Studying individuals in same-sex couples using longitudinal administrative data from Canadian tax records: Opportunities and challenges

Chih-lan Winnie Yang

Nicole Denier

Xavier St-Denis

Sean Waite

Contents

Studying individuals in same-sex couples using longitudinal administrative data from Canadian tax records: Opportunities and challenges

Chih-lan Winnie Yang[1]

Nicole Denier[2]

Xavier St-Denis[3]

Sean Waite[4]

Abstract

BACKGROUND
Quantitative research on the social, demographic, and economic outcomes of sexual minorities has long been hampered by data shortfalls, with most surveys and censuses limited by sample sizes and/or a lack of direct questions on sexual identity. The growing availability of administrative data presents an opportunity to fill some of these gaps.

OBJECTIVE
This article highlights the challenges and opportunities involved with using a novel administrative dataset – the Longitudinal Administrative Databank, which includes 20% of Canadian tax filers – to study sexual minority populations in Canada. We identify three sources of bias, propose strategies to adjust for this bias, and introduce a measure of "inferred sexual minority status" to improve the identification of sexual minorities in tax data.

RESULTS
Administrative tax data offers significant advantages, including a large sample size, high-quality income data for individuals and linked family members, a longitudinal design, and the ability to trace individuals' same-/different-sex partnership histories. Our adjustment strategies mitigate some biases in identifying same-sex couples, including underreporting, misclassification, and measurement errors. The estimated proportion of individuals in same-sex marriages closely aligns with Canadian census estimates from 2006–2021, while the proportion in same-sex common-law partnerships is

---

[1] McGill University, Montreal, Quebec, Canada. Email: chih-lan.yang@mail.mcgill.ca.
[2] University of Alberta, Edmonton, Alberta, Canada.
[3] Institut national de la recherche scientifique (INRS), Montreal, Quebec, Canada.
[4] University of Western Ontario, London, Ontario, Canada.

underestimated. Finally, our earnings gaps analyses highlight the utility of the inferred sexual minority status measure.

CONTRIBUTION

This article contributes to research on sexual minority data landscapes, offering new insights into the identification and measures of sexual minority populations using longitudinal administrative tax data. Our approach points to new opportunities for studying the long-term longitudinal income and family dynamics of sexual minority populations on the national level.

1. Introduction

Over the past few decades, the visibility of and rights accorded to gender and sexual minorities have expanded in many industrial democracies, in parallel with research on the social, health, and economic well-being of these population groups (Umberson et al. 2015; Velasco 2023). However, existing quantitative research, largely relying on health and social surveys and censuses, has several limitations. In this paper, we introduce a novel data source – the Longitudinal Administrative Databank (LAD) – based on administrative tax records in Canada, which allows us to longitudinally trace individuals who have ever been in a same-sex couple (SSC) or a different-sex couple (DSC) every year since 2000. As we argue, this large-scale, nationally representative longitudinal dataset unlocks new opportunities for scholars to study the income and family dynamics of sexual minority populations at the national level over an extended period. To begin with, we outline the data landscape of studies on sexual minorities before highlighting the strengths and limitations of the LAD.

Much quantitative research on sexual minority populations relies on two types of data, each with important shortcomings. First, survey data, most frequently from health and social surveys, may include questions on sexual orientation/identity and attraction, which provide a more direct way of analyzing various outcomes among those who identify as sexual (and/or gender) minorities (Badgett 2018; Carpenter and Eppink 2017; Waite and Denier 2015; Mize 2016; Waite, Pajovic, and Denier 2020). According to a 2019 report, only 15 OECD countries included at least one direct measure on the above aspects in nationally representative surveys carried out by official or public institutions, with the United States collecting this information in the greatest number of surveys (OECD 2019). For example, in the United States, the National Survey of Family Growth, the National Health and Nutrition Examination Survey, and the General Social Survey (GSS) all have questions on sexual orientation and same-sex sexual behaviors (Heslin and Alfier 2022; Mize 2016). In Canada, the Canadian Community Health Survey

(CCHS) and the GSS include these questions (Waite and Denier 2019). While surveys may have direct measures to capture different dimensions of sexuality (e.g., identity, attraction, behavior), they tend to have relatively small sample sizes for sexual minorities in each wave. Researchers usually need to pool multiple waves of repeated cross sections to obtain observations that render sufficient statistical power (Umberson et al. 2015). This cross-sectional design further limits researchers' ability to establish the temporal order of life events or to trace time-varying characteristics on the individual level (Elder 1994).

The second type of frequently used data is larger in scale, including nationally representative population censuses or large household or labor force surveys. While offering large sample sizes, this data commonly does not have direct measures of sexual identity, attraction, or behaviors. Instead, the sources offer partnership-based data or "couple data" (Waite and Denier 2019), where the identification of individuals in same-sex couples relies on coresidence with a partner of the same sex in the available cross section (Umberson et al. 2015). In other words, only partnered individuals can be selected for analyses, excluding currently single or otherwise non-cohabiting individuals. Moreover, as most of this data is cross-sectional, it does not allow for the identification of those who have partnership histories with both same-sex and different-sex partners, resulting in a potential bisexual erasure or ignoring fluidity in partnership practices, especially for women (Hu and Denier 2023; Mittleman 2023).

To be sure, several longitudinal surveys with direct questions about sexuality have become increasingly available and are being more frequently used by researchers. These include, for example, the National Longitudinal Study of Adolescent to Adult Health in the United States; Understanding Society – the UK Household Longitudinal Study; the Household, Income and Labour Dynamics survey in Australia; and the Canadian Longitudinal Study of Aging. However, these surveys often have small sample sizes for sexual minority subgroups by gender (e.g., gays, lesbians, bisexual men and women) (Joyner, Manning, and Bogle 2017; Sabia, Wooden, and Nguyen 2017) – especially after considering other intersectional identities (Umberson et al. 2015; Waite and Denier 2019) and attrition over time (Campbell, Perales, and Baxter 2020). Small sample sizes might further be constrained by a limited number of waves of longitudinal surveys (Joyner, Manning, and Bogle 2017; Sabia, Wooden, and Nguyen 2017).

Recently, scholars in other areas of social science have turned to increasingly available administrative data to address similar challenges. Comprehensive in providing official population-level records related to vital events, tax records, and social security and health records, administrative data is particularly beneficial for studies on segments of society or minority populations that are hard to reach, including sexual minorities (Card et al. 2010; Connelly et al. 2016). For example, studies on registered partnerships, marriages, dissolutions, and fertilities among SSCs in Scandinavian countries (e.g.,

Denmark, Sweden, Norway, and Finland[5]) are among the first to utilize population register data to paint comprehensive longitudinal portraits of SSCs in partnership and family dynamics (Aldén et al. 2015; Andersson et al. 2006; Andersson and Noack 2010; Boye and Evertsson 2021; Kolk and Andersson 2020; Möllborn and Kolk 2024; Noack, Seierstad, and Weedon-Fekjær 2005; van der Vleuten, Evertsson, and Moberg 2023). Similar register data from the Netherlands[6] has also been used to study same-sex unions (Jaspers, Mazrekaj, and Machado 2024; Kabátek and Perales 2021; Machado and Jaspers 2023). However, in some contexts, such as Denmark and Finland, only registered partners or married spouses can be identified, not cohabiting partners (Evertsson, Jaspers, and Moberg 2020). In Sweden, cohabiting partners can be identified through the new linkage with household data going back to 2011, but this includes only different-sex cohabiting couples (see Cantalini, Ohlsson-Wijk, and Andersson 2024; Möllborn and Kolk 2024).

In Canada and the United States, there is no national population register system designed to document and track the names, addresses, vital events, and other demographic data of the population. Instead, individual government agencies maintain most databases, which are compiled and merged through record linkage by statistical agencies. This administrative data has not been widely used to study sexual minority populations. In the United States, the small number of studies that drew on administrative data focused on estimating same-sex unions by marital status and geography (e.g., Badgett and Mallory 2014; Fisher, Gee, and Looney 2018) and earnings penalties (Downs et al. 2023).

In Canada, despite more than 30 series of administrative data sources and data linkages available for academic research, very few of them allow for the identification of sexual minorities. The main one is the T1[7] Family File (T1FF), a data source of *all* tax filers,[8] which includes a "same-sex couple flag" that identifies individuals in a common-law (cohabiting) or married SSC in every tax year since 2000 (more details below). T1FF is available to researchers through derived data products and record linkages such as the LAD, the Longitudinal Immigration Database (IMDB), and the GSS–T1FF linkage. Thus far, research that draws on these data sources often focuses on the general population instead of foregrounding sexual minorities (e.g., Hou, Magolis, and Haan 2017; Pugliese, Pelletier, and Le Bourdais 2023). To our knowledge, only one recent study has used the same-sex couple flag from the IMDB, which examines the socioeconomic profile of immigrants in SSCs between 2000 and 2020 in Canada (Stick, Leanage, and Arim 2024).

---

[5] Same-sex registered unions were first enacted in 1989 in Denmark, 1993 in Norway, 1995 in Sweden, and 2002 in Finland. Same-sex marriage was legalized in 2012, 2009, 2009, and 2017 in these countries, respectively.

[6] The Netherlands was the first country to legalize same-sex marriage, in 2001. It recognized same-sex registered partnership in 1998.

[7] The T1 general form is the primary form for personal income tax in Canada.

[8] Full T1FF data is not available to researchers for confidentiality reasons.

Tax-based administrative data remains an untapped resource that presents opportunities for studies of individuals in same-sex partnerships in Canada and elsewhere.[9]

In this paper we focus on identifying individuals who have ever been in an SSC in the LAD, a readily available but underutilized longitudinal dataset of tax filers in Canada. We first discuss research opportunities presented by the LAD regarding sexual minority populations. Next we highlight three potential sources of bias in estimates of individuals (ever) in SSCs – underreporting, misclassification, and measurement errors – and we outline adjustment strategies to reduce the impacts of these biases. We then compare pre- and post-adjustment measures and show that the downward bias of the estimated share of individuals who have ever been in an SSC can be reduced. We also compare the cross-sectional share of individuals in SSCs in the LAD to estimates from the Canadian census between 2001 and 2021. We find that the difference in estimates from the two sources is minimal in 2016 and that the LAD offers a nearly identical estimated share of individuals in same-sex marriages but underestimates the share in same-sex common-law unions. Finally we introduce the measure of "inferred sexual minority status"[10] based on individuals' partnership history and demonstrate its usefulness in an example of earnings gaps. We conclude by discussing research areas that can benefit from analyses based on the LAD and other administrative datasets in Canada and beyond.

2. Opportunities for research on sexual minorities with the LAD

The LAD has several unique strengths that allow researchers to expand knowledge of sexual minority populations. First, the LAD provides a large sample of individuals in SSCs. The sample in the LAD represents 20% of Canadian tax filers each year, which is about 7.6 million people in 2020 (Statistics Canada 2020b). In the analysis presented in Section 5 below, we estimate that about 0.65% of (approximately 253,590) individuals aged 18 or above were *ever* in an SSC between 2000 and 2021.

Second, the LAD has the quality of high indexicality – allowing for identifying and following micro trends (Kitchin 2013). The dataset has been updated annually since 1982 and is consistent in terms of key measures over time, such as year of birth, sex code, total income, province of residence, and family type. The longitudinal nature of the data can overcome many limitations of prior research using cross-sectional data and that which

---

[9] A Canadian study that relies on administrative data from the Uniform Crime Reporting Survey was able to identify victims of domestic violence who were in same-sex couples (see Whitehead, Dawson, and Hotton 2021). To our knowledge, this is the only other case where a Canadian administrative dataset (but not tax data) was used to study sexual minorities.

[10] We use the term *sexual minority status* to refer to the statistical minority status of "a person whose sexual identity, attraction, or behavior does not align with heterosexuality" (Reczek 2020: 33).

used longitudinal surveys with small sample sizes. Once a first-time tax filer is selected into the LAD, a longitudinal profile of the individual is created in subsequent tax years, with little attrition and missing data, unless the person stops filing[11] taxes (Statistics Canada 2023). For this reason, the LAD is apt for longitudinal analyses of within-individual changes over the life course and between-individual comparisons over time (Connelly et al. 2016). It can also facilitate age-period-cohort analyses concerning the progression and changes in the rights of and protections for sexual minorities.

For example, the HIV/AIDS epidemic in the 1980s disproportionately affected gay and bisexual men, men who have sex with men, and transgender populations in Canada and elsewhere, leaving significant marks in the collective memory of gender and sexual minorities (Hammack et al. 2018; Kinsman and Patrizia 2010). Institutionally, the Canadian Charter of Rights and Freedoms recognized sexual orientation as an equivalent ground on which claims of discrimination may be based in 1995 (Government of Canada 2022). The legal recognition of SSCs regarding social and tax benefits based on Bill C-23 in 2000 and marriage equality under the Civil Marriage Act in 2005 further impacted various domains of sexual minorities' lives (Chamie and Mirkin 2011; Kinsman and Patrizia 2010). The LAD and other relevant T1FF-derived datasets are the few sources that allow us to trace changes in the well-being of sexual minorities longitudinally in Canada.

