With Toronto Centre By-election approaching, polling results suggest NDP needs to appeal more to ethnic minorities, young voters.

Haili Su, Jiaxiang Miao, Xingyu Yu, Minghui Chen

2020-10-07

Executive summary

Petit Poll has been asked by the federal New Democratic Party (NDP) to survey Toronto Centre federal electoral district ('the riding') to understand whether people are likely to vote for NDP candidate in the upcoming by-election on October 26th. We surveyed 1,061 eligible voters of this by-election, who is a Canadian over the age of 18 and is living in Toronto Centre riding, and asked their options regarding this by-election and demographic information, including gender, age, ethnicity and highest educational level, in order to analyze how these factors effect personal political preferences. The survey method is simple random sampling without replacement (SRSWOR), and we reached potential respondents via phone calls and mailed letters due to the COVID-19 pandemic. We then created and stimulated the dataset supporting our later analysis based on the real-world population parameters. The supporting rates for NDP, Liberal Party of Canada (LPC) and Conservative Party of Canada (CON) in our stimulated results are 23.3%, 26.5%, and 21.3% respectively, indicates that the NDP candidate will have a chance to win the by-election this time while the Liberals and the Conservatives would be the strong contenders. The demographic session in our report shows that young adults under the age of 34, white Canadians and individuals without a secondary diploma have a slightly larger chance to support NDP than other demographic groups. However, no major differences for voting patterns among demographic groups are observed overall. Meanwhile, ethnical minorities and young voters under the age of 44, who are non-NDP-leaning voters, show high tendency to switch to NDP (i.e. likelihood to choose NDP instead >7 for Question 3 in our survey). We highlights these findings because these data suggests NDP should appeal more to these voters to maximize NDP's chances for winning. But, our stimulated results are not perfect because non-response problems, response errors and systematic errors, which lead to the lack of reliability and generality, may exist at some levels through our study period. To minimize these bias, we will improve our survey methods in the future studies by trying zoom (during the outbreak)/ in-person interviews, training our interviewers to have all essential skills for interviewing, and increasing incidence rates. Our future studies will be focused on the analysis of NDP-leaning voters' reasons for supporting NDP and how these reasons are influenced by their personal background. These studies will be planned to give NDP better suggestions on how the NDP candidate should appeal for potential voters.

Introduction

A by-election for the House of Commons seat representing the Toronto Centre federal electoral district ('the riding') is set to be held on Oct 26, 2020, after the former MP of Toronto Centre and Minister of Finance Bill Morneau's resignation. Since 2008, the New Democratic Party (NDP) candidates have been close runner-up in this riding, which is in the downtown of the largest city of Canada, with the highest population density among all 338 ridings (338Canada, 2020).

This survey is an opinion poll on Toronto Centre voters' preferences over which party to vote for. For people who tend to vote for the NDP, we are also asking how determined they are about this choice. For people who tend to vote for another party, are undecided, or do not intend to vote, we will ask how likely they will switch to voting for the NDP candidate. We believe the results of this survey can aid the NDP to better evaluate the party's chance of winning this by-election and which demographic group they should appeal to more.

For the purpose of data collecting, we will use Simple Random Sampling Without Replacement method and collect at least 1061 responses. A more detailed account of our methodology is provided in the following section. The details of our survey questions and design can be found in the appendix.

According to the 2016 census data, this riding has 73,675 Canadian citizens over 18, (33,970 women and 39,705 men). 49.7% of the riding population is of visible minority backgrounds while 1.5% are of indigenous backgrounds (not counted in the visible minority) (Statistics Canada, 2017). Based on the census data, we estimated population parameters. With these parameters and with the consideration of the electoral history of this riding and the prediction of 338canada.ca, we simulated 1061 responses to our survey and analyzed the simulated results to show what kind of insights our survey and our team of statisticians could provide to the NDP campaign. The simulation is based on the scenario where the Liberal Party leads in the poll with NDP being a close runner up. With the results, demographic information, and respondents' indication of their likelihood of switching to/from NDP, we make the suggestion that NDP has a stronger chance of winning this by-election than before, and should appeal more to ethnic minority and younger constituents in this riding.

We used R language (R Core Team, 2020) for simulation and data analysis, and the report is written in a R Markdown (Allaire et al., 2020) along with packages including tidyverse (Wickham et al., 2019), tibble (Müller & Wickham, 2020), ggplot2 (Wickham, 2016), knitr (Xie, 2020) and gridextra (Auguie, 2017). The Github repo link for code and dataset supporting our analysis is available here: https://github.com/MingHuiChen98/Petit-Poll.

Survey methodology

For the purpose of our survey, we identify the target population as all the eligible voters in this by-election, i.e. all the people who are 18 years or older Canadian citizens and reside within the Toronto Centre riding on the day of the election ($\hat{N} = 73,675$, estimated by the number of Canadian citizen over 18 in this riding in 2016 census data (Statistics Canada, 2017)). Our frame will be the voters who have registered as residing in this riding according to Elections Canada voter registry. Our sample population will be the eligible voters randomly sampled from the voter registry who completed our survey (n = 1061).

For our sampling method, we chose Simple Random Sampling Without Replacement (SRSWOR), which means we randomly sample 1061 individuals on the voter registry, with every voter having the same probability of being selected. This method, while basic, does give a sufficient representation of the population when sample size is large enough, and is easier to design given that by-election date is approaching. Also the list of voters that is our frame is close to include our total population, while our population size is moderate. Our minimum required sample size is calculated based on the proportion of women in this riding. According to 2016 census, about 46.1% of the adult citizen population are female. With margin of error being 3% and a 95% confidence, our minimum n is calculated by

$$1. \quad n = \frac{\hat{p}(1-\hat{p}) * z_{\alpha}^2/2}{d^2} = 1060.28$$

, thus we should have at least 1061 respondents.

Based on the method we chose, for any parameters, the population mean μ_y will be estimated by the sample mean $\bar{y}$:

$$2. \quad E(\bar{y}) = \mu_y$$

and our variance estimate for $\bar{y}$ will be:

$$3. \quad v(\bar{y}) = (1 - \frac{n}{\hat{N}})\frac{s_y^2}{n}$$

with s_y being our sample standard deviation. (Wu & Thompson, 2020)

Giving the situation of COVID-19 pandemic in Toronto, it would be unsafe to have field workers interviewing voters on the ground. Therefore we will try to match as much sampled voters as possible to a phone number in public records and try to reach them by phone. For the rest of the voters whose numbers are unable to reach or be found, we will send a letter to their address on the registry to invite them to participate in our polls.

In the situation of a non-response, we will randomly sampled another voter on the registry with the same address if the it is an apartment building, or the same first three digits of the postal code if otherwise. In the event of a person we sampled no longer being eligible due to lags in voter information update (e.g. moved or deceased), it will be treated the same as a non-response.

To protect the privacy of the respondents, their personal information will only be accessed when needed. For example, when searching for phone numbers, we have two group of staffs: one searching based on the person's name, and another searching by the address. The callers will only know the person's name and age and the number will be dialled by the system. In our database they will all be assigned an ID number for tracking.

