Employment and Social Development Canada / Emploi et Développement social Canada

Canada

Gender-Based Analysis Plus Exploratory Evaluation Study on Selected Labour Market Programs

Presentation to the TBS Strategic Research Network

ESDC Evaluation Directorate

December 2023

Presentation outline

- Background
- Objective and scope of the study
- Data sources and indicators
- Methodology
- Results
- Conclusion
- References

Background

ESDC has used “matching” methods to assess the effectiveness of its labour market programs

- “Matching” is robust, but only provide average impacts.
- Not possible to estimate the distribution of program impacts across participants.
- Difficult to conduct subgroup analyses on different intersecting factors of identity.

Recent developments in machine learning have been applied to evaluate labour market programs in Europe (Belgium and Switzerland):

- Machine learning was used to estimate granular incremental impacts at the individual level, thereby also uncovering “what works for whom” (Wager and Athey, 2018; Lechner, 2019).
- Causal Machine Learning Evaluation of Training in Belgium (Lechner, 2019)

Objective of the study

- Test the effectiveness of a novel machine learning method to estimate incremental program impacts according to different GBA Plus intersecting identity factors.

What is Gender-Based Analysis Plus (GBA Plus)?
• An analytical process used to assess the experience of different women, men and gender diverse people with regard to policies, programs and initiatives. • The 'plus' in GBA Plus acknowledges that GBA goes beyond biological (sex) and sociocultural (gender) differences. *Source : Women and Gender Equality Canada*[1]

[1] https://women-gender-equality.canada.ca/en/gender-based-analysis-plus/what-gender-based-analysis-plus.html

Scope of the study

- Use machine learning to produce incremental program impacts through a GBA Plus lens:
  - Produce results according to intersecting GBA Plus factors of identity (e.g., a subgroup of individuals who are men AND visible minority).

- Examine two active labour market programs:
  - Labour Market Development Agreements (LMDA); and
  - Opportunities Fund for Persons with Disabilities (OFPD).

Data sources and indicators

This study uses linked datasets of rich longitudinal administrative data.

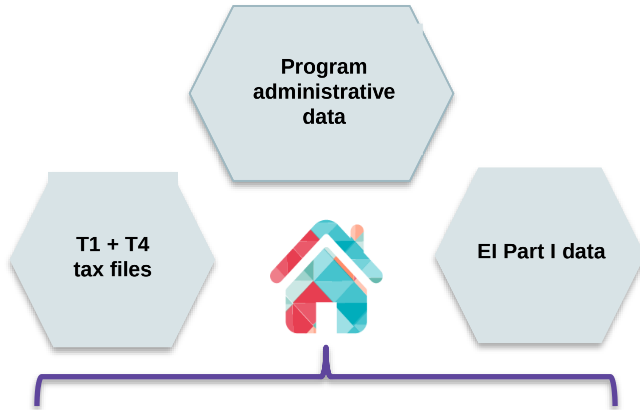
- **Participant groups:** Individuals who participated in LMDA and OFPD
- **Control groups:** similar individuals who did not participate in LMDA or OFPD:
  - For LMDA: Active EI claimants who did not participate in LMDA.
  - For OFPD: individuals with disabilities who participated in Employment Assistance Service.
- **Dependent variables (Main indicators):**
  - Incidence of Employment
  - Employment earnings
  - Dependence on income support
- **Over 75+ variables used:**

Socio-Demographics

- Age
- Gender
- Indigenous status
- person with disability
- Immigrant
- Visible minority
- Marital status
- Education level
- Province of Residence
- Urban/Rural indicator

Labour Market Characteristics

- Reason for job separation
- Industry (NAICS)
- New entrant to labour force
- Past interventions taken
- Past Earnings up to 5 years
- Past EI Benefit usage up to 5 years
- Past social assistance income up to 5 years
- Past tuition amounts up to 5 years



