Vrolijk, Ademir-Paolo – 9:13 AM
Are those 730k the total number of SMEs or the number of firms that received the treatment of BIGS funding?

A. **Feng, Yan –** 9:27 AM
   The core target population of the survey is 730,000 SMEs. A random representative sample size of 17,323 SMEs is taken from that population. The response rate of those 17,323 firms was 59.7 percent. Therefore, we end up with approximately 10,000 firms in the survey.

Yang, Doo DY [NC] – 9:33 AM
For PSM or entropy matching, did you select control groups matching those scores or scale the scores to make the control group comparable with the treatment group?

A. **Bousmah, Ibrahim** – 9:43 AM
   For PSM, I use the firm who have the closest distance of propensity score of each of the treated firms. For entropy balancing it is a different technique that achieve balancing through reweighting.
A. **Yang, Doo DY [NC]** – 10:20 AM
   Thanks, that was what I was curious about. The number of the control group, 467, matches after matching in PSM, while the number, 5414, remained the same in entropy matching.
   Would you see my understanding is correct?
   PSM allows to select control group that matches the treatment group, while the entropy allows to scale the covariates matching the treatment group.
A. **Bousmah, Ibrahim** – 10:40 AM
   Yes exactly, your understanding is correct :)

Rak, Anika (she, her | elle, la) (ISED/ISDE) – 9:33 AM
Could you describe propensity scoring in really simple terms please? :)

A. **Bousmah, Ibrahim** – 9:45 AM
   Propensity score is an index that estimate the likelihood of participating in a treatment based on the observed characteristics.
A. **Frenette, Marc (StatCan)** – 9:50
   To add to Ibrahim's response, if two individuals or two firms have similar propensity scores, this means that they had similar chances (or propensity) to participate in the program. In other words, the two groups are similar. In that case, a difference in outcomes will not be due to differences in the characteristics that were used in the propensity score index. This will then make us somewhat confident that the difference in outcomes is due to program participation. Our confidence will be higher with the more characteristics we can use in the analysis.
A. **Frenette, Marc (StatCan)** – 9:52 AM
   Since we hardly ever have access to al of the relevant characteristics that could affect the outcome, there could still be factors that could explain the difference in outcomes. This is the "matching on observables" issue: you can only match based on the data that you have, not on data you don't have, but could be relevant.
A. **Frenette, Marc (StatCan)** – 9:53 AM
   My final point on this is that matching (of any kind) is like a regression in the sense that we are "controlling for" (or "taking into account difference in") observed characteristics. However,

matching is more detailed than regression analysis. In this way, I like to refer to matching as "regression on steroids"

A. **Ward, Alicia** – 10:00 AM
   Thank you for the detailed explanation **:)**
A. **Frenette, Marc (StatCan) –** 10:04 AM
   I lied - I have another thought to add...there is a danger in applying matching estimators to small samples. If we set out to do propensity score matching by matching on the closest matches, we may force the model to match on bad matches. In other words, since the sample is so small, there may be no close matches, but if we don't account for this (say, by imposing minimum matching criteria), we could end up comparing individuals in the treatment and control that are very different.

Ward, Alicia – 9:50 AM
Which software tools are you using for propensity score matching? any packages for R you might recommend?

A. **Bousmah, Ibrahim** – 10:05 AM
   I am using the psmatch and teffect packages in STATA for propensity score, sorry I cannot recommend anything for R.
A. **Frenette, Marc (StatCan) –** 10:07 AM
   I can vouch that STATA is excellent, not only for matching, but for any statistical technique.
A. **Ward, Alicia** – 10:08 AM
   Thanks Marc, I've never used STATA but have lots of experience in R. How does it compare?
A. **Frenette, Marc (StatCan) –** 10:12 AM
   Since R is open source, I wouldn't be surprised if a routine has been developed to do matching. However, I have zero experience in R. Sorry, I'm from the SAS/STATA generation :)
A. **Bouwer, Charles** – 10:14 AM
   Propensity score matching with R: conventional methods and new features - PMC (nih.gov)
A. Torshizi, Mohammad (AAFC)
   I use Stata, but I have a paper in the handout that talks about how you can do this in R. I can't remember the name of the command that they use in R, but it's a package specifically for matching difference-in-difference with fixed effects.

