Who Identifies Outside the Gender Binary in Online Surveys? Evidence from the 2019 and 2021 Canadian Election Studies

Quinn M. Albaugh
University of Toronto

Peter John Loewen
University of Toronto

Allison Harell
Université du Québec à Montréal

Daniel Rubenson
Ryerson University

Laura B. Stephenson
University of Western Ontario

December 3, 2021

Abstract

A small but growing number of people identify as non-binary – that is, as neither men nor women. However, we know relatively little about the correlates of non-binary identity in online general population surveys, and we know even less about the impact of questionnaire design on the correlates of non-binary identity. We compare data from two large-sample online surveys – the 2019 and 2021 Canadian Election Study (CES) online panels. The 2019 CES uses a single three-category gender identity question, while the 2021 CES uses a two-step approach that measures both gender identity and transgender identity. We find that the 2019 CES generates inflated estimates of the percentage of non-binary people, which we explain using evidence that transgender men and women select the same response category. This measurement error distorts the relationship between age, sexual identity, and attitudes toward lesbians and gay men.

Note: This paper is a work in progress. The next step is to use multiple imputation to model 2019 CES respondents' values on the 2021 CES gender questions. We welcome suggestions on modelling non-binary identity, additional covariates to consider, and, particularly, the proposed methodology for the multiple imputation.

An increasing number of people identify as non-binary – that is, as neither men nor women.[1] Statistical agencies and survey researchers are increasingly starting to collect data on non-binary people by allowing individuals to report gender identities other than "man" or "woman." However, general population surveys usually do not include a large enough non-binary respondents to allow for subgroup analyses of this community. For that reason, much of the survey research on non-binary people relies on non-representative samples, particularly respondent-driven or community-based samples (James et al., 2016; Bauer, 2020*a,b*). Many of these studies actually focus on transgender people (or trans people) as a whole without consistently disaggregating non-binary people from transgender men (or trans men) and transgender women (or trans women).[2] As a result, we know relatively little about the demographic correlates of non-binary identity in general population samples.[3]

Given this increasing recognition of the need to collect data on trans and non-binary people, different researchers have started to recommend different approaches to measuring sex/gender. Some researchers recommend simply including response options that allow respondents to identify with other sex/gender categories beyond male/man or female/woman (Fraser, 2018; Fraser et al., 2020; Medeiros, Forest, and Öhberg, 2020), while others recommend "two-step approach," which first asks respondents about their gender identity and then asks a question about assigned sex or gender modality (Badgett et al., 2014; Westbrook and Saperstein, 2015).[4] The original rationale of the two-step approach was to make sure that trans men and women would be identifiable as trans given that they are likely to identify as male and female, respectively, in traditional sex and gender questions (Badgett et al., 2014). However, many possible additional sex or gender categories – such as "male/female/transgender" or even "male/female/other" – are likely to push some binary trans people to select options other than male/man or female/woman by preventing them from expressing their gender modality. To date, we have relatively few comparisons of the estimates results from different approaches.

---

[1]For the sake of parsimony, we use the term "non-binary" to refer to people who identify as neither men nor women. We recognize that some people who identify as neither men nor women may prefer other terms, such as agender, genderqueer, or bi-gender. Unfortunately, we cannot disaggregate these identities using the data presented here.

[2]For short, we sometimes refer to trans men and women as binary trans people.

[3]To our knowledge, Wilson and Meyer (2021) come the closest to recruiting a probability sample of non-binary people. However, their inclusion criteria may miss non-binary people who do not necessarily identify as LGBTQ. We discuss this study in more detail below.

[4]By gender identity, we mean an individual's self-categorization of their own gender. For example, "I am a woman" or "I am agender." By assigned sex at birth, we mean the category – generally male or female – assigned to children by societal institutions including law, medicine, and families at birth, usually based on genitalia. Finally, by gender modality, we mean "how a person's gender identity stands in relation to their gender assigned at birth" (Ashley, 2021, 1). For example, one's gender modality can be cisgender (not transgender), transgender, or something else.

In this article, we focus on three main questions. First, how do different approaches to measuring gender identity affect which respondents identify as non-binary in surveys? Second, what are the correlates of non-binary identity in general population online surveys? Third, how do the answers to the two previous questions affect how researchers should measure gender identity in online surveys?

We argue that a two-step approach is necessary to capture non-binary respondents accurately within online surveys. We illustrate the pitfalls of relying on a single close-ended gender question by comparing two-large sample online surveys that use different approaches – the 2019 Canadian Election Study (CES) online panel (N = 37,822) and 2021 CES online panel (N = 22,328). The 2019 CES online panel used a single three-option gender identity question (man/woman/other), while the 2021 CES adopted a series of gender questions that include a gender identity question with an explicit close-ended non-binary response option and a open-ended gender response option, a transgender identity question, and a Two-Spirit identity question.[5] Most correlates of non-binary identity – including education, income, country of birth, race, mother tongue, region, and religion – are similar across the two surveys. However, the 2019 CES "other gender" respondents differ from the 2021 CES non-binary respondents on age, sexual identity, and attitudes toward lesbians and gay men (LGs). We present evidence that the reason for these discrepancies is that a substantial number of binary trans people – that is, trans men and women – selected the "other" category on the 2019 CES. This evidence includes both comparisons of the 2019 CES other gender respondents with the 2021 CES trans and non-binary respondents and multiple imputation of the 2019 CES respondents' answers to the 2021 questions from pooling the two surveys together. Finally, we estimate the correlates of non-binary identity in the pooled dataset using multiple imputation.

In this article, we make three main contributions. First, we demonstrate that the 2019 CES three-option gender identity question captures not only non-binary people but also some trans men and women, which introduces substantial measurement error. This type of problem is common to surveys that rely on a single gender identity question rather than a two-step approach, although different gender identity questions may be more or less susceptible to this measurement error. Second, we offer guidance for future research on what responses to demographic and political questions to expect from non-binary respondents in general population surveys. Third, we make recommendations on best practices for questionnaire design for trans and non-binary respondents.[6]

---

[5]Two-Spirit is a term coined by Indigenous LGBTQ+ activists to represent Indigenous understandings of gender/sexuality in English. It does not neatly fit into Western notions of gender and sexuality as analytically separate (Pruden and Salway, 2020). Although we examine the Two-Spirit identity question, our preliminary analyses suggest that the behaviour of binary trans respondents but not Two-Spirit respondents is central in explaining the differences between the 2019 and the 2021 CES.

[6]These recommendations will be in the concluding section.

Questionnaire Design and Non-Binary Respondents

As statistical agencies and survey researchers have increasingly recognized that questionnaires that only include binary "sex" variables do not capture trans and non-binary respondents, they have begun to adopt a variety of different approaches to measuring sex and gender in survey research. These approaches fall broadly into two camps. The first camp argues for single-item measures of gender identity, while the second camp recommends a "two-step approach" that includes a gender identity question and an additional question on assigned sex at birth or gender modality. The single-item measures are likely to overstate the percentage of non-binary people within surveys by pushing binary trans people to identify as something or than male/man or female/woman to be visible, while the two-step approach addresses this problem by allowing them to express their trans identities as well as their identities as men and women.

Researchers have proposed several single-item measures of gender identity. One commonly-used question offers respondents three response options – male, female, and transgender. Trans and non-binary respondents generally find this question objectionable, for three reasons (Holzberg et al., 2017). First, it suggests that being trans and being a man or a woman are mutually exclusive – that is, that trans men and women are not *really* men or women, which is offensive and invalidating to trans men and women. Second, this question format suggests to trans and non-binary respondents that researchers are not familiar with basic concepts relevant to their lives. Third, this question pushes non-binary people to identify as transgender, even though many do not.[7] This type of question is likely to push some binary trans people and non-binary people to select the same response option, while other binary trans people select the "male" or "female" options to assert their gender identities. The result is a measure that does not capture non-binary people and has some measurement error in capturing both trans and non-binary people together.

Another approach is to use a question that provides "Other" rather than "Transgender" as the third response option (Medeiros, Forest, and Öhberg, 2020). This type of question has received criticism for being literally othering to non-binary respondents (Bauer et al., 2017). However, it is also possible that this question *also* mis-measures non-binary identity. Some trans men and women may see their gender modalities as so central to their experiences they would prefer to select "Other" rather than selecting male/man or female/woman without being able to indicate their gender modalities. If so, then it is not enough to simply add additional categories to a gender identity question (Westbrook and Saperstein, 2015).

The only single-item approach that is unlikely to have error in measuring non-binary identity is a single open-ended gender identity question (Fraser, 2018; Fraser et al., 2020).

---

[7]In the 2021 CES, only 35 percent of non-binary respondents also identify as transgender.

This is approach is less likely to mismeasure non-binary respondents because they are able to write in their gender identities using their preferred language, and trans men and women for whom their identities as trans are important to express can do so. For example, a trans man can write in "trans man" as an open-ended response. However, a single open-ended question may still under-count binary trans people who do not see their gender modality as central to their lives or who do not wish to disclose their gender modality to researchers.

The accepted solution to these potential problems with measuring trans and non-binary identities is to use the "two-step approach." Traditionally, the two-step approach has relied on asking separate questions for gender identity and assigned sex (Badgett et al., 2014; Westbrook and Saperstein, 2015). Under the traditional two-step approach, researchers identify trans respondents as any respondents whose gender identities do not correspond to their assigned sex. However, trans and non-binary respondents often view asking about assigned sex as invasive (Holzberg et al., 2017). In addition, this approach identifies trans people ascriptively rather than based on their identities, which is generally not the approach used for racial or sexual minority groups in political surveys.[8] An alternative approach to asking about assigned sex is to ask about gender modality – that is, to ask respondents if they are transgender. If identities are what are important to measure – for example, in explaining political behaviour (Achen and Bartels, 2016) – then a transgender identity question is more suitable than an assigned sex question. Although the two-step approach traditionally focuses on avoiding under-counting binary trans men and women, it may also help measure non-binary identity more accurately. If researchers provide trans men and women with opportunities to disclose their trans identities in a separate question from the gender identity question – ideally on the same page in an online survey – they will no longer face a dilemma of asserting their identities as trans (or their difference from cis men and women) or their identities as men and women (or their similarities to cis men and women).