The estimated total costs for our survey are CAD$42,440, with an average cost of 40 dollars per completed respondent, based on Lee (2002)'s article and wages in Toronto Centre riding. We will use Microsoft Forms, as a free resource for NDP employees, to design our survey questionnaire. As we will send our survey via both phone and mail, the incidence rates are set to be about 90% (i.e., out of 100 surveys we send, 90 individuals respond and complete the survey). Our plan is to hire six workers and pay them 26 dollars per hour for tracing and interviewing each selected individual. The whole process for one completed submission of our survey is expected to spend one interviewer about 1.15 hours. We will then ask for another 10,000 dollars to pay for the telephone and postage charges, data management and analysis, and the list of mailing address and phone numbers.

Results

As we mentioned in the introduction, we simulated the following results of a sample of 1061 respondents based on real-world population parameters. The codes of simulation, along with the specific weightings and their sources, can be found in the GitHub repo mentioned in the introduction section. The ethnicity categories are based on the 'Visible Minority' and 'Aboriginal Identity' section in the 2016 census (Statistics Canada, 2017). The results, regarding the voting patterns among demographic information, have been extracted and analyzed.

Table 1: Polling results for Toronto Centre By-Election on Oct 26, 2020

Party	n	Support %
CON	226	21.30
GRN	107	10.08
LPC	281	26.48
NDP	247	23.28
Not decided	2	0.19
Not Voting	93	8.77
Other/IND	105	9.90

Firstly, our simulated polling results, seen in table 1, show that the four main federal parties, the Liberals (LPC), NDP, the Conservatives (CON) and the Green (GRN) each has support rate of 26.48%, 23.28%, 21.30% and 10.08% respectively. Comparatively speaking, we can conclude that NDP has its advantage to win the election. However, LPC and CON will be strong contenders to NDP. It's also worth noticing that

8.77% of the voters said they are not going to vote, while 9.90% of the voters said they are voting for a non-major-party candidate, comparing to the extremely small proportion of undecided voter (0.19%).

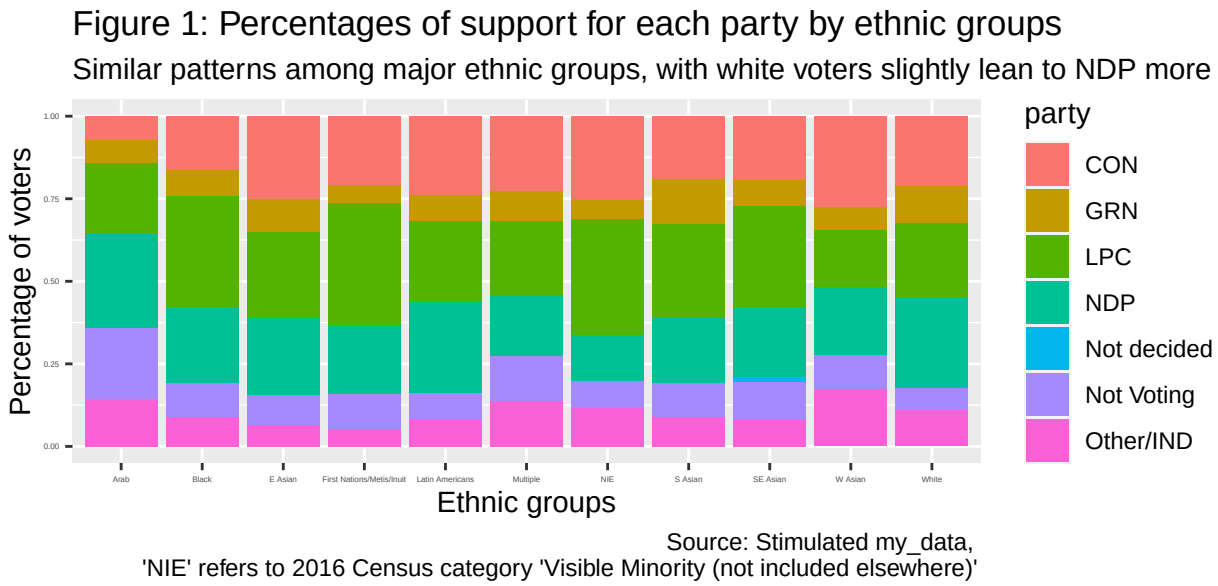
Does certain demographic group have different voting patterns than the others? Here are the voting patterns by different demographic groups:

Ethnicities

Table 2: Ethnical group representations in the sample

Ethnicity	n	Proportion %
Arab	14	1.32
Black	79	7.45
E Asian	123	11.59
First Nations/Metis/Inuit	19	1.79
Latin Americans	25	2.36
Multiple	22	2.07
NIE	86	8.11
S Asian	178	16.78
SE Asian	62	5.84
W Asian	29	2.73
White	424	39.96

Table 2 is an overall look at how each ethnic group is represented in our sample. As we can see, five ethnic groups (Black/African, East Asian, South Asian, Southeast Asian, and White/European), and 'NIE', which means not included elsewhere (i.e. in other ethnic groups), are represented with more than 3% in our samples. White voters, the largest voter group by ethnicity, compose 39.96% of our sample, following with South Asian (16.78%) and East Asian (11.59%) voters.



Here we investigated the distribution of party support by ethnic groups. From fig. 1 we can see among the ethnic groups that compose more than 3% in our sample, in this riding, voting patterns are similar. Although we can also see that white voters lean to NDP more (actual support rate 27.1%) with the Liberals, favored by other ethnic groups, tightly following (support rate 22.6%). The East Asian voters, while favoring the Liberals more, have the highest support for the Conservatives (the actual support rate in East Asian

voters are: 26% for the Liberals, 25.2% for the Conservatives & 23.6% for NDP, all within 3% margin of error). Groups that comprise less than 3% of our samples separately include voters with Arab, Indigenous, Latin American, West Asian, or multiple ethnic backgrounds who may have shown different voting patterns. However, the sample sizes are insufficient for conclusive analysis.

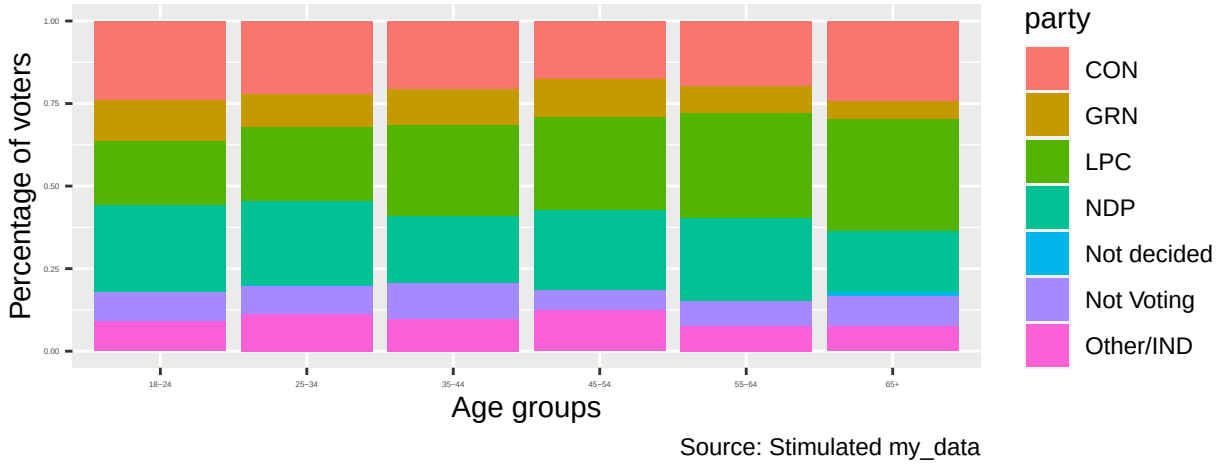
Age groups

Table 3: Age group representations in the sample

Age group	n	Proportion %
18-24	134	12.63
25-34	274	25.82
35-44	267	25.16
45-54	147	13.85
55-64	132	12.44
65+	107	10.08

Here in table 3 we can see the age group compositions in our sample. Overall this is a young riding with voters between 25-44 composed more than half of our sample, while senior voters (over 65) are 10.08% of the sample.

