

Proceedings of Statistics Canada Symposium 2024: The Future of Official Statistics

A Safe and Inclusive Approach to Disseminating Statistical Information about the Non-binary Population in Canada

by Claude Girard and France-Pascale Ménard

Release date: September 8, 2025



A Safe and Inclusive Approach to Disseminating Statistical Information about the Non-binary Population in Canada

Claude Girard and France-Pascale Ménard¹

Abstract

In 2022, Canada became the first country to release statistical information about its transgender and non-binary populations based on census data. Moreover, following a 2018 government-wide policy direction, Statistics Canada's surveys have been collecting and disseminating information about gender by default rather than sex at birth. Due to the small size of the transgender and non-binary populations, disseminating safe statistical information about them at detailed geographical levels poses a challenge.

The dissemination strategy adopted for the 2021 Census, which was subsequently adapted and recommended for surveys, is centered on a new 2-category gender variable (Men+, Women+) that includes non-binary persons. In this paper, the methodological considerations that have gone into creating and adopting this approach, which is deemed both inclusive of non-binary people and safe, are highlighted.

Key Words: Gender; Sex; Non-binary; Transgender; Census.

1. Context

In recent years in Canada, the public awareness and sensitivity towards 2SLGBTQI+² populations have grown substantially, and the legislative environment related to gender has changed (Statistics Canada, 2020). For instance, in 2017 the *Canadian Human Rights Act* and the *Canadian Criminal Code* were amended to protect individuals from discrimination and hate crime based on gender identity and expression. Also, during that period significant data gaps on gender diversity were identified.

In 2018, to support government operations, analysis and evidence-based decision-making, a government-wide policy direction (Government of Canada, 2019) stated that collecting accurate gender data should be the default for all federal organizations. Moving forward, this meant that sex at birth was only to be collected when necessary. Data evidence collected since suggests that previous survey interpretations of "sex" were conflating two distinct concepts: sex *at birth* and gender – see Knarr et al. (2024). These findings undermine the continuity argument provided by those favoring the collection and dissemination of information on sex at birth over gender.

The 2018 Survey of Safety in Public and Private Spaces was the first large-scale survey conducted by Statistics Canada to collect information about gender and sex at birth. Among other things, it found that 1 in 400 people (0.24%) in Canada aged 15 or older were transgender or non-binary, and that there were significant differences in victimization, discrimination and mental health by gender diversity status of individuals (Jaffray, 2020).

In 2022, Canada became the first country to disseminate statistical information about its transgender and non-binary populations based on census data. Statistics Canada protects the privacy and confidentiality of all individuals while providing relevant data for decision-making. The census dissemination strategy introduced a new 2-category gender variable (Men+, Women+) created from the 3-category gender variable (Men, Women, Non-binary persons) as shown in Figure 1-1, which was used to safely release gender information at the most granular geographic levels (Statistics

¹ Claude Girard, Statistics Canada, 150 Tunney's Pasture Driveway, Ottawa, Ontario, Canada K1A 0T6; (claude.girard@statcan.gc.ca); France-Pascale Ménard, Statistics Canada, 150 Tunney's Pasture Driveway, Ottawa, Ontario, Canada K1A 0T6

² In 2023 the Government of Canada adopted the use of the acronym 2SLGBTQI+ to refer to Two-Spirit, lesbian, gay, bisexual, transgender, queer and intersex people, and those who use other terms related to gender or sexual diversity.

Canada, 2022a). This information can be used by public decision-makers to support Gender-Based Analysis Plus (GBA+) and inform policies, programs, employers, and providers of healthcare, education and justice, and other services to better meet the needs of all persons living in Canada.

The 2021 Census strategy was subsequently adapted to surveys, leading to recommendations as to how they are to collect and disseminate gender-based statistical information. This paper focuses on the methodological aspects underlying the creation and implementation of the new 2-category gender variable, the cornerstone of these recommendations.

