

Shahid, Leyla - 9:29 AM

[Impacts of Gender Based Analysis Plus - Canada.ca](#)

Langlois, Marc (he/him, il) - 9:45 AM

Is it possible to present via PowerPoint live (for accessibility purposes)?

A. **MacVicar, Matthew (he/him, il/lui)** - 9:47 AM

Hi Marc, unfortunately this is not possible at this time, but we will absolutely take this feedback for future iterations of this workshop. Thank you so much!

Chartrand, Louis - 9:49 AM

What about a program whose goal is not to have impact? (E.g. a program to protect the public in the event of a pandemic)

A. **Frenette, Marc** - 5:27 PM

There is an intended impact in this example – it is to increase protection in case of a pandemic. Without the program, protection levels might be low. The outcomes could be mortality, illness, job loss, etc (all the things we experience during COVID). However, we may not be able to test the program until a new pandemic arises. In these cases, you will have to rely on intelligence and experience. Using data from COVID may be helpful, but I suspect no two pandemics will be alike.

Hadziavdic, Sadia (TC/TC) - 9:50 AM

Re. 'Was the program delivered as intended?' - how do you quantify policy intent / process in this? For all the good reasons, policy might be detached from real-life context.

A. **Yang, Doo DY [NC]** - 10:12 AM

I think it is more about the implementation and delivery related to the direct output. For example, if a program is targeting youths with barriers to employment, then the program can measure how many of them participated in the program.

A. **Frenette, Marc (StatCan)** - 5:30 PM

This refers to program fidelity. We were simply noting that some programs have as an objective that it be delivered in a certain way. For example, participants are informed by phone of a grant. If they were informed by email during the pilot test, then there was a low level of fidelity.

Tisi, Shirqille (ISED/ISDE) - 9:53 AM

Is impact measured on a scale or degree of change?

A. **Frenette, Marc (StatCan)** - 5:33 PM

It is always the difference in actual vs counterfactual outcomes (i.e. with an without the program).

El-Mallah, Sarah - 9:54 AM

How can we implement these concepts in the context of updating an employment equity policy - how can we assess impact on staff and the organization in this case?

A. **Frenette, Marc (StatCan)** - 11:33 AM

Without knowing more details on the specific policy, I could see this as being a good candidate for a difference-in-differences approach, making sure that there are no significant confounding factors. Defining and tracking an outcome (or outcomes), before, during, and after program

implementation will be important with a good data system. Finding an appropriate comparison group will also be important, but this could be tricky as it might involve inequity in program implementation between departments.

Catherine Mathieu - 9:55 AM

Can we get a copy of this presentation please

A. **MacVicar, Matthew (he/him, il/lui)** - 9:57 AM

The materials can be found here: [QIA Workshop - Fundamentals \(GCwiki\)](#)

Les matériaux peuvent être trouvés ici : [Atelier sur l'EQI - Éléments fondamentaux \(GCwiki\)](#)

Chartrand, Louis - 9:56 AM

I'm curious about that counterfactual group as well. In academia, there is a range of criteria for what makes a good quasi-experimental control group. What would be the threshold here? And who decides what is satisfactory?

A. **Frenette, Marc (StatCan)** - 5:08 PM

This question was asked during lecture 1, before we discussed specific approaches to determining counterfactual groups in lectures 2 and 3.

Hadziavdic, Sadia (TC/TC) - 10:04 AM

perhaps its time for a new study... on people who are data-literate and know how to read and analyse data and those who don't!

Adesokan, Goodness (ISED/ISDE) - 10:06 AM

Quick question: 4+sweetened sodas per day? week? or monthly?

A. **MacVicar, Matthew (he/him, il/lui)** - 10:10 AM

Here is a link to the article: [Sweetened Beverages, Coffee, and Tea and Depression Risk among Older US Adults - PMC \(nih.gov\)](#)

Safa Sel, Halil - 10:11 AM

Can you explain the difference between 'reverse causality' and 'Confounding' ?

A. **Chartrand, Louis** - 10:15 AM

Say you have a correlation between A and B. You might first think that A causes B ($A \rightarrow B$), but actually, it's the opposite ($B \rightarrow A$). That would be reverse causality.