Moreover, the LAD has at least two advantages relative to yearly repeated cross-sectional survey or census data. First, while the LAD does not offer a direct measure of sexual identity, it allows us to consider individuals' partnership practices and histories, assuming that coresidential partnership indicates, on balance, a certain degree of financial, social, and emotional interdependence in a joint household (Belleau, Piazzesi, and Seery 2020). Specifically, with key variables such as an individual code for the status of filers in the family, a code for family structure, and the spousal ID, the LAD unlocks opportunities to study transitions into and out of different relationship statuses (e.g., LaRochelle-Côté, Myles, and Picot 2012; Le Bourdais et al. 2016; Margolis et al. 2019), as well as the order and sequences – trajectories – of partnerships over the life course. For example, older cohorts of sexual minorities might be more likely to partner with a different-sex spouse earlier in life (Ophir, Boertien, and Vidal 2023).

Second, the longitudinal nature of the LAD is particularly important in overcoming limitations of census data, where sexual minorities are reflected only when they are partnered. Individuals who have a history of forming couples with partners of the same and the opposite sex – who can potentially be gay, lesbian, bisexual, or pansexual – are

---

[11] Filers may permanently stop filing taxes for various reasons, but it most often occurs following emigration or death. On rare occasions, individuals may have "gap years" in their longitudinal profiles, where a few years of information is missing before information appears again in later tax years. The reasons for gap years are not accounted for but can include temporary emigration or late filing.

invisible in census data, which has implications for inferences about specific groups (Waite and Denier 2019). For example, research shows that bisexual individuals face distinctive disadvantages in terms of health, well-being, and labor market outcomes (Badgett 2018; Feinstein and Dyar 2017; Mize 2016; Waite, Pajovic, and Denier 2020). Analyses of the 2008–2018 CCHS further reveal that sexual minorities, especially bisexual men and women and gay men, are more likely to be single than their heterosexual counterparts (Waite, Denier, and Pajovic 2021). Similarly, partnered individuals might differ from those currently single. Research on sexual orientation earnings gaps generally finds that wage gaps based on couple data tend to be larger than those based on data that includes individuals' self-reported sexual identity or attraction. This latter type of data source mostly includes individuals currently single (Aksoy, Carpenter, and Frank 2018; Mize 2016; Uhrig 2015; Waite, Pajovic, and Denier 2020).

Finally, the LAD is rich in high-quality income- and tax-related characteristics for both individuals and their linked family members at four levels of aggregation: self, spouse, parent, family and child (Statistics Canada 2023). The LAD relies on tax records, which are more reliable and less prone to missing data than are self-reported income data. It also offers longitudinal information in great detail at both the provincial and federal levels (Frenette, Green, and Picot 2004). Moreover, research has shown that in typical surveys, response rates are generally lower than tax filing rates and samples are less representative of the population in various income groups. Given the higher coverage rate (above 95% since 1992) of tax data, it is less likely to suffer from income-related response bias (Frenette, Green, and Picot 2004). For research investigating various sources of income and inequalities at different income levels over time, tax data can be more advantageous than surveys or censuses.

3. Identifying individuals (ever) in same-sex couples from the LAD in the absence of a direct measure

Despite the clear strengths of administrative data, identifying sexual minority populations remains a major challenge to its use. This section describes how individuals in SSCs are identified in tax data – critical background to our approach to inferring sexual minority status in administrative data. Governmental data processing practices, including how the LAD samples are constructed and how the same-sex couple flag is created, have significant implications for who can be identified as SSCs and introduce three potential biases in our estimates of individuals in SSCs. First, the differential data processes used to identify those in SSCs and DSCs contribute to the *underreporting* of individuals coresiding in same-sex unions. Second, the opposite-sex code assignment of imputed spouses contributes to the *misclassification* of SSC status. Finally, due to additional data

processing logics unrelated to the first two issues, there are potential *measurement errors* in the sex codes of individuals.

3.1 The LAD sample

The LAD is a random, 20% sample of all Canadian tax filers, including temporary residents. It is drawn from the T1FF and updated annually. The sample included about 3.2 million people in 1982, 5.5 million in 2014, and 7.6 million in 2020 (Statistics Canada 2020b), reflecting the growth of population and tax filers. In 1982 an initial sample of 20% of all tax filers was used to create the LAD. Since 1983, 20% of first-time tax filers have been sampled[12] and added to the LAD each year.[13] Sampled members of the LAD are linked across tax years by a unique identification number generated from the Social Insurance Number (SIN), which constitutes a longitudinal profile of each individual (Statistics Canada 2023).

Beyond individual identifiers, the LAD includes family identifiers as well as information on family structure and a filer's family status (spouse or child). Crucially, an individual coresiding with a partner in a conjugal relationship (commonly known as a cohabiting couple) for at least 12 continuous months is legally considered to be in a common-law union in Canada. Like married spouses, common-law partners[14] are required to report their conjugal status on their tax returns (Canada Revenue Agency 2023) and can identify their spouse/partner by providing the spouse/partner's SIN. This is similar to population registers in some European countries[15] (e.g., Sweden, Norway, the Netherlands, and Belgium) (Evertsson, Jaspers, and Moberg 2020; Festy 2007).

Overall, while filing rates may vary across population subgroups (e.g., non-filers are mostly individuals from the bottom quintile of the income distribution) (Abraham et al. 2001; Robson and Schwartz 2020), the population coverage rate of the T1FF has been as high as about 95% since 1992 (Frenette, Green, and Picot 2004). Approximately 93% of couples with children in the 2016 census had both spouses filing taxes (Harding, Laporte, and Olson 2019). However, non-filing might present a challenge because without a direct measure of sexual identity in tax data, we rely on observing individuals filing taxes as a couple to identify sexual minorities. Two scenarios are possible in the case of non-filing.

---

[12] Note that sampling is conducted at the individual level, meaning that in many cases a filer may be selected into the LAD while their spouse may not be, regardless of whether the spouse files taxes in the same year.

[13] For this reason, no backfilling of information is necessary, except in the case of individuals with more than two SINs (SIN changes) that are not immediately identified.

[14] Note that for Canadian tax data, both common-law partners and marital spouses are considered spouses. Thus references to the "spousal ID" or "spousal SIN" include common-law partners.

[15] In Denmark and Finland, only married couples and registered partnerships are included in population registers.

First, if a sampled LAD member does not file taxes in a given year, some of their key demographic variables, such as immigration status and sex code, can be imputed from their filing spouse's or parents' T1FF data or from information from an earlier year when they did file taxes (Statistics Canada 2023). At least one spouse in the couple must be a tax filer for both spouses' records to be included in the LAD. That is, both spouses cannot be imputed in the same year.[16] In this case, family-level variables and some imputation conducted by Statistics Canada mitigate some issues of non-filing (see Table A-1, couple types B-1, B-2, and E). The share of imputed spouses is on average about 2.18% of all records but declined from about 5% in 1982 to 2.57% in 2000 to 0.35% in 2020 (authors' own calculation). Second, if a non-filing spouse is not a sampled member of the LAD (See Table A-1, couple type D), no information about this non-filing non-sampled spouse is provided in the LAD, including their sex code.

3.2 The identification of individuals in same-sex couples – the same-sex couple flag

We rely on a derived family-level LAD variable, the same-sex couple flag (variable name SSFLG), to identify individuals in SSCs in a given tax year. This flag appears in the T1FF and is carried over to any dataset derived from it, including the LAD, and to T1FF data linkages such as the IMDB and the GSS. The flag is first available for 2000, corresponding to the year after same-sex common-law partners were legally recognized in Canada (Government of Canada 2022). It takes the value of 1 if an individual is flagged as being in an SSC in a given tax year and 0 if an individual is not in an SSC (including unpartnered individuals); the value is missing if the sex composition of the couple cannot be determined. However, relying on this variable without further adjustments might lead to biased estimates of individuals in SSCs because of the characteristics of LAD data and the way the same-sex couple flag is derived.

3.2.1 Underreporting

First, differential data processes used to identify individuals in SSCs and DSCs contribute to the underreporting of individuals coresiding in same-sex unions. Tax filers themselves

---

[16] On rare occasions, the record of a sampled LAD member who does not file taxes in a given year is not imputed. In this case, there will be no record for the said individual in the non-filing year. However, if they file again in a subsequent year, they will reappear in the LAD.

do not directly report that they are in an SSC. Instead the flag is created post hoc in the T1FF by Statistics Canada based on three criteria:[17]

1. Both spouses in a couple must have the same sex code.
2. Both spouses must file taxes. In other words, no spouse in the couple can be a non-filer, regardless of whether they are a sampled member of the LAD.
3. Both spouses must declare spousal SINs (discussed more blow).

Couples who do not meet all three criteria cannot be assigned a 1 in the same-sex couple flag; the value will be either 0 or missing. (See Table A-1; only couple types A and C can be assigned a 1 in the original same-sex couple flag.) Accordingly, the underreporting of individuals in SSCs stems from these more restrictive criteria for establishing SSCs than for DSCs. It occurs for couples when one spouse is non-sampled and non-filing (couple type D in Table A-1) or when both spouses file taxes but do not report each other's SIN. More specifically, the T1FF links tax filers to their common-law partners or married spouses in each tax year in two ways, each with a different level of certainty. With the highest degree of certainty, spouses can be matched by reporting the spousal SIN on the tax form. With a lower degree of certainty, if no spousal SIN is reported, spouses can be matched based on name, address, age, sex, and marital status (Statistics Canada 2023). A unique spousal ID is then assigned to the spouse matched to the individual for that tax year.

The tax authorities often mandate coresiding DSCs who file at the same address – but do not file as a couple – to correct their tax returns and file as a common-law couple. Statistics Canada matches spouses on non-SIN personal identifiers when a spouse is identified in a tax return but cannot be identified through the SIN. Neither of these two corrections is implemented for individuals in SSCs by Statistics Canada. Since researchers do not have access to dwelling-level data, we are unable to address this source of underreporting.

3.2.2 Misclassification

The issue of underreporting individuals in SSCs is further compounded by the second source of bias: the misclassification of SSCs as DSCs in some cases. One of the most impactful data processing issues of the T1FF, and thus the LAD, is that imputed spouses

---

[17] This differs from a more direct approach in the Canadian censuses from 2001 to 2016, including the 2011 National Household Survey, which asks whether one is in a male or a female SSC. The 2021 Canadian census does not ask if one is in a female or a male SSC but simply whether one is the husband or wife, or common-law partner, of the reference person (Statistics Canada 2020a).

(i.e., non-filing sampled members of the LAD) are algorithmically assigned the sex code opposite that of their filing spouse. As a result, an individual who is in fact in an SSC will not be flagged as so in years when a sampled spouse does not file taxes and becomes an imputed spouse (See Table A-1, couple types B-1, B-2, and E). This also means that in years when a non-filing sampled member of the LAD becomes an imputed spouse, their sex code might differ from that in the years when they filed taxes. This opposite sex code assignment translates into the misclassification of SSCs as DSCs in some cases, resulting in a downward bias in cross-sectional estimates of individuals in SSCs and an upward bias in estimates of those in DSCs.

3.2.3 Measurement errors

The same-sex couple flag, by definition, corresponds to the sex codes of the two spouses in the couple and is thus sensitive to how sex codes are recorded and processed in tax data. Of note, we use *sex code* instead of *sex* to refer to the variable in the LAD storing a binary code of the tax filer's sex on a yearly basis. While we recognize that there might be discrepancies between individuals' sex assigned at birth, gender identity, legal sex/gender marker change, and how the self-reported sex on T1 files (or the lack thereof) is processed by Statistics Canada, we address only post-processed sex codes in the data – the only information available to make inferences about individuals' sex and SSC status.

The LAD processing of sex codes might lead to measurement errors. Apart from the change introduced by tax filers themselves (e.g., when people complete a legal sex/gender marker change), there are technical reasons why sex codes might not be consistent across tax years. First, Statistics Canada encounters missing information on the sex of tax filers every year.[18] Second, on rare occasions "X" rather than "male" or "female" is reported as the sex. In both situations, Statistics Canada randomly imputes the sex code of an individual into either male or female. The direction of the bias resulting from this type of measurement error is not clear. While little can be done by end users conducting research with T1FF-derived data such as the LAD, Statistics Canada may be able to update the processing methodology of the missing and X categories to diminish measurement errors, especially given that many tax filers reporting their sex as such may be sexual or gender minorities.

---

[18] According to the authors' communication with Statistics Canada's Centre for Income and Socioeconomic Well-being Statistics, every year there is about 0.08% to 0.11% missingness in sex in tax files.

4. Strategies for reducing errors related to the sex code and the same-sex couple flag

Understanding this internal data processing logic informs strategies to address some of the sources of bias. While it is impossible to correct underreporting, it is possible to address misclassification and, to some extent, measurement errors in the sex code. We therefore seek to minimize the impact of misclassification of SSCs in years when a spouse's record is imputed by assuming consistency of individuals' SSC status within couples across tax years. We also seek to minimize the impact of measurement errors arising from post-processing technical errors rather than reasons related to sex/gender changes reported by tax filers. We do so by reducing the potential inconsistency in sex codes within individuals' records across tax years. These broad strategies include three rounds of adjustments performed in order. We implement these adjustments for every record (person-year) based on the 1982–2021 LAD for individuals who (1) were at least 18 years old at any point (i.e., birth cohorts 2003 or older) and (2) had at least one record in or after 2000 (those deceased before 2000 were dropped).[19] This age restriction is in line with the minimum age requirement for common-law partners in Canada. To conserve space, we describe the overall logics of each strategy and place technical details for interested readers in the Appendix.