Figure 2: Percentages of support for each party by age groups



From the figure 2, which is the distribution of party support by age, it is found that slightly more younger individuals tend to support NDP, especially in the age group of 18-24 (26.1%) and 25-34 (25.2%), while the least NDP support is in 65+, which is 18.7%. In terms of the Liberals and the Conservative, they got a huge proposition of supporters at the age between 25-34 and 35-44 as well. It seems like those two parties are still competitive in this riding. However, the voting patterns among age groups don't diverge greatly.

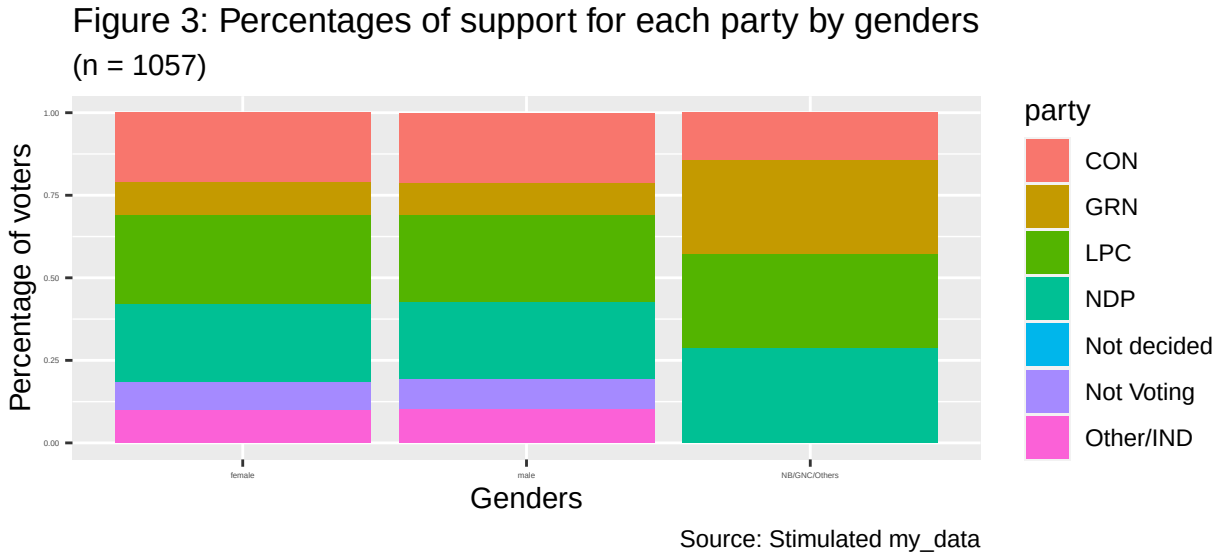
Genders

Table 4: Gender representations in the sample

Gender	n	Proportion %
female	471	44.39

Gender	n	Proportion %
male	579	54.57
NB/GNC/Others	7	0.66
Not to say	4	0.38

From table 4 we can see female voters compose 44.39% of our data, while male voters comprise 54.57%. Non-binary and Gender nonconforming voters, along with voters of other gender identities, compose 0.66% of our sample. 4 respondents (0.38%) chose not to disclose their gender identities.



There is no significant distinction in voting patterns between female and male voters (Figure 3), while the sample size of voters of other gender identities is far too small for any conclusive analysis.

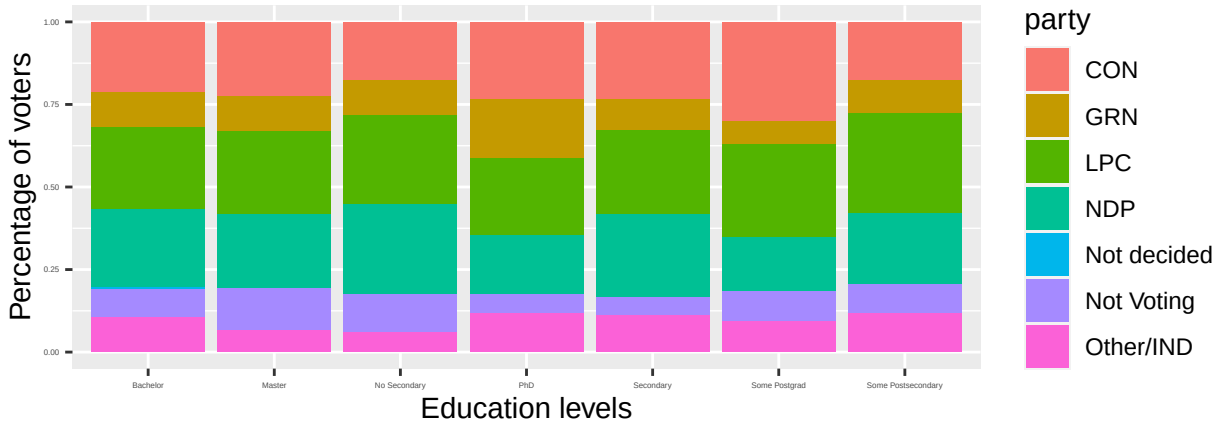
Education levels

Table 5: Education level representations in the sample

Education level	n	Proportion %
Bachelor	315	29.69
Master	151	14.23
No Secondary	85	8.01
PhD	17	1.60
Secondary	236	22.24
Some Postgrad	43	4.05
Some Postsecondary	214	20.17

In table 5 we see most of the voters in our sample have received at least some postsecondary education and people with bachelor's degree(s) are the largest voter group (20.69%) by education level. There is also a significant portion of voters who have a high school diploma but no postsecondary education (22.24%).

Figure 4: Percentages of support for each party by education levels
(n = 1061)

