# Technical Notes for COVID-19 infection in Toronto: Ethno-racial identity and income (September 30, 2020)

## Socio-demographic Questions:

Toronto Public Health (TPH) is asking questions on priority socio-demographic characteristics to people who have a probable or confirmed diagnosis of COVID-19. These questions were added to our case and contact investigation process on May 20, 2020. Socio-demographic data is entered into CORES (The Coronavirus Rapid Entry System) along with other case and contact information. The case and contact investigators who are asking these questions are trained on how to collect the information in a sensitive way, including answering questions on why the data are being collected and how they will be used.

The selected questions were developed by the City of Toronto's Data for Equity initiative. This initiative built on the work of early leaders in this field in Canada, including the 'We Ask Because we Care' initiative, which focused on Toronto area hospitals and Community Health Centres. Additional engagement with experts from the Ontario Anti-Racism Directorate (ARD), the Toronto District School Board, provincial child welfare, youth justice and social assistance programs, and other researchers and community consultations allowed the Data for Equity questions to be refined from earlier iterations to their current format, below.

1. People often describe themselves by their race or racial background. For example, some people consider themselves "Black", "White" or "East Asian".

Which race category best describes you? Please select one only.

* Arab, Middle Eastern or West Asian (examples: Afghan, Armenian, Iranian, Lebanese, Persian, Turkish)
* Black (examples: African, African-Canadian, Afro-Caribbean)
* East Asian (examples: Chinese, Japanese, Korean)
* First Nations (status, non-status, treaty or non-treaty), Inuit or Métis
* Latin American (examples: Brazilian, Colombian, Cuban, Mexican, Peruvian)
* South Asian or Indo-Caribbean (examples: Indian, Indo-Guyanese, Indo-Trinidadian, Pakistani, Sri Lankan)
* Southeast Asian (examples: Filipino, Malaysian, Singaporean, Thai, Vietnamese)
* White (examples: English, Greek, Italian, Portuguese, Russian, Slovakian)
* More than one race category or mixed race, please select all that apply:
  + Arab, Middle Eastern or West Asian
  + Black
  + East Asian
  + First Nations, Inuit or Métis
  + Latin American
  + South Asian or Indo-Caribbean
  + Southeast Asian
  + White
  + Not listed, please describe: \_\_\_\_\_\_
* Not listed, please describe: \_\_\_\_\_\_\_\_\_\_\_\_
* Prefer not to answer

2. What was your total household income before taxes last year? Your best estimate is fine. Please select one only.

* 0 - $29,999
* $30,000-$49,999
* $50,000-$69,999
* $70,000-$99,999
* $100,000-149,999
* $150,000 or more
* Don't know
* Prefer not to answer

3. Including yourself, how many people live in your household on a regular basis? Please select one only.

* 1
* 2
* 3
* 4
* 5
* More than 5, please specify: \_\_\_\_\_\_
* Prefer not to answer

## Assessing Low Income

The table below provides rough estimates to identify people living in low income households based on their reported household income and household size. For example, if a person reports their income as $30,000-$49,999 and their household size as 3, they are categorized as low income. If a person reports their income as $70,000-$99,999 and their household size as 3, they are not categorized as low income. Any case with a reported income of $0 - $29,999 was categorized as low income regardless of household size. Similarly, any case with a reported income in the ranges of $100,000-149,999 or $150,000 or more was categorized as not low income regardless of household size.

|  |  |
| --- | --- |
| **Income range** | **Notes** |
| 0 - $29,999 | Roughly aligns with:   * Low Income Measures (LIMs) for **one person households** * Annual salary of a person working full-time for minimum wage |
| $30,000-$49,999 | Roughly aligns with Low Income Measures (LIMs) for **2, 3, and 4 person households** |
| $50,000-$69,999 | Roughly aligns with Low Income Measures (LIMs) for **5, 6, 7, 8 and 9 person households** |
| $70,000-$99,999 | Roughly aligns with Low Income Measures (LIMs) for **10 person households** |
| $100,000-149,999 | Does not align with low-income |
| $150,000 or more | Does not align with low-income |

## Inclusion criteria

The socio-demographic questions on ethno-racial identity, income and household size were added to CORES on May 20th, 2020. Clients of TPH with confirmed or probable COVID-19 are paired with a case and contact investigator who ask them a series of questions to help manage their COVID-19 infection and prevent transmission. Data that can inform TPH's response to the pandemic are also asked, including the socio-demographic questions on ethno-racial identity, income and household size.

These questions were not asked to people living in retirement or long-term care homes. There were two main reasons for this. Firstly, case and contact investigators do not always liaise directly with retirement home and long-term care home clients. Information is sometimes collected from a family member or a member of staff. These questions of a personal nature are not appropriate to be answered via a third party. A second reason is that the questions on income and household size do not apply well to the context of a collective dwelling.

