Fundamental Concepts of Quantitative Impact Assessment: Module 3

Aneta Bonikowska, PhD

Social Analysis and Modelling Division

March 19, 2024

Statistics Canada Statistique Canada

Canada

Module 3

Randomized controlled experiments: The gold standard in QIA

Module 3 contents

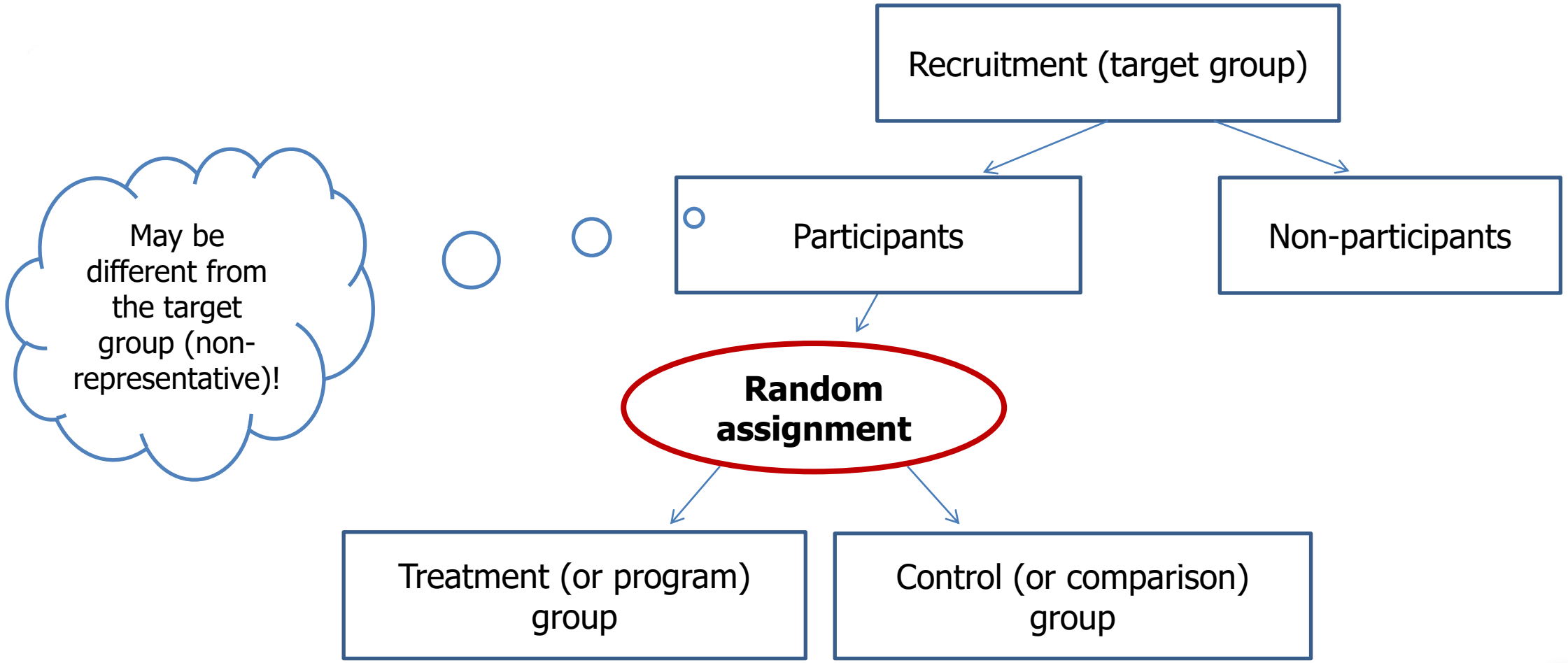
- Two types of experiments
- A basic experimental design
- Why experiment?
- Limitations of experiments
- Main points to remember

Two Types of Experiments

- Laboratory experiments
  - Policy is tested under laboratory conditions
  - (Very) controlled environment
  - Criticism:
    - may be difficult to generalize some findings to real life
    - In social science context, individual characteristics such as ability, effort, etc. matter for outcomes, but cannot be easily measured and therefore controlled for

- Field experiments
  - Policy is implemented in real life setting (in the field) among a small group, who may or may not know that they are part of an experiment
  - Also called a ‘demonstration project’
  - We will focus on this

A basic experimental design



Why Experiment?