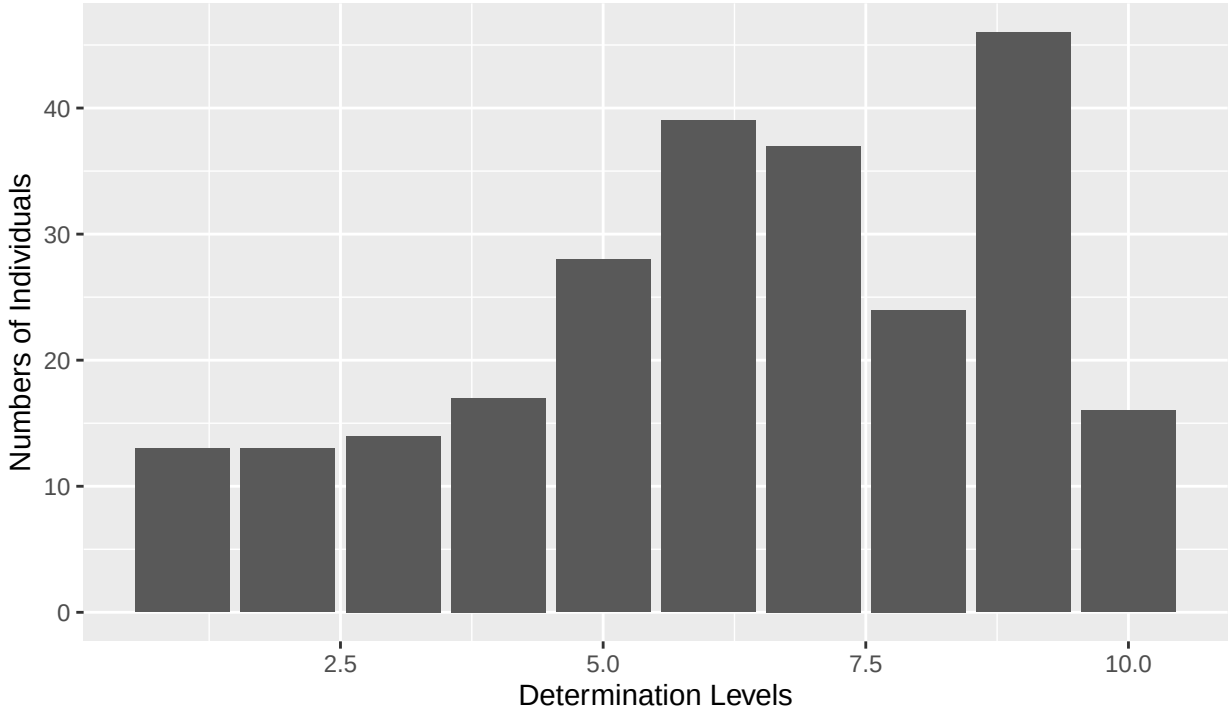


Source: Stimulated my_data

From figure 4, we compare the education level of different parties' supporters, people with some college education below bachelor level showed the highest support for LPC (30%), people with no secondary education diploma shows the highest support for NDP (27.1%) and least support for CON (17%), but what is worth noticing is that their support for the LPC is the same as for NDP.

Determination level of NDP voters

Figure 5: The determination level of NDP−leaning voters
(n = 247)



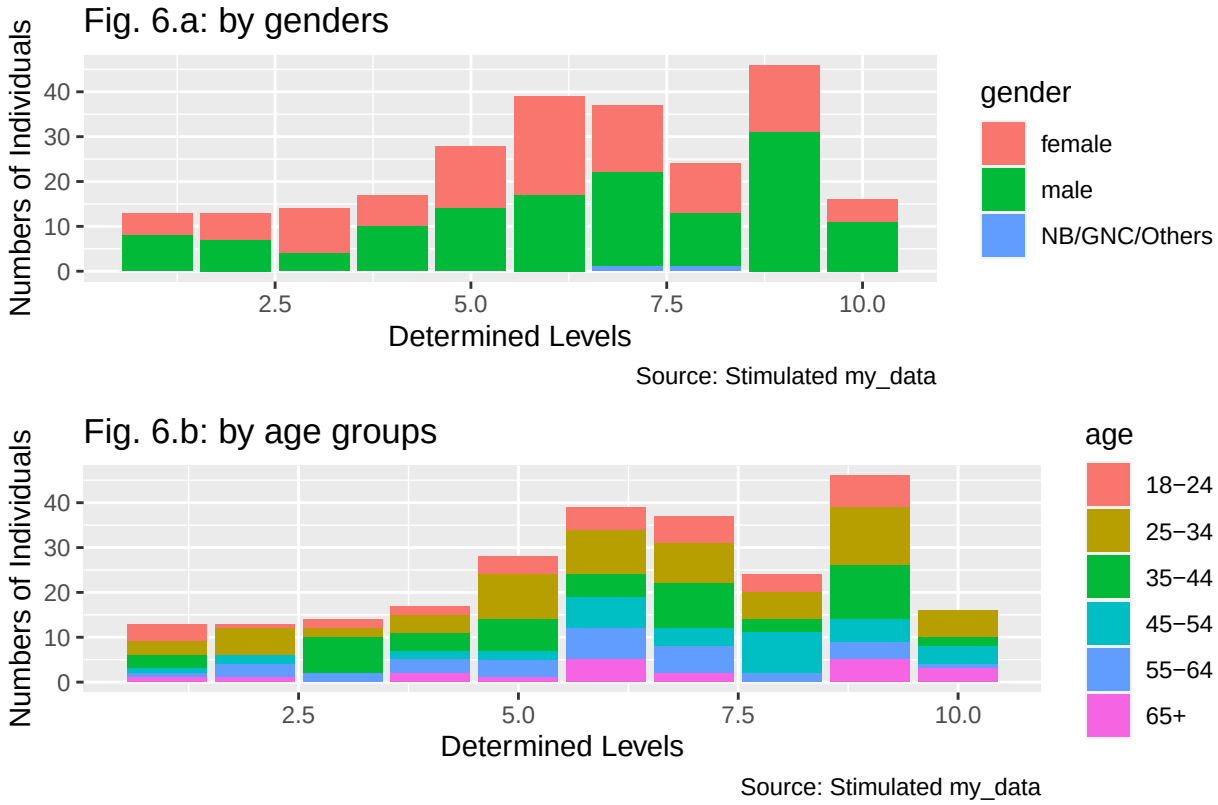
Source: Stimulated my_data

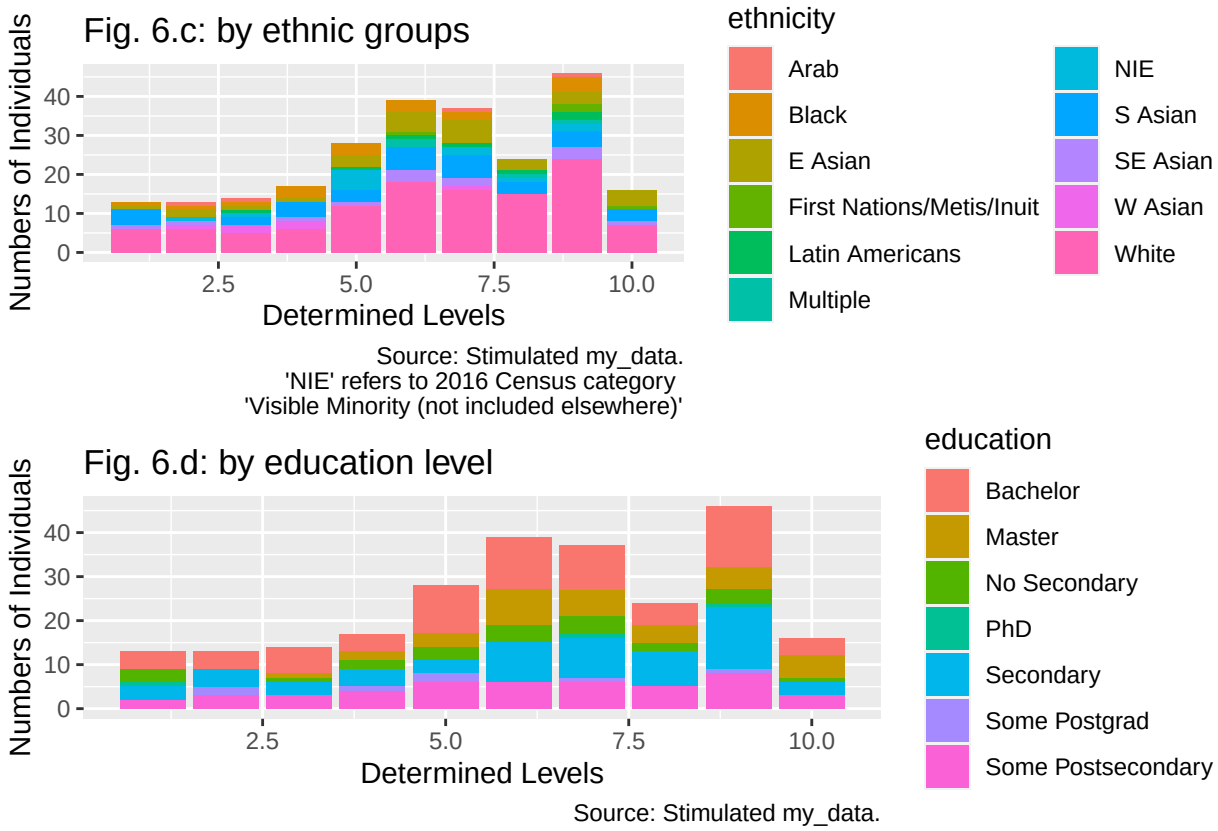
Table 6: Estimate mean and variance of levels of determination in NDP voters