Past Work on Correlates of Non-Binary Identity

Large-scale population surveys offer relatively little guidance on the correlates of non-binary identity. Even surveys that collect data on non-binary people often do not report results separately for non-binary people. For example, Statistics Canada's 2018 Survey of Safety in Public and Private Spaces – the first Statistics Canada survey to include questions that identify trans and non-binary respondents – only publishes results that pool "transgender and gender-diverse" respondents together.

---

[8]For example, it would be like measuring race by looking at skin colour rather than asking about identity or measuring sexuality by looking at attraction or behaviour rather than identity.

Given the limitations of large-scale population surveys, the clearest expectations on the correlates of non-binary identity come from alternative research designs. The first approach relies on non-probability community-driven samples, such as the 2015 U.S. Trans Survey (James et al., 2016) – or USTS, for short – or the 2019 Trans PULSE Survey in Canada (Bauer, 2020b). These surveys rely on recruitment through trans and non-binary organizations and networks. They also use . The second approach is to recruit a very large (hundreds of thousands) probability sample to take a screener survey and then select respondents who fit particular criteria to respond to a more in-depth questionnaire. The Williams Institute presents data on "non-binary LGBTQ adults" drawn from a pooled dataset of two Gallup surveys (Generations and TransPop) using this large-sample screener methodology conducted from 2016-2018 (Wilson and Meyer, 2021). The first includes respondents who identify as transgender. The second includes respondents who identify as lesbian, gay, bisexual, or queer (LGBQ) but not transgender. However, the pooled Generations/TransPop dataset does not include non-binary people who do not otherwise identify as LGBTQ.

These studies suggest several expectations about the correlates of non-binary identity. The USTS, Trans PULSE, and Generations/TransPop all suggest that non-binary people are disproportionately likely to be young and low-income and unlikely to be born in another country. The two community-driven surveys – the USTS and Trans PULSE – suggest that non-binary people are likely to have high educational attainment, particularly graduate or professional degrees. By contrast, the Generations/TransPop data suggest that non-binary people are disproportionately likely not to have any education beyond high school. Given the different sampling approaches, it is likely that community-driven surveys over-represent non-binary people with Bachelor's and, especially, graduate or professional degrees. In addition, all three studies suggest that non-binary people are disproportionately unlikely to identify as straight or heterosexual in comparison with the general public. Only two percent of non-binary USTS and Trans PULSE respondents identify as straight or heterosexual (James et al., 2016). Similarly, at most two percent of Generations/TransPop non-binary respondents identify as straight or heterosexual (Wilson and Meyer, 2021).[9] While respondent- or community-driven surveys and the Generations/TransPop survey may be less likely to reach non-binary people who identify as straight or heterosexual due to their recruitment and sampling frames, respectively, the percentage of straight and heterosexual respondents is so low it is unlikely that this pattern does not exist among non-binary people as a whole.

Given racial, ethnic, and geographic differences between Canada and the United States, we rely on Trans PULSE for expectations on the racial, ethnic, and geographic correlates of non-binary identity. The non-binary Trans PULSE respondents are dispro-

---

[9]Wilson and Meyer (2021) do not report the exact percentage, but straight or heterosexual respondents would presumably fall in the two percent who report "another [sexual] identity" in their report.

portionately likely to be white (79 percent) or Indigenous (7 percent). The Trans PULSE study does not report data by language of questionnaire or mother tongue. However, non-binary respondents were disproportionately unlikely to be from the four Atlantic provinces (Nova Scotia, New Brunswick, Prince Edward Island, and Newfoundland and Labrador) and, particularly, Quebec. The low number of non-binary respondents from Quebec suggests that native speakers of French (francophones) may be less likely to identify as non-binary.

There is, to date, no published research on political attitudes or behaviour strictly on non-binary people. Instead, research on LGBTQ political behaviour tends to rely on surveys such as the Cooperative Election Study that only have binary gender variables. Still, based on research on LGBTQ political behaviour, we generally expect that non-binary people will be more likely to support parties of the left, such as the NDP or the Greens, and less likely to support parties of the right, such as the Conservatives (Perrella, Brown, and Kay, 2012; Perrella, Andrea M. L., Brown, Steven D., and Kay, Barry, 2019). In addition, we expect non-binary people to express (in general) positive attitudes toward lesbians and gay men, given that they are all associated with the broader LGBTQ2S+ community and that non-binary respondents may be likely to interpret questions about lesbians and gay men as, in part, about the broader LGBTQ2S+ community.

Based on this review of past studies, we expect that non-binary people in Canada are disproportionately young, born in Canada, low-income, not holding a post-secondary degree, not heterosexual, white or Indigenous, living outside Atlantic Canada and Quebec, not francophone, NDP or Green partisans, and expressing positive attitudes toward lesbians and gay men.

Data and Methodology

We draw on two large-sample online surveys conducted during the 2019 and 2021 Canadian federal election campaigns, the 2019 CES (N = 37,822) and the 2021 CES (N = 22,328). Due to their large sample sizes, these surveys have relatively large trans, non-binary, and Two-Spirit sub-samples relative to traditional general population surveys. These two surveys share many common questions that measure standard demographic and political variables, including age, education, income, province/territory or region of residence, language of questionnaire, mother tongue, country of birth, citizenship status, religion, marital status, number of children, community size, party identification, vote intention, attitudes toward lesbians and gay men. In addition, we apply a common coding scheme for racial and sexual identity variables that draw on non-comparable questions. We describe the question wording and coding scheme for these variables in more detail in Appendix A. However, the 2019 and 2021 CES use different approaches to measuring gender identity, which we detail below. We use the combination of the two surveys to

8

estimate the correlates of non-binary identity and illustrate the differences in correlates and in survey response across the two approaches to measuring gender identity.

The 2019 CES uses a three-category close-ended question to measure gender identity. At the beginning of the first wave of the 2019 CES online panel, respondents see the following gender question:

Are you...?
1. A man
2. A woman
3. Other (e.g. Trans, non-binary, two-spirit, etc.)

This question is mandatory to proceed with the survey. Respondents must select one and only one of these answers. The "Other" category is unusual in explicitly mentioning trans, non-binary, and Two-Spirit respondents. Methodologically, this question risks measurement error in identifying non-binary people by incentivizing trans or Two-Spirit men and women to select the "Other" category so that they can be counted based on their trans or Two-Spirit identities. While this may seem like a problem that is unique to the 2019 CES, many gender questions used in practice, such as a three-option close-ended question that provides Male/Female/Transgender as response options, present similar problems of pushing trans men and women to choose between identifying as trans or as men or women. In most cases, we do not know for certain whether respondents who select the "Other" category in the 2019 CES actually identify with a gender identity that falls outside the gender binary – such as non-binary, genderqueer, agender, or bigender – as opposed to identifying as trans men or women or as Two-Spirit.

By contrast, the 2021 CES uses a two-step approach that includes both a question on gender identity and a question on gender modality immediately after the gender identity question on the same page of the online survey. In addition, the 2021 CES includes a separate question on Two-Spirit identity asked only of Indigenous respondents later in the survey, based on recommendations from Two-Spirit researchers ().

Are you...?
1. A man
2. A woman
3. Non-binary
4. Another gender, please specify:

Are you transgender?
1. Yes
2. No
3. Don't know/Prefer not to say

Are you Two-Spirit?
1. Yes
2. No
3. I don't know/Prefer not to answer

This combination of questions allows us to identify trans, non-binary, and/or Two-Spirit respondents based on those identities rather than lumping them into a single "Other" category. As a result, these questions are likely to provide more accurate estimates of the percentage and correlates of non-binary identity.

Table 1: Number and Percentage of Respondents, by Gender Identity, 2019 and 2021 CES

	2019 N	2019 Unweighted (%)	2019 Weighted (%)	2021 N	2021 Unweighted (%)	2021 Weighted (%)
Man	15,551	41.1%	48.4%	10,204	45.7%	48.4%
Woman	21,980	58.1%	51.3%	12,005	53.8%	51.3%
Other/Non-Binary	291	0.8%	0.3%	107	0.5%	0.3%
Missing	0	0.0%	0.0%	12	$< 0.1\%$	$< 0.1\%$

The 2019 and 2021 CES differ substantially in the percentage of respondents who identify as "Other" or "Non-Binary." Table 1 shows the number of respondents and the unweighted and weighted percentage of respondents who identify as men, women, and "Other" (in 2019) or non-binary (in 2021), along with the number and percentage of missing data. Both the 2019 and the 2021 CES include raked weights on age, gender, education, and province or territory of residence based on census data. For respondents who do not select "man" or "woman," the weights only take into account age, education, and province or territory of residence because the 2016 Census of Canada – the most recent one with published data – only measures "sex" as a binary variable. We recode the 36 open-ended gender identity responses to the 2021 gender identity question into other categories where possible. Of the 35, nine provide recognizable gender identity responses other than man or woman (such as "genderfluid," "agender," or "bi-gender"), 15 provide binary gender responses (usually male or female rather than man or woman, sometimes combined with political commentary on the "political correctness" of asking about gender rather than sex), and 11 respondents provide hostile or non-cooperative responses that do not allow us to code their gender identities. We recode the first group as non-binary, the second group as men or women based on their responses, and the third group as missing. We include a detailed list of open-ended gender responses and their coding into in other categories in Appendix A.

Table 1 shows some similarities between the 2019 and the 2021 CES in terms of the percentage of men, women, and non-binary people. Both the 2019 and the 2021 CES over-represent women and under-represent men relative to the 2016 Census in the unweighted percentages. However, the two surveys differ substantially in the percentage of respondents who select a category other than "man" or "woman" – 0.8 percent of respondents to the 2019 CES selected the "Other" response, while only 0.5 percent of 2021 CES respondents selected the non-binary response. This discrepancy in the unweighted percentages of "non-binary" respondents is surprising given considerable evidence that the percentage of people who identify as neither men nor women is increasing over time (Kidd et al., 2021). (See Appendix B for additional details.) As a result, this discrepancy is suggestive evidence that the 2019 CES gender identity question does not solely measure non-binary identity but includes some people who identify as trans and/or Two-Spirit but *not* non-binary.