It is worth noting that our strategies are not the only way of correcting errors and that other criteria can be created based on other sets of consideration. We encourage researchers to conduct sensitivity tests using the proposed measures and alternative ones relevant to their own research objectives when analyzing data based on same-sex couple flags, especially when the selected sample size is smaller. Further, issues of underreporting and, to a lesser extent, measurement error are more difficult to correct, in part because academic researchers have little say in tax data processing.

4.1 Adjustment strategy 1: Consistency in the sex code within individual records

The first round of correction addresses sex code consistency within individual records over the entire time an individual is present in the LAD from age 18 onward. We relied on both sex codes in the LAD main files and those in register files to create three sources of sex codes for further adjustments. The register file is created every year for up-to-date information on individuals present in the LAD (Statistics Canada 2023). Although the sex codes in the register file might still be inconsistent over time, they show less variation

---

[19] The LAD is a confidential dataset that can be accessed through a national network of university-based Research Data Centres in Canada. To gain access, project proposal approval and security clearance are required. For more information on data access, see https://crdcn.ca/.

within individual records and thus can serve as an additional source for adjustments when inconsistency arises (Statistics Canada 2023). We created three versions of the sex code and chose the source of the sex code that showed the highest degree of consistency within individual records over time. We treated sex code inconsistency for one tax year only as an error and recoded it as "sex code consistent." We ended the first round of adjustment with a variable categorizing: (1) individuals whose sex code is consistent; (2) those with one lasting sex code change; and (3) those whose sex code is undetermined. (See Appendix 2 for detailed steps and criteria.)

This adjustment strategy is a conservative one for two reasons. First, we did not make adjustments for those with a lasting sex code change, which may indicate that an individual has changed their sex/gender marker in legal documents. Second, we remained agnostic about those whose sex codes changed across multiple years back and forth (sex code undetermined). This might overlook individuals with some degree of gender fluidity, but we did not have sufficient information that allowed us to differentiate between post-processing errors and changes reflecting self-reported sex/gender due to gender fluidity.

4.2 Adjustment strategy 2: Consistency in the sex code and same-sex couple flag within couples

The second round of adjustment addressed same-sex couple flag consistency across years and within couples from 2000 onward. Because the yearly same-sex couple flag corresponds to the yearly sex code of individuals and that of their spouses, the same-sex couple flag might be incorrect when the sex code is incorrect. To address this problem, we assumed the SSC status was stable during the entire period when each unique pair stayed together, unless an individual or spouse was flagged as having a "lasting sex code change" in the previous step. In other words, once a couple is flagged as an SSC, we assume they stay an SSC the entire time they are together.

There are two reasons for this assumption. First, recall that due to the more restrictive criteria for establishing SSCs than DSCs, an SSC is more likely to be misclassified as a DSC than a DSC is likely to be misclassified as an SSC. Second, for most of the cases, the sex codes of the spouses are not directly accessible because they are not sampled members of the LAD. (See Table A-1 for couple type C.) Thus we can only infer the sex codes of the non-sampled spouses based on the same-sex couple flag (i.e., the non-sampled spouse's sex code is the same as that of the sampled spouse in years when they are flagged as an SSC). By assuming the SSC status to be stable (SSFLG is consistently 1), we also assumed the sex codes of the non-sampled spouses to be stable. The inferred stable sex codes of the non-sampled spouses were then used to correct the

previously erroneous sex codes of individuals during the relationship period. For sampled spouses, their sex codes were directly used (instead of being inferred).

After establishing the new sex codes for individuals, we again calculated if they fell into the category of (1) sex code consistent, (2) lasting sex code change, or (3) sex code undetermined. We updated the same-sex couple flag for those whose sex codes were not undetermined, using a combination of their own sex code and their spouse's inferred or direct sex code.

4.3 Adjustment strategy 3: Consistency in the sex code within individual records post-2000

Finally, the last round of adjustment concerns sex code status post-2000. Given that analyses of individuals in SSCs most likely start from 2000 – the first year for which the same-sex couple flag is available – researchers can opt to consider only post-2000 conditions in terms of sex code consistency. We identified whether the first and last sex code inconsistency took place before or in/after 2000. We then categorized individuals as follows: (1) sex code consistent; (2) sex code inconsistent pre-2000 but consistent post-2000; (3) lasting sex code change post-2000; and (4) sex code undetermined. One can opt to include different groups in the analyses. We then created a final version of the sex code and thus a final version of the same-sex couple flag based on the above categories.

After implementing strategies 1–3, we explore the extent to which our adjustments minimize SSC identification bias, differences in partnership history before and after implementation, whether our adjusted estimates of individuals (ever) in SSCs mirror those from the census, and, lastly, the impact of our adjustments on sexual minority earnings gaps.

5. Descriptive statistics before and after adjustments

5.1 Sex code consistency before and after adjustments

Recall that the three biases outlined above are linked to governmental data processing issues related to the sex code, so in Table 1 we show descriptive statistics of selected characteristics by sex code consistency status based on data from the 1982–2021 LAD. In doing so, we show the impact of implementing the adjustments described above on the sample composition. We highlight that individuals initially categorized as having ever been in an SSC for at least one tax year are also more likely to have inconsistency in the sex code across years relative to those who have never been in an SSC. We also show

that this disproportionate impact on potential sexual minorities can be reduced by implementing the adjustments outlined above.

The left panel of Table 1 shows the proportion of individuals having an inconsistent sex code for at least one tax year based on the original sex code and the same-sex couple flag. The right panel shows the post-adjustment distribution. These statistics are again based on all sampled LAD members (1982–2021) who were at least 18 at any point (i.e., individuals born in or before 2003) and had at least one record in or after 2000. We find that overall, 0.74% of individuals have had an inconsistent sex code in their longitudinal profile for at least one tax year since age 18. The share is reduced to 0.05% post-adjustment, mostly comprised of those with an inconsistent sex code pre-2000 but a stable sex code after 2000, followed by those whose sex codes remain undetermined post-2000 (missing). The share of those with a lasting sex code change post-2000 is too small to be shown based on Statistics Canada disclosure rules. Overall, about 0.69% of all LAD sampled members were initially classified as having an inconsistent sex code but were reclassified as having a consistent sex code post-adjustment (weighted N ≅ 248,108 individuals). In other words, 93% of individuals who had an inconsistent sex code pre-adjustment were deemed to have a consistent sex code post-adjustment.

Breaking down the distribution by selected characteristics, we see that those whose first observed sex code is female are slightly more likely to have sex code inconsistency (0.78 vs. 0.69%). However, the share of those with a consistent sex code is about the same for both sexes (99.95%) post-adjustment. There are also subgroup variations in the proportion of those having had an inconsistent sex code regarding other characteristics (e.g., birth cohort, first observed province).

Importantly, individuals who have ever been in an SSC for at least one tax year pre-adjustment are much more likely to have an inconsistent sex code, with 12.65% falling into this category compared to only 0.66% among those never in an SSC. The post-adjustment proportion of those having an inconsistent sex code is reduced to 0.3% for those ever having been in an SSC, including 0.08% of those with an inconsistent sex code pre-2000 but a stable sex code post-2000 and 0.22% of those whose sex code remains undetermined post-2000.[20] Specifically, among those who have never been in an SSC, including those who have only been in DSCs, those never coupled, and those whose same-sex couple flag is always missing, 0.72%, 0.49%, and 0.81% of these individuals have had an inconsistent sex code, respectively. The post-adjustment share is reduced to 0.05%, 0.03%, and 0.07%, respectively. Among those who have been in SSCs only,

---

[20] The category "self/spouse sex code undetermined" is created for analytical purposes: when a person and/or their spouse does not have a consistent sex code even after adjustments, there is a lower degree of certainty regarding the type of couple they have been in. We dropped this group of individuals in empirical analysis. "Always missing" means an individual has been in a couple but the SSFLG is never 0 or 1. Thus the type of couple cannot be determined based on the original same-sex couple flag.

1.91% and 0.08% have had an inconsistent sex code before and after adjustments, respectively.

Table 1: **Descriptive characteristics of individuals by sex code (SC) consistency status, pre- and post-adjustment (birth cohorts 2003 and older and ever had records in/after 2000; N = 38,957,810)**

	Pre-adjustment (%)		Post-adjustment (%)		
	SC consistent	SC inconsistent at least one tax year	SC consistent	SC inconsistent pre-2000, stable post-2000 + lasting SC change post-2000[a]	SC undetermined post-2000
Total	99.26	0.74	99.95	0.05	0.00
Sex code*					
Male	99.31	0.69	99.95	0.05	0.00
Female	99.22	0.78	99.96	0.04	0.00
Missing[b]	100.00	0.00	100.00	0.00	0.00
Partnership history					
Never in SSC	99.34	0.66	99.94	0.06	0.00
Ever in SSC	87.35	12.65	99.70	0.08	0.22
Partnership history (refined)					
DSC only	99.28	0.72	99.95	0.05	0.00
SSC only	98.09	1.91	99.92	0.08	0.00
DSC & SSC	75.34	24.66	99.96	0.04	0.00
Never coupled	99.51	0.49	99.97	0.03	0.00
Always missing	99.19	0.81	99.93	0.07	0.00
Self/spouse SC undetermined	--	--	5.96	3.31	90.73
Immigration status					
Non-immigrant	99.27	0.73	99.96	0.04	0.00
Immigrant	99.23	0.77	99.95	0.05	0.00
Cohort					
< = 1919	99.34	0.66	99.87	0.12	0.01
1920–1929	99.64	0.36	99.94	0.05	0.01
1930–1939	99.66	0.34	99.96	0.03	0.01
1940–1949	99.62	0.38	99.95	0.05	0.00
1950–1959	99.36	0.64	99.90	0.10	0.00
1960–1969	98.77	1.23	99.89	0.11	0.00
1970–1979	98.73	1.27	99.97	0.03	0.00
1980–1989	99.15	0.85	100.00	0.00	0.00
> = 1990	99.68	0.32	100.00	0.00	0.00

Table 1:   (Continued)

	Pre-adjustment (%)		Post-adjustment (%)		
	SC consistent	SC inconsistent at least one tax year	SC consistent	SC inconsistent pre-2000, stable post-2000 + lasting SC change post-2000[a]	SC undetermined post-2000
Province/territory*					
Ontario	99.15	0.85	99.95	0.05	0.00
Quebec	99.55	0.45	99.96	0.04	0.00
British Columbia	99.08	0.92	99.96	0.04	0.00
Western provinces and territories	99.27	0.73	99.95	0.05	0.00
Eastern provinces	99.34	0.66	99.94	0.06	0.00
Census metropolitan area (CMA)*					
Non-CMA	99.42	0.58	99.94	0.06	0.00
Montreal, Toronto, Vancouver	99.13	0.87	99.95	0.05	0.00
Other CMA	99.31	0.69	99.96	0.04	0.00
N (weighted subtotal)	38,670,965	286,845	38,939,005	18,120	685
N (weighted total)	38,957,810		38,957,810		

Note: Number of observations rounded to the nearest 5 based on confidentiality rules from Statistics Canada.
* First observed value for the variable.
-- Category not created pre-adjustment.
[a] Distribution of "lasting sex code change post-2000" suppressed and collapsed into "sex code inconsistent pre-2000, stable post-2000" due to cell size that does not meet the minimum count to be released.
[b] "Missing" in sex code is extremely rare and is consistent across years. These individuals do not have any sex code assigned to them and are thus separated from other groups.

Crucially, the share of those with an inconsistent sex code pre-adjustment is as high as 24.66% for those who have been in both DSCs and SSCs, which is significantly reduced to 0.04% post-adjustment. This is because many of those who originally seem to have been in both types of couples are in fact misclassified as having been in DSCs (SSFLG = 0) in years when they did not file taxes (becoming an imputed spouse). A closer examination of the data, however, shows that they are with the *same* same-sex spouse (the same spousal ID) instead of being partnered with another person of a different sex. Overall, this shows that it is suboptimal to use the original same-sex couple flag directly without examining the consistency of individuals' longitudinal profiles in the sex code. While these errors might not be impactful in analyses of a large sample of sexual minorities, results might be biased when certain subgroups are selected, such as those who have been in both DSCs and SSCs.

