# Fundamental Concepts of Quantitative Impact Assessment: Module 2

Aneta Bonikowska, PhD

Social Analysis and Modelling Division

**March 19, 2024**

Statistics Canada Statistique Canada

Canada

# Objectives of the course

- Introduce the Quantitative Impact Assessment (QIA) approach and related concepts
  - Seeks to estimate the **causal** impact of a policy/program on outcomes of interest (e.g. individual wages or employment, firm sales)
  - QIA relies on numerical data
- Demonstrate that correlation ≠ causation
- Introduce common methods used to estimate causal impacts

# Modules

**Module 1:** Policy context

**Module 2:** Quantitative Impact Assessment (QIA): An overview

**Module 3:** Randomized controlled experiments: The gold standard in QIA

**Module 4:** Common approaches to QIA

# Module 2

# Quantitative impact assessment (QIA): An overview

# Module 2 contents

- Where does QIA fit in a broader program evaluation?
- The fundamental problem in QIA (which we may never solve!)
- QIA concepts

# Where does QIA fit in a broader program evaluation?

- Three parts to program evaluation:
  - Implementation/process
  - Impact
  - Cost-benefit

# Where does QIA fit in a broader program evaluation?

Implementation/process:

- Fidelity: Was the program delivered as intended?
- Take-up: What percentage actually participated in the program?

- Example: A fire-safety awareness program
  - Brochures are delivered to houses
  - Follow-up:
    - Were the brochures delivered to the right homes and on time? (fidelity)
    - How many households read the brochure? (take-up)

# Where does QIA fit in a broader program evaluation?

Impact:

- Did the program **change** the lives/behaviour of participants?

  E.g.:
  - Installed fire detectors on each floor?
  - Fewer fire-related fatalities?
  - Feel safer as a result?

- QIA is concerned exclusively with the impact

- Note: Qualitative/mixed methods are not covered here

# Where does QIA fit in a broader program evaluation?

Cost-benefit:

- Not an accounting concept (i.e. not just dollars and cents)
- Was the program worth it from a societal point of view?

# Where does QIA fit in a broader program evaluation?

Cost-benefit - comprehensive:

- Costs
  - Direct costs of the program (producing, printing and delivering brochures)
  - Opportunity costs (what other programs/services were foregone as a result of this new program? Fewer firefighters available for calls?)
- Benefits
  - Impact (fewer fires)
  - Spillovers (fewer days missed at work benefits the firm and the economy as a whole)
  - Some benefits hard to price (feeling of safety? value of a life?)

# Where does QIA fit in a broader program evaluation?

- We will focus on tools to facilitate impact assessment:
  - Did the program change behaviour or the lives of participants?

# The fundamental problem in QIA (which we may never solve!)

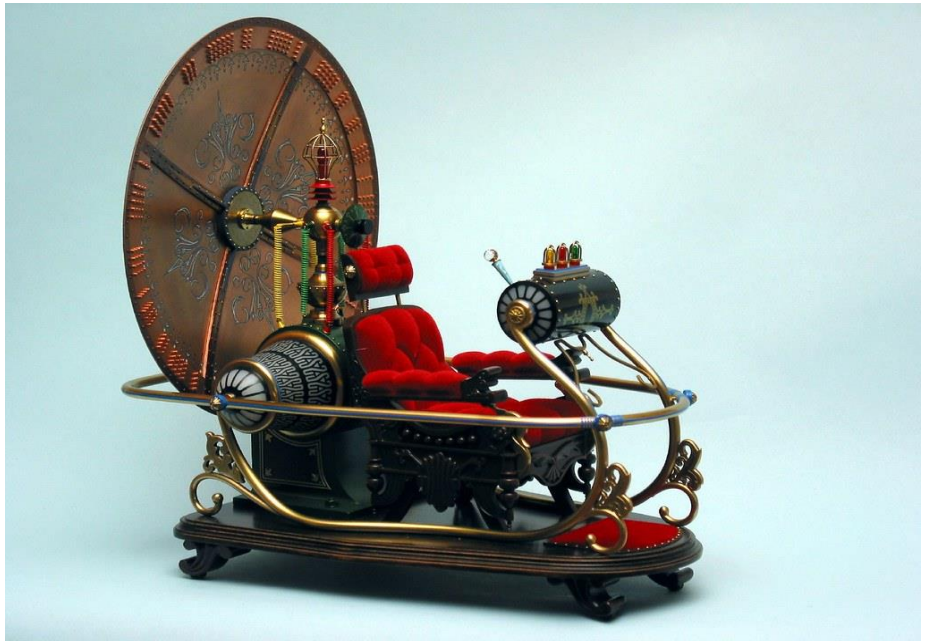
- A program is successful if it has changed the lives or the behaviour of participants