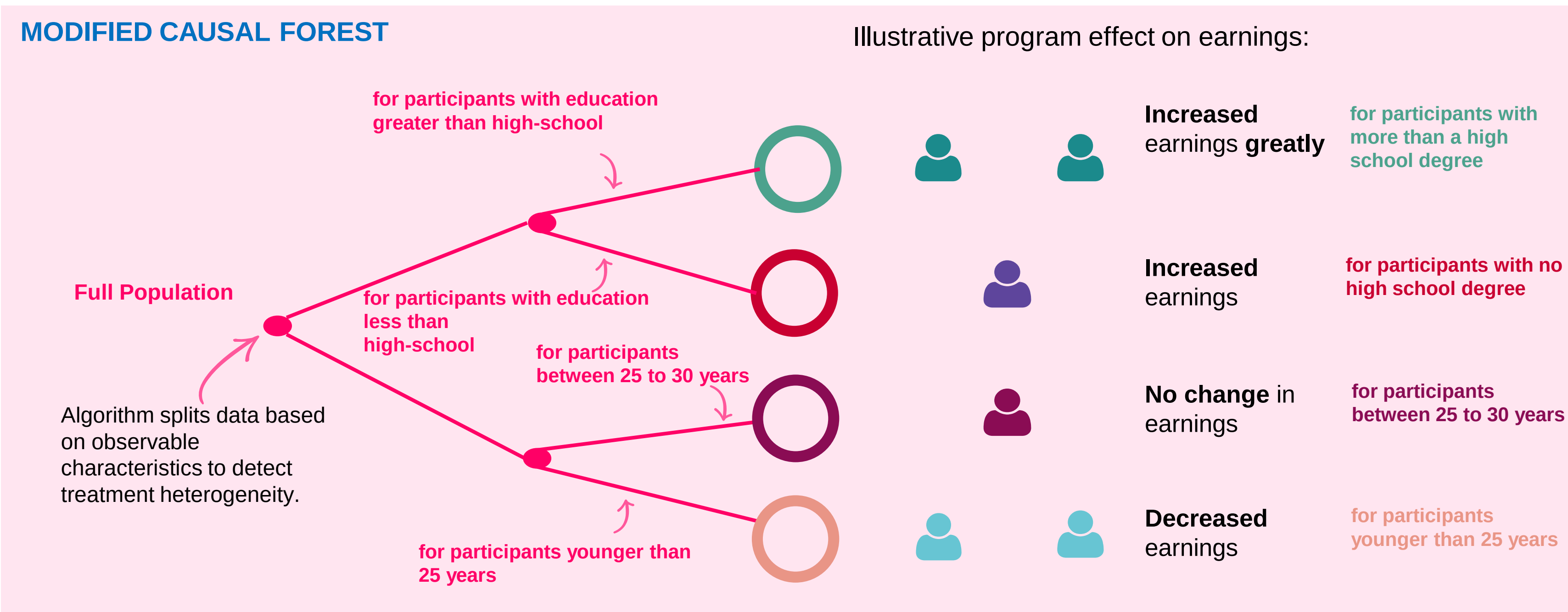
LMDA
2010-2012 cohort, followed up to 2018

OFPD
2011-2012 cohort, followed up to 2018

Methodology

The study uses the Modified Causal Forest (MCF):

- It is a supervised causal machine learning algorithm that builds an ensemble of decorrelated trees, learns the characteristics from the data and estimates the program impacts (Lechner, 2019).

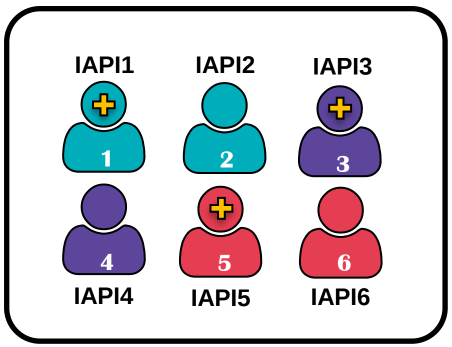


Methodology (cont.)

The MCF method produces 3 levels of program impacts, two of which were not available previously with other methods (Lechner, 2019):

Previously unavailable

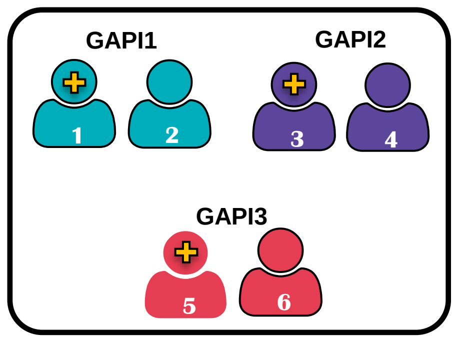
Individualized Average Program Impact (IAPI)



Measures the impact of a program on an individual with a given set of characteristics or profile. Represents causal program impact at the *finest level of granularity.*

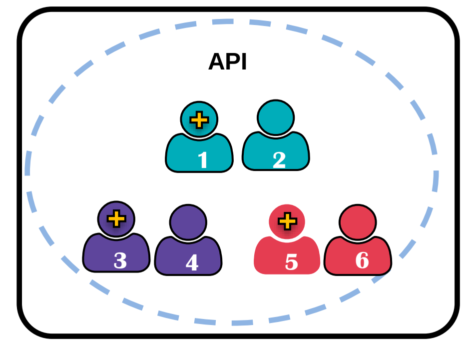
Previously unavailable

Grouped Average Program Impact (GAPI)



The GAPI can be estimated by aggregating and weighting the IAPIs over specific subgroups. Unlike traditional subgroup analyses, GAPIs can be *compared across groups.*

Average Program Impact (API)