mean	variance
6.267207	0.9980214

We also survey the determination level of NDP-leaning voters. From fig. 5 we can see most of these voters have a higher (>6) level of determination on their choice. Table 6 shows the estimated mean and variance of NDP voter's general determination level calculated by equations 2 and 3 in the methodology section.

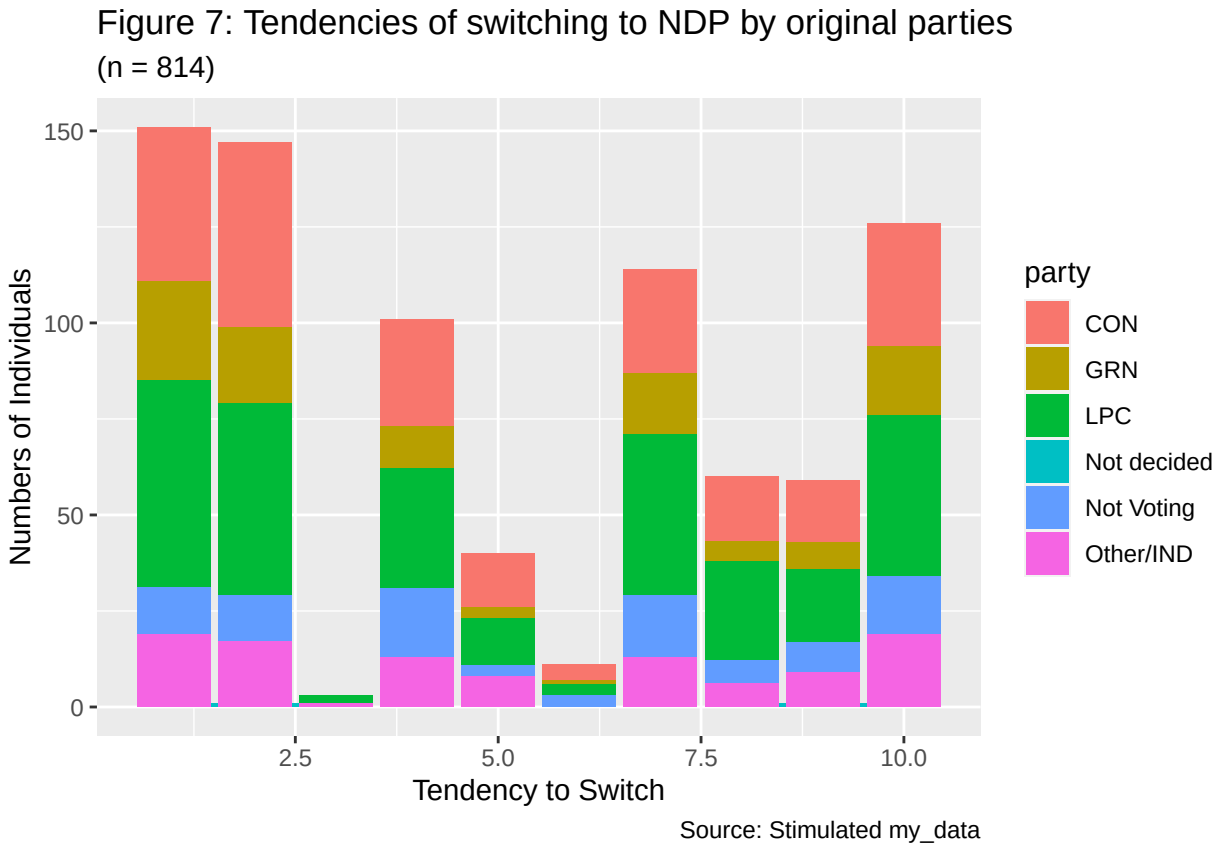
Figure 6: The determination levels of NDP voters by demographic groups





Meanwhile, none of the surveyed demographic information, age, highest education level, ethnicity, and gender, shows a voting pattern on the levels of their confidence to choose NDP as the final decision at the time of this by-election (Figure 6).

Likelyhood of switching to NDP for non-NDP-leaning voters



From Fig. 7 we can see the tendencies of other-party-leaning voters switching to NDP are unevenly distributed: these voters either show almost no interests or high interests in NDP. Among those who are more likely (>7) to change their votes, most of them are Liberal-leaning, but it's the same for those who have almost no interest in NDP (<2). It is probably due to the fact that this riding generally favors the Liberal. However, we can see a higher portion of Conservative-leaning voters in the <2 likelihood groups. Here table 7 shows the estimated mean and variance of the said likelihood of each party's voters, calculated by equation 2, 3 in the methodology section:

Table 7: Mean, Sample variance and estimated variance on other party voters' tendencies to switch to NDP

Party	Mean	Sample Variance	Estimated Variance
CON	5.004425	10.759980	6.90e-06
GRN	4.915888	12.058896	1.62e-05
LPC	5.177936	11.082511	5.70e-06
Not decided	5.500000	24.500000	1.75e-03
Not Voting	5.623656	9.845956	1.52e-05
Other/IND	5.314286	11.429121	1.57e-05

From table 7 we can see that the average likelihood for other voters to switch to NDP is around 5, regardless of origin party. Although the sample variance is large, the adjusted estimated population variance is very small.

Overall observations

Overall, it seems a typical NDP voter will be a white, young person, between 18-34, with no secondary education diploma. However there's no major distinction in voting patterns among demographic groups.

In terms of NDP-leaning voter's determination, most of them are relatively determined to vote for the NDP, with an average level of determination of 6 (1-10 scale, $v(\bar{y}) = 1.00$, table 6). Meanwhile, the likelihood of other party-leaning voters switching to NDP is generally neutral with most of the average likelihood around 5 (table 7). Though the sample suggests it's because of the even distribution between the two extremes, estimated population variances suggest otherwise.

