



You Have Figured Out Your Intervention and Outcomes

What's Next? Experimental and Quasi-experimental Design Options

Guillaume Beaulac, PhD

Fellow en sciences du comportement, Bureau du Conseil privé
Behavioural Scientist, Global Affairs Canada

guillaume.beaulac@international.gc.ca



Government
of Canada

Gouvernement
du Canada

Canada

To read more on design options

This presentation will not go into the details of each design option – for more:



*Measuring impact
by design*

*Modèles de mesure
des impacts*



FR: <https://www.canada.ca/fr/centre-innovation/services/rapports-ressources/modeles-mesure-impacts.html>

EN: <https://www.canada.ca/en/innovation-hub/services/reports-resources/measuring-impact-design.html>

1

General considerations

Before
you start

Establish your framework

What is the change you hope to see?

What is your intervention?

What are the touchpoints?

What are your data sources?

What is the baseline?



La mesure des impacts

Certaines données sont plus faciles d'accès que d'autres.

Autant que possible, mesurez le comportement *directement* (observation).

Sinon, indirectement (données administratives, mesures indirectes).

En troisième lieu, mesurez *l'intention d'agir*.

Avoir un bon portrait de la situation

De meilleures données *avant* l'intervention permet d'avoir une idée plus solide des impacts.



Retarder pour recueillir des données ?

Déterminer l'ampleur du problème, discuter des résultats attendus, mais aussi de problèmes potentiels.

Pre-mortem, red teaming.

Basic notions

Control group and treatment/experimental group:

Britannica: “Ideally, the control group and the experimental groups are identical in every way except that the experimental groups are subjected to treatments or interventions believed to have an effect on the outcome of interest while the control group is not.”

2

**Designs that are
not experimental**

Il n'est pas toujours possible de faire de l'expérimentation

Il est tout de même possible d'avoir des approximations des impacts – et c'est souvent une analyse qui est effectuée.

Deux options principales existent:

Benchmarking with aggregate data

Compare to a regional or national average.

It can be useful to *approximate* a baseline if you do not have one.

It can also be useful in your analysis if you do not have proper comparison or control groups.

Low validity (test group will likely be different), but it can be a starting point.

Before/after comparisons (pre/post)

Measure the outcome before and after your intervention.

Easy to implement; it can indicate an effect but will not be robust to any extent.

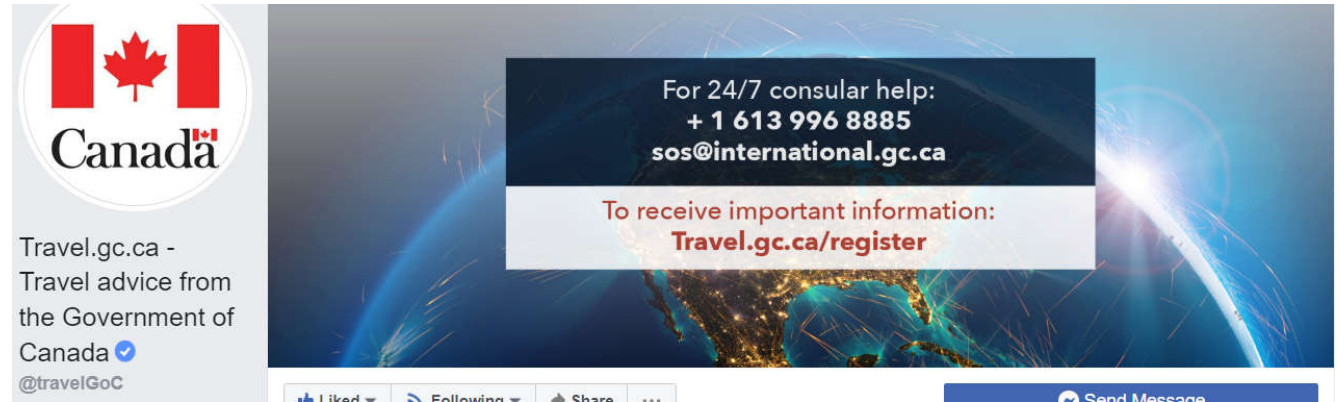
Workshop 1: Why? What might go wrong with such an analysis?

Before/after comparisons (pre/post)

From *Measuring Impact by Design*:

“Assumption: Nothing except for the intervention changed between the two measurements – this never holds in practice.”

Example



Registration of Canadians Abroad

We want to raise awareness around this program, detailed communications plan, using the website, events, Facebook, Twitter, etc.