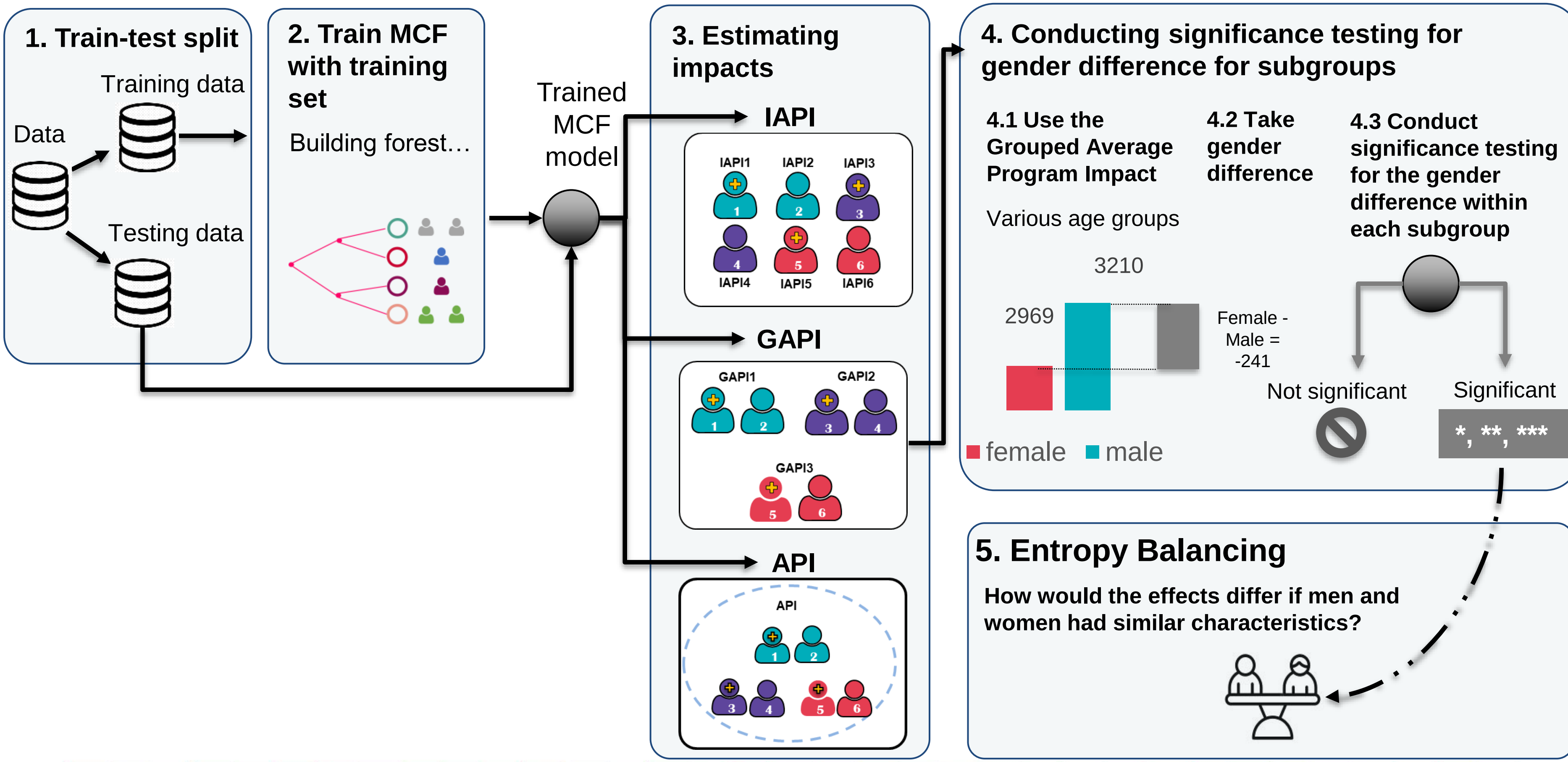


Represents the population average program impact.

Note: ✚ Indicates that the individual is a participant

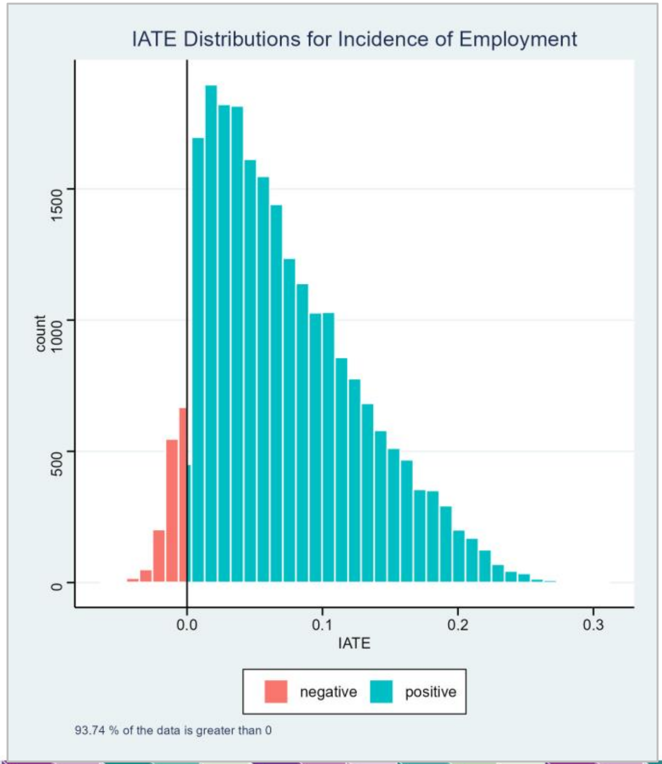
Methodology (cont.)