The vast majority of the cases (N=7,402) with valid data on ethno-racial identity, income and household size had a reported or classification date of May 20th or later. However, there were 752 cases reported prior to May 20th, 2020 who had valid socio-demographic data due to their case still being open when the questions were added to CORES. Their information was also included in the analysis.

## Analysis Notes

### Calculating Rates and proportions

Proportions (shares) were calculated using the formula below:

|  |  |
| --- | --- |
| Number of cases in a category  *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* | x 100 |
| Total number of cases |

At the population level, proportions were calculated using the formula below:

|  |  |
| --- | --- |
| Total number of people in a socio-demographic category as per the 2016 census[[1]](#footnote-1),[[2]](#footnote-2)  *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* | x 100 |
| Total number of people as per the 2016 census1,2,3 |

Rates were calculated using the formula below:

|  |  |
| --- | --- |
| Number of Cases  *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* | x 100,000 |
| Total number of people in that socio-demographic category as per the 2016 census1,2,[[3]](#footnote-3) |

Rate ratios were calculated using the formula below:

|  |  |
| --- | --- |
| Rate in sociodemographic category  *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* |  |
| Rate in all other groups in sociodemographic category |

* For proportions, those who identified as mixed race or of another race were included in the analysis but not displayed on the dashboards.
* For the purpose of calculating the rates and proportions all cases with invalid data (i.e. with answers such "Not asked", "Prefer not to answer" or "Don't know") were removed from the analysis.
* Due to low numbers and lack of an appropriate denominator, cases classified as transgender, other or unknown were removed from the analysis but are included in the overall counts.
* As we are not currently presenting data on Indigenous identity, data on these cases were removed for the analysis on rates and proportions for ethno-racial groups. In addition, those with indigenous identity were also removed from the census counts.
* We are not releasing any data on Indigenous identity at this time because consultation with representatives from the Indigenous community is ongoing
* Engagement with Indigenous partners is an important to ensure that TPH's data collection and knowledge translation strategies for Indigenous data align with the principles of OCAP® (Ownership, Control, Access, and Possession).
* As part of the current data collection process, TPH will be seeking guidance from Indigenous leaders on how Indigenous data should be collected, used, and reported.

## Characteristics of COVID-19 cases with valid socio-demographic data (up to Sept 30) compared to cases with missing data[[4]](#footnote-4)

The following section provides a description of the cases with valid data on ethno-racial identity, income groups and household size. It also compares this sample of cases to a) cases reported before May 20th without these data and b) cases reported after May 20th with missing data for these questions. This analysis reflects data collected up until September 30th. It will be updated again once there is a sufficient addition of cases to yield differences in the findings.

### Cases with reported COVID-19 infection with valid socio-demographic data

In total, 8154 cases were included in this sample. Fifty-one (51%) percent of the sample was female. The average age of the cases was 39.3. The most common exposure for this sample was close contact (61.1%), followed by community (21.8%) and the vast majority of these cases were sporadic in nature (85.9%). A small proportion of these cases were hospitalized (5.6%) or flagged as severe (5.6%) and less than 1% had a fatal outcome (Table 1).

### Cases with reported COVID-19 infection prior to May 20th who were not asked the socio-demographic questions

There were 5,678 cases included in this sample. It includes all of the cases with a reported and classification date prior to May 20th, 2020 who were not asked the socio-demographic questions. A very small proportion (~5%) of these cases had a reported or classification date after May 20th but were entered in the tool used at TPH prior to CORES (iPHIS). They were also included as part of this sample, as it would not be possible to collect data on ethno-racial identity, income groups and household size from them.

Compared to the sample with valid socio-demographic data, the sample without sociodemographic data had a higher average age (45.5 vs. 39.3) (Table 1). This sample also had higher proportion of cases who had travel as an exposure (7.4% vs. 5.4%) or had an unknown or missing exposure (3.7% vs. 1.4%). There was also a higher proportion of cases in this sample who were outbreak related (28.9% vs 14.1%), had ever been hospitalized (15.2% vs. 5.6%) or were flagged as severe cases (15.4 vs. 5.6%) (Table 1).

### Cases reported after May 20th with missing data

For all three socio-demographic questions of interest, the most common reason why data was not used for analysis is because it was not asked (Table 2-4). This percentage ranged from 22.6% for the question on household size to 27.0% for the question on household income (Table 2-4). Reasons for this include the following:

* Individuals who are very ill and don't have the strength to speak for an extended period of time during daily follow up calls
* Language barriers that may exist with clients
* Case investigators may have difficulty establishing a relationship with some clients in order to ask these questions.
* Case and contact staff may be new and may require further training in order to gain comfort with asking these questions

Roughly five percent (4.7%) of cases preferred not to answer the question on household size, roughly seven percent (7.1%) preferred not to answer the question on ethno-racial identity and 19.5% preferred not to answer the question on household income (Table 2-4). An additional 14.8% did not know their household income (Table 2).