- We would like to run a group of people through life twice: once going through a program, once without – but we can't, instead…
- Random assignment *essentially* solves the fundamental problem in program evaluation:
  - Treatment and control group are the same (on average) at the beginning of the experiment (luck assigned them, not personal choice)
  - Baseline survey (prior to experiment) needed to confirm similarity of characteristics between treatment and control group
- **Impact = Treatment group outcome – Control group outcome**

Why Experiment?

- Random assignment removes the choice in selecting treatment (policies or programs)
- As a result, control group serves as a good counterfactual to treatment group (what would have happened had the treatment group not been offered the treatment)
- Random assignment is often referred to as the ‘gold standard’ in QIA

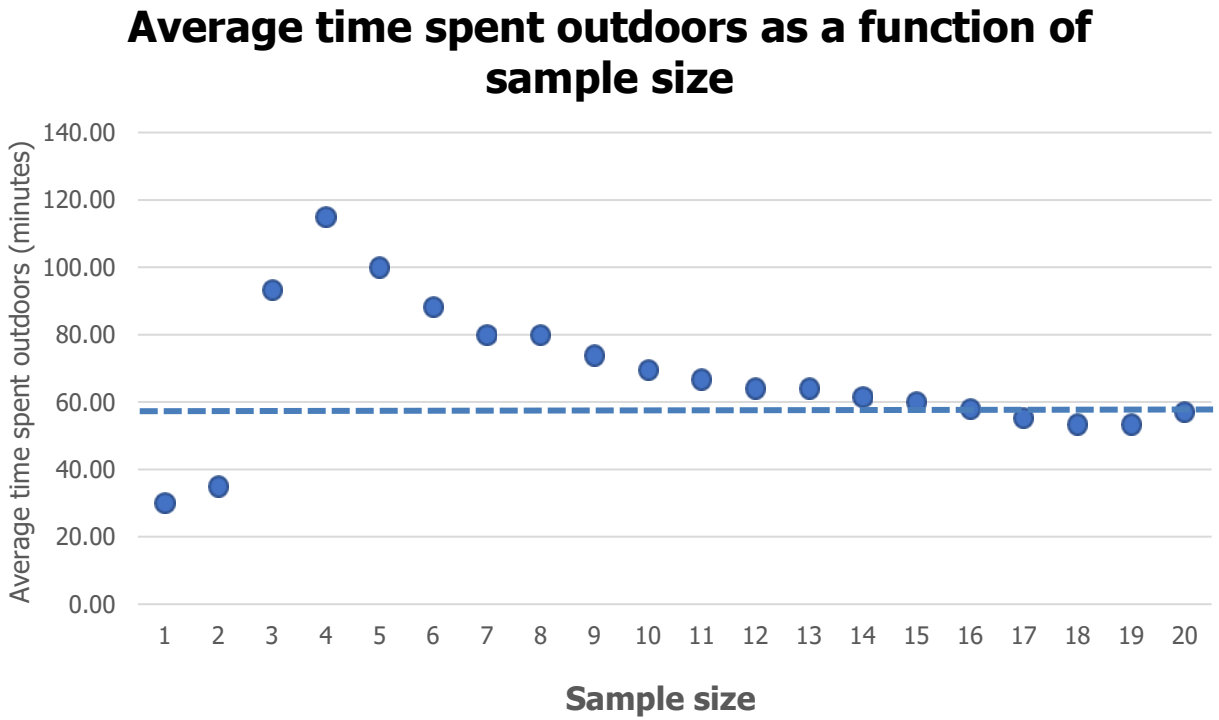
All that glitters… Limitations of experiments

- Sample size
- Internal/external validity
- Intention-to-treat
- Attrition within control group
- Dropout bias
- Substitution bias
- Spill-over effects
- Ethical concerns
- Costs

Sample size

- Random assignment is expected to create identical treatment and control groups
- But what if the sample size is small, as it is in some experiments?
- With small samples, anything can happen (i.e. treatment and control group may be different)
- Solution is to take a larger sample

Sample size



- Small samples produce unreliable estimates, regardless of population size
- Can match treatment and control groups according to certain characteristics; e.g. ensure same sex, age, and education breakdowns in both groups (through stratification)

Internal/external validity

- **Internal validity**: For the participants who agreed to the experiment, random assignment will generate valid results (as long as the sample is large enough and the experiment was well designed and implemented)

- **External validity**: The results also apply to the target group or to other groups

Internal/external validity

- Field (social) experiments have been described as internally valid, but not necessarily externally valid
  - i.e. social experiments are valid for those who accepted (the opportunity to receive) treatment (but may not have taken it)

- To provide evidence of external validity:
  - Describe participants well (compared to target group or other groups)
  - Sub-group analysis (are results the same for each type of program participant?)