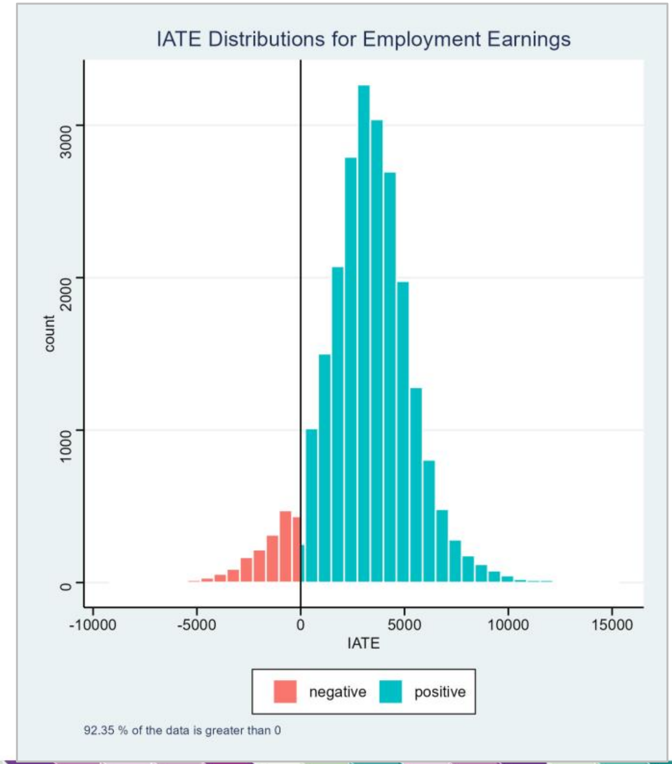
Using the results from the MCF, the methodology includes significance testing and entropy balancing to assess gender differences.



Examples of results – LMDA Skills Development Distribution of IATEs

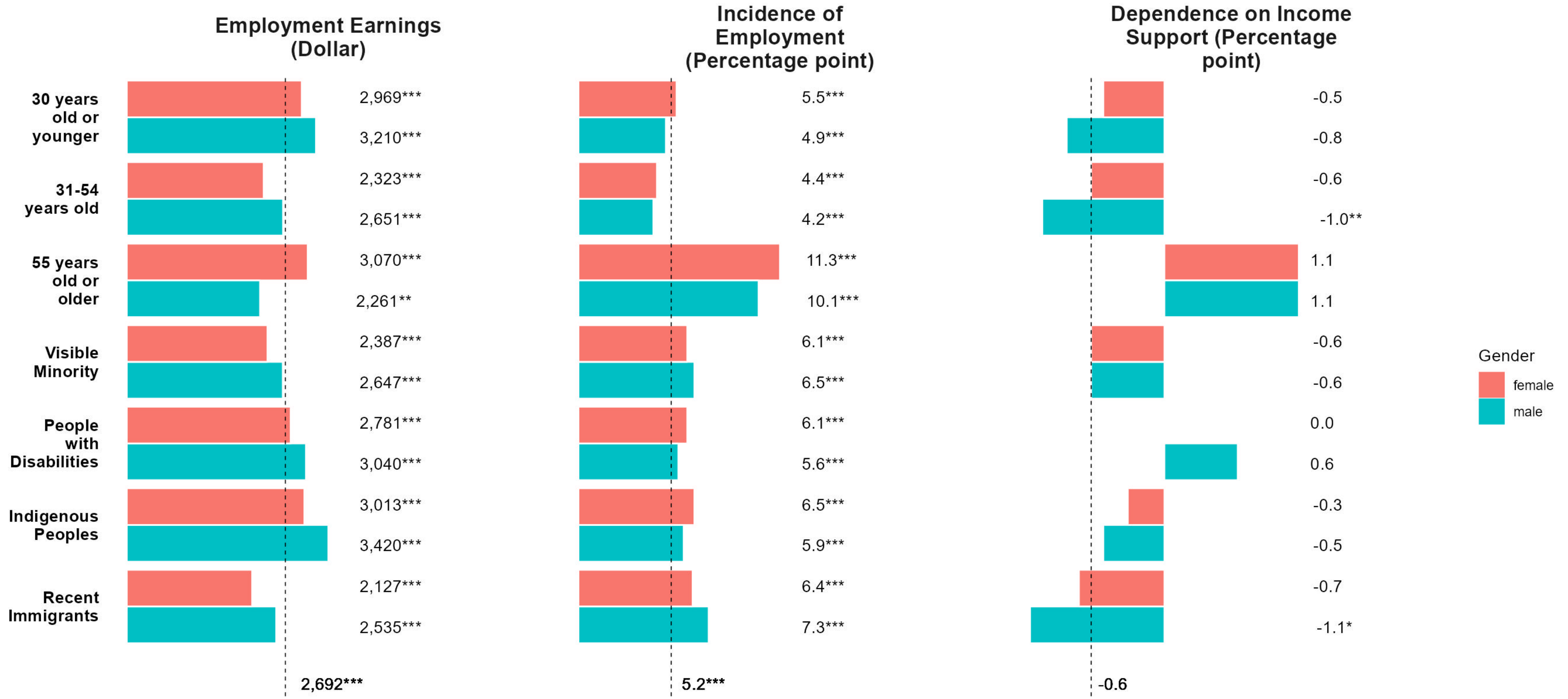
- The incremental impacts revealed that there is limited heterogeneity in program impacts.
- The majority of active EI claimant participants in SD benefited from it.
- The results indicate that:
  - 94% of participants experienced an increase in the incidence of employment
  - 92% of participants increased their employment earnings