Table 2: 2019 Gender Identity Responses for Respondents Who Mention Non-Binary, Genderqueer, Transsexual, Transgender, or Two-Spirit Identity in Their Sexual Identity Open-Ended Responses

Open-Ended Sexual Identity Response Term	Gender Identity Responses
Non-Binary (1)	Other (1)
Genderqueer (2)	Woman (1), Other (1)
Transsexual (2)	Man (1), Woman (1)
Transgender (4)	Other (4)
Two-Spirit (2)	Other (2)

There is additional evidence in the 2019 CES that respondents who identify as trans or Two-Spirit men or women may select the "Other" category. In the 2019 CES, several respondents give gender identity, gender modality, and/or Two-Spirit identity responses in the sexual identity question. Table 2 displays the responses to the 2019 gender identity across categories of respondents who gave gender identity, gender modality, and/or Two-Spirit identity responses in the open-ended sexual identity response category. One respondent wrote in "non-binary" and selected the Other category, as intended. Two respondents wrote in "genderqueer" or a variant thereof. Of these two respondents, one selected Woman and the other selected Other. Two respondents identified as "transsexual," and they both selected binary gender responses. All four respondents who identified as transgender selected the "Other" option, even the ones. One of the four transgender respondents wrote "hetero transgender male" in the sexual identity open-ended response but selected "Other" in the 2019 CES gender identity question. Finally, two respondents wrote in "Two-Spirit" as a sexual identity response, and both of these respondents selected "Other" in the 2019 CES gender identity question. While these open-ended re-

sponses are not necessarily representative of all trans, non-binary, or Two-Spirit respondents, they illustrate that respondents who take these identities as central selected the Other gender identity response on the 2019 gender identity question – even if they identified as men or women in an open-ended response later in the survey.

Our analysis proceeds in three steps. First, we identify the correlates of "non-binary" identity using the 2019 CES "Other" respondents and the 2021 non-binary respondents. While we do not believe the 2019 CES "Other" respondents are necessarily all non-binary people, this comparison helps illustrate the limitations in using the 2019 CES question to draw inferences about non-binary respondents. Second, we pool the 2019 and 2021 CES together to conduct multiple imputation of the 2019 respondents' responses to the 2021 CES gender identity, transgender identity, and Two-Spirit identity questions based on a common set of demographic and political variables. We provide additional details on the multiple imputation below. Finally, we use the multiple imputation to estimate the correlates of non-binary identity in the pooled 2019-2021 dataset.

Correlates of "Non-Binary" Identity in 2019 and 2021 CES

We begin by comparing the demographics 2019 CES "Other" respondents and the 2021 CES non-binary respondents with the 2019 Trans PULSE Survey's non-binary sub-sample ($N = 1,327$) and the general population in the 2016 Census of Canada. The Trans PULSE non-binary sub-sample provides a baseline expectation for non-binary people, while the Census provides a comparison of the demographics with the general population. Table 3 displays the percentage of each sample that falls into different categories of age, education, income, country of birth, race, mother tongue, region, and sexual identity. We use these variables because they are available in a comparable form in at least three of the four datasets. (Mother tongue is not available for the Trans PULSE sample, while sexual identity is not available in the Census.) Most of these categories are mutually-exclusive and collectively-exhaustive, but the Census does not provide a white category, only Indigenous ("aboriginal identity") and racialized ("visible minority") categories. Importantly, the age results are not fully comparable because Trans PULSE and the Census include individuals ages 14-19 in the "Under 20" category, while the 2019 and 2021 CES only include individuals ages 18-19. Finally, we run Kolmogorov-Smirnov tests to examine how likely it is that the 2019 CES Other respondents and 2021 CES non-binary respondents come from the same distribution. In Table 3, we indicate variables that are significantly different ($p < 0.05$) between these two sub-samples using an asterisk (*).

While the data from all three sub-samples in Table 3 suggest that non-binary people are generally young in comparison with the general population, the 2019 and 2021 data differ in the magnitude of these age differences. Only 31 percent of Canadians ages 14+ are under age 35, but 49 percent of 2019 CES Other respondents, 69 percent of 2021

Table 3: Demographic Comparison of 2019 CES Other Gender Respondents, 2021 CES Non-Binary Respondents, Trans PULSE Non-Binary Respondents (2019), and General Population Ages 15+ (2016 Census Data)

Variable	Categories	2019 CES %	2021 CES %	Trans PULSE %	2016 Census %
Age*	Under 20	6	7	11	7
	20-24	16	29	26	8
	25-34	27	31	41	16
	35-49	23	18	17	24
	50-64	17	10	4	26
	65+	11	6	<1	20
Sexual Identity*	Straight or Heterosexual	49	12	2	–
	Not Straight or Heterosexual	51	88	98	–
Education	Below High School	11	8	2	18
	High School	19	19	27	26
	College/University	55	58	48	47
	Graduate/Professional	15	16	23	8
Income	Under 30,000 CAD	31	33	53	18
	Over 30,000 CAD	69	67	47	82
Country of Birth	Canada	89	87	87	78
	Outside Canada	11	13	13	22
Race	Indigenous	11	11	7	5
	Racialized	12	15	14	22
Mother Tongue	English	66	62	–	56
	French	8	9	–	21
	Both English and French	22	27	–	<1
	Neither English nor French	3	3	–	22
Region	Atlantic	7	3	6	6
	Quebec	15	18	13	23
	Ontario	41	39	34	38
	West	37	39	46	32
	Territories	1	1	<1	<1

* indicates that the 2019 CES Other respondents and the 2021 CES non-binary respondents are unlikely to be from the same underlying distribution using a Kolmogorov-Smirnov test ($p < 0.05$).

CES non-binary respondents, and 78 percent of Trans PULSE non-binary respondents are under age 35. The differences in age distributions between the 2019 CES Other respondents and the 2021 CES non-binary respondents are substantial enough that a Kolmogorov-Smirnov test indicates that they are unlikely to come from the same underlying distribution. This result fits with the explanation that the 2019 CES gender identity question captures binary trans people as well as non-binary people, given that the 2021 CES indicates that binary trans people are generally older than non-binary people.

The most striking divergence in the 2019 CES is that the "Other" gender respondents are much more likely to say that they are straight or heterosexual than the 2021 non-binary respondents (by a factor of four) and the 2019 Trans PULSE non-binary respondents (by a factor of almost 25). A Kolmogorov-Smirnov test on sexual identity indicates that the 2019 CES Other respondents and the 2021 CES non-binary respondents are highly unlikely to come from the same underlying distribution. The simplest explanation of this discrepancy is that the 2019 CES "Other" respondents include a substantial number of trans men and women who select the "Other" category because it explictly mentions trans people. In the 2021 CES, 69 percent of binary trans people report their sexual identity as straight or heterosexual, while only 12 percent of non-binary people do. If we assume that the 2019 "Other" respondents are either binary trans people or non-binary people, we can use data from a restricted sample of binary trans and non-binary respondents in the 2021 CES and Bayes' Rule to generate a "back of the envelope" calculation of the the percentage of binary trans respondents, which suggests that 60 percent of the 2019 CES "Other" respondents may actually be binary trans people. (See Appendix D.)

The remaining variables in Table 3 show consistent patterns across the three surveys. All three surveys suggest that non-binary people are less likely to have not finished high school and more likely to have graduate and professional degrees than the general population. This gap may be a result of generational differences in educational attainment. Similarly, all three surveys suggest that non-binary people are more likely to be from low-income households than the general population, although the percentage of respondents in low-income households is smaller in the 2019 and 2021 CES than in the Trans PULSE data. As with education, these income results may be attributable to non-binary people being younger than the general population, given that young adults tend to have lower average household incomes than the general population. All three surveys suggest that non-binary people are less likely (about half as likely) to be born outside Canada than the general population. All three surveys suggest that non-binary people are more likely to identify as Indigenous and less likely to identify as part of a racialized minority group than the general population. While the Trans PULSE Survey reports do not provide data on mother tongue among non-binary respondents, the 2019 and 2021 CES generally suggest that non-binary people are more likely to speak English or both English and French as first languages than the general population and are much less likely to speak only French or neither English nor French as first languages. This result is likely attributable

in part to age, country of birth, and racial backgrounds, given that native speakers of both English and French are generally younger on average and that native speakers of neither English nor French are more likely to be born outside Canada and/or racialized. However, they also suggest that non-binary people are much less likely to be unilingual francophones.[10] The results for region generally suggest that non-binary people are less likely to live in Quebec and more likely to live in Ontario or Western Canada relative to the general population, which is consistent with the observation that non-binary people are less likely to be unilingual francophones.

Based on the age and sexual identity results, we suspect that the resemblance between the 2019 CES "Other" respondents and the 2021 CES non-binary respondents exist because many of the correlates of identifying as non-binary are similar to the correlates of identifying as a binary trans person, as suggested in data from community-driven surveys that include both binary trans and non-binary people (James et al., 2016; Bauer, 2020*b*; Navarro et al., 2021).

Next, we present a series of figures that plot the estimated percentage of non-binary people given different demographic variables. Due to the small size of the non-binary population, we use Wilson confidence intervals rather than the more commonly-used Wald confidence intervals. By construction, Wilson confidence intervals always produce values between 0 and 1, while Wald confidence intervals can fall outside of the range between 0 and 1. (As a consequence, Wilson confidence intervals tend to be asymmetric around the estimate with longer bounds on the side of the estimate closer to 0.5 and shorter bounds on the side of the estimate further away from 0.5.) Past work on confidence intervals for small proportions indicates that Wilson confidence intervals typically produce better estimates than normal approximations of confidence intervals for small proportions (Newcombe, 1998; Vollset, 1993; Wilson, 1927). These results include weights for age, education, and province.

Figure 1 displays the estimated weighted percentages identifying as non-binary for age, sexual identity, education, income, country of birth, racial identity groups (white, Indigenous, and racialized), mother tongue, and region with 95 percent Wilson confidence intervals separately for the 2019 other gender respondents and the 2021 non-binary respondents. The weights include age, education, and province. By construction, the weights do not include respondents from the territories (Nunavut, the Northwest Territories, and Yukon). We include a horizontal line at the weighted percentage of other gender or non-binary respondents in each year (0.30 in 2019 and 0.25 in 2021). We set the scale of each panel so that the 2019 and 2021 estimates are comparable.