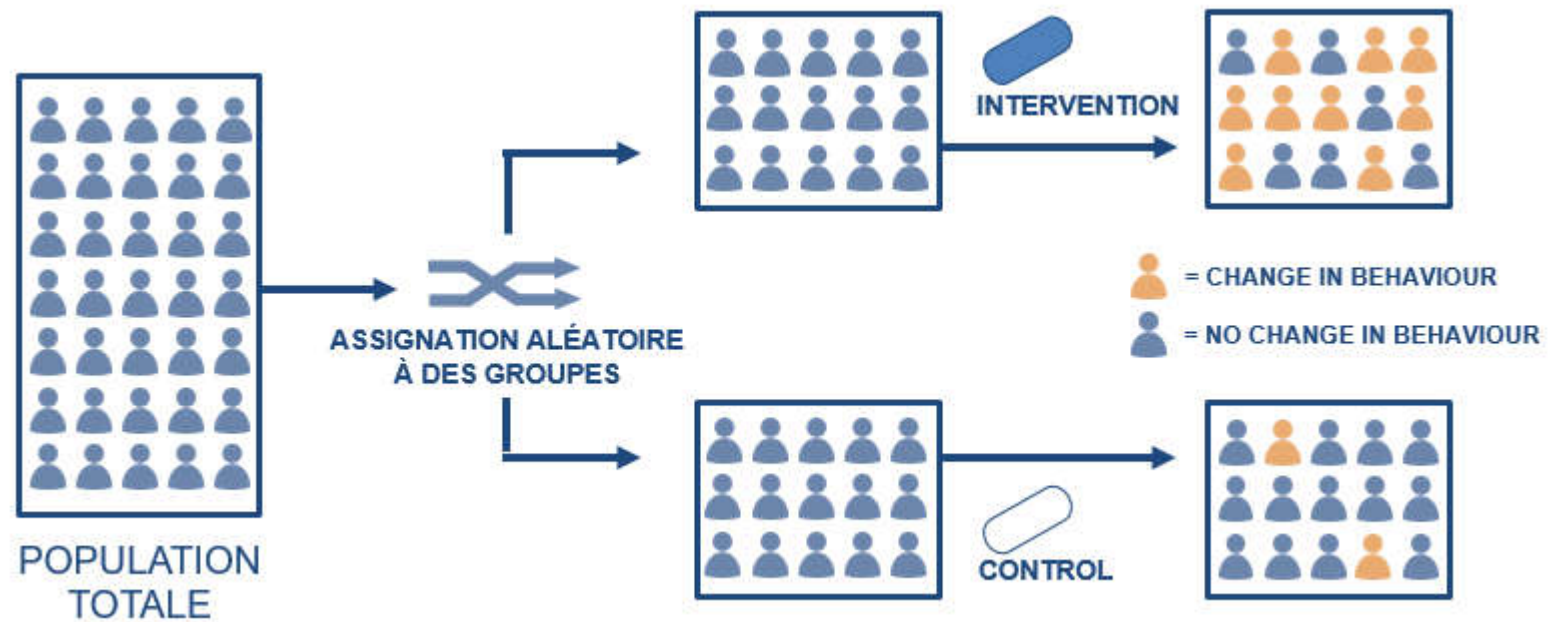
COVID-19, worldwide consular operation.

3

Randomized Controlled Trial

RCTs as the
“gold standard”

Les ECRs



- ▣ Permet de **déduire un lien de causalité**.
- ▣ Fournit une compréhension quantitative de l'**impact** d'une intervention.
- ▣ **Minimise les risques et les coûts** par la réalisation d'essais pilotes à plus petite échelle avant la mise en œuvre.

Features of a good RCT

Large population that can be divided into two groups randomly.

Random assignment helps make the groups “balanced”.

Ideally the experimenter does not know which group is which.

Features of a good RCT

Pay attention to attrition (dropouts).

Double check how your intervention is distributed.

Be careful that there is no contamination (spill-over) between the groups.

Unless it does not really matter ...

Contamination / Spill-over effects

Think about your assumptions: would contamination matter?
What are you measuring?



Questions: Essais contrôlés randomisés

Everything you always wanted to know about RCTs (but were afraid to ask).

=> sli.do

4

D'autres options ?

Autres designs
expérimentaux et
quasi expérimentaux

Si un ECR n'est
pas une option ...

S'il y a de l'opposition aux ECRs ...

Parfois, il peut être difficile de faire accepter un essai contrôlé randomisé (ECR), malgré les meilleurs arguments.

Il est cependant possible d'utiliser d'autres approches qui peuvent permettre d'estimer l'impact de l'intervention.

Cibler les objections ...

Objections sur l'éthique

Quelle est la source de l'objection ?

Il est possible de faire valoir l'importance de tester/valider ou de faire un projet pilote plutôt qu'une simple implémentation.