Examples of results – LMDA Skills Development

Incremental impacts by gender and by other subgroups, 5-year post-participation period, annual averages



- All gender subgroups experienced an increase in their employment earnings and incidence of employment.
- Two groups saw statistically significant decreases in their dependence on income support:
  - Male participants aged between 31 and 54 years
  - Male recent immigrants

<u>Notation for significance levels</u>:

*** 1% level

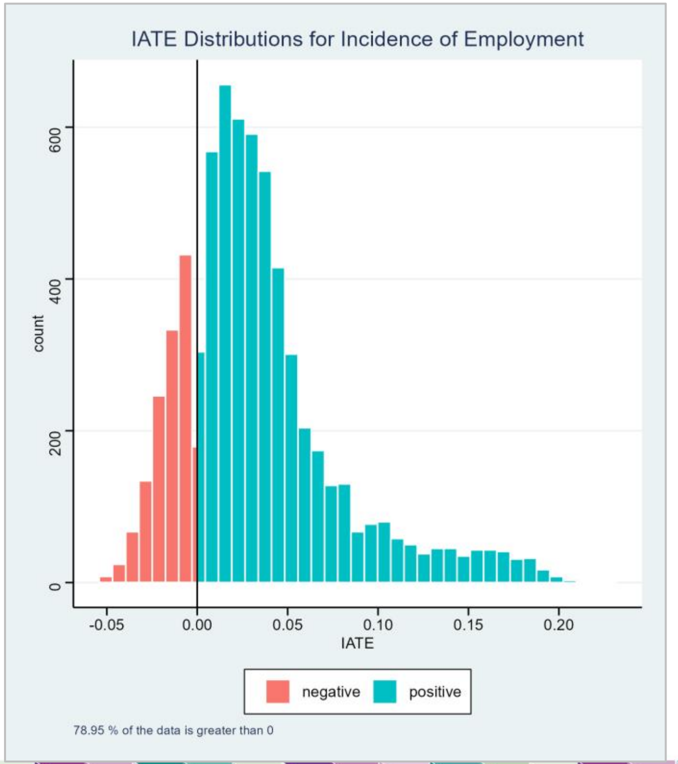
** 5% level

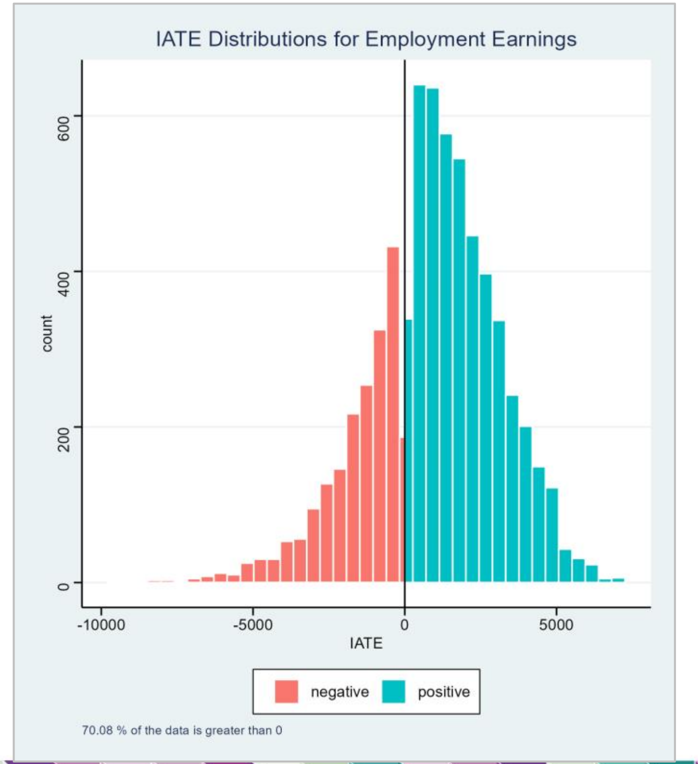
* 10% level

The overall average treatment effect on the participants annotated as the dashed line.