Confounding involves another variable (say, C). You might have, for example, $B \leftarrow C \rightarrow A$: C causes both A and B. That would be confounding.

A. **Yang, Doo DY [NC]** - 10:15 AM

Reverse causality:

Drinking soda causes depression. Or depression causes to drink more soda.

Confounding:

Using a computer earns more money.

Sitting at desk earns more money. Then

Office jobs earn more money.

A. **Frenette, Marc (StatCan)** - 5:35 PM

These are great answers

Connors, Marianna M [NC] - 10:13 AM

On treatment effects - Is the point that we prefer to measure treatment-on-treated effect (rather than the intention-to-treat effect)?

A. **Chartrand, Louis** - 10:23 AM

I don't think so, I think it's more of a way to define a control group

A. **Luke, Jonathan** - 10:27 AM

Depends on which impact you care the most about. If you want to know how the treatment impacts people who receive it, you need treatment on treated. If you want to know how a program that tries to deliver the treatment (and is not always going to be successful) you want to know intention to treat.

A. **Frenette, Marc (StatCan)** - 10:31 AM

It would be nice to successfully estimate treated effects on the treated. This would inform the broader population of the potential impacts, if they took the treatment. However, it is difficult to account for the "selection biases" that occur when people, who are offered a treatment, choose to take (or not take) this treatment. Intention to treat is much easier to estimate because we can control who is offered the treatment (i.e. there is no selection biases in being offered a treatment).

A. **Connors, Marianna M [NC]** - 10:50 AM

Thanks Marc! (I'm at Atiq's Learning Branch at ESDC). You may recall our program funds tutoring/mentoring etc. Because we target, our youth "self select" (or are selected), so there isn't a group who are offered our service but don't take it up. Therefore I think (?) I want to show via QIA that our supports are helping youth served (compared to those not served - granted I can't control for other services 'unserved' youth may access outside our program). IE - treatment effect on treated (by comparing to those untreated by our service). Yes?

A. **Frenette, Marc (StatCan)** - 11:17 AM

Marianna - maybe send me an email...I'd like to give it more thought before offering advice/opinions

A. **Connors, Marianna M [NC]** - 11:18 AM

Amazing. Looks like my colleague Stacey (who works on the Global Skills Opportunity) has the same line of thinking about 'treated' groups. We'll collab and send you something! thanks!

Lawrence, Lola (StatCan) - 10:14 AM

Could you give few examples on how quantitative outcome is different from qualitative outcome?

A. **Chartrand, Louis** - 10:20 AM

My guess is that it's not the outcome that is quantitative of qualitative, but rather the measurement. Quantitative measurement is convenient: easier for comparisons, or to insert in a cost-benefit analysis. But not everything is quantitatively measurable, and usually, it's not a good idea only to value quantifiable impacts.

<https://issues.org/limits-of-data-nguyen/>

A. **Frenette, Marc (StatCan)** - 5:12 PM

Qualitative analysis digs deeper into narratives, even if samples may not be drawn scientifically and inferences cannot be made on populations. We did not touch upon this approach in the session.

Bertoia, Carl (HC/SC) - 10:14 AM

Just a few observations from my evaluation experience. Often there is no quantitative data to use. And programs will state that they are not responsible for attribution.

A. **Frenette, Marc (StatCan)** - 10:21 AM

In some cases, StatCan can help with that. Session 2 (Data) will go over ways that policy departments can create the data they need to answer program evaluation questions.

Procca, Andrew – 10:23 AM

In my studies, I've been pleased when I received a 10% response rate.

In your experience what is a good or average response rate???

A. **Chartrand, Louis** - 10:35 AM

I think it depends on the context. There might be reasons to believe your 10% is still representative. On the other hand, sometimes you have a relationship with your population (e.g. maybe they benefit from the program, and they would be willing to help) and 10% would look low, and suggest that something in delivery was problematic, and might be excluding some groups or biasing your sample

A. **Frenette, Marc (StatCan)** - 10:57 AM

10% is quite low, certainly by StatCan standards, but it is in line with public opinion polls. The important point is to understand who responded and who didn't. How are they different? and then make corrections through sampling weights (this is critical). As an example, if young people were particularly unlikely to respond to your survey, then the few young people that did respond need to represent the young people who did not respond. We do this by giving those few young responders more "weight" (i.e, we count them more times in our analysis, compared to older responders). this isn't perfect because some factors that determine survey response is not observed, and we are left with responses biases.