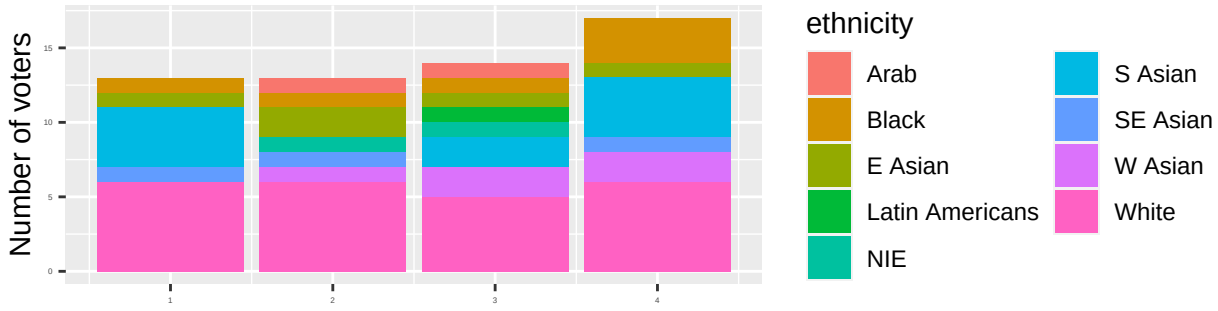
Discussion

338Canada.com categorized Toronto Centre as "LPC safe" with 56.2% of the projected popular vote on Oct 4, 2020, while 24.7% for the NDP (338Canada, 2020). Our simulated poll results, however, show a tight race with 26.48% of votes for the Liberal and 23.28% for the NDP (table 1). Since our margin of error is 3%, this means with 95% confidence, NDP and the liberal have almost the same chance of winning the popular vote. While the Liberals still hold the upper hand on the popular vote projection, the survey results are a stark contrast to the double-digit lead the Liberals enjoyed from May to early July. Allegations of nepotism and conflicts of interest surrounding the WE Charity may not have brought down the government's vote share in the way opposition parties would have hoped, but it did take a toll on the Liberals' numbers. In our survey, we try to use Simple Random Sampling Without Replacement (SRSWOR) method to collect at least 1061 respondents trying to figure out the rate of how many support NDP or other parties. We found in terms of ethnicity, white voters are slightly more likely to vote for the NDP, and in terms of age groups, younger voters are leaning slightly more for the NDP. However, there seems to be no major differences in voting patterns among demographic groups.

How about the reluctant NDP supporters? Here in fig 8, we can see the ethnic composition of those NDP supporters with a determination level less than 5:

Figure 8: More low−determination NDP supporters are ethnical minority

Ethnic group distribution of NDP voters with determination level < 5



Source: Stimulated my_data.
'NIE' refers to 2016 Census category
'Visible Minority (not included elsewhere)'

Although according to our survey, most of the NDP supporters are quite determined about their choice, among those not that determined, most of them are of ethnic minority background (though it shouldn't be interpreted as minority NDP supporter are generally less determined). This, along with the fact that white voters showed a stronger tendency of voting for NDP, may suggest that NDP needs to find a better way to appeal to the minority population in Toronto Centre.

How about those more likely to switch to the NDP? Here fig 9 shows the age group composition of those non-NDP voters who indicate their likelihood of switching to NDP greater than 6:

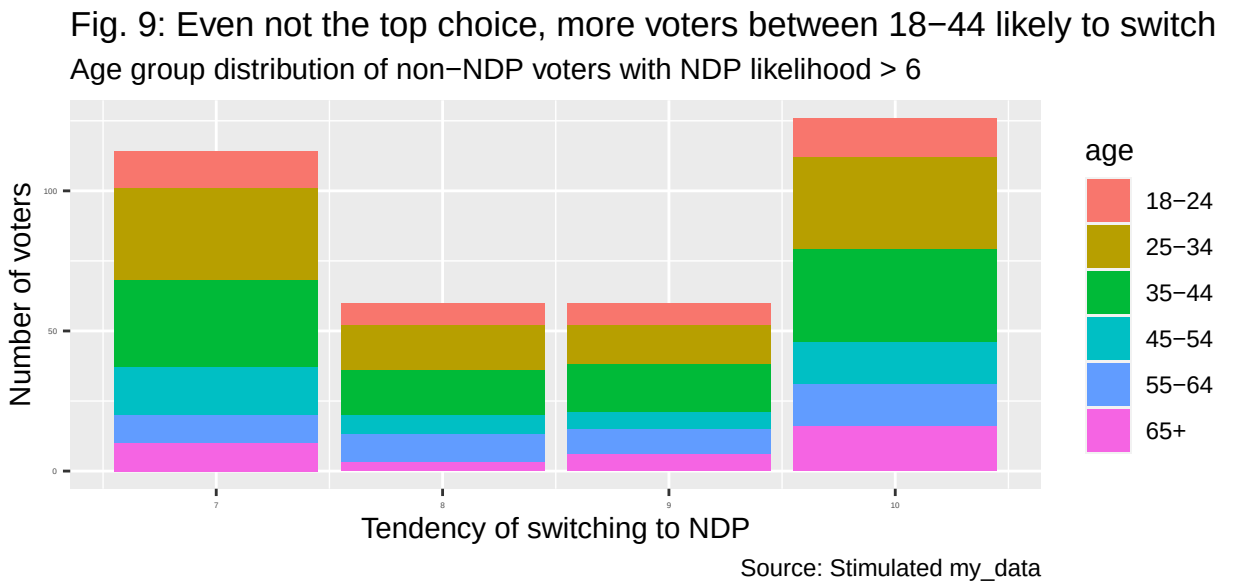


Figure 9 shows that even when NDP is not their top choices, more of those who are leaning to switching to the NDP are among the 18-44 age group. Younger voters are already the strongest base for NDP (fig 2). These data suggest that NDP should invest more campaign resources on younger voters, and convince those in the 18-44 age group who are on the verge of switching to NDP.

In conclusion, NDP has a much stronger chance of winning in this by-election according to the polling results. Our stimulated polling results also suggest the demographic group that the NDP should invest more resources on, including ethnical minorities and younger voters.

Weakness and areas for future

After our investigation, we have the following weaknesses that should be avoided and minimized on the further studies. Firstly it lacks reliability. Without face to face connection due to COVID-19 pandemic, it is actually hard for our interviewers to determine whether the respondents provide true information and opinions. Besides, people might feel offend or unhappy to answer their real thoughts by psychological effects. There inaccurate responses lead to response errors in our analysis. Secondly, it is also lack of generality. Since we conducted the survey throughout the phone, some potential respondents probably missed the phone or did not get access to the phone. Even if we sent a letter by mail for the other group of selected voters, it was also possible that they might miss the letter. Analysis bias due to non-responses problems may also exist as certain individuals might feel uncomfortable to answer some questions and refused to answer the survey. Lastly, our interviewers might be misled by some respondents and recorded the wrong opinions for their survey due to various factors (e.g. languages). All of above could lead to the non-sampling errors by providing non-response problems, response errors, and interviewer bias. In the future, we will improve our survey method by trying increase incidence rates and motivating their willingness to respond our survey by providing some benefits. Also, we would do zoom interviews and train our interviewers in terms of both response problems and interviewer bias.

Our area of future would be focused on NDP-leaning voters. It will be an interesting study if we can get to know the reasons why they support NDP so much. In other words, we can focus on this part and find the relationship between them and those groups that they do not support NDP. This could help with getting higher supporting rates and possibility of winning the election for NDP in the future.