- To know if this has definitely happened, we must be able to observe an outcome (e.g. fire fatalities) in two different states (with and without the program), for the same people, at the same time

# The fundamental problem in QIA (which we may never solve!)

- Why do we need to observe outcome **with and without the program**?
  - If fatality rate is zero with or without the program, then the program did not work
- Why the **same people**?
  - If flyers were delivered to a neighbourhood that already had fire detectors on each floor, and not in neighbourhoods that did not have detectors, it will look like the program worked
- Why at the **same time**?
  - It may take some time to observe different outcomes (a change in the fire fatality rate)
  - People and circumstances change over time
  - Neighbourhood gentrifies
  - People age and become more concerned with fire safety

# The fundamental problem in QIA: foolproof solutions…



Time machine

Parallel universes

# The fundamental problem in QIA (which we may never solve!)

- We would like to be able to observe an outcome in two different states (with and without the program), for the same people, at the same time
- This will only be possible once we build a time machine or make contact with parallel universes
- Until then, we have to settle for imperfect methods

# The fundamental problem in QIA (which we may never solve!)

- But we can come close to estimating the true impact (very close in fact)
- We will go through different approaches that succeed to varying degrees
- First, some important concepts...

# Outcome vs Impact

- Example #1:
  - Program = Tax break on R & D spending by firms
  - Outcome = Spending on R & D by firms after program implementation
  - Impact = Change in R & D spending as a result of the tax break

# Outcome vs Impact

- Example #2:
  - Program = Tutoring program for low-achieving high school students
  - Outcome = Math test scores after tutoring program
  - Impact = Change in math score as a result of the tutoring program

# Outcome vs Impact

- Impact = Outcome (with program) – Outcome (without program)

# Actual vs. counterfactual outcomes

- Any outcome that we observe is an **actual outcome** (e.g. fatalities following the fire safety program)
- Ideally, we want another actual outcome (the one observed had the city not had the program)
- ...the difference in the two would give us the true impact

# Actual vs. counterfactual outcomes

- We can't observe both actual outcomes (with and without the program) since no one (or no firm) can follow a program and not follow a program at the same time

- So we settle for the **counterfactual outcome**:

  – Our best estimate of the outcome that would have prevailed had the program not been in place

- *QIA is focused almost entirely on finding a good counterfactual outcome!!!*

# Actual vs. counterfactual outcomes

- The counterfactual outcome relies on finding ‘similar people’ to those who were offered the program
- ‘Similar’ means on average similar in terms of a broad range of characteristics that could determine the outcome
- These ‘similar people’ could be:
  - those in the program group observed before the program,
  - other people who were not in the program group,
  - or a combination of the two

# Actual vs. counterfactual outcomes

- Since we have to settle for a counterfactual outcome, at best, the impact will be an estimate

(Estimated) impact =

Actual outcome (with program) – Counterfactual outcome (no program)

# Treatment effect

- What policy analysts call a ‘program’ is often referred to as a **treatment** by analysts involved in QIA
- Consequently, an ‘impact’ is often referred to as a **treatment effect**
- Term comes from medical literature: participants are patients in need of a treatment (program)