5.2 Partnership history before and after adjustments

Table 2 provides the distribution within the partnership history category by sex consistency status before (upper panel) and after (lower panel) adjustments. These statistics offer the proportion and the number of individuals aged 18 or older who were ever in an SSC between 2000 and 2021. Based on the original sex code and same-sex couple flag (upper panel), 0.62% of individuals (weighted N ≅ 241,880) have ever been in an SSC, including 0.33% who have been in SSCs only and 0.29% who have been in both SSCs and DSCs. Among individuals who have ever had an inconsistent sex code, 10.66% (weighted N ≅ 30,590) have ever been in an SSC, which is mostly driven by those who have been in both SSCs and DSCs (9.81%), despite the fact that there are more individuals who have been in SSCs only than those who have been in both SSCs and DSCs. The post-adjustment share (lower panel) of individuals having ever been in an SSC increased by 0.03 percentage points (from 0.62% to 0.65%, an increase of about 11,710 individuals). Notably, the share of those who have been in both types of couples decreased from 0.29% to 0.21%, while the share of those who have been only in SSCs increased from 0.33% to 0.44%. The share of individuals having been in DSCs only is virtually the same pre- and post-adjustment (about 73.97%); the same is true of those never coupled (about 25.23%).

Overall, we show that the original LAD estimate of individuals in SSCs is downwardly biased, owing to coding and data processing decisions arising from the operationalization of the same-sex couple flag. Our adjustments address the issue of misclassification in the sex code, correct the same-sex couple flag, and rectify some downward bias for those who have been in SSCs only and upward bias for those who have been in both SSCs and DSCs. However, the pre- and post-adjustment estimates of the overall share of individuals inferred to have sexual minority status (ever in an SSC) are very close. This means that the most important bias corrected by our adjustment strategies seems to concern the distribution *within* the sexual minority population: Those who were initially classified as "ever been in both DSCs and SSCs" were reclassified as "ever been in SSCs only."

Table 2: Distribution of partnership history by sex code consistency status, pre- and post-adjustment (birth cohorts 2003 and older and ever had records in/after 2000; N = 38,957,810)

Partnership history	Pre-adjustment					
	SC consistent		SC inconsistent at least one tax year		Total	
	%	N	%	N	%	N
Never in SSC	**99.45**	**38,459,680**	**89.34**	**256,260**	**99.38**	**38,715,940**
DSC only	73.98	28,607,800	72.27	207,310	73.96	28,815,110
Never coupled	25.28	9,777,535	16.85	48,340	25.22	9,825,875
Always missing	0.19	74,345	0.21	610	0.19	74,955
Self/spouse SC undetermined	--	--	--	--	--	--
Ever in SSC	**0.55**	**211,290**	**10.66**	**30,590**	**0.62**	**241,880**
SSC only	0.32	125,285	0.85	2,445	0.33	127,730
DSC & SSC	0.22	86,005	9.81	28,145	0.29	114,150
Total % (weighted N)	100	38,670,970	100	286,850	100	38,957,820

Partnership history	Post-adjustment							
	SC consistent		SC inconsistent pre-2000, stable post-2000 + lasting SC change post-2000[a]		SC undetermined post-2000		Total	
	%	N	%	N	%	N	%	N
Never in SSC	**99.35**	**38,685,585**	**99.06**	**17,945**	**0.00**	**685**	**99.35**	**38,704,225**
DSC only	73.97	28,803,240	81.42	14,750	0.00	0	73.97	28,817,990
Never coupled	25.23	9,824,125	17.28	3,130	0.00	0	25.23	9,827,260
Always missing	0.15	58,175	0.22	40	0.00	0	0.15	58,215
Self/spouse SC undetermined	0.00	45	0.14	25	100.00	685	0.00	760
Ever in SSC	**0.65**	**253,420**	**0.94**	**170**	**0.00**	**0**	**0.65**	**253,590**
SSC only	0.44	172,860	0.75	135	0.00	0	0.44	172,995
DSC & SSC	0.21	80,560	0.19	35	0.00	0	0.21	80,595
Total % (weighted N)	100	38,939,005	100	18,115	100	685	100	38,957,815

Notes: Percentages might not sum to 100 due to rounding. Number of observations might not be consistent due to rounding.
Number of observations (N) are weighted. The unweighted N in the LAD is roughly one-fifth (20%) of the weighted N. Number of observations rounded to the nearest 5 based on confidentiality rules from Statistics Canada.
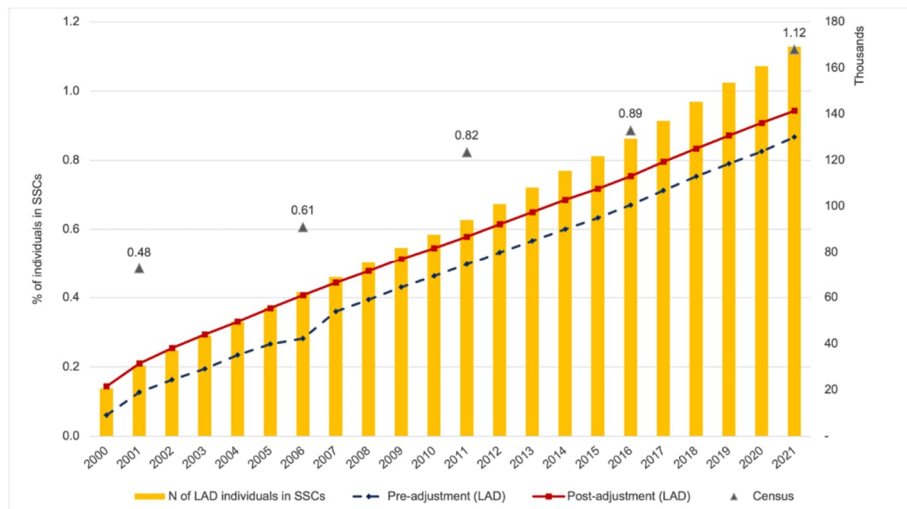-- This category is not created before adjustments.
[a] Distribution of "lasting sex code change post-2000" collapsed into "sex code inconsistent pre-2000, stable post-2000" due to cell size that does not meet the minimum count to be released.

5.3 Cross-sectional share of individuals in same-sex couples, pre- and post-adjustment, compared to Canadian census

Apart from the distribution of LAD members by their longitudinal partnership history between 2000 and 2021 in Table 2, we further compare the LAD estimates with estimates based on five waves of the Canadian census[21] to demonstrate the downward bias regarding individuals in SSCs in the LAD,[22] as well as the differences between pre- and post-adjustment estimates. Based on individuals who were 18 or older in any given tax year, Figure 1 shows the cross-sectional share of individuals in SSCs (as a share of all coupled individuals). Figure 2 further breaks down the proportion by marital status. (See Table A-6.)

Figure 1: **Proportion of individuals in same-sex couples as a share of all coupled individuals, pre- and post-adjustment, LAD 2000–2021 (age 18 and older) and Canadian census (age 15 and older)**

Source: Estimates based on Table A-6.

---

[21] In 2011 the census was replaced with the National Household Survey, which is comparable with other waves of the Canadian census

[22] A few differences between census data estimates and LAD estimates are worth noting. First, the Canadian census captures coresiding same-sex and different-sex partners on the reference date, regardless of whether they have been cohabiting for at least one year, which differs from the legal requirement for common-law couples in tax data. Second, Canadian census estimates are based on those aged 15 and older, whereas the LAD estimates are based on those aged 18 and older. However, given the small share of individuals in couples between age 15 and age 17, this difference should be negligible.

As shown in Figure 1, overall, less than 1% of partnered individuals are in SSCs every year between 2000 and 2021 in the LAD, but the pre-adjustment share has increased steadily, from 0.06% in 2000 to 0.87% in 2021. Post-adjustment measure also shows a consistent upward trend but generally a higher share in each year by about 0.08 percentage points, ranging from 0.14% in 2000 to 0.94% in 2021. This upward trend is similar to that found in census data. While post-adjustment estimates are lower than those in the census, the 2016 LAD estimate, 0.75%, is the closest to the census estimate (0.89%) for corresponding years, followed by the most recent census year, 2021[23] (0.94% in the LAD vs. 1.12% in the census).

Figure 2: **Proportion of individuals in same-sex couples by marital status as a share of all coupled individuals, post-adjustment, LAD 2000–2021 (age 18 and older) and Canadian census (age 15 and older)**

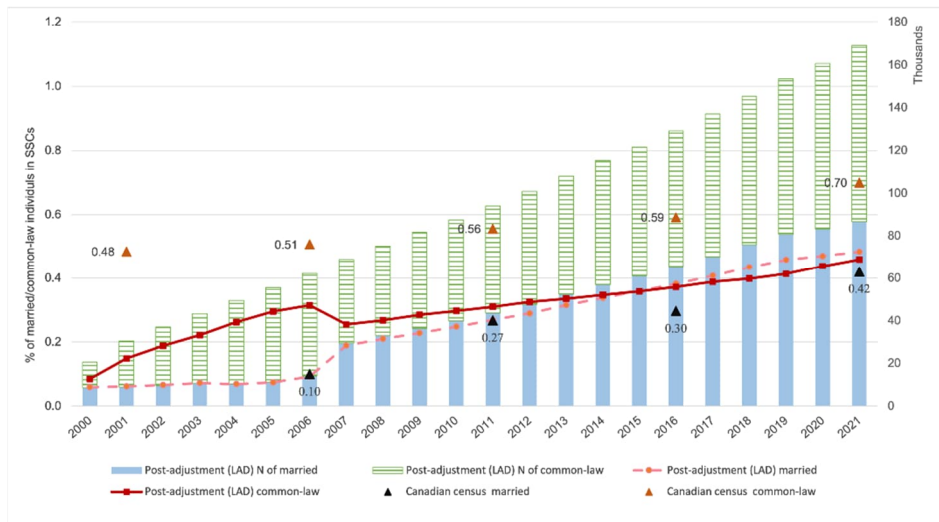*Source*: Estimates based on Table A-6.

---

[23] Note that the share of individuals in SSCs in the 2021 census captures same- and different-gender couples, which differs from previous years, where the numbers reflect same- and different-sex couples. Specifically, this is a result of the modifications of R2P1 (relationship to person 1) response options in 2021 (due to the change from "sex" in 2016 to "sex at birth" in 2021) and changes in Statistics Canada's editing and imputation processing methods in 2021.

Figure 2 further breaks down the proportion of individuals in SSCs by marital status (married vs. common-law) for post-adjustment LAD estimates[24] and census estimates. It shows that the post-adjustment LAD share of married individuals in SSCs is nearly identical to the estimates from the 2006 and 2011 census (0.10% and 0.27%, respectively) but that in more recent years, the LAD estimates of individuals in married SSCs surpassed the census estimates. By contrast, the LAD estimates of individuals in common-law SSCs were persistently lower than the census estimates between 2000 and 2021. In other words, the underestimation of individuals in SSCs in the LAD relative to the census can be overwhelmingly attributed to the underestimation of individuals in common-law SSCs. This is likely due in part to the more stringent criteria for the identification of common-law unions in tax records relative to the census and the underreporting bias discussed earlier.

6. Inferred sexual minority status: an example of earnings gaps between individuals ever in different-sex and same-sex couples

In the absence of a direct measure in the LAD, we develop an indirect backward-looking measure – inferred sexual minority status – which retrospectively assigns sexual minority status to individuals since 2000 based on their full partnership history. Given the features of the LAD mentioned above, this measure allows us to partially address bisexual erasure and the lack of information on the sexuality of unpartnered individuals in cross-sectional data (with the caveat that our measure is not a direct measure of sexual identity). As mentioned, 2000 is the first year for which the same-sex couple flag is available. We leverage this information to establish partnership history by examining longitudinally whether an individual has ever been in a certain type of couple within a fixed observation window (such as an age range or period). Accordingly, this yields four time-invariant categories: individuals who have ever been in (1) DSCs only, (2) SSCs only, or (3) both DSCs and SSCs, plus (4) those never coupled. Researchers may choose their own criteria related to time period (e.g., observed in a partnership between 2005 and 2010), age range (e.g., age 20–35), and the number of (consecutive) years in a couple (e.g., at least two consecutive years in an SSC), depending on their research objectives.

---

[24] According to Statistics Canada's confidentiality rules, pre-adjustment statistics by marital status cannot be released due to empty cells of married individuals in SSCs between 2000 and 2006 (prior to nationwide marriage equality in July 2005). However, in the pre-adjustment LAD, many individuals in SSCs who were partnered between 2000 and 2006 were incorrectly categorized as being in married DSCs, so post-adjustment estimates of married individuals in SSCs likely reflect this classification as well as individuals' self-reported marital status.

In this section, we use analysis of earnings gaps to demonstrate the usefulness of this measure and to help make sense of potential differences in estimates generated by different data sources and measures (pre- and post-adjustment). For the sake of comparison, we adopt sample selection criteria similar to those used in Waite and Denier (2015), based on the 2006 Canadian census. We show two sets of results. The first analysis follows the approach adopted in Waite and Denier (2015), based on only currently coupled individuals. The second analysis shows results using the inferred sexual minority status measure and its interaction with current marital status. The latter analysis demonstrates how the LAD allows us to go beyond couple data typically drawn upon in the census and take into consideration the importance of current partnership status, including common-law, married, and single. In both analyses, we show results based on the sex code and same-sex couple flag pre- and post-adjustment, demonstrating how data processing and the operationalization of the same-sex couple flag can expand or contract sexual minority group membership.