Vrolijk, Ademir-Paolo - 10:24 AM

Can you suggest methods to estimate the size of the control group given a treated group of a specific size (say 10, 100, and 1000 treated individuals)?

A. **Chartrand, Louis** - 10:29 AM

G*Power is a great tool to calculate sample size:

[Universität Düsseldorf: G*Power \(hhu.de\)](https://www.psychologie.uni-duesseldorf.de/lehre/studienleistungen/psychometrie/gpower/)

[Unité de Recherche Clinique et Évaluative \(ulaval.ca\)](https://www.ulaval.ca/recherche/clinique-et-evaluative/)

Champagne, Luc L [NC] - 10:27 AM

la taille utile de l'échantillon dépend de la population totale!

A. **Frenette, Marc (StatCan)** - 10:44 AM

Désolé Luc, mais je ne suis pas d'accord. Un petit échantillon (disons 3) ne suffira pas pour des analyses significatives, que la taille de la population soit de 20 ou de 1 000 000 d'habitants. N'importe quoi peut arriver lorsque nous retirons seulement 3 personnes d'une population.

Pereira, Brian - 10:30 AM

Sample size needed is based on the expected size of the effect. Conducting a power analysis with different estimated effect sizes can help you determine what your target sample should be.

A. **Frenette, Marc (StatCan)** - 10:38 AM

This is correct. One caveat I would add is that, in my experience, things tend to fall apart with very small samples (despite the fact that statisticians have created small sample distributions to guide our statistical tests). I don't tend to favour such tests. StatCan has a lot of large data sets, so fortunately, data users often have the opportunity to leverage large samples for their analyses.

A. **Pereira, Brian** - 11:47 AM

I don't disagree but there are scenarios where small samples can find validated effects. There are many reasons for this but one clear example is type of intervention. I used to do animal research with addictive substances. Injecting an addictive substance is a very heavy handed type of intervention. You won't need very large samples to find large effects and to some extent medium-small sized effects. That is why lots of research on animals use small samples (along with other considerations like ethics). I don't fully support this and there are many reasons that this is still a problem. Regardless, you can still find real effects with small samples. Even in human research you can have small samples depending on the type of intervention or how many trials are given to people. Even in the case of program evaluation, if you have a strong intervention (expecting a large effect) then you may not need 400 (that was the number said earlier as a rule of thumb). If your intervention is weak (expecting a small effect, i.e. lots of communications style interventions) then you might need more than 400. This gets more complicated if you want to look at interactions. Lets say you wanted to compare your intervention across different employee tenures. You may still not be powered to find a real effect with a 400 sample that interacts with intervention and specific characteristics

There are many considerations to consider that can impact whether or not your statistical analysis will find something (with or without an intervention design). Overall, your statistical analysis will never hurt from having large samples.

Gracia, Romie (PHAC/ASPC) - 10:31 AM

Les questions/remarques soulevées dans ce chat sont forts pertinents. Croyez-vous qu'il soit possible d'obtenir une copie des questions et des réponses par écrit après la séance?

A. **Roberge, Julie (she/her, elle)** - 10:34 AM

Une transcription des questions et des commentaires sera fournie aux participants dans les semaines à venir.

Cabedoce, Alexandra AC [NC] - 10:40 AM

With the spill-over effect, does that mean that your sample would still include all students and then you would try to keep minimum contact between the treatment group (that would attend the presentation) and the control group (that would not attend) even after the presentation was given?