The estimates in Figure 1 largely confirm the expectations from Table 3, but some of

---

[10]More research is necessary to explain this pattern. It is notable that non-binary people challenge the embedded assumption in French grammar that is possible to categorize everyone into masculine and feminine gender categories.

Figure 1: Weighted Percentage Other Gender/Non-Binary with 95 Percent Wilson Confidence Intervals, by Demographic Variables, 2019 and 2021 CES
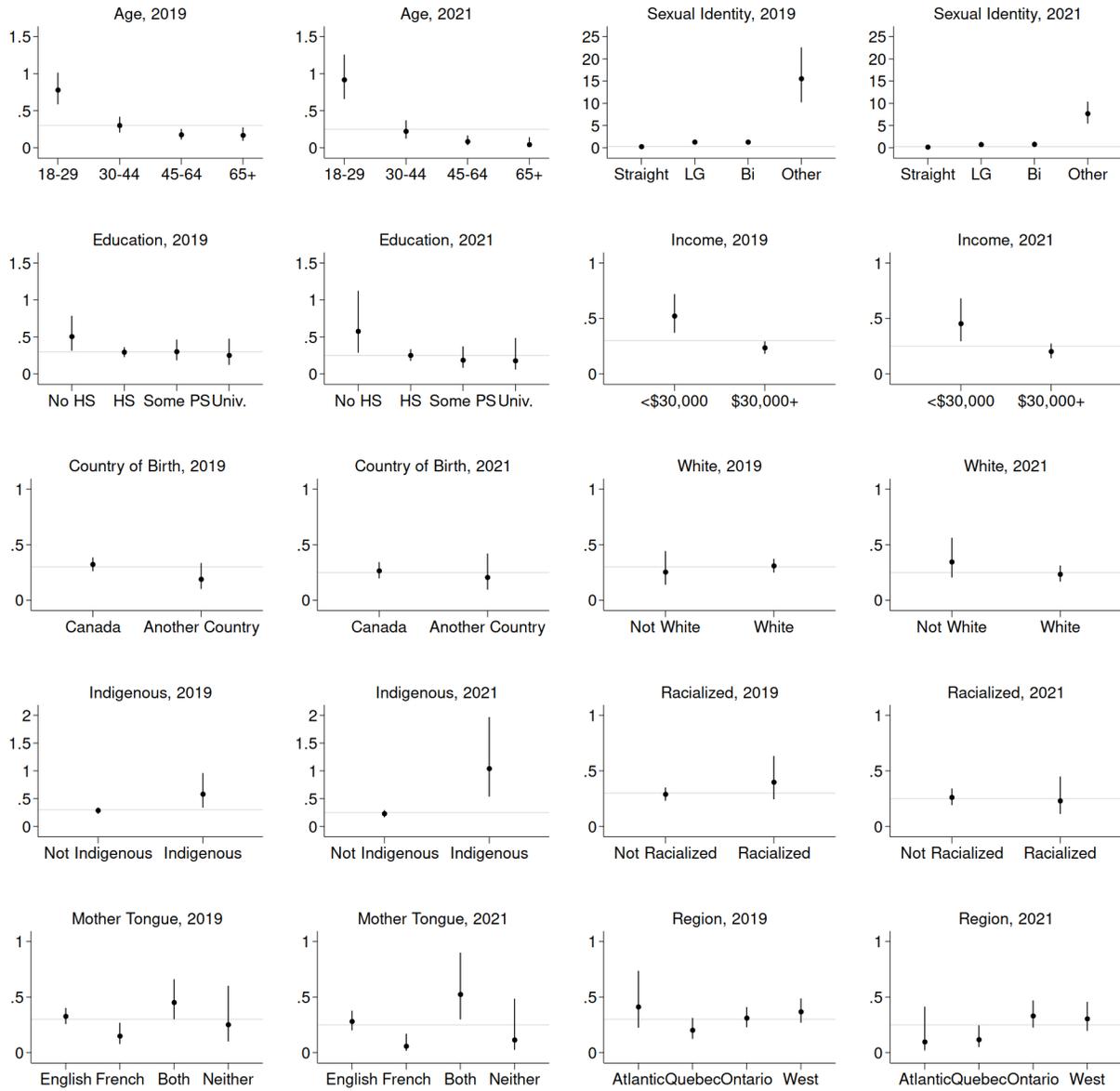
the differences between other gender or non-binary respondents and the general population appears to be due to the survey mode. Both surveys suggest that non-binary people are more likely to young, lesbian, gay, bisexual, or other sexual identities, Indigenous, and native speakers of both English and French and less likely to be native speakers of French (and not English) and residents of Quebec than men and women. However, the results for education and income are actually the reverse of what we would expect from Table 3. Non-binary people are less likely to have finished high school than men and women in both the 2019 and the 2021 CES. Similarly, non-binary people are more likely to be low-income (under $30,000$) than men and women in both the 2019 and the 2021 CES. These differences from Table 3 exist because the 2019 and 2021 CES online panels generally over-represent respondents of higher educational attainment and income.

Figure 2: Weighted Percentage Other Gender/Non-Binary with 95 Percent Wilson Confidence Intervals, by Religion, 2019 and 2021 CES
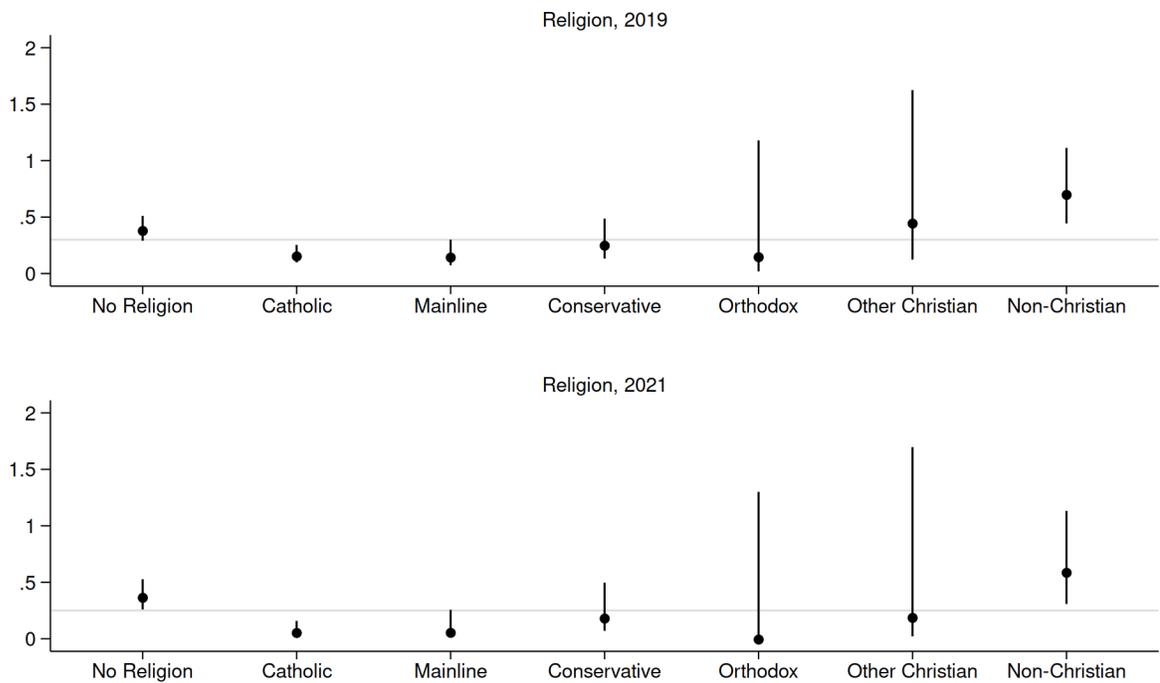
In Figure 2, we construct a similar plot of the weighted percentage of respondents who identify as non-binary for religion. We plot seven categories for religion: Atheist, Agnostic, or No religion; Catholic; Mainline Protestant; Conservative/Evangelical Christian;

Eastern Orthodox; Other Christian; and Non-Christian Religions.[11] We do not plot two other categories in our data ("Spiritual, not religious" and "Other") because these categories have small other gender/non-binary sub-samples that skew the scale of the graph, particularly given that two percent of both groups identify as non-binary. In both 2019 and 2021, other gender/non-binary respondents are more likely to identify as atheist, agnostic, or no religion; non-Christian; spiritual, "not religious"; and other religion. They are less likely to identify with all Christian religions, although these results are only significant for Catholicism or Mainline Protestant denominations.

We turn our attention to attitudes toward lesbians and gay men (LGs). Unfortunately, the only items available are in the post-election wave, which has substantial survey attrition. The post-election wave has two relevant items – a feeling thermometer toward "gays and lesbians" (0-100) and "how much should be done for gays and lesbians?" (much more, somewhat more, about the same as now, somewhat less, much less), which we rescale from 0-1, where 0 is "much less" and 1 is "much more" (that is, 0 is negative and 1 is positive toward LGs).[12] In Figure 3, we plot each respondent's feeling thermometer rating of "gays and lesbians" by gender identity and year. Here, we do not use multiple imputation to deal with missing data only pairwise deletion. We jitter the values of the feeling thermometer along the x-axis to make it easier to see the individual points. We include a horizontal line at the mean for each gender identity category. Figure 3 indicates that, in 2019, women and other gender respondents are more positive toward LGs than men are. By contrast, in 2021, non-binary people are more positive than women, who are more positive than men. A two-sided t-test indicates that 2019 other gender respondents do not rate LGs significantly differently from women ($d = 2.0$, $p = 0.55$), while the 2021 non-binary respondents are significantly more positive toward LGs than women are ($d = 13.8$, $p = 0.00$). More importantly, a two-sided t-test indicates that the 2021 non-binary respondents have significantly more positive ratings of LGs than the 2019 other gender respondents ($d = 9.3$, $p = 0.03$). Again, we can explain this discrepancy between the 2019 and the 2021 results using the information that binary trans people provide lower feeling thermometer responses toward lesbians and gay men than non-binary people do ($d = -11.1$, $p = 0.01$). These results are substantively similar using responses to "How much should be done for gays and lesbians?" However, the difference between binary

---

[11]Mainline Protestant refers to the "big five" Protestant denominations in Canada – Anglican, Baptist, Lutheran, Presbyterian, and United Church. Of these, Baptists are the most likely to be theologically conservative, while the others tend to be theologically liberal. Conservative/Evangelical Christian refers not only to evangelitcal or born again Christians but also a number of other Christian religious groups that maintain conservative beliefs about gender and sexuality, including Jehovah's Witnesses, Mormons, the Salvation Army, and so forth.