6.1 Analysis 1: Earnings gaps between individuals currently in same-sex and different-sex couples

In the first analysis, we use data from the 2007 LAD, the first year in which married SSCs[25] are captured in the data, comparable to the 2006 census, selecting individuals aged 25–64 who were in either same-sex or different-sex couples.[26] Our dependent variable is log annual employment earnings, with the key independent variable being gender and couple type (men in DSCs; men in SSCs; women in DSCs; women in SSCs). We ran analyses first with all coupled individuals, with men in DSCs as the reference category, and then separated men and women, with individuals in DSCs as the reference category. For each sample, we show a baseline model that controls for only age, age squared, and couple type, and a model with additional controls.[27]

Table 3 shows results based on all coupled individuals (Models 1 and 2), coupled men (Models 3 and 4), and coupled women (Models 5 and 6), using pre- (upper panel) and post- (lower panel) adjustment measures. Consistent with Waite and Denier (2015),

---

[25] Same-sex marriage was legalized nationwide in Canada on July 20, 2005. The 2007 LAD data reflects individuals' conjugal status for the 2006 tax year.

[26] Other sample inclusion criteria include: aged 25–64; non-immigrant; not self-employed (working for wages and salaries); annual earnings of at least $1,000. While Waite and Denier (2015) also exclude visible minorities and indigenous people, the LAD does not have such information.

[27] These controls include province, census metropolitan area, industry code, marital status, and presence of children. In Waite and Denier (2015), fully adjusted models also control for education, Mincer proxy, weeks worked, part-time status, and occupation, which unfortunately are not available in the LAD.

men in SSCs earn less than men in DSCs (about 9.5% less), followed by women in SSCs (about 19%–21% less) and finally women in DSCs (about 41%–42% less).

Earnings gaps between men in SSCs relative to men in DSCs can be partly explained by sociodemographic characteristics introduced in Model 4-1. Pre- and post-adjustment estimates are consistent. In models based on coupled women only, women in SSCs experience an earnings advantage relative to women in DSCs, with the gap reduced by about 16% (pre-adjustment) and 14% (post-adjustment) after adding controls (Models 5-2 and 6-2). Notably, estimates among women in SSCs are quite sensitive to pre- and post-adjustment measures in that pre-adjustment measures underestimate the earnings advantage of women in SSCs in baseline models by 3 percentage points (Models 1-1 and 2-1) and overestimate their advantage over women in DSCs by 4 percentage points (Models 5-1 and 5-2).

Table 3: **OLS regression estimates of earnings gaps pre- and post-adjustments, coupled individuals (age 25–64, 2007 LAD)**

	Pre-adjustment					
	All coupled individuals		Coupled men		Coupled women	
	Baseline	+ Controls	Baseline	+ Controls	Baseline	+ Controls
	Model 1-1	Model 2-1	Model 3-1	Model 4-1	Model 5-1	Model 6-1
Gender and couple type (ref. = men in DSC)						
Men in SSC	−0.10	−0.10				
	(0.02)	(0.02)				
Women in DSC	−0.54	−0.53				
	(0.00)	(0.00)				
Women in SSC	−0.21	−0.24				
	(0.02)	(0.02)				
Couple type (ref. = DSC)						
In SSC			−0.10	0.02	0.33	0.21
			(0.02)	(0.02)	(0.02)	(0.01)
Constant	−1.41	−1.31	−1.55	−1.14	−1.83	−2.05
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Weighted N	1,197,055	1,197,055	593,055	593,055	604,000	604,000
R squared	0.12	0.24	0.04	0.18	0.03	0.17

Table 3:   (Continued)

	Post-adjustment					
	All coupled individuals		Coupled men		Coupled women	
	Baseline	+ Controls	Baseline	+ Controls	Baseline	+ Controls
	Model 2-1	Model 2-2	Model 3-2	Model 4-2	Model 5-2	Model 6-2
Gender and couple type (ref. = men in DSC)						
Men in SSC	−0.10	−0.10				
	(0.02)	(0.01)				
Women in DSC	−0.55	−0.53				
	(0.00)	(0.00)				
Women in SSC	−0.25	−0.27				
	(0.01)	(0.01)				
Couple type (ref. = DSC)						
In SSC			−0.10	0.00	0.30	0.19
			(0.02)	(0.01)	(0.01)	(0.01)
Constant	−1.41	−1.31	−1.56	−1.14	−1.82	−2.05
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Weighted N	1,197,050	1,197,050	592,745	592,745	604,310	604,310
R squared	0.12	0.24	0.04	0.18	0.03	0.17

Notes: Robust standard errors in parentheses. Estimates are statistically significant (at least *p* < 0.05) except for those in gray. Baseline models adjust for age and age squared. Models with controls adjust for age, age squared, province, census metropolitan area, industry code, marital status, and presence of children. Number of observations rounded to the nearest 5 based on confidentiality rules from Statistics Canada.

Overall, these estimates suggest a larger earnings gap between partnered sexual minority men and women relative to men in DSCs and their heterosexual counterparts compared to findings in Waite and Denier (2015), even after controlling for relevant characteristics. Waite and Denier (2015) find an earnings gap between men in SSCs and men in DSCs of about 4.9% and find an earnings advantage for women in SSCs over women in DSCs of about 8.2% (compared to about 21% based on the LAD). These differences likely reflect the absence of certain controls (such as occupation and education) – one of the key limitations of administrative data.

6.2 Analysis 2: Earnings gaps by inferred sexual minority status and current marital status

For the second analysis, we use our measure of inferred sexual minority status by including individuals who were single at the time of data collection (census year or tax year) but were partnered at one point between 2000 and 2021. This measure allows us to retroactively assign sexual minority status to individuals while taking into account their

current marital status (married, common-law, or single). As mentioned, researchers may modify their own criteria regarding how to categorize individuals. Here we adopt the least restrictive criteria: any individuals who were in a couple for at least one year between 2000 and 2021 since age 18. We exclude individuals who were never coupled between 2000 and 2021 to ensure variability in marital status.

We adopt the same sample restriction criteria used in Analysis 1.[28] Table 4 shows regression results based on inferred sexual minority status instead of the cross-sectional conjugal status variable used in Analysis 1 (Table 3). To highlight earnings differences between currently partnered and single individuals, we added an interaction term between inferred sexual minority status and current marital status. Again, pre- and post-adjustment measures are used, shown in the left and right panels, respectively. For ease of interpretation, we visualize results in Figure 3 based on Models 3 and 4.

Results show that beyond the gender composition of the couple at the time of data collection, both inferred sexual minority status and current marital status matter when it comes to earnings gaps. This analysis uncovers two patterns that could not be captured in census data. First, individuals who have been in *both* DSCs and SSCs exhibit varied earnings disadvantages relative to the reference group, highlighting significant heterogeneity within sexual minority subgroups. For example, men and women who have been in both SSCs and DSCs earn about 7.7% and 27%–29% less, respectively, than men who have been in DSCs only, with the gap being slightly reduced after the inclusion of controls (Models 2-1 and 2-2). Pre-adjustment estimates are similar. In comparison, results based on currently coupled individuals in Analysis 1 (Table 3) show larger gaps for women in SSCs relative to men in DSCs, whereas the penalty for women who have been in both SSCs and DSCs in Analysis 2 is about double that of women who have been in SSCs only, some of whom may have been included among women in SSCs in Analysis 2.

Second, there are significant variations by current marital status, as shown in Figure 3, based on interaction models. Among men, while all currently single men experience earnings disadvantages relative to their currently partnered peers, it is those who have been in DSCs only but are currently single who are the most disadvantaged (earning about 27% less than currently married men who have only been in DSCs). Meanwhile, while all currently married sexual minority women experience earnings disadvantages relative to their currently unmarried peers, those who have been in DSCs only are consistently the most disadvantaged across marital status compared to any other group, earning about 44% less than currently married men who have been in DSCs only.

---

[28] Age 25–64; non-immigrant; not self-employed (working for wages and salaries); annual earnings at least $1,000.

Table 4: **OLS regression estimates of earnings gaps pre- and post-adjustment, based on inferred sexual minority status (age 25–64, 2007 LAD)**

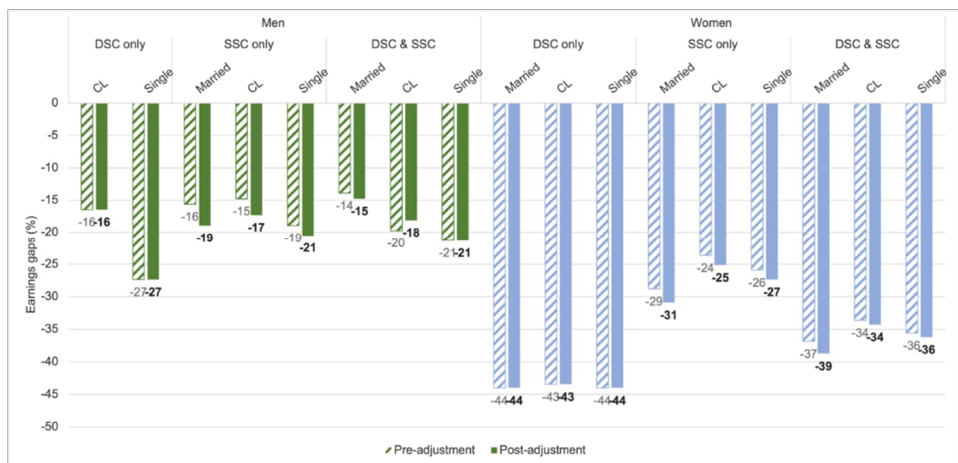|	Pre-adjustment				Post-adjustment			
	Baseline	+ Controls	Baseline + marital status x ISMS	Baseline + controls + marital status x ISMS	Baseline	+ Controls	Baseline + marital status x ISMS	Baseline + controls + marital status x ISMS
	Model 1-1	Model 2-1	Model 3-1	Model 4-1	Model 1-2	Model 2-2	Model 3-2	Model 4-2
Gender and ISMS (ref. = men in DSC only)								
Men in SSC only	−0.05	−0.02	−0.10	−0.17	−0.07	−0.06	−0.17	−0.21
	(0.01)	(0.01)	(0.05)	(0.05)	(0.01)	(0.01)	(0.03)	(0.03)
Men in SSC & DSC	−0.08	−0.07	−0.13	−0.15	−0.08	−0.06	−0.13	−0.16
	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
Women in DSC only	−0.50	−0.48	−0.59	−0.58	−0.51	−0.48	−0.60	−0.58
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Women in SSC only	−0.14	−0.14	−0.25	−0.34	−0.17	−0.18	−0.31	−0.37
	(0.01)	(0.01)	(0.04)	(0.03)	(0.01)	(0.01)	(0.03)	(0.02)
Women in SSC & DSC	−0.32	−0.30	−0.44	−0.46	−0.34	−0.30	−0.49	−0.49
	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)
Marital status (ref. = married)								
Common-law		−0.08	−0.24	−0.18		−0.08	−0.25	−0.18
		(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)
Single		−0.14	−0.33	−0.32		−0.14	−0.33	−0.32
		(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)
Gender and ISMS x marital status (ref. = men in DSC only x married)								
Men in SSC only x common-law			0.21	0.19			0.26	0.20
			(0.05)	(0.05)			(0.04)	(0.04)
Men in SSC only x single			0.24	0.28			0.30	0.30
			(0.05)	(0.05)			(0.04)	(0.03)
Men in SSC & DSC x common-law			0.16	0.11			0.18	0.14
			(0.04)	(0.04)			(0.05)	(0.05)
Men in SSC & DSC x single			0.25	0.23			0.24	0.24
			(0.04)	(0.03)			(0.05)	(0.04)
Women in DSC only x common-law			0.19	0.19			0.19	0.19
			(0.00)	(0.00)			(0.00)	(0.00)
Women in DSC only x single			0.29	0.32			0.29	0.32
			(0.00)	(0.00)			(0.00)	(0.00)
Women in SSC only x common-law			0.27	0.25			0.30	0.26
			(0.04)	(0.04)			(0.03)	(0.03)
Women in SSC only x single			0.32	0.36			0.36	0.37
			(0.04)	(0.04)			(0.03)	(0.03)
Women in DSC & SSC x common-law			0.27	0.23			0.30	0.25
			(0.04)	(0.03)			(0.04)	(0.04)
Women in DSC & SSC x single			0.33	0.34			0.37	0.36
			(0.03)	(0.03)			(0.04)	(0.04)
Constant	−1.34	−1.12	−1.04	−1.10	−1.34	−1.12	−1.04	−1.10
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Weighted N	1,475,520	1,475,520	1,475,520	1,475,520	1,475,570	1,475,570	1,475,570	1,475,570
R-squared	0.10	0.23	0.12	0.24	0.23	0.23	0.24	0.24

Notes: ISMS = inferred sexual minority status. Robust standard errors in parentheses. Estimates are statistically significant (at least *p* < 0.05) except for those in gray. Baseline models adjust for age and age squared. Models with controls adjust for age, age squared, province, census metropolitan area, industry code, marital status, and presence of children. Number of observations rounded to the nearest 5 based on confidentiality rules from Statistics Canada.