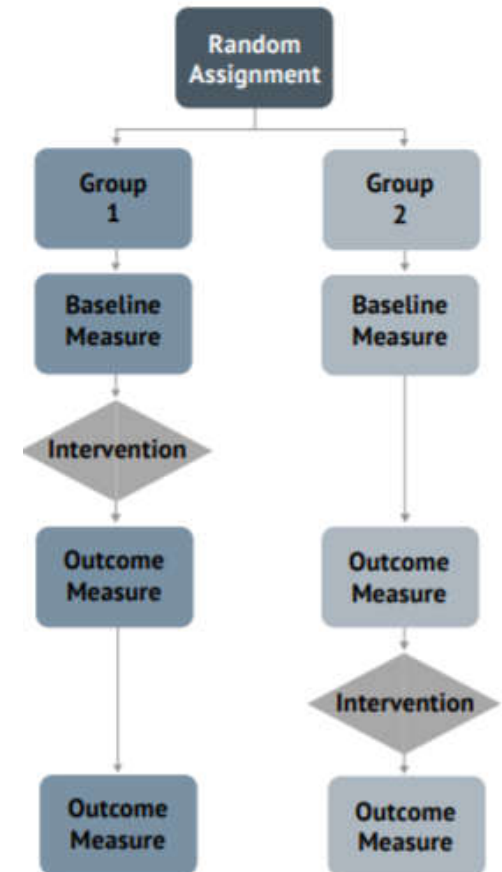
Cependant, parfois il s'agit parfois d'un désir que tout le monde reçoive l'intervention le plus vite possible.

Phased introduction

Everyone *eventually* receives the intervention, but at different times.

When the worry is that not everyone would receive a given intervention, a phased introduction design can work well.

You can also adjust the intervention.



Phased introduction

These are repeated measures. The control group receives the intervention *later* rather than not at all.

Randomization is still an important part of this, and the worry about contamination should be kept top-of-mind.

This can be just as strong a design as a RCT.

Phased introduction without randomization (quasi experiment)

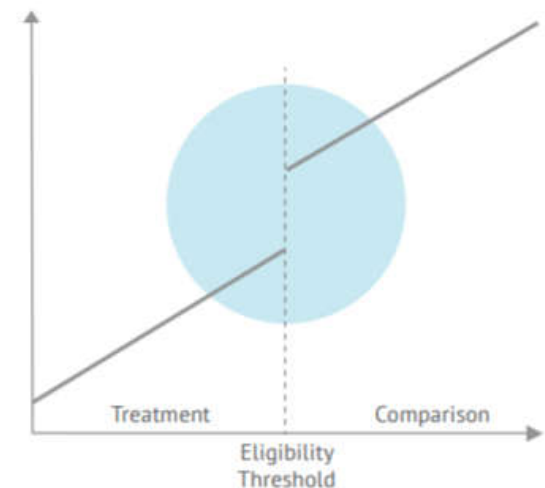
Sometimes randomization is not possible.

GAC and its missions would be an example here since the groups would not be equivalent / comparable.

Repeated measures over rounds of introduction of an intervention can still yield interesting results.

Regression discontinuity designs (RDD – quasi experiment)

In some programs, a threshold determines eligibility.



Evaluation: a score determines if you are in or out of a program.

Age cut-offs, income levels, etc.

The difference between a 3.49 and 3.51 GPA is small, but can have a big effect.

Regression discontinuity designs (RDD – quasi experiment)

In those cases, the assumption is that the values close to the threshold will form comparable groups where one receives the intervention and the other one does not.

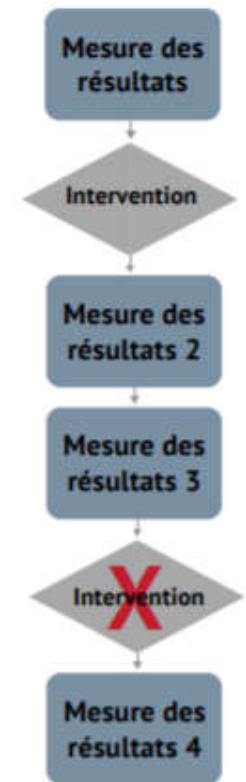
Comparing these groups can potentially indicate the impact of a program.

Modèle de traitement supprimé / interrompu (quasi expérimentation)

Qu'arrive-t-il si un groupe qui reçoit un traitement arrête de le recevoir ?

Cette approche ne peut fonctionner que dans des cas précis, mais peut aider avec l'absence d'un groupe témoin.

Pour l'analyse d'effets immédiats, souvent liés à des règles / des frais.



5

Présenter et défendre son projet

Équilibre entre la rigueur de l'expérimentation et la faisabilité

Dur constat:

Il vaut parfois mieux faire une expérimentation qui soit moins rigoureuse que de ne pas en faire du tout.

Évaluer ce qui est possible, évaluer l'ouverture, défendre le projet ...

Mais aussi accepter les compromis.

6

Discussion

Quelles sont vos expériences ?

Qu'est-ce qui a bien fonctionné ?

Did something go horribly wrong?