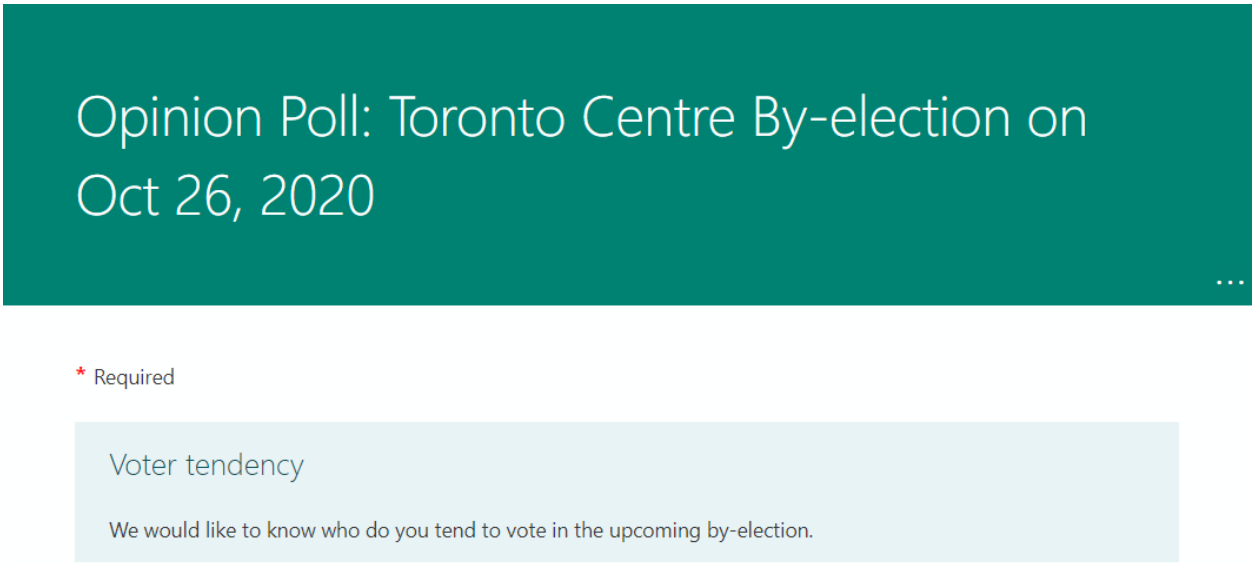
Appendix

The details of our survey questionnaire are provided in this section.

Survey

Here is the link for our survey questionnaire: https://forms.office.com/Pages/ResponsePage.aspx?id=JsKqeAMvTUuQN7RtVsVSEPBBuir1EfRAm1OEhOL5TIJUOURDTUxQVDU5V1MyOENQNjVJU0FKVFVFSS4u.

We designed our survey to have a title and brief description for introducing our purpose in the study.



As Question 1 is a screening question, respondents who choosing the option "no" would not get access to the rest of survey.

1

Are you, or will you be on Oct 26, 2020, an eligible voter in this by-election?
(An eligible voter is a Canadian citizen AND 18 years or older on Oct 26, 2020 AND residing in Toronto Centre riding (The official Elections Canada map of the riding boundaries is shown here). An eligible voter will need to register in Elections Canada in order to vote.) *

- ○ Yes
- ○ No

Next

Question 2 is designed to collect the voter tendency for party selections.

2

If the election is being held today, which party's candidate are you intending to vote for? *

- ○ Green Party of Canada
- ○ New Democratic Party (NDP)
- ○ I am not intending to vote
- ○ Other/Independent
- ○ Conservative Party of Canada
- ○ Liberal Party of Canada
- ○ I am not sure

Next

For individuals choosing NDP as their determined party in Question 2, they will get a Question 3 asking how determined they are in voting for the NDP candidate.

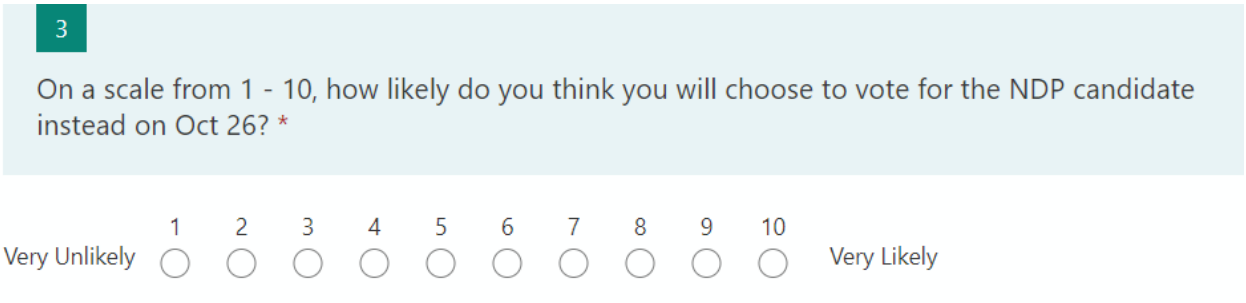
3

On a scale from 1 - 10, how determined are you in voting for the NDP candidate? *

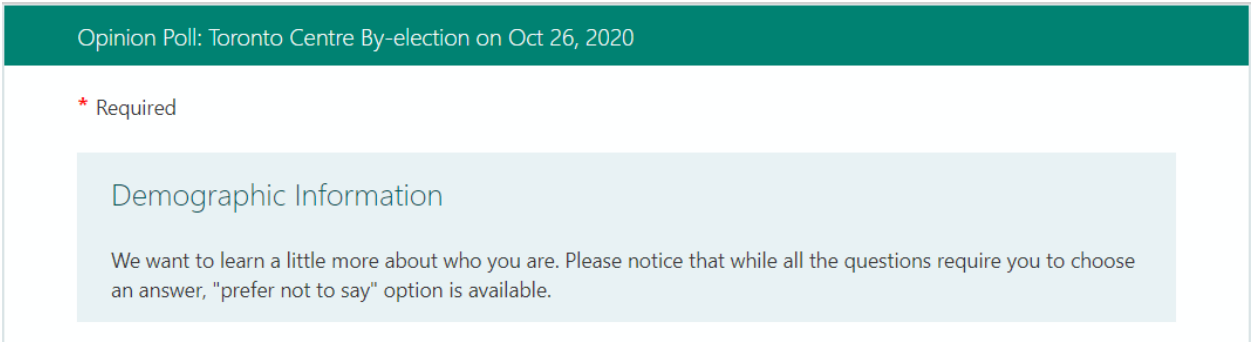
	1	2	3	4	5	6	7	8	9	10	
Very Undetermined	○	○	○	○	○	○	○	○	○	○	Very Determined

Next

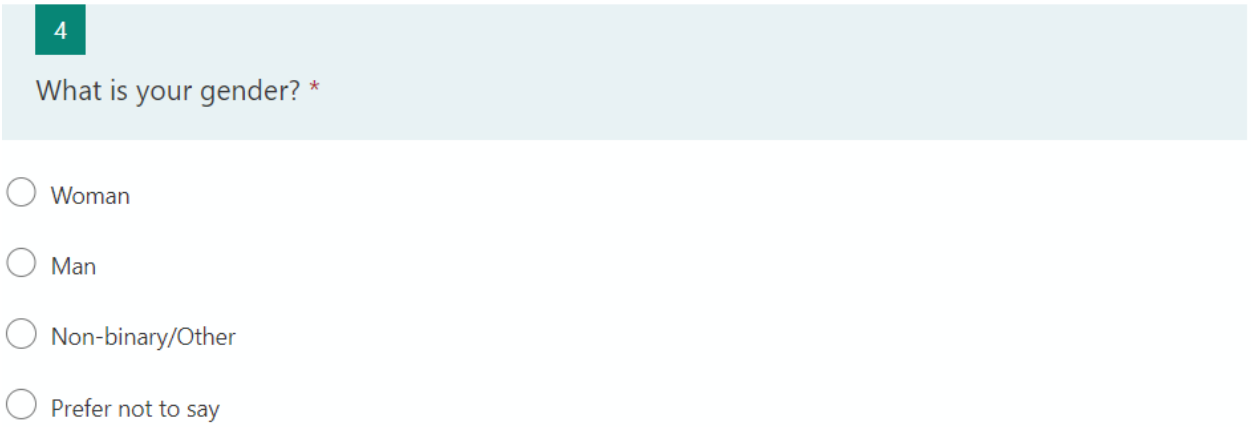
The below is Question 3 for individuals who do not choose NDP in Question 2.