A. **Chartrand, Louis** - 10:43 AM

This could be a way to address this. Another way might be to start with similar communities that already have limited contact (e.g. two different schools with similar populations in different cities)

Anonymous User (Guest) - 10:46 AM

Just a comment: Incentives can present issues as well if done online with the public for example (folks signing up more than once, bots, etc)

Bryant, Stacey SB [NC] - 10:47 AM

In terms of the "intention-to-treat", I know it said it is tempting to just look at those in the treatment group who actually took the treatment but that only the offer of treatment was randomly assigned. Why would you not want to look only at those who took the treatment rather than all those offered the treatment (if you can definitely tell who took it - e.g. it is an employment support program) rather than also including those in the intention to treat group who didn't take part in the program? Is this because there could be a third variable impacting who chose to take up the treatment? But by including those who didn't take up the treatment as well as those who did with the counterfactual group, aren't you also not really comparing the treated and untreated?

A. **Chartrand, Louis** – 11:00 AM

I guess a part of that is that we're evaluating programs, not treatments.

A. **Luke, Jonathan** - 11:07 AM

First, you may not be able to estimate the treatment on treated impact, if you do not observe whether assigned individuals actually took up the treatment. But you may also be more interested in the intent to treat effect as it often is a closer estimate of your overall program impact since programs often only control who they try to impact.

A. **Bonikowska, Aneta (StatCan)** – 3:16 PM

In the slides, we're talking about evaluating the impact of the program itself, and here we're thinking about programs that in real life would not be expected to have 100% take-up (although the higher the better and take-up is itself something of interest in the broader program evaluation process). By running an experiment, with the help of randomization, you are essentially creating two 'states of the world' - one in which a program was rolled out, that's the state the treatment group lives in (some of the individuals in the treatment group will choose to take the treatment, others won't, just like in real life, and yes, third factors will contribute to the decision); and a second state in which the program was never rolled out, that's the state the control group lives in. We are thus able to compare outcomes of participants who lived in a world where the program was introduced and outcomes of participants who lived in a world where the program never happened - this comparison gives us an estimate of the causal impact of rolling-out the program on outcomes of interest in our target population of interest; and it's because randomization allowed us to observe outcomes of (almost) identical groups of people,

at the same time, in two different states of the world (which is as close as we can get to observing the same people, in two different states, at the same time). Hope this helps.

Ortlieb, Kayla (HC/SC) – 10:48 AM

Could you not also offer the treatment at a later date to the control group?

A. **Luke, Jonathan - 11:04 AM**

Yes, that's a common approach to addressing equity concerns

A. **Frenette, Marc (StatCan) - 11:09 AM**

Yes, but expensive with large experiments (especially if a grant is the treatment). Very common in medical experiments where participants alternate being on the treatment and not.

Yousra Nafie (CSPS-EFPC) – 10:50 AM

Thanks Marc and Aneta for the amazing lecture

Understanding the critical role of sample size, how can one ensure the experiment's duration aligns with ethical guidelines while also being adequate to capture the intended results?

A. **Frenette, Marc (StatCan) - 11:06 AM**

Participants need to be well informed on the duration and other aspects of the experiment. This should be done before they agree to participate. This will avoid ethical issues. Importantly, and as I mentioned live, approval needs to be obtained to link their data to other data sources (if StatCan data is targeted here, it would be helpful to reach out to StatCan for advice since we do have precedents on this).

Li, Lulu - 10:51 AM

Thank you Aneta for the presentation! I am curious about how randomized controlled experiments (RCE) are utilized in program evaluation. Will any real-life example(s) of this specific method (RCE) be provided in session #4?

A. **Roberge, Julie (she/her, elle) - 10:55 AM**

Hello Lulu, the agenda for Session 4 can be found here: [Quantitative Impact Assessment Workshop/Case Studies II - wiki \(gccollab.ca\)](#)

A. **Frenette, Marc (StatCan) - 11:01 AM**

StatCan organized session 3, and experiments are not used. However, the presenters use "quasi-experimental" approaches, which we will touch upon after the break. The point of going through RCT (experiments) is pedagogical - it shows that even in a well-designed, controlled experiment, there are still biases. Many of the concepts Aneta covered with experiments applies to any analysis (attrition, spillovers, etc.)