Analysis 2 also suggests some salient differences between results based on pre- and post-adjustment measures, especially for men and women who have been in SSCs only. As shown in Figure 3, pre-adjustment estimates generally underestimate earnings

disadvantages relative to the reference group, potentially because some individuals who have been in SSCs only post-adjustment were likely previously classified as having been in both DSCs and SSCs.

Together, this analysis shows that inferred sexual minority status can reveal hidden sources of variability in earnings gaps, including an individual's diverse partnership history and current marital status. However, it is worth noting that the LAD has a limitation of lacking key measures such as education and occupation. This issue can be overcome once survey linkages with tax data become more available.

Figure 3:   Inferred sexual minority status earnings gaps relative to men who have been in different-sex couples only, fully adjusted models, pre- and post-adjustment (age 25–64, 2007 LAD)

Note: CL= common-law.
Source: Earning gaps percentages based on results in Table 4, Models 4-1 and 4-2.

7. Conclusion and future research

The past two decades have witnessed growing interest in the LGBTQ+ population among policymakers and social science researchers. While census and survey data has been foundational to most quantitative research on sexual and gender minorities in many countries, the increasing availability of administrative data offers new ways to broaden the scope of queer demography and LGBTQ+ studies. This article discusses how administrative tax data, such as the LAD in Canada, can overcome key limitations in

prior research, including sample size, cross-sectional design, and the exclusion of segments of the sexual minority population. This large-scale, nationally representative longitudinal data provides promising opportunities for studying the income and family dynamics of sexual minority populations. At the same time, we suggest that governmental data processing and operationalization can introduce bias in population estimates and the conclusions we draw about the underlying population, which in this case results in underreporting, misclassification, and measurement errors. Thus we caution against an uncritical use of the original sex code and same-sex couple flag. Our descriptive portraits of LAD samples based on pre- and post-adjustment measures highlight the impacts of these biases especially on sexual minorities and the extent to which they can be reduced using our adjustment strategies. We also show that the post-adjustment cross-sectional LAD estimates of the share of individuals in SSCs are close to those based on the 2016 and 2021 census and that the LAD estimates for individuals in married SSCs have been highly consistent with census estimates between 2006 and 2021. Finally, the earnings gaps example demonstrates the usefulness of the measure "inferred sexual minority status," which allows us to consider individuals' diverse partnership histories and their current marital status at the same time. Specifically, this approach captures earnings gaps for those who have been in both SSCs and DSCs, as well as those who are currently single – categories unavailable in cross-sectional couple data, such as the census.

While the LAD has been used for studies on access to higher education (Finnie and Pavlic 2013), individual- and family-level income and partnership dynamics (Le Bourdais et al. 2016; Margolis et al. 2019), mobility and poverty (Zhang 2014, 2021), policy impacts (Genest-Grégoire et al. 2023; LaRochelle-Côté, Myles, and Picot 2012), among others, it remains an untapped source for studies of sexual minority populations. Admittedly, there are trade-offs in using administrative data, including the lack of several sociodemographic characteristics, such as race/ethnicity and education, and a direct measure of sexual identity. On balance, however, administrative tax data like the LAD has opened a door for studies that can examine the intersection of inferred sexual minority status, partnership status, and key outcomes. For example, studies of various outcomes can be situated in the context of individuals' union formation and dissolution dynamics, considering the gender composition of unions. At the same time, given rich data on various types of incomes, such as earnings, governmental transfers, and pensions, studies of sexual minorities and inequalities can significantly benefit from the LAD when it comes to examining areas such as detailed patterns of stratification processes over time, the effectiveness and impacts of policies, and the geographic distribution of services provided and needed.

With Statistics Canada seeking to increase the use and availability of administrative data, linkages with survey data and other vital statistics have been established in the Social Data Linkage Environment (Statistics Canada 2022). Recently, the linkage of the

LAD to the Discharge Abstract Database (DAD) (selected waves) allows detailed analyses of the intersection of socioeconomic status and health/hospitalization outcomes, such as death, transfer and discharge, long-term care, and surgery (Statistics Canada 2022). Given prior studies on sexual orientation–based health disparities (e.g., Feinstein and Dyar 2017; Liu and Reczek 2021), this new linkage presents a promising avenue for studying areas such as the intersection of inferred sexual minority status, partnership dynamics, and health and well-being. Research programs like this can have critical implications for relevant policy development and the delivery of health services (Doiron et al. 2013).

In closing, we encourage researchers studying the life course dynamics of sexual minorities to take advantage of the unique strengths of administrative tax data, specifically its size, rich income-related information, and longitudinal design. We also encourage interested readers to adopt and test the strategies laid out in this paper and to reflect on the implications of various measures and methods used. These research practices can contribute to the literature on various substantive areas and facilitate further dialogues regarding how studies on the LGBTQ+ community can be improved.

8. Acknowledgments

References

Abraham, E., Rivard, M., Giles, P., and Lathe, H. (2001). Results of the tax permission questionnaire in the Survey of Labour and Income Dynamics. (Catalogue no. 75F0002MIE-01002). Ottawa: Statistics Canada. https://www150.statcan.gc.ca/n1/pub/75f0002m/75f0002m2001002-eng.pdf.

Aksoy, C.G., Carpenter, C.S., and Frank, J. (2018). Sexual orientation and earnings: New evidence from the United Kingdom. *ILR Review* 71(1): 242–272. doi:10.1177/0019793916687759.

Aldén, L., Edlund, L., Hammarstedt, M., and Mueller-Smith, M. (2015). Effect of registered partnership on labor earnings and fertility for same-sex couples: Evidence from Swedish register data. *Demography* 52(4): 1243–1268. doi:10.1007/s13524-015-0403-4.

Andersson, G. and Noack, T. (2010). Legal advances and demographic developments of same-sex unions in Scandinavia. *Zeitschrift für Familienforschung/Journal of Family Research* Sonderheft 22: 87–101.

Andersson, G., Noack, T., Seierstad, A., and Weedon-Fekjær, H. (2006). The demographics of same-sex marriages in Norway and Sweden. *Demography* 43(1): 79–98. doi:10.1353/dem.2006.0001.

Badgett, M.V.L. (2018). Left out? Lesbian, gay, and bisexual poverty in the U.S. *Population Research and Policy Review* 37(5): 667–702. doi:10.1007/s11113-018-9457-5.

Badgett, M.V.L. and Mallory, C. (2014). Patterns of relationship recognition for same-sex couples: Divorce and terminations. Los Angeles, CA: The Williams Institute, UCLA School of Law. https://escholarship.org/uc/item/1hp9k879.

Belleau, H., Piazzesi, C., and Seery, A. (2020). Conjugal love from a sociological perspective: Theorizing from observed practices. *Canadian Journal of Sociology* 45(1): 23–46 doi:10.29173/cjs29434.

Boye, K. and Evertsson, M. (2021). Who gives birth (first) in female same-sex couples in Sweden? *Journal of Marriage and Family* 83(4): 925–941. doi:10.1111/jomf.12727.

Campbell, A., Perales, F., and Baxter, J. (2020). Sexual minority women in longitudinal survey research: Is attrition a problem? *Archives of Sexual Behavior* 49(5): 1443–1461. doi:10.1007/s10508-020-01669-z.

Canada Revenue Agency (2023, January 4). Marital status. https://www.canada.ca/en/revenue-agency/services/tax/individuals/topics/about-your-tax-return/tax-return/completing-a-tax-return/personal-address-information/marital-status.html.

Cantalini, S., Ohlsson-Wijk, S., and Andersson, G. (2024). Cohabitation and marriage formation in times of fertility decline: The case of Sweden in the twenty-first century. *European Journal of Population* 40(1): 15. doi:10.1007/s10680-024-09703-9.

Card, D.E., Chetty, R., Feldstein, M.S. and Saez, E. (2010). Expanding access to administrative data for research in the United States. *American Economic Association, Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas.* doi:10.2139/ssrn.1888586.

Carpenter, C.S. and Eppink, S.T. (2017). Does it get better? Recent estimates of sexual orientation and earnings in the United States. *Southern Economic Journal* 84(2): 426–441. doi:10.1002/soej.12233.

Chamie, J. and Mirkin, B. (2011). Same-sex marriage: A new social phenomenon. *Population and Development Review* 37(3): 529–551. doi:10.1111/j.1728-4457.2011.00433.x.

Connelly, R., Playford, C.J., Gayle, V., and Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research* 59: 1–12. doi:10.1016/j.ssresearch.2016.04.015.

Doiron, D., Raina, P., Fortier, I., and Linkage Between Cohorts and Health Care Utilization Data: Meeting of Canadian Stakeholders workshop participants (2013). Linking Canadian population health data: Maximizing the potential of cohort and administrative data. *Canadian Journal of Public Health* 104(3): e258–e261. doi:10.17269/cjph.104.3775.

Downs, B., Foster, L., Nesbit, R., and Sandler, D.H. (2023). Same-sex couples and the child earnings penalty. (Working Papers Article 23–25). Washington, DC: Center for Economic Studies, U.S. Census Bureau. https://ideas.repec.org//p/cen/wpaper/23-25.html.

Elder, G.H. (1994). Time, human agency, and social change: Perspectives on the life course. *Social Psychology Quarterly* 57(1): 4–15. doi:10.2307/2786971.

Evertsson, M., Jaspers, E., and Moberg, Y. (2020). Parentalization of same-sex couples: Family formation and leave rights in five northern European countries. In: Nieuwenhuis, R. and Van Lancker, W. (eds.). *The Palgrave handbook of family*

policy. New York, NY: Springer International Publishing: 397–428. doi:10.1007/ 978-3-030-54618-2_16.

Feinstein, B.A. and Dyar, C. (2017). Bisexuality, minority stress, and health. *Current Sexual Health Reports* 9(1): 42–49. doi:10.1007/s11930-017-0096-3.

Festy, P. (2007). Enumerating same-sex couples in censuses and population registers. *Demographic Research* 17(12): 339–368. doi:10.4054/DemRes.2007.17.12.

Finnie, R. and Pavlic, D. (2013). Background characteristics and patterns of access to postsecondary education in Ontario: Evidence from longitudinal tax data. Toronto: The Higher Education Quality Council of Ontario. https://heqco.ca/wp-content/uploads/2020/03/Final-LAD-ENG.pdf.

Fisher, R., Gee, G., and Looney, A. (2018). Same-sex married tax filers after Windsor and Obergefell. *Demography* 55(4): 1423–1446. doi:10.1007/s13524-018-0684-5.

Frenette, M., Green, D.A., and Picot, G. (2004). Rising income inequality in the 1990s: An exploration of three data sources. (Analytical Studies Branch Research Paper Series Catalogue no. 11F0019MIE – 219). Ottawa: Statistics Canada. https://publications.gc.ca/Collection/Statcan/11F0019MIE/11F0019MIE200421 9.pdf.

Genest-Grégoire, A., Dadjo, J., Robson, J., and Schwartz, S. (2023). What proportion of tax returns could the Canada Revenue Agency complete*?* (SSRN Scholarly Paper 4428147). New York: SSRN. doi:10.2139/ssrn.4428147.

Government of Canada (2022). Rights of LGBTI persons. Ottawa: Government of Canada. https://www.canada.ca/en/canadian-heritage/services/rights-lgbti-per sons.html.

Hammack, P.L., Frost, D.M., Meyer, I.H., and Pletta, D.R. (2018). Gay men's health and identity: Social change and the life course. *Archives of Sexual Behavior* 47(1): 59–74. doi:10.1007/s10508-017-0990-9.

Harding, A., Laporte, C., and Olson, E. (2019). Accessing the Canada learning bond: Meeting identification and income eligibility requirements (Catalogue no. 75F0002M). Ottawa: Statistics Canada. https://www150.statcan.gc.ca/n1/en/ pub/75f0002m/75f0002m2019007-eng.pdf?st=0ltyRcFm.

Heslin, K. and Alfier, J. (2022). Sexual orientation differences in access to care and health status, behaviors, and beliefs: Findings from the National Health and Nutrition Examination Survey, National Survey of Family Growth, and National Health

Interview Survey. *NCHS National Health Statistics Reports* 171. doi:10.15620/cdc:115982.

Hou, F., Magolis, R., and Haan, M. (2017). Estimating parental leave in Canada using administrative data. (Catalogue no. 11-633-X). Ottawa: Statistics Canada. https://www150.statcan.gc.ca/n1/pub/11-633-x/11-633-x2017009-eng.htm.

Hu, Y. and Denier, N. (2023). Sexual orientation identity mobility in the United Kingdom: A research note. *Demography* 60(3): 659–673. doi:10.1215/00703370-10769825.

Jaspers, E., Mazrekaj, D., and Machado, W. (2024). Doing Genders: Partner's Gender and Labor Market Behavior. *American Sociological Review* 89(3): 518–541. doi:10.1177/00031224241252079.

Joyner, K., Manning, W., and Bogle, R. (2017). Gender and the stability of same-sex and different-sex relationships among young adults. *Demography* 54(6): 2351–2374. doi:10.1007/s13524-017-0633-8.

Kabátek, J. and Perales, F. (2021). Academic achievement of children in same- and different-sex-parented families: A population-level analysis of linked administrative data from the Netherlands. *Demography* 58(2): 393–418. doi:10.1215/00703370-8994569.