Figure 1-1
The two-category gender variable part of the census dissemination strategy

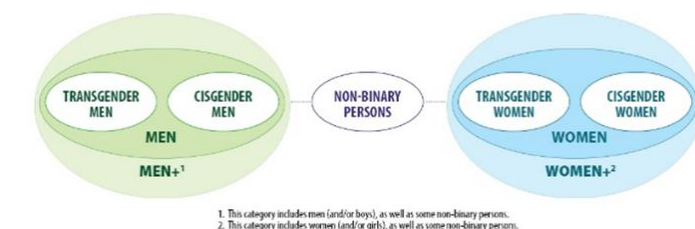


Figure 1.1 – The two-category gender variable part of the census dissemination strategy

2. Collecting and Disseminating Census Gender Information: An Overview

In Canada, the census of population occurs every 5 years and was last conducted in May 2021. After engaging with experts in the field of gender identity and conducting qualitative tests with non-binary, transgender and cisgender individuals as well as a national census test, the census question about sex was replaced in 2021 by the new gender and sex *at birth* questions (Statistics Canada, 2020). Transgender individuals were not asked whether they are transgender but were rather identified based on the answers they provided to the gender and sex at birth questions.

In April 2022, the 2021 Census revealed (Statistics Canada, 2022b, 2022c) that 0.33% of the Canadian population aged 15 years or older and living in a private household were transgender (0.19%) or non-binary (0.14%).

To allow for subnational counts to be safely released about gender given the size of the transgender and non-binary populations as well as their increased risks of experiencing violent victimization, a new 2-category gender variable (Men+, Women+) was derived from the census gender data: all transgender and cisgender men were assigned to Men+, and all transgender and cisgender women to Women+, while records associated with non-binary individuals were randomly recoded between Men+ and Women+.

The unpredictability inherent in this recoding process is a crucial element. Indeed, through the randomness one deliberately surrenders the precise control over the outcome that a deterministic approach would permit. This safe approach, discussed further in Section 4, avoids conducting residual suppression under the traditional disclosure control framework and is therefore deemed more inclusive and respectful of non-binary people.

An example of gender information that was released by Statistics Canada using the new 2-category gender variable pertains to immigration (Statistics Canada, 2022d); the accompanying note is standard with such releases. While the labels “Men+” and “Women+” are to be used in a paper’s charts, figures and infographics, it is recommended to use “Men” and “Women” in the main text for better readability, for instance “*Among immigrants with a foreign degree, overqualification was higher among women (28.3%) than among men (23.1%)*”.

For the upcoming 2026 Census, Statistics Canada has been consulting with the Canadian population, experts and stakeholders on gender (Statistics Canada, 2023). The Agency has finished conducting extensive qualitative and quantitative testing, notably to assess the impact of modifying the gender response categories to include “man” and

“woman”, and “boy” and “girl” for those younger than 15 years, and reversing the order of the 2021 Census questions as to now have date of birth asked first, followed by gender and then sex at birth.

3. Recommendations on Collecting and Disseminating Gender Information

In line with the government-wide policy direction and the precedent set by the Census, gender now stands at the forefront of surveying activities at Statistics Canada, with sex at birth being collected and disseminated only when deemed necessary. By themselves, longitudinal comparability and the existence of biological factors related to sex do not provide adequate justification for using sex *at birth* instead of the gender variable in statistical analyses. More nuanced considerations are necessary when deciding which of these variables to include in the analysis. Indeed, studies have shown that gender too can be strongly linked to biological factors such as obesity (Taormina and Iwamoto, 2023) and the emergence of breast cancer in men because of hormonal therapy (de Blok et al., 2019). Nonetheless, Statistics Canada continues to use sex at birth to measure key demographic indicators such as total fertility rates (Statistics Canada, 2024).

The dissemination of statistical gender information by surveys ought to be based on the 2-category gender variable (Men+, Women+), with the exception perhaps of the highest geographic levels, and on any other statistical disclosure control measure deemed necessary. In most cases, because of the limited coverage of the non-binary population offered by the survey’s sample, this may mean using the 3-category gender variable only for national-level estimates and relying on the 2-category gender variable everywhere else.