[12]Although the item refers to "gays and lesbians, we generally write "lesbians and gay men" rather than "gays and lesbians" because the established practice in the LGBTQ2S+ movement is to treat gay as an adjective, rather than a noun, and to list lesbians before gay men to address gender inequalities within the movement. We only use the wording "gays and lesbians" when referencing the exact question wording.

Figure 3: Feeling Thermometer Toward "Gays and Lesbians," by Gender Identity, 2019 and 2021 CES
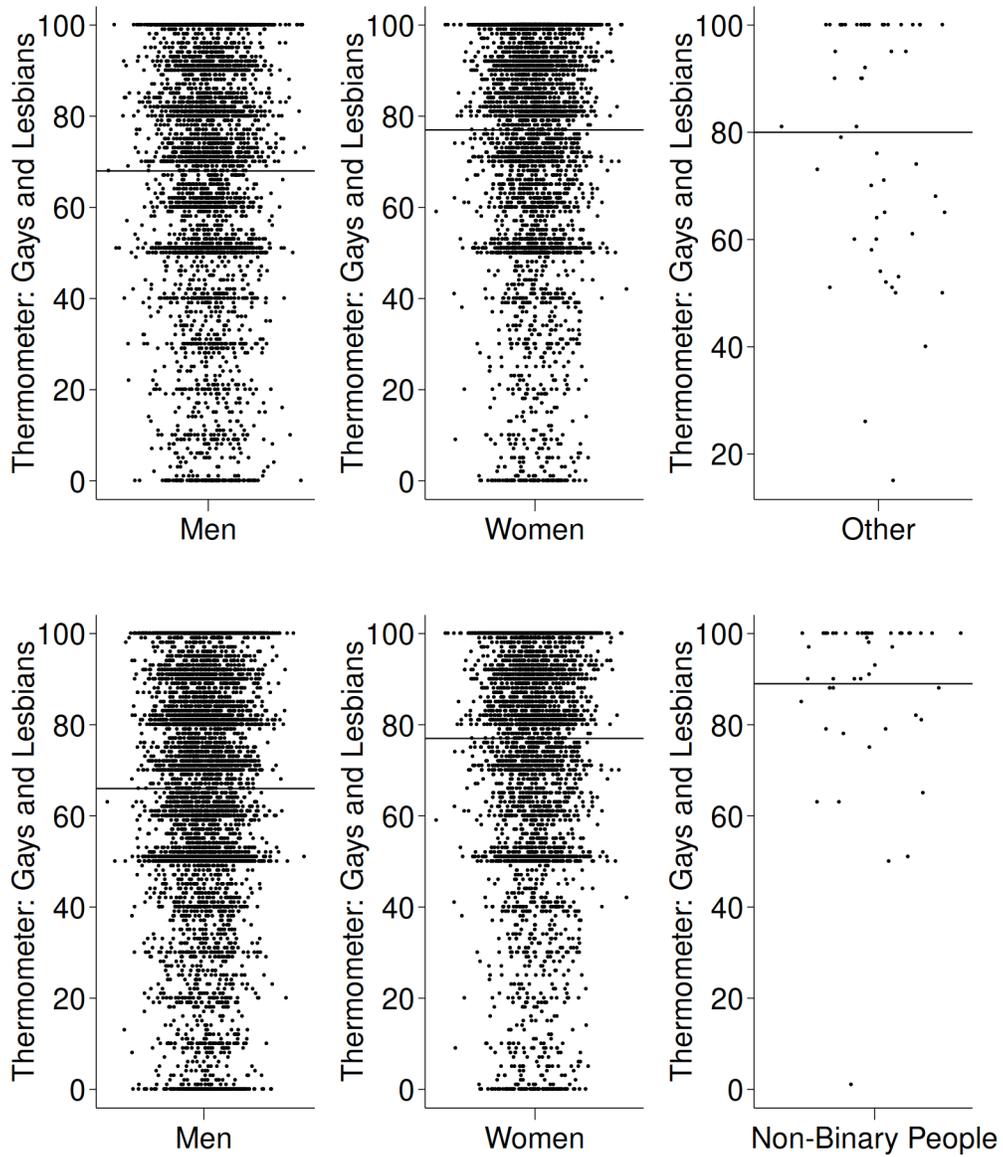
trans and non-binary respondents in 2021 falls short of statistical significance.

Figure 4: Predicted Percentage "Other" Gender from Weighted Logistic Regressions Without and With Sexual Identity, 2019 CES
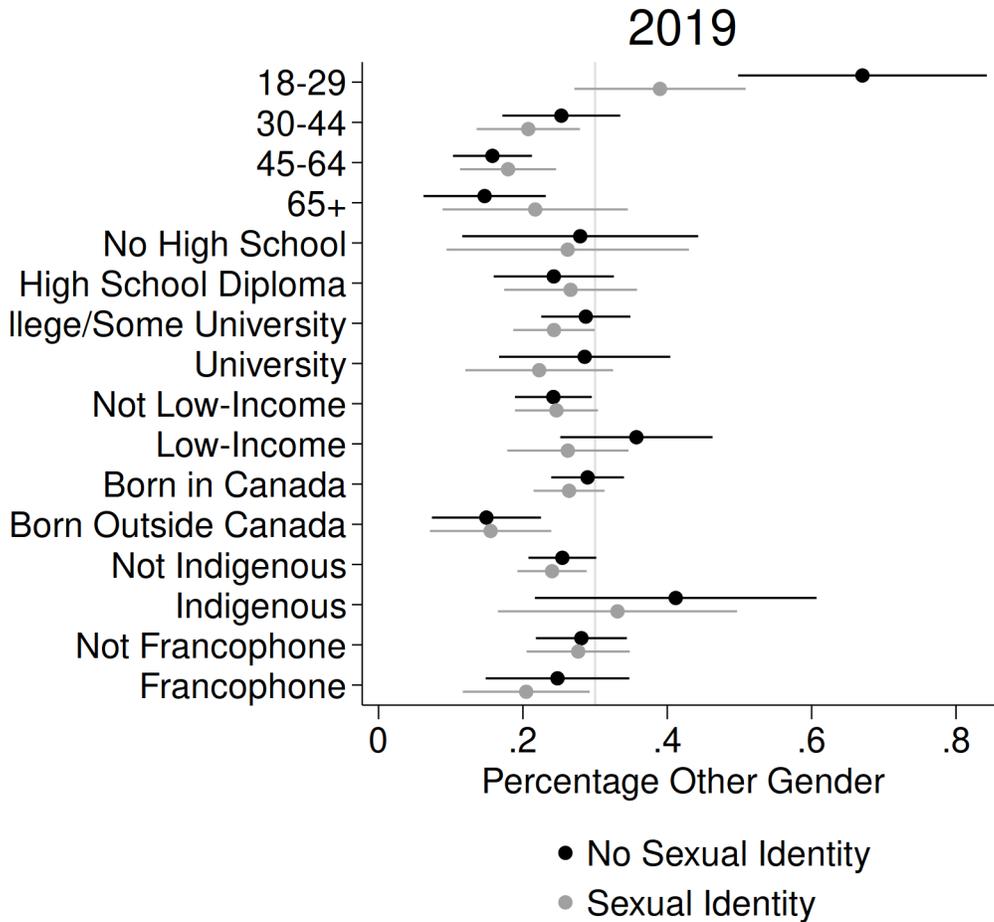
Finally, we model other gender/non-binary identity as a function of demographic variables. In Model 1, we present estimates of the percentage of other gender respondents in 2019 and non-binary people in 2019 across age, education, income, country of birth, Indigenous identity, French mother tongue, and province or territory of residence using weighted logistic regression. We do not include additional racial identity and mother tongue variables due to multicollinearity with country of birth, and we include education and province primarily because we weight along these variables. In Model 2, we add in sexual identity. In Figures 4 and 5, we display coefficient plots for all the variables except