Kinsman, G. and Patrizia, G. (2010). *The Canadian war on queers*. Vancouver: UBC Press. doi:10.59962/9780774816298.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267. doi:10.1177/2043820613513388.

Kolk, M. and Andersson, G. (2020). Two decades of same-sex marriage in Sweden: A demographic account of developments in marriage, childbearing, and divorce. *Demography* 57(1): 147–169. doi:10.1007/s13524-019-00847-6.

LaRochelle-Côté, S., Myles, J., and Picot, G. (2012). Income replacement rates among Canadian seniors: The effect of widowhood and divorce. *Canadian Public Policy / Analyse de Politiques* 38(4): 471–495. doi:10.3138/CPP.38.4.471.

Le Bourdais, C., Jeon, S.-H., Clark, S., and Lapierre-Adamcyk, É. (2016). Impact of conjugal separation on women's income in Canada: Does the type of union matter? *Demographic Research* 35(50): 1489–1522. doi:10.4054/DemRes.2016.35.50.

Liu, H. and Reczek, R. (2021). Birth cohort trends in health disparities by sexual orientation. *Demography* 58(4): 1445–1472. doi:10.1215/00703370-9357508.

Machado, W. and Jaspers, E. (2023). Money, birth, gender: Explaining unequal earnings trajectories following parenthood. *Sociological Science* 10: 429–453. doi:10.15195/v10.a14.

Margolis, R., Choi, Y., Hou, F., and Haan, M. (2019). Capturing trends in Canadian divorce in an era without vital statistics. *Demographic Research* 41(52): 1453–1478. doi:10.4054/DemRes.2019.41.52.

Mittleman, J. (2023). Sexual fluidity: Implications for population research. *Demography* 60(4): 1257–1282. doi:10.1215/00703370-10898916.

Mize, T.D. (2016). Sexual orientation in the labor market. *American Sociological Review* 81(6): 1132–1160. doi:10.1177/0003122416674025.

Möllborn, S. and Kolk, M. (2024). Same-sex marriage in Sweden 1995–2021: A research note on changing patterns in gender, parenthood, and socioeconomic composition. (Stockholm Research Reports in Demography). Stockholm: Stockholm University. doi:10.17045/sthlmuni.26779633.v1.

Noack, T., Seierstad, A., and Weedon-Fekjær, H. (2005). A demographic analysis of registered partnerships (legal same-sex unions): The case of Norway. *European Journal of Population / Revue Européenne de Démographie* 21(1): 89–109. doi:10.1007/s10680-005-3626-z.

OECD (2019). Society at a glance 2019 (summary in English): OECD social indicators. Paris: OECD Publishing. doi:10.1787/soc_glance-2019-en.

Ophir, A., Boertien, D., and Vidal, S. (2023). Sexuality and demographic change: Documenting family formation trajectories and cohort change in the LGB population. *Demography* 60(5): 1581–1605. doi:10.1215/00703370-10968468.

Pugliese, M., Pelletier, D., and Le Bourdais, C. (2023). Separation and savings in tax-favored retirement accounts among Canadian men and women. *Population Research and Policy Review* 42(4): 65. doi:10.1007/s11113-023-09812-z.

Reczek, C. (2020). Sexual- and gender-minority families: A 2010 to 2020 decade in review. *Journal of Marriage and the Family* 82(1): 300–325. doi:10.1111/jomf.12607.

Robson, J. and Schwartz, S. (2020). Who doesn't file a tax return? A portrait of non-filers. *Canadian Public Policy* 46(3): 323–339. doi:10.3138/cpp.2019-063.

Sabia, J.J., Wooden, M., and Nguyen, T.T. (2017). Sexual identity, same-sex relationships, and labour market dynamics: New evidence from longitudinal data

in Australia. *Southern Economic Journal* 83(4): 903–931. doi:10.1002/soej.121 81.

Statistics Canada (2020a). Updated content for the 2021 census of population: Family and demographic concepts, and activities of daily living. (Catalogue no. 98-20–0001). Ottawa: Statistics Canada. https://www12.statcan.gc.ca/census-recensement/2021/ref/98-20-0001/982000012020001-eng.cfm.

Statistics Canada (2020b). The Daily – Canada's population estimates: Age and sex, July 1, 2020. (Catalogue no. 11-001-X). Ottawa: Statistics Canada. https://www150.statcan.gc.ca/n1/daily-quotidien/200929/dq200929b-eng.htm.

Statistics Canada (2023). Longitudinal administrative data dictionary 2021. (Catalogue no. 12-585-X). Ottawa: Statistics Canada. https://www150.statcan.gc.ca/n1/pub/12-585-x/12-585-x2023001-eng.pdf.

Statistics Canada (2022). Linkage of the Longitudinal Administrative Databank (1982–2016) to the Discharge Abstract Database (1994/1995–2016/2017). Ottawa: Statistics Canada. https://www.statcan.gc.ca/en/statistical-programs/document/4107_D1_V1.

Stick, M., Leanage, A., and Arim, R. (2024). Socioeconomic profile of working-age immigrants in same-sex couples in Canada from 2000 to 2020. (Economic and Social Reports Catalogue no. 36-28-0001). Ottawa: Statistics Canada. doi:10.25318/36280001202400600005-eng.

Uhrig, S.C. (2015). Sexual orientation and poverty in the UK: A review and top-line findings from the UK household longitudinal study. *Journal of Research in Gender Studies* 5(1): 23–72.

Umberson, D., Thomeer, M.B., Kroeger, R.A., Lodge, A.C., and Xu, M. (2015). Challenges and opportunities for research on same-sex relationships. *Journal of Marriage and the Family* 77(1): 96–111. doi:10.1111/jomf.12155.

van der Vleuten, M., Evertsson, M., and Moberg, Y. (2023). Joint utility or sub-optimal outcomes? Household income development of same-sex and different-sex couples transitioning to parenthood in Denmark, Finland, Norway, and Sweden. *Journal of Family Issues* 45(8): 2049–2076. doi:10.1177/0192513X231194305.

Velasco, K. (2023). Transnational backlash and the deinstitutionalization of liberal norms: LGBT+ rights in a contested world. *American Journal of Sociology* 128(5): 1381–1429. doi:10.1086/724724.

Waite, S. and Denier, N. (2015). Gay pay for straight work: Mechanisms generating disadvantage. *Gender and Society* 29(4): 561–588. doi:10.1177/089124321558 4761.

Waite, S. and Denier, N. (2019). A research note on Canada's LGBT data landscape: Where we are and what the future holds. *Canadian Review of Sociology = Revue Canadienne de Sociologie* 56(1): 93–117. doi:10.1111/cars.12232.

Waite, S., Denier, N., and Pajovic, V. (2021). Who's hitched? Lesbian, gay, bisexual, and heterosexual partnering in Canada. *Canadian Studies in Population* 48(4): 403–439. doi:10.1007/s42650-021-00059-7.

Waite, S., Pajovic, V., and Denier, N. (2020). Lesbian, gay and bisexual earnings in the Canadian labor market: New evidence from the Canadian Community Health Survey. *Research in Social Stratification and Mobility* 67: 100484. doi:10.1016/j.rssm.2020.100484.

Whitehead, J., Dawson, M., and Hotton, T. (2021). Same-sex intimate partner violence in Canada: Prevalence, characteristics, and types of incidents reported to police services. *Journal of Interpersonal Violence* 36(23–24): 10959–10988. doi:10.1177/0886260519897342.

Zhang, X. (2014). What can we learn about low-income dynamics in Canada from the Longitudinal Administrative Databank? (Income Research Paper Series. Catalogue no. 75F0002M – 002). Ottawa: Statistics Canada. https://www150.statcan.gc.ca/n1/pub/75f0002m/75f0002m2014002-eng.htm.

Zhang, X. (2021). Low-income persistence in Canada and the provinces. (Income Research Paper Series. Catalogue no. 75F0002M). Ottawa: Statistics Canada. https://www150.statcan.gc.ca/n1/pub/75f0002m/75f0002m2021004-eng.htm.

Appendices

Table A-1:   Record type by sample and filing status of individuals and spouses

LAD observations	Sampled filing individual				Sampled imputed (non-filing) individual	
	A	B-1	C	D	B-2	E
Sample status of the spouse	✓	✓	✗	✗	✓	✗
Filing status of the spouse	✓	✗	✓	✗	✓	✓
Imputation status of the spouse	Filer	Imputed	Filer	Not imputed	Filer	Filer
Spousal ID	✓	✓	✓	✗	✓	✓
Original SSFLG can be 1	✓	✗	✓	✗	✗	✗
Post-adjustment SSFLG can be 1	✓	✓	✓	✗	✓	✓

Technical details for strategies of adjusting for biases in the identification of SSCs

Adjustment strategy 1

1. Using the annual LAD register files as an additional source, we created a new version of the sex code (*sex_adj*)[29] by replacing the original sex code in the LAD main files (*sex_lad*) with the sex code from the register files (*sex_reg*) in years when the record was imputed. This yielded three versions of the sex code.

2. We then sorted the data by tax year and laid out the sequence of the sex code for each individual (see Table A-2). We calculated the total number of instances where the sex code in a given year was inconsistent with the sex code in the majority of years for all three sources of the sex code. The results were categorized into three groups: (1) sex code consistent (e.g., column 1 in Table A-2); (2) sex code inconsistent for only one year (e.g., column 2 or 3 in Table A-2); and (3) sex code inconsistent for more than one year (either separately or consecutively) (e.g., column 4 or 5 in Table A-2).

---

[29] We include italicized variable names in parentheses.

Table A-2: Example of within-person sex code consistency across tax years

	1	2	3	4	5
	SC consistent	SC inconsistent for only one year	SC inconsistent for only one year	SC inconsistent for more than one year	SC inconsistent for more than one year
1996	F	F	F	F	F
1997	F	F	F	F	M
1998	F	F	F	F	F
1999	F	F	F	F	F
2000	F	M	F	M	M
2001	F	F	F	M	M
2002	F	F	F	F	F
2003	F	F	F	F	F
2004	F	F	M	F	F

Note: F = female; M = male. SC = sex code.

3. With the three sources of the sex code (*sex_lad*, *sex_reg*, and *sex_adj*) and the three categories mentioned in point 2, we created a total of nine categories, as shown in Table A-3. We applied the following criteria to correct potential errors:

    a) When an individual's sex code from any source fell into the category "sex code consistent" (groups 1, 4, and 7), we adopted that version of the sex code, where the sex code was consistent across all years, regardless of the source. In this case, the individual was considered sex code consistent.

    b) When an individual fell into any category of "sex code inconsistent for only one year" (groups 2, 5, and 8), we adopted the sex code that appeared in all years except the single inconsistent year and considered this individual to be sex code consistent.[30]

    c) After completion of steps 3a) and 3b), the remaining individuals were those for whom all three sources of the sex code indicated "sex code inconsistent for more than one year" (groups 3, 6, and 9). These individuals were further divided into two broad categories:

        1) Sex code inconsistent for more than one year but remaining the same for the consecutive years thereafter: We consider this group of individuals to have experienced a "lasting sex code change."[31]

        2) Sex code inconsistent multiple times across years: The sex code for these individuals could not be determined and was recoded as missing.

---

[30] In cases where more than one source shows that a person has an inconsistent sex code for one year only, we prioritized the source of *sex_adj* followed by *sex_reg* and finally *sex_lad*.

[31] Again, since there might be inconsistency in the sex code from the three sources (i.e., the year in which the sex code differs from that in the previous year may vary by source of sex code), we prioritized the source of *sex_adj* followed by *sex_reg* and finally *sex_lad*.

4. After completion of step 3, all individuals were assigned an adjusted version of the sex code, referred to as *sex_adj2*, except for those whose sex code could not be determined. Based on the adjusted sex code (*sex_adj2*), we created a variable to categorize individuals into one of the following categories: (1) sex code consistent; (2) lasting sex code change; (3) sex code undetermined.

5. Finally, we created another version of the sex code (*sex_adj3*) that was the same as the previous version except that individuals whose sex code (*sex_adj2*) could not be determined were assigned a missing value in *sex_adj3*. In other words, *sex_adj3* included only individuals whose sex code was consistent or who had a lasting sex code change.

Table A-3: **Groups of individuals by sex code consistency from three sex code sources**

	SC consistent	SC inconsistent for only one year	SC inconsistent for more than one year
Sex_lad	1	2	3
Sex_reg	4	5	6
Sex_adj	7	8	9

Note: SC = sex code.

Adjustment strategy 2

After the first round of adjustment using different sex code sources (LAD main files and register files), we were left with individuals whose sex code (*sex_adj3*) remained undetermined – specifically, those with inconsistent sex codes across multiple years. To explore the possibility of further adjusting sex codes for this group, we examined the consistency of the same-sex couple flag since 2000. Using the spousal ID, the original SSFLG, and the spouse's sex code (if available), we conducted a second round of adjustments, first for the SSFLG and then for individuals' sex codes. We followed the steps outlined below:

1. Assume stability in each unique pair (based on individual ID and spousal ID) with regard to SSC status (SSFLG consistently = 1 or 0): Once a couple was flagged as an SSC in a given year, we assumed they remained an SSC for the entire relationship, unless the individual or their spouse was categorized as having a lasting sex code change in the previous step. Conversely, if a couple was never flagged as an SSC, they were assumed to remain a DSC throughout the entire relationship. Based on this assumption, we created a new SSFLG variable

(*ssflg_new*) that reflected this stability within each unique couple pair (see Table A-4, columns 6 and 7, for examples).