A. **Bonikowska, Aneta (StatCan) - 1:28 PM**

The last few slides in Lecture 2 of today's session talk about an actual RCE that was conducted in Canada in the 1990s (the Self-Sufficiency Project) and researchers today are still working on evaluating the long-term impacts of the program that was being tested in that RCE. There are reports and studies published on this, if you're interested in finding out more about it.

Greene, Graham (ISED/ISDE) - 11:19 AM

Sorry if someone has already asked this. Will the slides be available after the workshop?

A. **MacVicar, Matthew (he/him, il/lui)** - 11:21 AM

Hi Graham, you can find the agenda and materials at the following link: [Quantitative Impact Assessment Workshop/Fundamentals - wiki \(gccollab.ca\)](https://wiki.gccollab.ca/Quantitative%20Impact%20Assessment%20Workshop/Fundamentals)

Vézina, Rosalie R [NC] - 11:31 AM

Can we qualify these methods as 'quasi-experimental'?

A. **Frenette, Marc (StatCan)** - 11:37 AM

Difference-in-differences, discontinuity estimators and instrumental variables are quasi-experimental approaches. As Aneta will probably mention soon, matching is more like a regression on steroids. It is a correlative tool on its own; however, it can be used to operationalize a quasi-experimental approach. We don't go through these here because it would require a strong background in econometrics.

Bryant, Stacey SB [NC] - 11:39 AM

Is matching always done at the individual level?

A. **Frenette, Marc (StatCan)** - 11:50 AM

In general, the more detail you have, the better. So matching at the individual level is always best. Any statistical approach can be applied at the group level, but you end up throwing away information (the intra-group information). There might be practical reasons why you would want to favour a group-level approach (e.g. you only have group level data), but my rule of thumb is to always leverage the most detailed data possible.

Rabbitt, Christopher - 11:48 AM

I know this is just an example, but wouldn't the CI on that regression line be ludicrous? $\beta = 0$? Any guidance on that?

A. **Frenette, Marc (StatCan)** - 11:51 AM

Use more data than the fictitious example shown in the chart 😊

This is the answer to many statistical problems

Barisonzi, Marcello (he, him, his | il, le, lui) (StatCan) - 11:49 AM

Will you be sharing references on how to implement these techniques in the major analytical tools (Python, R, Stata, SAS)?

A. **Frenette, Marc (StatCan)** - 11:52 AM

StatCan and TBS have discussed the possibility of creating a course for this specific purpose. If you think this is useful, let us know.

A. **Ward, Alicia (she; her | elle; la)** - 11:54 AM

Courses on R basics would be very useful for the general population. It is user-friendly with great analysis tools and lots of built-in programs, but some basics need to be understood in order for the general population to understand.

A. **Barisonzi, Marcello (he, him, his | il, le, lui) (StatCan)** - 11:54 AM

Yes, a course would be very appreciated, but as a first step a list of available packages and methods would do.

A. **Rabbitt, Christopher** - 11:55 AM

Or even Excel. It's got issues, but has the advantage of being pretty ubiquitous, and has an interface that most are familiar with.

A. **Frenette, Marc (StatCan)** - 5:16 PM

Excel would be very limited for the advanced stuff. Stata is excellent. R and Python are flexible (and free), but I am not a user, so I can't comment specifically on their appropriateness for this kind of work.

Vanderkooy, Anna -SRD-KXD - 11:54 AM

We use results-based management in my department and consistently track changes in outcomes, with no counterfactual, then use that to conclude if a project/ program worked. What do you think of this?

What alternately should be done?

A. **Chartrand, Louis** - 12:00 PM

I'd be uncomfortable with that. But I know many depts do that.

I think you really should look for counterfactuals.

Besides, it might protect you the day external factors make your outcomes look worse.

A. **Vanderkooy, Anna -SRD-KXD** - 12:08 PM

I agree. What would you recommend, under RBM, for reporting/eval if a counterfactual isn't possible? Just monitor outputs and activities and know that you can't make conclusions on impact?

A. **Chartrand, Louis** - 12:23 PM

Like, absolutely impossible? Remember, for example, that while it's not perfect, "same cohort, another time" can count as a counterfactual. You'd probably want to try to control for as many potential confounds as possible by tracking relevant variables and doing things like matching, regressions mediator/moderator analysis.