Then, we move to the section of demographic information.



Question 4 is for gender.



Question 5 is for age.

5

What is your age? *

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- Age 65 and older
- Prefer not to say

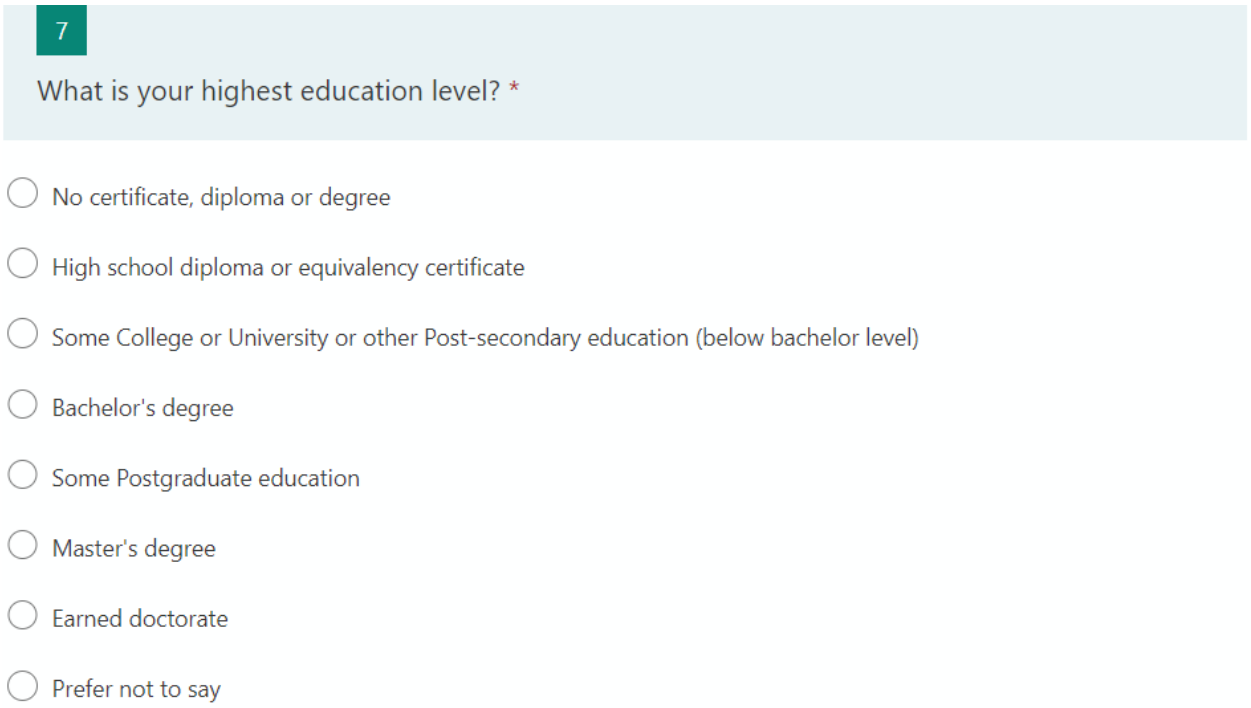
Question 6 is for ethnicity.

6

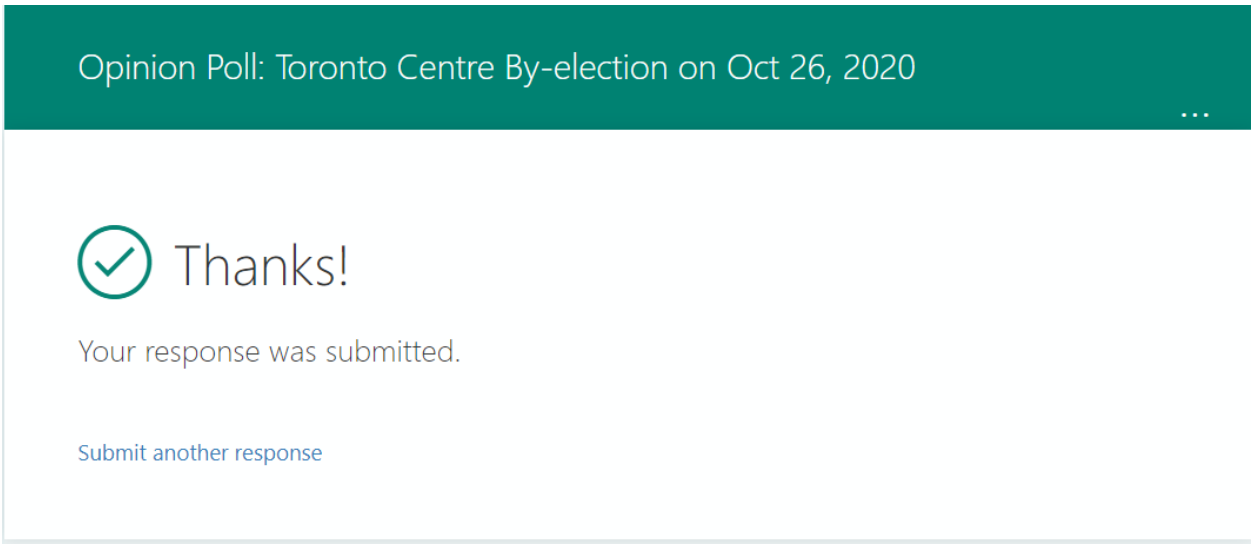
Which group best describes your ethnicity? *

- ○ Visible Minorities (not included elsewhere)
- ○ Prefer not to say
- ○ West Asian
- ○ White/European
- ○ First Nations/Métis/Inuit
- ○ Black/African Canadian
- ○ East Asian
- ○ Latin American
- ○ South Asian
- ○ Arab or North African
- ○ Multiple ethnic origins
- ○ Southeast Asian

Question 7 is for highest education level.



We want to say thanks to all respondents after they submit the survey.



References

- Allaire, JJ and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2020). rmarkdown: Dynamic Documents for R. R package version 2.4. URL, https://rmarkdown.rstudio.com.
- Auguie, Baptiste (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3.
- Mie- Yun Lee (2002). Conducting surveys and focusing groups. Entrepreneur Asia Pacific, Marketing.

URL, https://www.entrepreneur.com/article/55680.
- Müller, Kirill and Hadley Wickham (2020). tibble: Simple Data Frames. URL, https://tibble.tidyverse.org/, https://github.com/tidyverse/tibble.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL, https://www.R-project.org/.
- Statistics Canada (2017). Toronto Centre [Federal electoral district], Ontario and Canada [Country] (table). Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. URL, https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E (accessed October 7, 2020).
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
- Wickham et al. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- Wu, Changbao and Mary E. Thompson (2020). Sampling Theory and Practice. Springer.
- Xie Y (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.30, URL, https://yihui.org/knitr/.
- 338Canada.com (2020). Popular vote projection - Toronto Centre. URL, https://338canada.com/35108e.htm.