Pereira, Brian – 10:10 AM
How long is a typical evaluation process from start to finish?

A. Torshizi, Mohammad (AAFC)
   I would say for us, anywhere between three to six months, from the time the clients give us a list of their participants. Usually, the program managers come to us with a list of participants in a specific program. We take that list to StatCan, they find those participants in the Business Register and then extract the data for them and data for firms that are in the same NAICS codes as those participants but did not participate. They then provide us with that data, and we go from there. From the time we receive a list of participants to the time we have a report it will be between three and six months, if things go well.

Al-Azzam, Mohammad MS [NC] – 10:11 AM
Would that possible to share the Stata Do file with the folks?

A. **Torshizi, Mohammad (AAFC)**
I'm not sure, because it's different for every project. I can give them to you but I don't think you'll be able to figure it out because there are so many filters going on within it. I'm happy to help you write the codes for any specific part of the project, but my own codes will not be useful to anyone. This is in part because I often write my own codes even when there is a package available.
It can also be difficult to install things in Stata on the StatCan network. Some of the packages are not available, but if there's something specific you're interested in, email me and I can give you parts of my code that I think are understandable, or I can tell you where you can find a code that you'll need, and how to make adjustments for your specific project. The codes are always project-specific, and they vary so much that you may get lost in the minutiae of things that are not related to propensity score matching, two-way fixed effects etc.

Bouwer, Charles – 10:56 AM
Good example of causal forests implementation: Causal Machine Learning for Econometrics: Causal Forests | by Haaya Naushan | Towards Data Science

Al-Azzam, Mohammad MS [NC] – 11:03 AM
Can the causal forest still be applied without a the existence of a policy intervention?

A. **Liu, Yu Hsien YHL [NC]** – 11 :15 AM
"treatment" can be defined differently. For example, use of new medicine on patients in the medical field.

Arsenault, Christiane C [NC] – 11:12 AM
Gender-Based Analysis Plus Exploratory Evaluation Study on Selected Labour Market Programs - Canada.ca

Lawrence, Lola (StatCan) – 11:14 AM
These are very great case studies and applications in program and policy interventions. For this particular case study, are there any limitations using small sample size or the effect of groupings?

A. **Arsenault, Christiane C [NC]** – 11:34 AM
Small sample sizes may limit the generalizability of findings to larger populations. In this case study, we have thousands of participants in Opportunities Funds for Persons with Disabilities program and tens of thousands of participants in the Labour Market Development Agreement program.
Groupings could potentially be affected by small sample size if the confidence intervals are wide due to small number of individuals in the group with extreme estimates.
It's important to acknowledge these limitations and interpret the results cautiously, considering the potential impact on the validity and reliability of the findings.

Liu, Yu Hsien YHL [NC] – 11:31 AM
MCF incorporates techniques to mitigate selection bias through its algorithm design and estimation strategy. Here are some ways MCF deals with selection bias:

- **Propensity score adjustment:** MCF can incorporate propensity scores, which estimate the probability of receiving the treatment based on observed covariates, to balance the distribution of covariates between the treatment and control groups. By adjusting for propensity scores, MCF can reduce selection bias by ensuring that treated and untreated units are comparable in terms of observed characteristics.
- **Weighting methods:** MCF can use weighting methods, such as inverse probability weighting (IPW), to give more weight to observations from underrepresented groups in order to balance the distribution of covariates between treated and control units. This helps to mitigate selection bias by accounting for the differential probabilities of treatment assignment.
- **Stratification:** MCF can divide the sample into strata based on observed covariates and estimate treatment effects separately within each stratum. By stratifying the sample, MCF can account for differences in treatment effects across different subpopulations, thereby reducing the impact of selection bias.
- **Cross-validation:** MCF typically employs cross-validation techniques to assess the performance of the model and ensure its generalizability. By splitting the data into training and validation sets, MCF can evaluate how well the model generalizes to new data and identify potential sources of bias in the estimation process.

Tohon, Aurelas AB [NC] – 11:34 AM
Overall, MCF employs various techniques, including propensity score adjustment, weighting methods, stratification, and cross-validation, to address selection bias and produce more accurate estimates of treatment effects. These methods help to ensure that the estimated effects reflect the true causal impact of the treatment, even in the presence of selection bias.