Intention-to-treat

- We can't force the treatment group to take its 'medicine'
- By comparing treatment and control group outcomes, we estimate the 'intention-to-treat' impact (or the impact of the offer of treatment)
- We can't estimate the impact of taking the treatment (*treatment-on-the-treated* effect)

Intention-to-treat

- **Warning**: tempting to just look at those in the treatment group who actually took treatment
  - Taking the treatment was not randomly assigned
  - Only the offer of the treatment was randomly assigned, so we can only compare treatment and control groups

Is intention-to-treat so bad?

- In the real world, programs usually can only be *offered* to individuals
- Some will take up the treatment, others won't
- Just like an experiment!

Attrition within control group

- Experiments usually have follow-up surveys
- In any longitudinal survey, ‘attrition’ is an issue
- It is problematic with experiments if the attrition rate is different for the treatment and control groups:
  - Control group may have fewer reasons to keep participating (they didn’t receive the treatment)
- Solution is to spend a lot on respondent relations and follow-up calls, or (if possible) use administrative data (e.g. income tax files) to measure outcomes

Dropout bias

- Treatment group may stop experimental treatment, possibly in favour of non-experimental treatment
  - E.g. felt job training program wasn't working, so sought help elsewhere
- Dropout bias reduces experimental treatment, but intention-to-treat can still be estimated
- In short, not a problem since they would do this in real life
  - We estimate the effect of the offer of the treatment

Substitution bias

- Treatment group may reduce non-experimental treatment take-up
  - E.g. Job training was working so well that they stopped another course they were already taking
  - Not a problem (they would do this is real life)

- Control group may seek other forms of treatment (reaction to inequity cause by randomization – “I’ll show you”)
  - This would not happen in real life (unless there is randomization)
  - The estimated impact is biased towards zero
  - Important to monitor (follow-up surveys or administrative data)

Spill-over effects

- A treatment may spill-over to the control group if the two are in close contact (e.g. an information session offered to *some* students in a school)
  - Not likely in real life – similar groups in close contact usually both either qualify or not for a government program
- Biases impact towards zero
- Solution is to minimize contact between treatment and control groups

Ethical concerns

- Experiments can be seen as denying program services to the control group. However:
  - Programs are not always beneficial. *A priori*, benefits are unknown, hence the experiment
  - Pareto improvement argument: no one is worse off, some are (potentially) better off
  - Some real-life programs limit enrolment e.g. Quebec public daycares, BC French immersion programs (is that more fair than random assignment? Probably not, so may as well use the best method for evaluation purposes)
  - In demonstration projects, informed consent is provided by program participants (they agree to the experiment, and understand their chances of being offered the treatment or not)
  - Cost considerations from taxpayers' perspective: potential for costly bill for ineffective programs that weren't properly tested

Costs

- Social experiments have been criticized for being too costly
- Although costs are higher than in non-experimental studies, it is often impossible to test a new policy until years later, when data from surveys and administrative sources are available
- Only by performing a pilot test can we know if the policy works
  - if so, may as well do it scientifically (random assignment)
- Not assessing the impact of programs may be costlier in the end than running experiments
  - Pilot projects are much smaller than full program roll-out
  - If program impact is low or zero (or negative, i.e. detrimental), a pilot project could save far more than the cost of the experiment