Figure 5: Predicted Percentage Non-Binary from Weighted Logistic Regressions Without and With Sexual Identity, 2021 CES
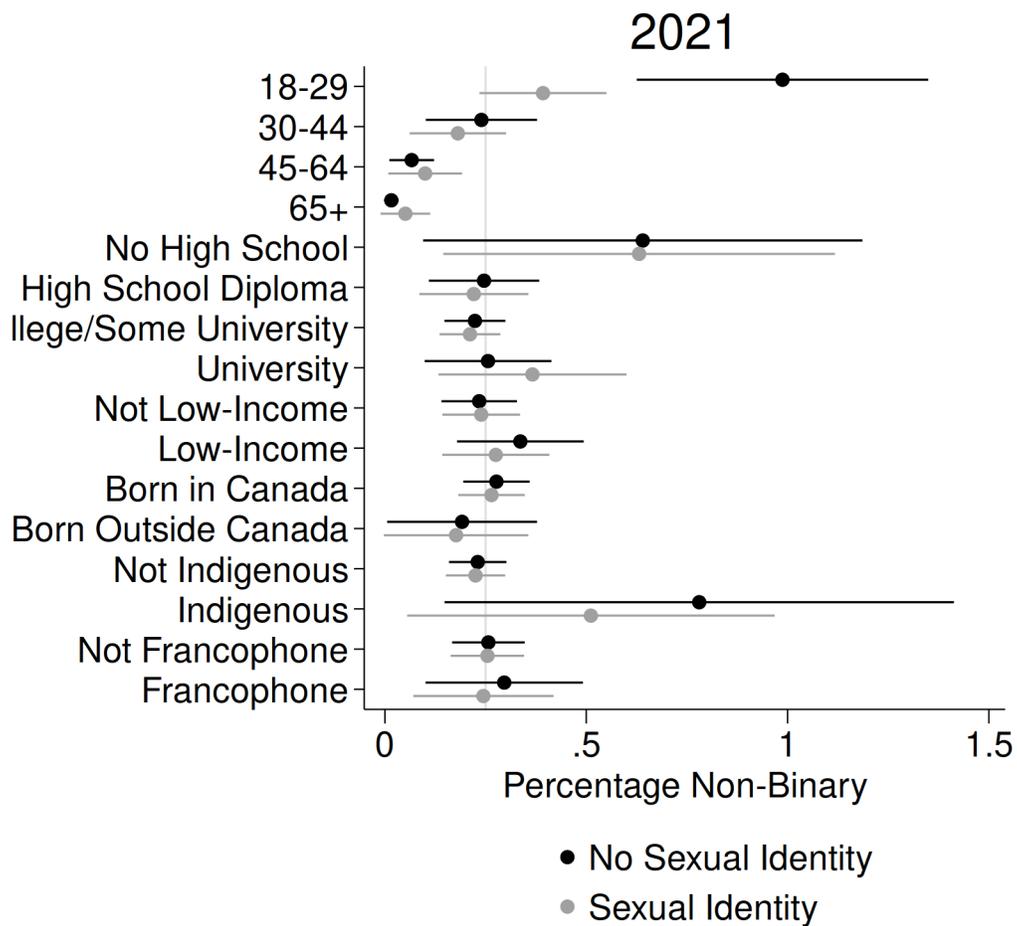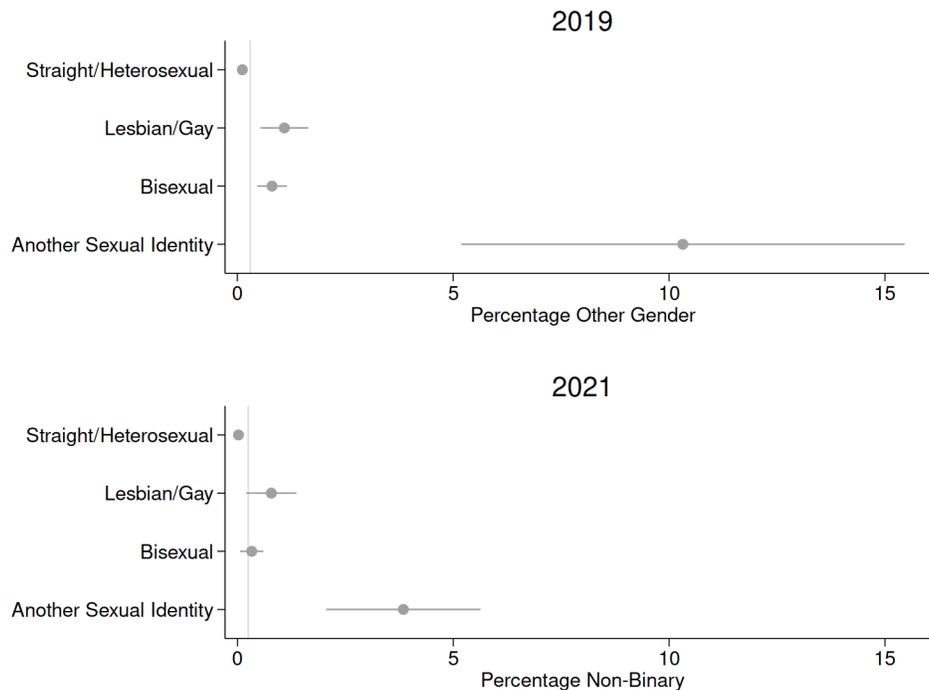
Figure 6: Predicted Percentage Other Gender/Non-Binary by Sexual Identity from Weighted Logistic Regressions with Other Demographic Controls, 2019 and 2021

for sexual identity and province of residence for the 2019 and 2021 CES. We show the results for sexual identity in both years separately in Figure 6 because the sexual identity variables do not fall on the same scale as the remaining variables.

Figures 4 and 5 show some broad similarities. Figure 4 shows that, in 2019, age is strongly predictive of responding as "Other," though this relationship becomes attenuated in Model 2 (with sexual identity included). There are minimal differences in selecting "Other" across education across both Model 1 and Model 2. Low-income respondents are more likely to select "Other" in Model 1, but this result disappears in Model 2. In both models, respondents born in Canada are more likely to select "Other" than those born in other countries. In both models, Indigenous respondents are more likely to select "Other," but there are too few Indigenous respondents to have much certainty over these estimates. The negative relationship between having French as first language and select "Other" increases when adjusting for sexual identity. While the results are generally similar in Figure 5, the relationship between age and non-binary identity is even stronger in 2021. Most of the other remaining variables show no clear, though the predicted percentages of non-binary people are notably higher (over 0.5 percent) for respondents who have not

22

completed high school and Indigenous respondents.

Figure 6 shows that the most substantial differences between the 2019 and the 2021 models are in the percentage of other gender/non-binary people by sexual identity. In 2019, the predicted percentage other gender is 0.1 for straight/heterosexual respondents, 1.1 for LG respondents, 0.8 for bisexual respondents, and 10.3 for respondents who report any other sexual identity. In 2021, the predicted percentages non-binary are 0.02 for straight/heterosexual respondents, 0.8 for LG respondents, 0.3 for bisexual respondents, and 3.8 for respondents who report any other sexual identity. The 2019 results inflate the estimates of the percentage of non-binary people across sexual identities – by a factor of five for straight/heterosexual respondents, a factor of 1.4 for LG respondents, a factor of 2.7 for both bisexual respondents and respondents of other sexual identities. The inflated estimates for straight/heterosexual respondents are likely due to the inclusion of binary trans people in the 2019 CES other gender category.

These models are substantively similar using unweighted logistic regression, penalized logistic regression (Firth, 1993), and unweighted and weighted rare events logistic regression (King and Zeng, 2001). Penalized and rare events logistic regression both address concerns about biased coefficients in cases of "rare events" (a small percentage of observations take the value of 1 on the outcome variable), as in the case in modelling non-binary identity.[13]

Next Steps: Multiple Imputation of 2019 Respondents' Responses to the 2021 CES Questions

As a next step, we plan to construct counterfactual estimates of how the 2019 respondents would have responded to the 2021 gender identity and trans identity questions if they had participated in the 2021 CES by treating the ambiguity in the 2019 CES as a missing data problem. We address this missing data problem using multiple imputation with chained equations. We assume that the gender identity and trans identity data are missing at random, provided we make certain assumptions about the 2019 respondents.

We plan to use two main models. In the first model, we use the entire pooled 2019 and 2021 dataset. For the 2019 respondents, we treat the 2021 gender identity variable as missing only for 2019 respondents who are neither men nor women *and* do not provide gender identity information in the open-ended responses to the sexual identity question. If 2019 respondents self-identify as non-binary, genderqueer, or other gender identities

---

[13]In subsequent drafts, we plan to use Firth's penalized logistic regression to present the main results. However, while the coefficients of penalized logistic regression estimates are unbiased in cases of rare events, the predicted probabilities are biased toward 0.5. As a result, we need to implement a bias-correction to the predicted probabilities from the models (Puhr et al., 2017).

than man or woman, we code them as non-binary. Similarly, we code respondents who identify as trans men and women as men and women, respectively, in this recoded gender identity variable. That is, we assume that (1) men and women will respond as men and women, respectively, to both the 2019 and the 2021 gender identity questions and (2) respondents who provide gender identity responses to the open-ended sexual identity response categories fall into the gender identity, transgender identity, and Two-Spirit identity categories they mention in their open-ended response. For the 2021 respondents, we use the 2021 gender identity variable as is, which has 12 missing values. We will model missing data gender identity and transgender identity jointly with other variables that have missing data and are plausbily related to these variables.

In the second model, we restrict the sample to the 2019 CES other gender respondents and the 2021 CES binary trans and non-binary respondents. There are several variables that discriminate between binary trans respondents and non-binary respondents in bivariate analyses of the 2021 data, including age, education, income, mother tongue, country of birth, sexual identity, martial status, and federal party identification. We plan to use these variables (and others) to impute whether each respondent is a binary trans person or a non-binary person. This approach makes the assumption that every 2019 CES other gender respondent is either a binary trans person or non-binary. As in the first model, we will model whether the respondent is a binary trans person or a non-binary person jointly with other variables that have missing data and can potentially explain this outcome of interest.

In both models, we assume that the relationships between predictor variables, gender identity, and trans identity identity are similar over time between 2019 and 2021. This assumption is almost certainly not exactly correct. However, we believe that the field dates of the surveys are close enough to make these relationships comparable.

We then intend to replicate the analyses of the correlates of non-binary identity using the multiply imputed data.

Appendix A: Description of Variables

Table 4: Description of Variables Based on Shared Questions, 2019 and 2021 CES

Variable	Question Wording	Coding
Age	To make sure we are talking to a cross section of Canadians, we need to get a little information about your background. First, in what year were you born? [If respondent is born 18 years before the election:] How old are you? 1. 17 [screened out] 2. 18	We use three different codings of age: raw age, a four-category option (1 = 18-29, 2 = 30-44, 3 = 45-64, 4 = 65+), and a six-category option (1 = 18-19, 2 = 20-24, 3 = 25-34, 4 = 35-49, 5 = 50-64, 6 = 65+). We use the four-category variable in regressions to model non-binary identity and the six-category variable to compare the demographics of non-binary people in the 2019 and the 2021 CES with Trans PULSE and the 2016 Census of Canada.

(Continued on next page)

Table 4 – Continued from previous page

Variable	Question Wording	Coding
Education	What is the highest level of education that you have completed? 1. No schooling 2. Some elementary school 3. Completed elementary school 4. Some secondary/high school 5. Completed secondary/high school 6. Some technical, community college, CEGEP, College Classique 7. Completed technical, community college, CEGEP, College Classique 8. Some university 9. Bachelor's degree 10. Master's degree 11. Professional degree or doctorate 12. Don't know/Prefer not to answer	We use two main codings of education. For regressions, we use a standard education coding (1 = Below High School, 2 = High School Diploma, 3 = Some College or University, 4 = Bachelor's or Higher). For comparison with the Census and Trans PULSE, we code a four-category education variable (1 = Below High School, 2 = High School Diploma, 3 = College or University, 4 = Graduate or Professional Degree).