Examples of results – LMDA Targeted Wage Subsidies Distribution of IATEs

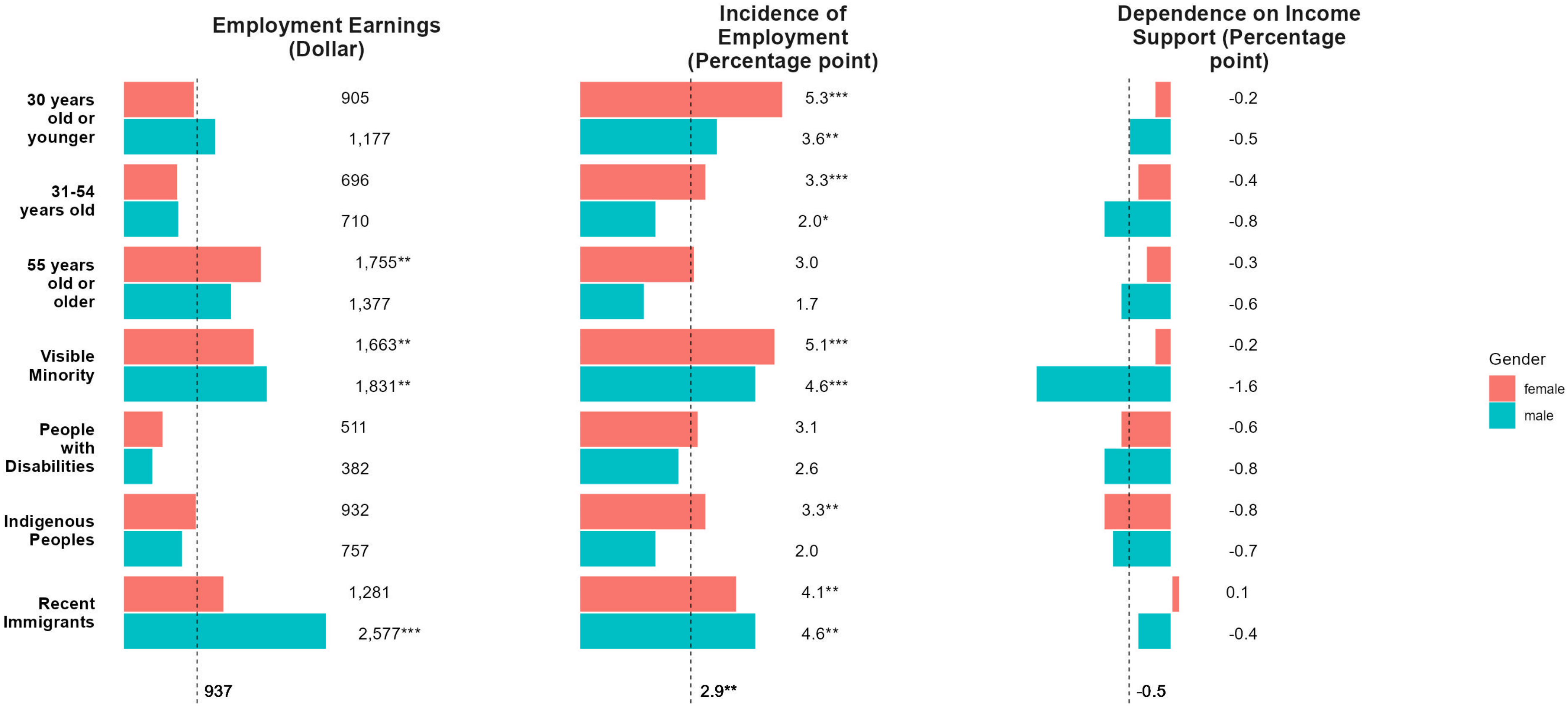
- The incremental impacts revealed that there is limited heterogeneity in program impacts.
- The majority of active EI claimant participants in TWS benefited from it.
- The results indicate that:
  - 79% of participants experienced an increase in the incidence of employment
  - 70% of participants increased their employment earnings





Examples of results – LMDA Targeted Wage Subsidies

Incremental impacts by gender and by other subgroups, 5-year post-participation period, annual averages



- Overall, all gender subgroups increased their incidence of employment. Two groups saw a larger increase in their incidence of employment and employment earnings:
  - Both female and male participants who were visible minorities
  - Male recent immigrants

Notation for significance levels:

*** 1% level

** 5% level

* 10% level

The overall average treatment effect on the participants annotated as the dashed line.