Example: Making work pay better than social assistance

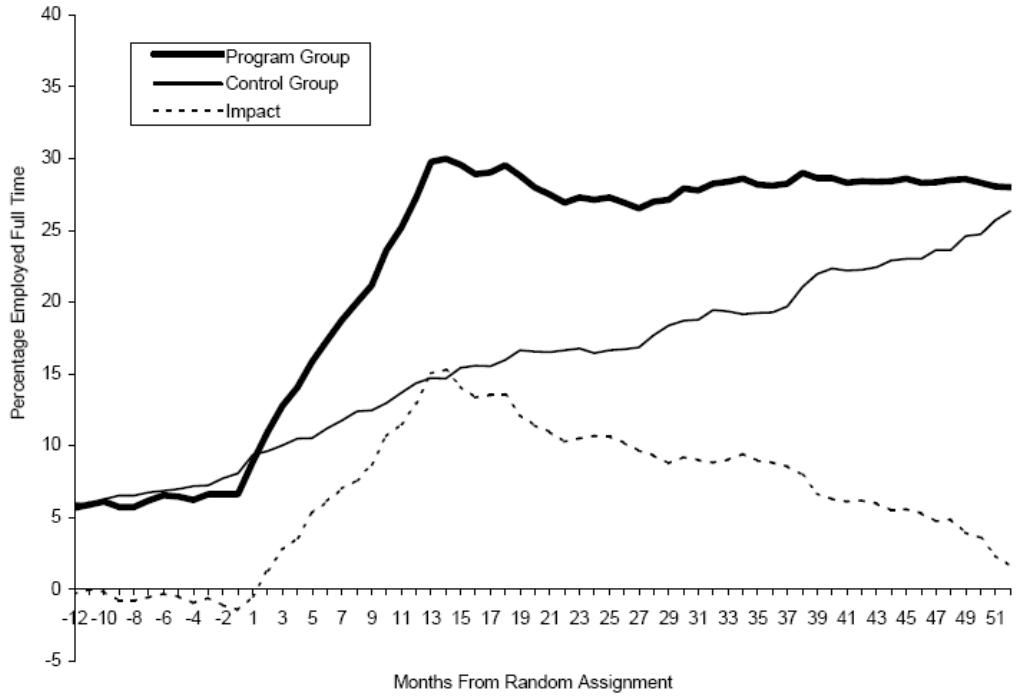
- In 1980s /1990s, social assistance (SA) paid about as much as minimum wage
- Many SA recipients could not expect to earn more than the minimum wage if they left SA → so why look for a job?
- The Self-Sufficiency Project (SSP) made work pay better than SA, in hopes of making people less dependent on SA
- Implemented by Social Research and Demonstration Corporation (SRDC)
- Lone-parent SA recipients (mainly single mothers) could be eligible to receive an earnings supplement to make work more attractive
- Recipients study: participants must have been on SA for at least 1 year (long-term SA recipients)

Example: Making work pay better than social assistance

- Participants were randomly assigned to either the treatment or the control group
- The treatment group was offered an earnings supplement if they found a job with at least 30 hours per week within one year
- Supplement paid out for up to 3 years, only in months where the participant worked full-time and did not collect SA
- The earnings supplement essentially doubled pre-tax income for minimum wage earners
- Implemented in NB and BC in the 1990s, with several follow-up interviews

Example: Making work pay better than social assistance

Figure ES.1: Percentage Employed Full Time, by Months From Random Assignment



Sources: Calculations from baseline survey data and 18-month, 36-month, and 54-month follow-up survey data.
Note: "Employed full time" is defined as working 30 hours or more in at least one week during the month.

1. Randomization worked - treatment and control groups had very similar employment rates before experiment started (other characteristics, too)
2. During the program period, employment higher among treatment group than control group
3. About 1/3 of the treatment group took up the supplement
4. % employed increased steadily with time among control group, even though they received no subsidy
5. Outcomes of treatment and control group converged soon after program ended

Source: Michalopoulos et al. 2002. Making Work Pay: Final Report on the Self-Sufficiency Project for Long-Term Welfare Recipients. Ottawa: Social Research and Demonstration Corporation.

Take-away points

- In QIA, the impact of a policy or program is defined as the difference between the actual outcome (with program) and a counterfactual outcome (the outcome we could expect without the program)
- Random assignment reduces the differences between treatment and control groups arising from choice, self-selection; in that sense, it provides the best control group and counterfactual outcome
- Randomized experiments are not without problems
- In most QIA contexts, random assignment is not present, and constructing a suitable control group and counterfactual outcome requires more thought – luckily, there are methods and tools that can help…

Thank You!

Aneta.Bonikowska@statcan.gc.ca

Stay connected!

StatsCAN app

Eh Sayers podcast

StatsCAN Plus

The Daily

Website

Surveys and statistical programs

Data service centres

My StatCan

Questions? Contact us: infostats@statcan.gc.ca

Statistics Canada—
Your National Statistical Agency

Delivering insight through data for a better Canada