In total 2,824 cases had no information that could be used for the analysis on ethno-racial identity, household size and income groups. Cases were included in this sample if they had a reported or classification date on or after May 20th, 2020 and had not been asked any of the questions or had responded in a way that did not render their data possible to use (i.e. "Prefer not to answer" or "Don't know"). Cases who had a disposition of "pending" and were not asked any of the questions were excluded as investigators might have a chance to ask later on in the interview process.

Compared to the sample with data on ethno-racial identity, household size and income groups, the sample with missing information had a higher proportion of cases that were male (53.3% vs. 48.2%) or were classified as probable cases (10.7% vs. 7.0%) (Table 1). In addition, a higher proportion of this sample had unknown or missing exposure (19.4% vs. 1.4%), were ever hospitalized (8.5% vs. 5.6%) or flagged as a severe case (8.8% vs. 5.6%) (Table 1). The sample with missing information also had a slightly higher proportion of cases from the Old Toronto Community Council area (22.7% vs. 18.6%), and a slightly lower proportion of cases from Scarborough (16.1% vs. 19.6%) and Etobicoke (17.9% vs. 20.2%).

Due to a rapid increase in case volume towards the end of September, Toronto Public Health temporarily paused the collection of these data from October 1 to October 8. Although data collection resumed on October 9, the percent of cases with an episode date in late September who do not have valid socio-demographic data is higher than the percent overall.

### Table 1. Characteristics of people with reported COVID-19 with valid socio-demographic data (SDD sample), cases before May 20 without socio-demographic data (Historical sample) and cases after May 20 missing socio-demographic data (Missing sample)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Value** | **SDD sample**  **N (%)** | **Historical sample**  **N (%)** | **Missing Sample**  **N (%)** |
| Sex | Male | 3929 (48.2) | 2697 (47.4) | 1505 (53.3) |
|  | Female | 4171 (51.2) | 2915 (51.4) | 1297(45.9) |
| Age | Average age | 39.3 | 45.5 | 40.0 |
| Classification | Confirmed | 7587 (93.0) | 5110 (90.0) | 2521 (89.3) |
|  | Probable | 567 (7.0) | 568 (10.0) | 303 (10.7) |
| Exposure | Close contact | 4979 (61.1) | 3380 (59.5) | 1430 (50.6) |
|  | Community | 1776 (21.8) | 843 (14.9) | 519 (18.4) |
|  | Healthcare | 606 (7.4) | 535 (9.4) | 153 (5.4) |
|  | Institutional | 241 (3.0) | 291 (5.1) | 107 (3.8) |
|  | Travel | 437 (5.4) | 419 (7.4) | 68 (2.4) |
|  | Unknown/Missing | 115 (1.4) | 210 (3.7) | 547 (19.4) |
| Outbreak vs. sporadic | Sporadic | 7001 (85.9) | 4040 (71.1) | 2433 (86.2) |
|  | Outbreak | 1153 (14.1) | 1638 (28.9) | 391 (13.8) |
| Ever Hospitalized | Yes | 456 (5.6) | 861 (15.2) | 240 (8.5) |
|  | No | 7698 (94.4) | 4817 (84.8) | 2584 (91.5) |
| Severe case | Yes | 460 (5.6) | 873 (15.4) | 248 (8.8) |
|  | No | 7694 (94.4) | 4805 (84.6) | 2576 (91.2) |
| Outcome | Recovered | 8085 (99.2) | 5472 (96.4) | 2763 (97.8) |
|  | Active cases | 7 (0.1) | 0 (0.0) | 7 (0.2) |
|  | Fatal | 62 (0.8) | 206 (3.6) | 54 (1.9) |
| Episode Month | January | 0 (0) | 2 (0) | 0 (0) |
|  | February | 0 (0) | 13 (0.2) | 0 (0) |
|  | March | 25 (0.3) | 1555 (27.4) | 13 (0.5) |
|  | April | 483 (5.9) | 3142 (55.3) | 239 (8.5) |
|  | May | 2229 (27.3) | 964 (17.0) | 783 (27.7) |
|  | June | 1614 (19.8) | 1 (0.0) | 364 (12.9) |
|  | July | 711 (8.7) | 1 (0.0) | 87 (3.1) |
|  | August | 742 (9.1) | 0 (0) | 117 (4.1) |
|  | September | 2350 (28.8) | 0 (0) | 1221 (43.2) |