Surveys are to collect detailed gender information leading to the classification (Men, Women and Non-binary persons) even if they do not expect the data to be disseminated other than through the 2-category gender variable. Indeed, it is both ethical and methodologically sound to offer reporting options beyond the categories on which statistical results will be disseminated; it is more respectful and engaging, and it also helps mitigating nonresponse.

4. Implementing the Random Recoding of Gender Data to Men+ and Women+

For each project, the random recoding of gender to Men+ and Women+ is carried out by subject-matter areas with the support of their methodology team and, when needed, in consultation from the gender expertise team.

The random recoding can be based on the Men/Women distribution found in the data at some predetermined level, or on a model. In Figure 4-1 the top table shows the **fictitious** example of the 3-category gender data distributed within two regions, with 3+2=5 records associated with non-binary persons (in red) to be randomly recoded. Hence, each of the three records in red in region 2 had, for instance, 34.8% chances of being recoded to Men+. The bottom table in Figure 4-1 shows one realization of the random recoding based on the regional Men/Women distributions, which resulted in two of the three records from region 2 recoded to Men+ alongside the region’s 40 transgender and cisgender men.

Whether the random recoding is based on a distribution or a model, the sex at birth variable, which is strongly correlated with gender, is not to be used as an auxiliary variable. While resorting to sex at birth would make statistical sense, it is more prudent and respectful in this context to do without.

The random recoding can conveniently be implemented using the imputation *methods* and *systems* available to them. For example, donor imputation with the regions as classes was used to get the outcome shown in Figure 4.1. However, gender random recoding is *not* imputing, as it is performed on valid data for disclosure protection purposes *and not* on missing information to enhance data quality.

Figure 4-1
One Realization of the Random Recoding

Gender (3)	Region 1	Region 2	Total
Men	30 (30 / 80 = 37.5%)	40 (40 / 115 = 34.8%)	70
Women	50 (50 / 80 = 62.5%)	75 (75 / 115 = 65.2%)	125
Non-binary Persons	2	3	5
Total	82	118	200

↓

Gender (2)	Region 1	Region 2	Total
Men+	30+2 (32 / 82 = 39.0%)	40+2 (42 / 118 = 35.6%)	74
Women+	50+0 (50 / 82 = 61.0%)	75+1 (76 / 118 = 64.4%)	126
Total	82	118	200

Figure 4.1 – One realization of the random recoding

5. On the Role Played by Documentation to Foster Transparency and Accountability

While seasoned teams routinely document standard statistical outputs, they may underestimate the importance of thoroughly documenting policy-shaping recommendations. The implementation of standardized gender dissemination practices across the Agency was greatly facilitated by the comprehensive supporting documentation released alongside the recommendations. Although creating process documentation is hardly considered exciting, we feel this was crucial in securing us the support and cooperation of both survey partners and users.

Consistently, during the consultation meetings, we observed that when tackling newly raised concerns by applying the guidelines and best practices outlined in our existing documentation, it significantly enhanced our credibility and strengthened our relationship with users. Even though we could not foresee every contingency, our thorough preparation and proactive approach instilled confidence in stakeholders, reassuring them that the unfolding initiative was meticulously planned and not improvised.

The supporting documentation fosters transparency and users' accountability, one of the United Nations Economic Commission for Europe Fundamental Principles of Official Statistics (UNECE, 1992). It notably includes a 17-pager main document, a growing list of Frequently Asked Questions (FAQ) and a decision tree informing gender dissemination which we briefly discuss next.

In addition to the recommendations themselves, the main document describes the key factors that were taken into consideration. We found that partners and users were more inclined to comply with the recommendations once they better understood their *raison d'être*. This can easily be overlooked by the working group members who fail to realize that the guiding principles they now take for granted were once elusive and unclear. And while partners and users are not looking to reinvent the wheel, they may nonetheless want to retrace first our main steps before setting out for the path that we have cleared for them.