Table 4 – Continued from previous page

Variable	Question Wording	Coding
Income	What was your total household income, before taxes, for the year [2018/2020]? Be sure to include income from all sources, to the nearest thousand dollars. *For example, if your household had a total before-tax income of $71,336 in 2020, you would enter 71000.* We don't need the exact amount; does your household income fall into one of these broad categories? 1. No income 2. $1 to $30,000 3. $30,001 to $60,000 4. $60,001 to $90,000 5. $90,001 to $110,000 6. $110,001 to $150,000 7. $150,001 to $200,000 8. More than $200,000 9. Don't know/Prefer not to answer	We use two income variables. For bivariate analyses of the correlates of non-binary identity, we use an eight-category income variable (0 = No Income; 2 = Under 30,000; 3 = 30,001 to 60,000; 4 = 60,001-90,000; 5 = 90,001-110,000; 6 = 110,001-150,000; 7 = 150,001-200,000, 8 = 200,000+).

Table 4 – Continued from previous page

Variable	Question Wording	Coding
Province/Territory	In which province or territory are you currently living? 1. Alberta 2. British Columbia 3. Manitoba 4. New Brunswick 5. Newfoundland and Labrador 6. Northwest Territories 7. Nova Scotia 8. Nunavut 9. Ontario 10. Prince Edward Island 11. Quebec 12. Saskatchewan 13. Yukon	We recode the provinces in order from east to west, then the territories from west to east, following the numerical scheme used by Elections Canada (1 = Newfoundland and Labrador, 2 = Nova Scotia, 3 = Prince Edward Island, 4 = New Brunswick, 5 = Quebec, 6 = Ontario, 7 = Manitoba, 8 = Saskatchewan, 9 = Alberta, 10 = British Columbia, 11 = Yukon, 12 = Northwest Territories, 13 = Nunavut).
Region	See Province/Territory	We code respondents' region based on their province or territory of residence (1 = New Brunswick, Newfoundland and Labrador, Nova Scotia, Prince Edward Island; 2 = Quebec; 3 = Ontario; 4 = Alberta, British Columbia, Manitoba, Saskatchewan).
User Language	N/A	A binary variable that indicates the language of the questionnaire selected by each respondent (0 = English, 1 = French).

Table 4 – Continued from previous page

Variable	Question Wording	Coding
Mother Tongue	Which language(s) did you learn as a child and still understand today? (Select all that apply) 1. English 2. French 3. Indigenous language (please specify) 4. Arabic 5. Chinese, Cantonese, Mandarin 6. Filipino/Tagalog 7. German 8. Indian, Hindi, Gujarati 9. Italian 10. Korean 11. Pakistani, Punjabi, Urdu 12. Persian, Farsi 13. Russian 14. Spanish 15. Tamil 16. Vietnamese 17. Other (please specify) 18. Don't know/Prefer not to answer	For regression analysis, we rely on a series of dummy variables indicating whether the respondent is an anglophone, a francophone, or an allophone (a native speaker of another language). For descriptive analyses, we construct a seven-category mother tongue variable (1 = English Only; 2 = French Only; 3 = Another Language Only; 4 = Both English and French; 5 = Both English and Another Language; 6 = Both French and Another Language; 7 = English, French, and Another Language). Finally, for comparison with Trans PULSE and the Census, we use a four-category variable based on whether each respondent is a native speaker of Canada's official languages (1 = English, 2 = French, 3 = Both English and French, 4 = Neither English nor French).

Table 4 – Continued from previous page

Variable	Question Wording	Coding
Born Outside Canada	Were you born in Canada? 1. Yes 2. No 3. Don't know/Prefer not to say	We code a binary variable (0 = Born in Canada, 1 = Born outside Canada). We recode "Don't know/Prefer not to say" to missing.
Non-Citizen	Are you a... 1. Canadian citizen 2. Permanent resident 3. Other [screened out]	We code a binary variable (0 = Citizen, 1 = Not a citizen).

Table 4 – Continued from previous page

Variable	Question Wording	Coding
Religion	Please indicate your religion, if you have one? 1. None/Don't have one/Atheist 2. Agnostic 3. Buddhist/Buddhism 4. Hindu 5. Jewish/Judaism/Jewish Orthodox 6. Muslim/Islam 7. Sikh/Sikhism 8. Anglican/Church of England 9. Baptist 10. Catholic/ Roman Catholic/ RC 11. Greek Orthodox/ Ukrainian Orthodox/Russian Orthodox/Eastern Orthodox 12. Jehovah's Witness 13. Lutheran 14. Mormon/Church of Jesus Christ of the Latter Day Saints 15. Pentecostal/ Fundamentalist/Born Again/Evangelical 16. Presbyterian 17. Protestant 18. United Church of Canada 19. Christian Reformed 20. Salvation Army 21. Mennonite 22. Other (please specify) 23. Don't know Prefer not to answer	We recode religion into a nine-category variable (1 = Atheist, Agnostic, or No Religion; 2 = Catholic; 3 = Mainline Protestant; 4 = Evangelical/Conservative Christian; 5 = Eastern Orthodox; 6 = Other Christian; 7 = Non-Christian Religions; 8 = Spiritual, Not Religious; 9 = Other). We not only recode the close-ended responses but also the open-ended responses into these categories.

Table 4 – Continued from previous page

Variable	Question Wording	Coding
Married	Are you presently married, living with a partner, divorced, separated, widowed, or have you never been married? 1. Married 2. Living with a partner 3. Divorced 4. Separated 5. Widowed 6. Never Married 7. Don't know/Prefer not to answer	We code a binary variable (0 = Not Married, 1 = Married)
Children	How many children, if any, do you have? [Number entry] Don't know	We code a binary variable (0 = No Children, 1 = One or More Children).
Party Identification	In federal politics, do you usually think of yourself as a: 1. Liberal 2. Conservative 3. NDP 4. Bloc Québécois [for Quebec respondents only] 5. Green 6. Another party (please specify) 7. None of these 8. Don't know/Prefer not to answer	We recode this variable into an eight-category variable (1 = Liberal, 2 = Conservative, 3 = NDP, 4 = Bloc québécois, 5 = Green, 6 = People's Party, 7 = Other, 8 = None). We identify People's Party identifiers using the open-ended responses. We recode "Don't know/Prefer not to answer" to missing.

Table 4 – Continued from previous page

Variable	Question Wording	Coding
Gay/Lesbian Feeling Thermometer	How do you feel about the following groups in Canada? Set the slider to any number from 0 to 100, where 0 means you really dislike the group and 100 means you really like the group. If you do not know, or prefer not to answer, please click → Gays and lesbians [Slider from 0-100 with starting at 50]	We use the raw thermometer values.
More Done	How much do you think should be done for gays and lesbians? 1. Much more 2. Somewhat more 3. About the same as now 4. Somewhat less 5. Much less 6. Don't know/Prefer not to answer	We rescale the variable from 0-1, where 0 = Much less, 0.25 = Somewhat less, 0.5 = About the same as now, 0.75 = Somewhat more, and 1 = Much more. We recode "Don't Know/ Prefer not to answer" as missing.

Table 4 – Continued from previous page

Variable	Question Wording	Coding
Community Size	Do you live in... 1. A rural area or village (less than 1000 people) 2. A small town (more than 1000 people but less than 15K) 3. A middle-sized town (15K-50K people) not attached to a city 4. A suburb of a large town or city 5. A large town or city (more than 50K people) 6. Don't know/Prefer not to answer	We use the categories as is, except we recode "Don't know/Prefer not to answer" to missing.

Table 5: Description of Variables Based on Different Questions, 2019 and 2021 CES

Variable	Question Wording, 2019	Question Wording, 2021	Coding
Sexual Identity	Do you consider yourself to be: 1. Heterosexual 2. Homosexual 3. Bisexual 4. Other (open-ended) 5. Don't know 6. Prefer not to say	Which of the following best represents how you think of yourself? 1. Straight or heterosexual 2. Gay or lesbian 3. Bisexual 4. Queer 5. Something else (open-ended) 6. I am not sure yet 7. I don't know what this question means 8. Prefer not to answer	For the sake of comparability, we construct a four-category sexual identity variable: 1. Straight or heterosexual 2. Gay or lesbian 3. Bisexual 4. Another sexual identity (queer, pansexual, asexual, etc.)

Table 5 – Continued from previous page

| Variable	Question Wording, 2019	Question Wording, 2021	Coding
Race	Please select all that apply. Please click the forward arrow → below once you are done. 1. Aboriginal/First Nations 2. British 3. Chinese 4. Dutch 5. English 6. French 7. French Canadian 8. German 9. Hispanic 10. Indian 11. Inuk/Inuit 12. Irish 13. Italian 14. Métis 15. Polish 16. Québécois 17. Scottish 18. Ukranian 19. Other 1 (please specify) 20. Other 2 (please specify) 21. Don't know/ Prefer not to answer	Do you identify as any of the following? (Please select all that apply) 1. Arab 2. Asian 3. Black 4. Indigenous (e.g., First Nations, Métis, Inuit, etc.) 5. Latino/Latina 6. South Asian (e.g., East Indian, Pakistani, Sri Lankan, etc.) 7. Southeast Asian (e.g., Vietnamese, Cambodian, Laotian, Thai, etc.) 8. West Asian (e.g., Iranian, Afghan, etc.) 9. White 10. Other (please specify) 11. None of the above 12. Prefer not to answer	Since individuals can have multiple racial backgrounds, we construct three separate binary variables (White, Indigenous, Racialized). In 2019, we code British, Dutch, English, French, French Canadian, German, Irish, Italian, Polish, Scottish, and Ukrainian as white; Aboriginal/First Nations, Inuk/Inuit, and Métis as Indigenous; and Chinese, Hispanic, and Indian as racialized. In 2021, we code White as white; Indigenous as Indigenous; and the remaining close-ended responses as Racialized. We also code open-ended responses into these categories. In all cases, we recode respondents who only select "Don't know/Prefer not to answer" as missing.