A. **Frenette, Marc (StatCan)** - 5:19 PM

Without knowing the specifics of your programs, I would also highly encourage you to find some counterfactual. What you are doing is using a first-difference (Aneta covered this). It is a start, but it is highly subject to confounding factors, like Louis mentioned. Can you pitch the idea of "piloting" the program in a certain area (as long as there is a similar area that could serve as a counterfactual)? You would want to think about potential spillovers in this case.

Prystay, Linda (TC/TC) - 11:53 AM

Can you provide an example when a regression would not be linear?

A. **Ward, Alicia (she; her | elle; la)** - 11:58 AM

Logistic regression in the case of categorical binary variable for example as opposed to measuring a continuous change in linear

A. **Bonikowska, Aneta (StatCan)** - 12:32 PM

Also, in a linear regression (OLS) of earnings, you could include, as control variables, age and age-squared, if you believe (or see in the data) that earnings increase with age but at a different rate, rather than a constant rate. Earnings could rise faster among younger workers and slower among older workers, and in fact decline among those nearing retirement. A quadratic in age would capture this nonlinear relationship between earnings and age.

Chartrand, Louis - 11:54 AM

I won't be able to attend session 3, but Marc's comments make me think I really want to see it. Will this be recorded and available to all as well?

A. **MacVicar, Matthew (he/him, il/lui)** - 11:58 AM

Hi Louis, Session 3 and all other sessions will be recorded and shared with all registrants in the coming days.

A. **Chartrand, Louis** - 11:59 AM

Thanks Matthew! Looking forward to it!

Sibley, Krissy - 11:56 AM

Thank you very much for an informative morning!

Turcotte, Sarah -KESB [She,Her | Elle] - 11:57 AM

Does StatsCan have data for populations in other countries to support international programs?

A. **Frenette, Marc (StatCan)** - 5:23 PM

We provide out Canadian data to international organizations to help with their international comparisons. We also, on an ad hoc basis, download readily available international data to do studies that have an international comparative component to it. However, these data are easily downloadable by anyone (e.g. the US Census public use files).

Pereira, Brian - 12:01 PM

You can get self-selection with or without incentives. What is important is having consistent application (incentives to everyone vs not).

Laliberté, Danièle DL [NC] - 12:01 PM

Incentives can also been seen as financial compensation for the time spent.

Petriu, Dorin - 12:07 PM

Would it be possible to save this Q&A and make it available to the participants? There is a lot of good info in here that would be a pity to just have disappear into the bit bucket.

A. **Roberge, Julie (she/her, elle)** - 12:09 PM

Hi Dorin. A record of the questions and comments will be shared with participants in the coming weeks.

Yang, Doo DY [NC] - 12:14 PM

The methods presented here today is basically a snapshot or a point-in-time comparison rather than longitudinal assessment, isn't it?

A. **Bouwer, Charles** - 12:15 PM

most are longitudinal requiring data before and after intervention

A. **Bonikowska, Aneta (StatCan)** - 12:24 PM

If you have longitudinal data on the outcomes of interest, you can apply these methods to study the impact at different points in time

A. **Yang, Doo DY [NC]** - 12:55 PM

Thank for your answers. Then can I understand like this? The analyses look like longitudinal study as they require before and after data, but QIA is mainly a comparison at a single time point.

Those methodology, however, are applicable for longitudinal studies.

A. **Bonikowska, Aneta (StatCan)** - 4:02 PM

Let's put it like this: the methods used to estimate causal impacts (i.e. used in QIA) typically estimate the impact at a specific point in time, but you can repeat the analysis for several different points in time, if you have the data to do that. If you have longitudinal data for several years after the program, then you can estimate the impact of a program on earnings in the 1st year after the program, 2nd year, 3rd year, etc. and you can do that in one fell swoop in a regression model. You can even estimate the average impact across years 1 to 3 or some other interval, if that's what you're interested in. Some methods will require longitudinal data to estimate an impact in even just one point in time though, like the different-in-differences approach which requires data from before the program's implementation, as well as after. Hope this helps.