The main document also contains useful complementary information intended to increase official statistics literacy of partners and users, such as how the 5-Safe framework (Green and Ritchie, 2023) can be used to navigate Statistics Canada's various data access solutions³ when disseminating gender information. As shown in Figure 5-1, we chose to represent the intensity of the various disclosure control measures used by a data access solution as mercury in the thermometer associated with each of the 5-Safe's levers. The larger thermometer indicates the overall intensity

³ In Figure 5-2, utility is depicted from the perspective of a researcher whose analytical needs go beyond the array of tables and infographics made available on the web by the Agency. Note that a master file is *not* a data access solution; it is only shown here for comparative purposes. Data synthesis is currently explored to create 'smart synthetic files' that could match the utility of more tightly controlled data files (UNECE, 2023). The Confidentiality Classification Tool (CCT) is a simple, lean and self-administered tool developed by Statistics Canada, further raising practitioners' awareness to the confidential nature of its data holdings (Duggan et al., 2018) and (Thomas, 2020).

reached. Regardless of the combination of levers used to form a data access solution, the total intensity must reach the ‘Safe’ mark.

Figure 5-1
The Levers of the 5 Safes Framework

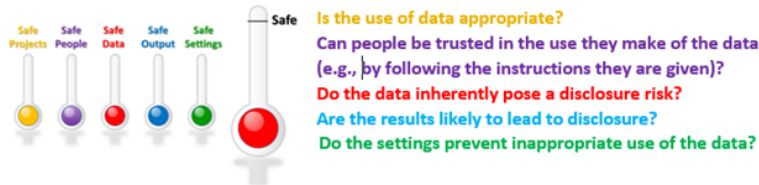


Figure 5.1 - The levers of the 5 Safes Framework

In a way, the thermometers are the signature of a data access solution. For example, the large thermometer of a Public-Use Microdata File would be of a single color, red, because the extensive anonymization measures taken to produce it all pertain to the “Safe Data” lever. In contrast, the multicolor signature of Statistics Canada’s Research Data Centres (RDCs) shown in Figure 5-2 reveals that *all* five levers of the 5-Safe framework are used to prevent the released statistical information from revealing the personal information gathered.

Figure 5-2
Social Data Access at Statistics Canada and the 5 Safes

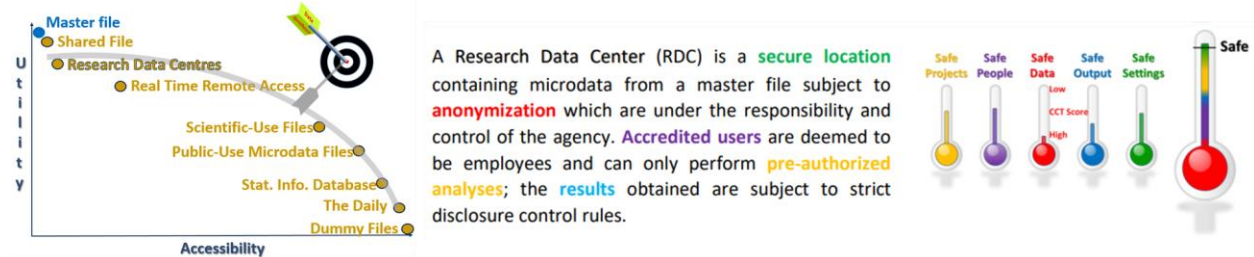


Figure 5.2 – Social Data Access at Statistics Canada and the 5 Safes

Note the low intensity in Safe Data leading to a RDC file, which reflects the minimal anonymization changes made to a master file as to retain most of its utility. For instance, a RDC file is expected to contain both the original three-category *and* the new randomly recoded two-category gender variables for authorized analysts to ponder. However, in the end, releasing gender information must obey the strict dissemination protocol laid out for RDCs users through the Safe Outputs lever; for instance, only one of the two gender variables is to be used; also, users are not allowed to derive their own 2-category gender variable from the detailed gender data and must use the one already in the dataset.