Appendix B: Coding of Open-Ended Gender Identity Responses, 2021 CES

Table 6: Recoding of Open-Ended Gender Responses, 2021 CES

Recoding	Open-Ended Responses
Man (7)	"Male" (2), "Genetic Male" (1), "Masculin" ["male" in French] (1), "Transman" (1), "There are only two valid genders, as listed in Genesis 5:2, and mine is male." (1), "I am a biological male in accordance with Genesis 1:27" (1)
Woman (8)	"Female" (3), "Woman" (1), "Femme" ["woman" in French] (1), "Femne" [typo of "woman" in French] (1), "Femme queer" ["queer woman" in French] (1), "I am offended that you are asking for 'gender,' an ideological position, instead of 'sex,' an immutable physical trait. I am a woman, which is not a gender, it is an adult human of the female sex." (1)
Non-Binary (9)	"Aucun" ["none [of the above]" in French] (2), "Two Spirited" (1), "Genderfluid" (1), "IDK" (1), "Bi-gender" (1), "Gender apathetic" (1), "Agender" (1), "Gender-queer" (1),
Missing (11)	"No" (2), "Moon helicopter" (1), "Funny clown" (1), "Dodge Ram" (1), "31" (1), "Pokemon" (1), "Monkey" (1), "Licorne à pois" ["spotted unicorn" in French] (1), "Human" (1), "Spiritual being having a human experience" (1)
Total (35)	

Appendix C: The Percentage of Non-Binary People in Survey Samples, 2015-2021

Figure 7: Scatterplot of Respondent Shares, 2015 LPP, 2019 CES, and 2021 CES
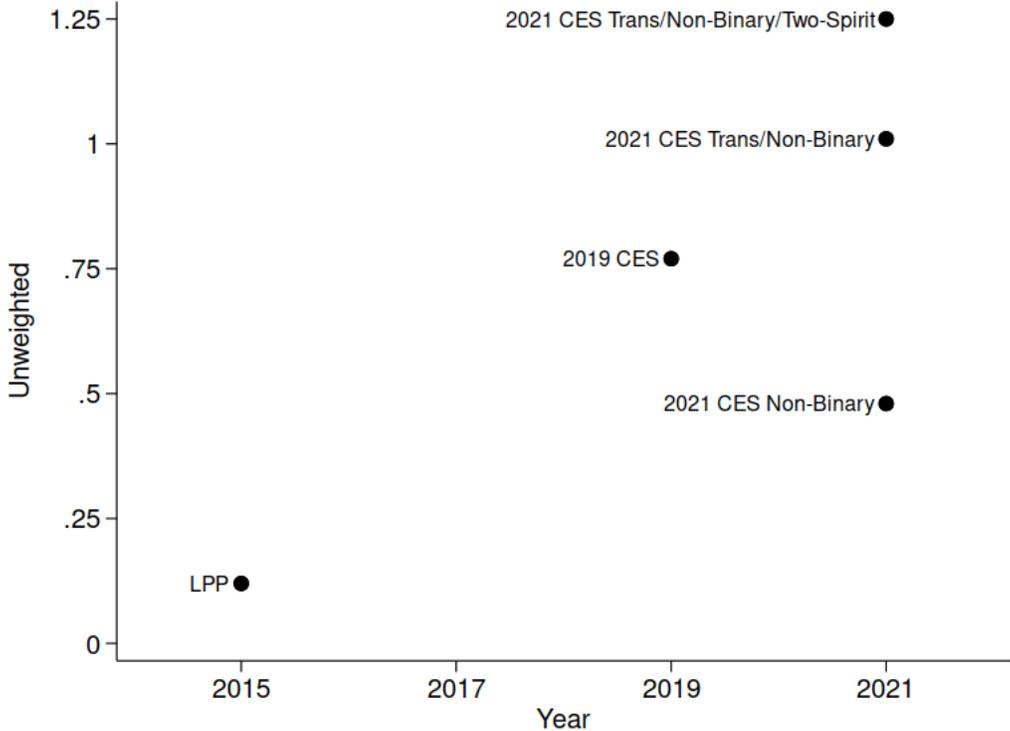
Appendix D: Using Bayes' Rule to Estimate the Percentage of Binary Trans People in the 2019 CES "Other" Respondents

Let B be a dummy variable that indicates binary trans respondents (as opposed to non-binary respondents) and S be a dummy variable that indicates straight or heterosexual respondents. We begin with Bayes' Rule:

$$Pr(B|S) = \frac{Pr(S|B) * Pr(B)}{Pr(S)} \tag{1}$$

We can reformulate this equation into the following equation:

$$Pr(B) = \frac{Pr(B|S) * Pr(S)}{Pr(S|B)} \tag{2}$$

Then we plug in information from the 2019 CES:

$$Pr(B) = \frac{Pr(B|S) * 0.48}{Pr(S|B)} \tag{3}$$

Then we plug in the bivariate relationship estimates for $Pr(B|S)$ and $Pr(S|B)$ from a restricted sample of binary trans and non-binary respondents the 2021 CES, as a potential comparison group to the 2019 CES "Other" respondents:

$$Pr(B) = \frac{0.86 * 0.48}{0.68} = 0.6 \tag{4}$$

This yields the results that 60 percent of the other gender respondents are likely to be binary trans people, rather than non-binary people.

References

Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.

Ashley, Florence. 2021. "'Trans' is My Gender Modality: A Modest Terminological Proposal." In *Trans Bodies, Trans Selves*. Second ed. Oxford, UK: Oxford University Press.

Badgett, M. V. Lee, Kellan E. Baker, Kerith J. Conron, Gary J. Gates, Alison Gill, Emily Greytak, and Jody L. Herman. 2014. *Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys*. Los Angeles, CA: The Williams Institute.

Bauer, Greta. 2020*a*. *Health and Health Care Access for Trans & Non-Binary People in Canada: National, Provincial, and Territorial Results*. London, Ontario: Trans PULSE Canada.

Bauer, Greta. 2020*b*. *Non-Binary People and Identity Documents: A Report Prepared for the Saskatchewan Human Rights Commission*. London, Ontario: Trans PULSE Canada Team.

Bauer, Greta R, Jessica Braimoh, Ayden I Scheim, and Christoffer Dharma. 2017. "Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations." *PloS one* 12(5): e0178043.

Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80(1): 27–38.

Fraser, Gloria. 2018. "Evaluating Inclusive Gender Identity Measures for Use in Quantitative Psychological Research." *Psychology & Sexuality* 9(4): 343–357. Publisher: Taylor & Francis.

Fraser, Gloria, Joseph Bulbulia, Lara M Greaves, Marc S Wilson, and Chris G Sibley. 2020. "Coding Responses to an Open-Ended Gender Measure in a New Zealand National Sample." *The Journal of Sex Research* 57(8): 979–986.

Holzberg, Jessica, Renee Ellis, Matt Virgile, Dawn Nelson, Jennifer Edgar, Polly Phipps, and Robin Kaplan. 2017. *Assessing the Feasibility of Asking about Gender Identity in the Current Population Survey: Results from Focus Groups with Members of the Transgender Population*. Washington, DC: Bureau of Labor Statistics.

James, Sandy E., Jody L. Herman, Susan Rankin, Mara Keisling, Lisa Mottet, and Ma'ayan Anafi. 2016. *The Report of the 2015 U.S. Transgender Survey*. Washington, DC: National Center for Transgender Equality.

Kidd, Kacie M., Gina M. Sequeira, Claudia Douglas, Taylor Paglisotti, David J. Inwards-Breland, Elizabeth Miller, and Robert W. S. Coulter. 2021. "Prevalence of Gender-Diverse Youth in an Urban School District." *Pediatrics* 147(6): e2020049823.

King, Gary, and Langche Zeng. 2001. "Logistic regression in rare events data." *Political analysis* 9(2): 137–163. Publisher: Cambridge University Press.

Medeiros, Mike, Benjamin Forest, and Patrik Öhberg. 2020. "The Case for Non-Binary Gender Questions in Surveys." *PS: Political Science & Politics* 53(1): 128–135.

Navarro, Jose, Nathan Lachowsky, Rebecca Hammond, Drew Burchell, Frederic S. E. Arps, Charlie Davis, Jacq Brasseur, Shaz Islam, Bretton Fosbrook, Hanna Jacobsen, Mattie Walker, Carol Lopez, Ayden Scheim, and Greta Bauer. 2021. *Health and Well-Being Among Non-Binary People*. Number 7 *in* "Trans PULSE Report" London, Ontario: Trans PULSE Canada.

Newcombe, Robert G. 1998. "Two-sided confidence intervals for the single proportion: comparison of seven methods." *Statistics in medicine* 17(8): 857–872. Publisher: Wiley Online Library.

Perrella, Andrea M. L., Brown, Steven D., and Kay, Barry. 2019. "Profile of the Lesbian, Gay, and Bisexual Electorate in Canada." In *Queering Representation: LGBTQ People and Electoral Politics in Canada*. Vancouver: UBC Press , 51–79.

Perrella, Andrea M.L., Steven D. Brown, and Barry J. Kay. 2012. "Voting Behaviour among the Gay, Lesbian, Bisexual and Transgendered Electorate." *Canadian Journal of Political Science* 45(1): 89–117. Edition: 2012/05/01 Publisher: Cambridge University Press.

Pruden, Harlan, and Travis Salway. 2020. "'What and who is Two-Spirit?' in Health Research.".

Puhr, Rainer, Georg Heinze, Mariana Nold, Lara Lusa, and Angelika Geroldinger. 2017. "Firth's logistic regression with rare events: accurate effect estimates and predictions?" *Statistics in medicine* 36(14): 2302–2317. Publisher: Wiley Online Library.

Vollset, Stein Emil. 1993. "Confidence intervals for a binomial proportion." *Statistics in medicine* 12(9): 809–824. Publisher: Wiley Online Library.

Westbrook, Laurel, and Aliya Saperstein. 2015. "New Categories Are Not Enough: Rethinking the Measurement of Sex and Gender in Social Surveys." *Gender & Society* 29(4): 534–560.

Wilson, Bianca D. M., and Ilan H. Meyer. 2021. *Nonbinary LGBTQ Adults in the United States*. Los Angeles, CA: The Williams Institute.

Wilson, Edwin B. 1927. "Probable inference, the law of succession, and statistical inference." *Journal of the American Statistical Association* 22(158): 209–212. Publisher: Taylor & Francis.