2. Infer the sex code of the spouse: Using the new SSFLG (*ssflg_new*, column 7 in Table A-4), we inferred the sex code of the spouse to fill in the sex code for an individual previously categorized as undetermined (*sex_adj3* = missing). By definition, the sex code of the spouse is the same as the individual's if the couple is an SSC and vice versa. Two approaches were adopted depending on whether the spouse was a sampled member of the LAD:

   a) Spouse sampled: The sex code of the spouse was available (i.e., the spouse had a sex code in the original LAD main files and register files) and could be used directly. Note that sampled spouses also underwent the first round of adjustment; thus the most updated sex code (*sex_adj3*) of the spouse was used.

   b) Spouse not sampled: The sex code of the spouse was not available. In this case, we inferred the spouse's sex code based on the first year in which the couple was flagged as an SSC in the original SSFLG (column 8, Table A-4). In that year, the spouse's sex code was assumed to be the same as the individual's sex code. We assumed the spouse's sex code remained stable throughout the relationship and thus inferred the spouse's sex code based on the initial year the couple was flagged as an SSC (column 9, Table A-4).

3. Assign a sex code to individuals whose adjusted sex code (*sex_adj3*) was previously missing: Using the inferred sex code of the spouse – either from the sex code of sampled spouses or the newly established one for non-sampled spouses – we assigned a sex code to individuals whose sex code had been categorized as missing (i.e., their sex code was inconsistent for more than one year) in years when they were partnered. If the couple was flagged as an SSC (*sslfg_new* = 1), the individual's sex code was assigned to be the same as their spouse's and vice versa (column 10, Table A-4). Note that the individual's missing sex code (*sex_adj3*) was replaced only in years when they had a spouse. In years when the individual was single, we used the adjusted sex code from the previous round of correction (*sex_adj2*) (column 11, Table A-4).

4. Reclassify individuals' adjusted sex code (*sex_adj3*): After filling in individuals' sex codes (*sex_adj3*), we reexamined these new sex codes longitudinally and created a sex code consistency variable with the following categories: (1) sex code consistent; (2) sex code inconsistent for only one year; (3) lasting sex code change; (4) sex code undetermined (column 12, Table A-4). As in the first round of adjustment, individuals in the second category were considered sex code consistent. Finally, the sex codes of individuals whose sex codes could not be determined were recoded as missing.

5. Establish new SSFLGs based on the adjusted sex codes (*sex_adj3*) of individuals and their spouses:
   a) If both the adjusted sex code (*sex_adj3*) and the spouse's sex code were available (i.e., both spouses were sampled), we used their sex codes to populate the same-sex couple flag. Spouses with the same sex code were coded as SSCs, and spouses with different sex codes were coded as DSCs.
   b) If the spouse's sex code was not available (i.e., the spouse was not sampled) and was inferred in step 2, we used the value in *ssflg_new* to populate the same-sex couple flag (*ssflg_adj3*), where 1 = SSC and 0 = DSC or single.

Table A-4:   Example of adjusting the same-sex couple flag and sex code (when spouse was not sampled)

1	2	3	4	5	6
Tax year*	*LIN__I*	*sex_adj2*	*sex_adj3*	*sex_adj3_sp*	*sex_consistency_var*
Tax year**	Individual ID	Adjusted SC	Adjusted SC	Adjusted SC of spouse	SC consistency var based on *sex_adj3*
1994	29	M	.	.	SC undetermined
1995	29	F	.	.	SC undetermined
1996	29	F	.	.	SC undetermined
1997	29	M	.	.	SC undetermined
1998	29	M	.	.	SC undetermined
1999	29	F	.	.	SC undetermined
2000	29	F	.	.	SC undetermined
2001	29	F	.	.	SC undetermined
2002	29	F	.	F	SC undetermined
2003	29	F	.	F	SC undetermined

1	7	8	9	10
Tax year*	*ssflg_adj3*	*sex_consistency_final*	*sex_adj4*	*ssflg_adj4*
Tax year**	SSFLG based on *sex_adj3*	SC consistency var based on *sex_adj4*	Final SC	SSFLG based on *sex_adj4*
1994	.	SC inconsistent pre-2000, stable post-2000	M	.
1995	.	SC inconsistent pre-2000, stable post-2000	F	.
1996	.	SC inconsistent pre-2000, stable post-2000	F	.
1997	.	SC inconsistent pre-2000, stable post-2000	M	.
1998	.	SC inconsistent pre-2000, stable post-2000	M	.
1999	.	SC inconsistent pre-2000, stable post-2000	F	.
2000	.	SC inconsistent pre-2000, stable post-2000	F	0
2001	.	SC inconsistent pre-2000, stable post-2000	F	0
2002	.	SC inconsistent pre-2000, stable post-2000	F	1
2003	.	SC inconsistent pre-2000, stable post-2000	F	1

Note: SC = sex code.
* The first row of the table shows variable names in the data (original or constructed).
** The second row of the table explains the content of these variables.

Adjustment strategy 3

Finally, we reexamined the sex codes of individuals (*sex_adj3*) who were categorized as having a lasting sex code change or as sex code undetermined in the *sex code consistency_var*. Specifically, we identified whether the sex code inconsistency occurred before 2000 or in/after 2000. (See Table A-5 for an example.)

1. We describe the steps for two different groups:
    Group 1: Individuals who had a lasting sex code change
    For this group, we identified the year in which the sex code in year *t* differed from that in year *t-1*. Individuals were then classified into one of two categories: (1) lasting sex code change occurred pre-2000; (2) lasting sex code change occurred in or after 2000 (inclusive).
    Group 2: Individuals whose sex code was undetermined
    We examined the sex codes of this group of individuals (*sex_adj3* = missing) based on the previous version (*sex_adj2*). We identified the multiple years in which the sex code in year *t* differed from that in year *t-1* and highlighted the last year of sex code inconsistency (column 3, Table A-5). If the last year of inconsistency occurred before 2000, we categorized this group as (1) sex code inconsistent pre-2000 but stable post-2000 (inclusive). If the inconsistency occurred in or after 2000, individuals were categorized as (2) sex code undetermined.
2. We categorized all individuals taken together into the following categories: (1) sex code consistent; (2) sex code inconsistent pre-2000, stable post-2000; (3) lasting sex code change post-2000; (4) sex code undetermined post-2000. Note that the second category combined individuals who had a lasting sex code change pre-2000 with those whose last year of sex code inconsistency occurred pre-2000. We referred to this final variable as *sex_code_consistency_final* (column 8, Table A-5), which served as the basis for generating the final version of the sex code: *sex_adj4*. This version of the sex code was largely the same as *sex_adj3*, with a few exceptions:
    a) Since some individuals previously classified as sex code undetermined in *sex_adj3* were now categorized into the second category (sex code inconsistent pre-2000, stable post-2000) (columns 4 and 8, Table A-5), their sex codes were updated from missing to *sex_adj4* (column 9, Table A-5).
    b) The final sex code (*sex_adj4*) of individuals classified in the fourth category (sex code undetermined) was coded as missing.
3. With the updated sex code, we updated the same-sex couple flag (*ssflg_adj3*) accordingly. The final version, *ssflg_adj4* (column 10, Table A-5), largely mirrored the previous version (*ssflg_adj3*). However, following the same logic described in step 5 under 'Adjustment strategy 2', whenever a valid sex code (*sex_adj4*) was

established, a corresponding *ssflg_adj4* was created. Similarly, when the sex code was missing, *ssflg_adj4* was also coded as missing.

Table A-5:   Example of adjusting the sex code and same-sex couple flag pre-/post-2000

1	2	3	4	5	6	7	8	9
Tax* year	*LIN__I*	*LIN__P*	*sex_lad*	*sex_lad_sp*	*ssflgi*	*ssflg_new*	*sex_sp1*	*sex_sp2*
Tax** year	Individual ID	Spousal ID	Original SC	Original SC of spouse	Original SSFLG	New SSFLG	Inferred SC of spouse (step 1)	Inferred SC of spouse (step 2)
1998	35	.	F	.	.			
1999	35	.	F	.	.			
2000	35	.	F	.	0	0		
2001	35	42	F	.	1	1	F	F
2002	35	42	M	.	0	1		F
2003	35	42	F	.	1	1	F	F
2004	35	42	M	.	0	1		F
2005	35	42	M	.	0	1		F
2006	35	42	F	.	1	1	F	F
2007	35	42	F	.	1	1	F	F

1	…	10	11	12	13
Tax* year	…	*sex_adj3*	*sex_adj3*	*sex_code_consistency_var*	*ssflg_adj3*
Tax** year	…	Updated SC (step 3)	Updated SC (step 4)	SC consistency var based on *sex_adj3*	SSFLG based on *sex_adj3*
1998	…		F	SC consistent	.
1999	…		F	SC consistent	.
2000	…		F	SC consistent	0
2001	…	F	F	SC consistent	1
2002	…	F	F	SC consistent	1
2003	…	F	F	SC consistent	1
2004	…	F	F	SC consistent	1
2005	…	F	F	SC consistent	1
2006	…	F	F	SC consistent	1
2007	…	F	F	SC consistent	1

Note: SC = sex code.
* The first row of the table shows variable names in the data (original or constructed).
** The second row of the table explains the contents of these variables.

Table A-6:   **Proportion (%) of individuals in same-sex couples as a share of all coupled individuals, pre- and post-adjustment, LAD 2000–2021 (age 18 or older) and Canadian census (age 15 or older)**

Tax year	Pre-adjustment (LAD) All individuals in SSCs[a]	Post-adjustment (LAD)					Census		
		Married	Common-law	Married/common-law[b]	All individuals in SSCs	Total weighted N in SSCs	Married	Common-law	All individuals in SSCs
2000	0.06	0.06	0.08	0.00	0.14	20,600			
2001	0.13	0.06	0.15	0.00	0.21	30,520	-	0.48	0.48
2002	0.16	0.07	0.19	0.00	0.26	37,225			
2003	0.19	0.07	0.22	0.00	0.29	43,425			
2004	0.23	0.07	0.26	0.00	0.33	49,505			
2005	0.27	0.07	0.30	0.00	0.37	55,580			
2006	0.28	0.09	0.31	0.00	0.41	62,325	0.10	0.51	0.61
2007	0.36	0.19	0.25	0.00	0.44	68,975			
2008	0.39	0.21	0.27	0.00	0.48	75,280			
2009	0.43	0.23	0.28	0.00	0.51	81,815			
2010	0.46	0.25	0.30	0.00	0.55	87,590			
2011	0.50	0.27	0.31	0.00	0.58	94,125	0.27	0.56	0.82
2012	0.53	0.29	0.33	0.00	0.62	100,980			
2013	0.57	0.31	0.34	0.00	0.65	108,085			
2014	0.60	0.34	0.35	0.00	0.69	115,430			
2015	0.63	0.36	0.36	0.00	0.72	121,740			
2016	0.67	0.38	0.37	0.00	0.75	129,295	0.30	0.59	0.89
2017	0.71	0.41	0.39	0.00	0.80	137,125			
2018	0.75	0.43	0.40	0.00	0.83	145,365			
2019	0.79	0.46	0.41	0.00	0.87	153,660			
2020	0.83	0.47	0.44	0.00	0.91	160,835			
2021	0.87	0.48	0.46	0.00	0.94	169,240	0.42*	0.70*	1.12*
Total	0.49	0.26	0.31	0.00	0.58	2,028,120	0.23	0.57	0.80

Notes: SSC = same-sex couple.
Proportion calculated as the number of individuals in SSCs divided by the total number of individuals in couples. Number of observations rounded to the nearest 5 based on confidentiality rules from Statistics Canada.
[a] Pre-adjustment estimates cannot be shown due to empty cells of individuals in married SSCs prior to 2007, according to Statistics Canada's confidentiality rules.
[b] A few individuals' conjugal status cannot be determined due to how spouses are linked in LAD data processing. This situation typically corresponds to unions where one partner/spouse is deceased.
* The 2021 census differentiates between same-sex and same-gender couples. The number in the table is based on individuals in same-gender couples rather than same-sex couples. Statistics by marital status for same-sex couples are not available for the 2021 census.
Source: authors' own calculation based the LAD and on the 2001, 2006, 2016, and 2021 Census of Population and the 2011 National Household Survey, https://www12.statcan.gc.ca/census-recensement/2006/as-sa/97-553/table/t2-eng.cfm; https://www150.statcan.gc.ca/n1/daily-quotidien/220713/cg-b004-eng.htm; https://www150.statcan.gc.ca/n1/daily-quotidien/220713/t002b-eng.htm; Statistics Canada Catalogue no. 95F0300XCB2001003; no. 97-552-XCB2006007; no. 98-400-X2016027; no. 98-500-X, issue 2021002; and no. 98-316-X2021.