Examples of results – LMDA Targeted Wage Subsidies Entropy Balancing

Overall, we found no gender differences in the program impacts

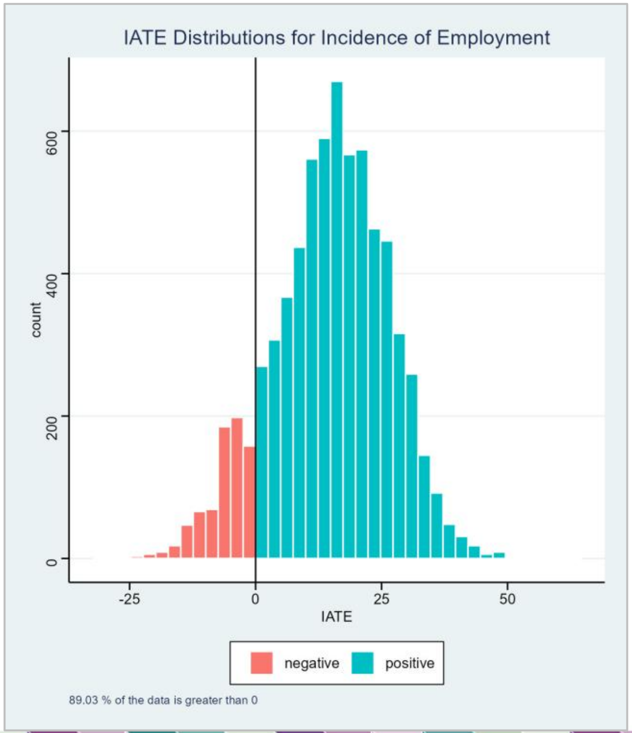
- Initial results for TWS suggested gender differences between men and women who were recent immigrants
  - Men increased their employment earnings by $1,296 more than women (statistically significant at 1%), which suggests a difference in program impacts.
- But after controlling for their socio-demographic characteristics, the differences became non-statistically significant, suggesting no difference in program impact.

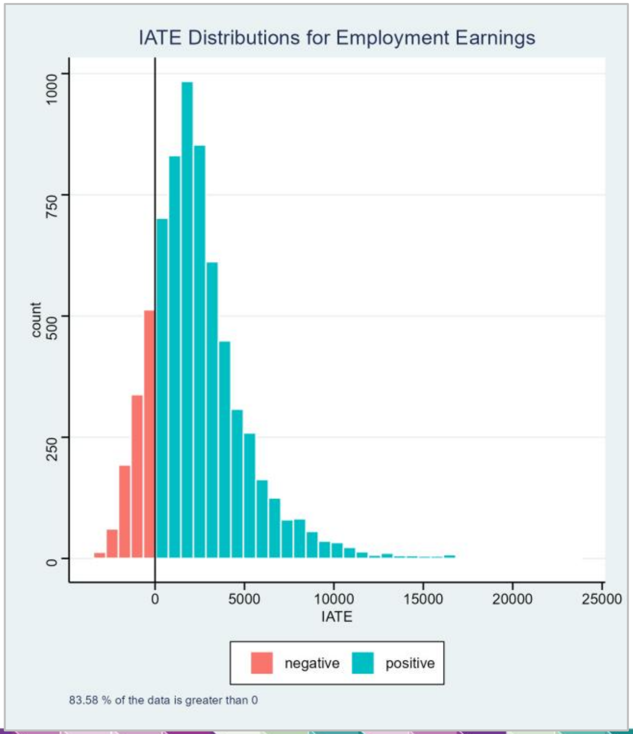
	Employment earnings (dollars)
Without controlling for socio-demographic characteristics	-1,296**
When men have similar socio-demographic characteristics as women	-328

Notation for significance levels: *** 1% level, ** 5% level, * 10% level.