### Table 2. Frequency and percent for values for race variable

|  |  |  |
| --- | --- | --- |
| **Value** | **N** | **Percent** |
| Arab, Middle Eastern or West Asian | 725 | 6.6 |
| Black | 1763 | 16.1 |
| East Asian | 277 | 2.5 |
| Latin American | 622 | 5.7 |
| More than one race category or mixed race | 120 | 1.1 |
| South Asian or Indo-Caribbean | 1383 | 12.6 |
| Southeast Asian | 984 | 9.0 |
| White | 1591 | 14.5 |
| Other | 22 | 0.2 |
| Not Asked | 2680 | 24.4 |
| Prefer not to answer | 779 | 7.1 |

### Table 3. Frequency and percent for values for income variable

|  |  |  |
| --- | --- | --- |
| **Household income** | **N** | **Percent** |
| $0 - $29,999 | 1092 | 10.0 |
| $30,000-$49,999 | 1010 | 9.2 |
| $50,000-$69,999 | 666 | 6.1 |
| $70,000-$99,999 | 648 | 5.9 |
| $100,000-$149,999 | 472 | 4.3 |
| $150,000 or more | 353 | 3.2 |
| Don't know | 1629 | 14.8 |
| Not Asked | 2967 | 27.0 |
| Prefer not to answer | 2141 | 19.5 |

### Table 4. Frequency and percent for values for household size variable

|  |  |  |
| --- | --- | --- |
| **Household Size** | **N** | **Percent** |
| 1 | 1194 | 10.9 |
| 2 | 1708 | 15.5 |
| 3 | 1489 | 13.6 |
| 4 | 1576 | 14.4 |
| 5 | 1028 | 9.4 |
| More than 5, please specify | 995 | 9.1 |
| Not Asked | 2478 | 22.6 |
| Prefer not to answer | 510 | 4.7 |

## Definitions

* **Cases:** Includes both confirmed and probable COVID-19 cases reported to Toronto Public Health
* **Classification date**: The date on which a decision is made to classify the case with the selected classification value (e.g. the date on which a case was determined to be 'confirmed')
* **CORES:** COVID-19 Rapid Entry System. TPH's in-house tool for collecting data on COVID-19 cases. It was developed to complement data entry in iPHIS. The questions ethno-racial identity and income were added on May 20th, 2020.
* **Disposition**: Value reflecting the current state of the investigation (pending, complete, lost-to-follow-up, untraceable, duplicate, does not meet definition, entered in error)
* **Exposure:** The most likely way that cases acquired their COVID-19 infection. Only the most likely exposure for each case is reported. Exposures that occurred up to 14 days before symptoms start are potential acquisition sources, and can include:
  + Travel: Travel outside of Ontario.
  + Close contact with a case: Was in close contact with a confirmed or probable COVID-19 case (e.g. reside in the same household).
  + Institutional setting: Institutional settings includes, but not limited to: long-term care homes, acute care hospitals, complex case hospitals, special care facilities, retirement homes, rehabilitation hospitals.
  + Healthcare setting: Healthcare settings includes, but not limited to: family physician, dentist, ophthalmologist, sports doctor.
  + Community: Cases with no reported travel outside of Ontario, no known close contact with a COVID-19 case, and no reported infection acquired in an institutional or healthcare setting.
* **iPHIS:** integrated Public Health Information System. Tool used by TPH to collect data on COVID-19 cases prior to the introduction of CORES.
* **Outbreak cases**: Outbreak associated cases include persons with COVID-19 within a defined group or setting. For the purpose of this analysis, they would refer to healthcare setting (e.g. hospitals) and residential or congregate settings (e.g., homeless shelters, group homes). They can also be in workplaces and other settings.
* **Reported date**: The date on which the case was reported to the health unit. When the case is reported by a laboratory, this is the date on which the result was received, and may be later than the test reported date if the test result was reported outside of business hours.
* **Severe cases**: Cases who were hospitalized (including those who had been admitted to ICU, intubated) or were deceased were flagged as severe.
* **Sporadic cases**: Sporadic cases (as contrasted to outbreak cases) refer to all other cases in members of the general population.

For more information on the data or if you need help with interpreting this document, please contact us at [seu@toronto.ca](mailto:seu@toronto.ca) or 416-338-7600.

## Acknowledgement

We would like to acknowledge the Wellesley Institute for their helpful advice and input on interpreting and presenting the findings.

1. Census data on ethno-racial groups was obtained from Statistics Canada, 2016 Census of Canada [↑](#footnote-ref-1)
2. Census data on income groups was obtained from a customized data request to the Statistics Canada [↑](#footnote-ref-2)
3. Census data for counts on Indigenous identity was accessed from the Household & Family TGP of the Aboriginal identity population, 2016 Census, Target Group Profiles. [↑](#footnote-ref-3)
4. [↑](#footnote-ref-4)