# Treatment effect

- Impacts or effects will vary by person (heterogeneous effects)
- It is much easier to estimate a treatment effect for a group of individuals, we term ‘average treatment effect’

# Different types of treatment effects/impacts

- In the fire safety program, the treatment (the brochure) was delivered to all households in a neighbourhood...
- ...but not everyone reads the brochure (not everyone takes the treatment)
- The *intention* was to treat everyone, but only some got treated

# Different types of treatment effects/impacts

- **Intention-to-Treat Effect** = Impact on those intended to be treated (e.g. all those offered a flyer)

- **Treatment-on-Treated Effect** = Impact on those who took the treatment (only those who read the flyer – i.e. took their medicine!)

# Different types of treatment effects/impacts

- We usually can't dictate who takes the program (i.e. who takes their medicine), so the intention-to-treat effect is more relevant for policy. It is also easier to estimate (as we will see).
- The treatment-on-treated effect is still important since it speaks to the potential of the program. Policy analysts can use it to advertise potential impact of program in an attempt to increase program take-up.

# Different types of treatment effects/impacts

- "**Local treatment effects**" (or Local Average Treatment Effects – LATE) are more relevant for program expansion or contraction

- Example: Mammography screening program to detect breast cancer – mammography available for free to women aged 50+ and strongly recommended

# Different types of treatment effects/impacts

Mammography screening example:

- Policy issue:
  - Early detection is important
  - But it may be costly and not worth it if the detection rate is low

# Different types of treatment effects/impacts

Mammography screening example:

- If screening is free to all women 50+, impact may be high (i.e. potentially lots of diagnoses)
- Tempting to use impact for 50+ to justify expanding program to include 40- to 49-year-old women, but potentially misleading
- We don’t know impact on 40-49 group, but more relevant is impact on 50-year-olds vs 49-year-olds (impact on those at the margin) – LATE

# Different types of treatment effects/impacts

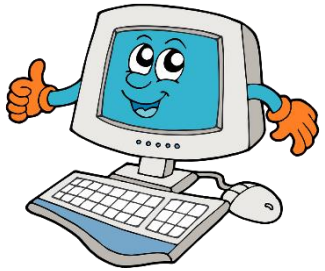
- LATE can be calculated in situations where:
  - the treatment is binary (e.g. take the screening test or not), and
  - the treatment assignment is binary (e.g. age 50+ or not)

- LATE is calculated using a subgroup of the population referred to as ‘compliers’, i.e. only those who would never take the test before turning 50, and would definitely take the test after turning 50
  - the estimated impact may not be representative of the rest of the population

# Why are correlations not enough?

**Question:** Did changes in technology cause increased wage inequality by making some workers more productive?

**Study design:** Do workers who use computers at work earn more than those who do not?



- Krueger (1993) finds that workers who use computers on the job earn 10% to 15% more than nonusers after controlling for standard worker attributes
- Is this an estimate of the return to computer skills?
- DiNardo and Pischke (1997) find a similar wage premium for workers who use a pencil at work, a calculator, a telephone, or who sit while working

→ Spurious correlation between computer use on the job and wages – some other (third) factor(s) is driving both wages and computer use on the job

# Why are correlations not enough?

**Question:** Does consumption of sugary drinks lead to depression?

**Study design:** Relate consumption in year 1 to depression diagnosis over subsequent 10 years

- Those who drink 4+ sweetened sodas had a 30% greater risk of depression (U.S. National Institutes of Health 2013)
- Does that mean that drinking soda causes depression?
- What if people who feel/are depressed try to self-medicate with comfort food, like soda?

⟹ Reverse causality

# Why are correlations not enough?

- Correlations calculated from **observational data** can give a misleading impression of the true impact of an intervention on outcomes of interest due to:
  - **Self-selection** (related to the outcome of interest)
  - **Confounding factors**
- A computer will always calculate a correlation
- It is up to the researcher/evaluator to know how to identify a causal impact from available data and when a correlation can be given a causal interpretation

# 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