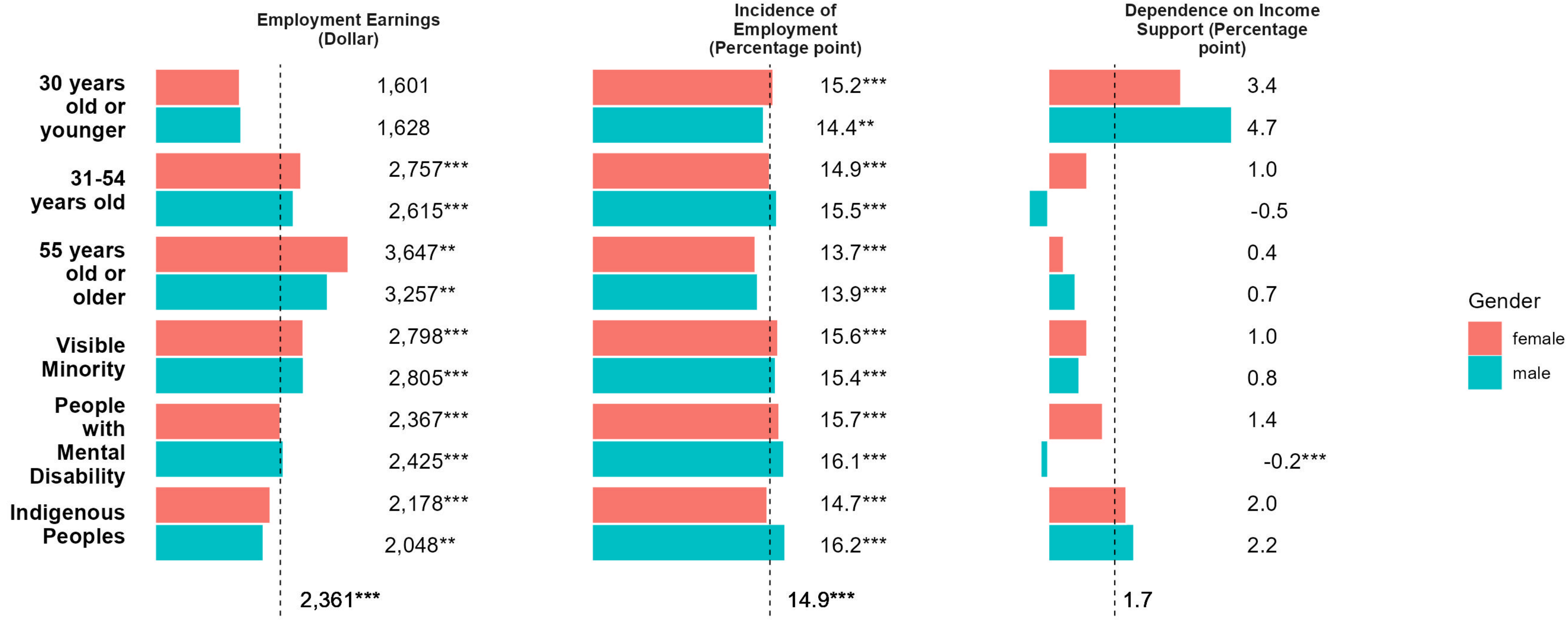
Examples of results – Opportunities Fund for Persons with Disabilities
Distribution of IATEs

- The incremental impacts revealed that there is limited heterogeneity in program impacts.
- The majority of participants in the Opportunities Fund (i.e., SD or TWS interventions participants) benefited from it.
- The results indicate that:
  - 89% of participants experienced an increase in the incidence of employment
  - 84% of participants increased their employment earnings





Examples of results – Opportunities Fund for Persons with Disabilities Incremental impacts by gender and by other subgroups, 5-year post-participation period, annual averages



- All subgroups of OFPD participants in SD and TWS experienced, on average, an increase in their employment earnings and their incidence of employment.
- The subgroups that saw the greatest improvement in employment earnings are both female and male participants aged over 54 years.

<u>Notation for significance levels</u>:

*** 1% level

** 5% level

* 10% level

The overall average treatment effect on the participants annotated as the dashed line.

Limitations

- This study was limited to the information available in administrative data:
  - Biological sex was used as a proxy for gender and data was not available for some GBA Plus factors of identity.
- Pre-existing differences might exist between participants and non-participants that were not measured during the matching process:
  - For example: ability, health, and motivation to seek employment.
- Results are not directly comparable between programs:
  - This analysis used comparison groups built by program intervention.
- The study does not capture participation in multiple interventions:
  - By using Action Plan Equivalents, the analysis attributed the longest intervention as the principal intervention in the unit of analysis.

Conclusion

The machine learning method was successful in generating robust results for key program interventions:

- Overall, results align with previous evaluations and provide a new level of granularity to examine program impacts through a GBA Plus lens.
- Note that complementary qualitative research and analysis would be required to contextualize these results. This could be done as part of future program-specific evaluation cycles.

However, machine learning is only possible with a large number of observations:

- Robust results were achieved by leveraging rich linked datasets, but machine learning did not produce statistically significant results for some interventions that had a relatively smaller number of participants.
- The propensity score matching method used in past evaluations remains a reliable tool that can perform well with smaller datasets.

Conclusion (cont.)

The study illustrates how:

- Quality participant-level data can be leveraged to examine program impacts through a GBA Plus lens.
- Machine learning can be more efficient than traditional methods to provide results at a granular level.
- Results can help understand the distribution of impacts on various groups and inform policy development and support program design from the perspective of “what works best for whom”.

Going forward:

- When only smaller datasets are available, the traditional matching method will remain the preferred method for conducting net impact analysis.
- Machine learning could be used in future evaluation cycles as a new line of evidence to explore differentiated impacts on subgroups when feasible. This method has the potential to provide new insights from a GBA Plus perspective.

References

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260.

Lechner, M. (2019). Modified causal forests for estimating heterogeneous causal effects.