  - data from SDO
  - the clinical and care "path" the year before the surgery


Mobilità passiva residenti per procedura cardiochirurgica indagata presso struttura extraregionale
27%
73%
2016
27% residenti ricorrono a struttura extraregionle per un totale di euro 8.000.059


Regioni maggiormente interessate dal flusso della mobilità passiva
3%
16%
50%
31%
Toscana
Lombardia
Piemonte
altro


Operati in struttura extra regionale che nell'anno precedente hanno effettuato una VISITA AMBULATORIALE cardiovascolare o cardiochirurgica in struttura ligure
50,5%


Operati in struttura extra regionale che nei 6 mesi precedenti hanno effettuato un RICOVERO cardiovascolare o
cardiochirurgica in struttur ligure
39,5%

The evaluation

- The reform had considered ex ante a system of monitoring and evaluation
- Too early for impact: an outcome evaluation with reference to a specific case involving the surgical cardiovascular area
- The next aim: an impact evaluation analysis to measure the medium-long term broader effects of the reform on the society

Workshop "Reti sanitarie tra volontarismo e prescrizione. Ricerche scientifiche ed esperienze a confronto", Firenze 28 -29 settembre 2018 (organizzato da Regione Toscana e Agenzia regionale sanità della Toscana)

Filippo Ansaldi, Walter Locatelli - **Regione Liguria** Marta Giachello, Cinzia Panero, Angela Testi - **Università di Genova**

- ***Governance delle reti sanitarie: i Dipartimenti Interaziendali Regionali (DIAR) nel nuovo sistema sanitario ligure***

Conclusion

- An open research field

- Impact evaluation should be established ex ante as a final stage of a policy or a program

Some References (1/3)

- Abadie A., Diamond A. and Hainmueller J. (2015). Comparative politics and the synthetic control method. American Journal of Political Science, 59(2), 495-510.
- Angrist J. D. and Pischke J. S. (2008). Mostly harmless econometrics: an empiricist's companion. Princeton: Princeton university.
- Angrist J.D. and Pischke J.S. (2014). Mastering' metrics: The path from cause to effect. Princeton University Press.
- Baker J.L. (2000). Evaluating the impact of development projects on poverty: A handbook for practitioners. World Bank Publications.
- Holland, P. W. (1986). Statistics and causal inference. Journal of the American statistical Association, Vol. 81, N°396, pp. 945-960.
- Khandker S. R., Koolwal G. B. and Samad H. A. (2009). Handbook on impact evaluation: quantitative methods and practices. World Bank Publications.
- Lee D. S. and Lemieux T. (2010). Regression discontinuity designs in economics. Journal of economic literature, 48(2), 281-355.
- OECD (2010). Glossary of Key Terms in Evaluations and Results Based Management. OECD, 2002, re-printed in 2010.
- Palumbo M. (2001). Il processo di valutazione: decidere, programmare, valutare (Vol. 3). Milano, FrancoAngeli.
- Purdon S., Lessof C., Woodfield K. and Bryson C. (2001). Research methods for policy evaluation. DWP Social Research Division Working Paper, 2.
- Thistlethwaite D. L. and Campbell D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. Journal of Educational psychology, 51(6), 309.
- Basu S., Meghani A. and Siddiqi A. (2017). Evaluating the health impact of large-scale public policy changes: classical and novel approaches. Annual review of public health, 38, 351-370.

Some References (2/3)

Review refereces

- Baicker K., Taubman S. L., Allen H. L., Bernstein M., Gruber J. H., Newhouse J. P., … and Finkelstein A. N. (2013). The Oregon experiment—effects of Medicaid on clinical outcomes. *New England Journal of Medicine*, *368*(18), 1713-1722.
- Barletta V., Profili F., Gini R., Grilli L., Rampichini C., Matarrese D. and Francesconi, P. (2016). Impact of Chronic Care Model on diabetes care in Tuscany: a controlled before-after study. The European Journal of Public Health, 27(1), 8-13.
- Bernal N., Carpio M. A. and Klein T. J. (2017). The effects of access to health insurance: evidence from a regression discontinuity design in Peru. Journal of Public Economics, 154, 122-136.
- Bruhn C. A., Hetterich S., Schuck-Paim C., Kürüm E., Taylor R. J., Lustig R., … and Weinberger D. M. (2017). Estimating the population-level impact of vaccines using synthetic controls. Proceedings of the National Academy of Sciences, 114(7), 1524-1529.
- Filippini M., Ortiz L. G. and Masiero, G. (2013). Assessing the impact of national antibiotic campaigns in Europe. The European journal of health economics, 14(4), 587-599.
- Friebel R., Hauck K. and Aylin P. (2018). Centralisation of acute stroke services in London: Impact evaluation using two treatment groups. Health economics, 27(4), 722-732.
- Johar M. (2009). The impact of the Indonesian health card program: a matching estimator approach. Journal of health economics, 28(1), 35-53.

Some References (3/3)

- Lépine A., Lagarde M. and Le Nestour A. (2015). Free primary care in Zambia: an impact evaluation using a pooled synthetic control method. Available at SSRN 2520345.
- Lu M. (1999). The productivity of mental health care: an instrumental variable approach. The Journal of Mental Health Policy and Economics, 2(2), 59-71.
- Malkin J. D., Broder M. S. and Keeler E. (2000). Do longer postpartum stays reduce newborn readmissions? Analysis using instrumental variables. Health services research, 35(5 Pt 2), 1071.
- Moscoe E., Bor J. and Bärnighausen T. (2015). Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. Journal of clinical epidemiology, 68(2), 132-143.
- Mukhopahyay T. and Behl A. S. (2010) Impact Evaluation of Health Micro Insurance Through Randomized Controlled Trials.
- Palmer M., Mitra S., Mont D. and Groce, N. (2015). The impact of health insurance for children under age 6 in Vietnam: a regression discontinuity approach. Social Science & Medicine, 145, 217-226.
- Wagstaff A. and Yub S. (2005). Do Health Sector Reforms Have Their Intended Impacts? The World Bank's Health VIII Project in Gansu Province, China.
- Wing C., Simon K. and Bello-Gomez, R. A. (2018). Designing Difference in Difference Studies: Best Practices for Public Health Policy Research. Annual review of public health, 39, 453-469.

Thank you for your attention