Through the FAQ, we were able to react swiftly to unforeseen issues that were encountered by users and partners ensuring that the documentation made available is ever green. For example, despite the main document warning users against mistaking random recoding for imputation, confusion still arose in practice when both needed to be performed because of gender partial nonresponse: is one to impute missing gender data before, during or after “imputing” gender data to get Men+ and Women+? Through the FAQs, we were able to clarify things by quickly developing and releasing a 2-step procedure to address this issue. It requires first imputing the missing gender data (to Men, Women and Non-binary persons – the three valid gender categories) which is then followed by the random *recoding* of records associated with non-binary individuals (either directly reported or as the result of the imputation just performed) to Men+ and Women+. Not only did the procedure provide practitioners with a clear methodology for handling gender-specific partial nonresponse, but it also made an important conceptual distinction between random recoding and imputation.

Due to its comprehensive nature, the main document lacks specific guidance for partners and users in common scenarios. To bridge this gap, we developed a user-friendly decision tree, providing a visual outline of the various gender dissemination options available – see figure 5-3. (The numbers attached to some of the nodes in the tree below refer to accompanying footnotes that provide more details, which are not shown here.)

Figure 5-3
The Gender Dissemination Avenues

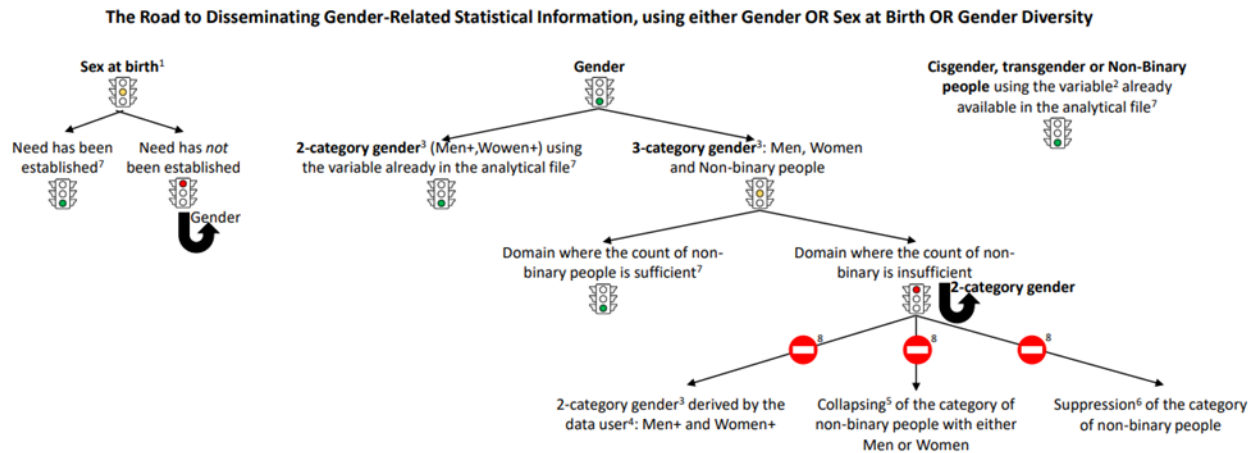


Figure 5.3 – The gender dissemination avenues

6. Concluding Remarks

The dual mission of National Statistical Organizations (NSOs) is nicely conveyed by Duncan et al. (2011) who write “[A NSO] is serving two masters [providing high quality information and protecting confidentiality], each with conflicting interests and concerns” (p.12).

Therefore, in practice NSOs must decide on the relative importance of the utility of the statistical information released and the protection of the underlying person-level information that was collected. But finding a suitable balance between the two is challenging in practice, especially when breaking new ground as it is the case by disseminating statistical information about the very small and victimization-prone gender diverse populations. This duality also helps to understand the various reactions expressed regarding the dissemination approach focused on the new gender variable with two categories, ranging from those who consider it too restrictive to those who advocate for even more restraint.

The recommendations put forward for surveys at Statistics Canada centered on the use of the new 2-category gender variable, which were born from the dissemination strategy of the 2021 Census of Population, are quite strict. Having prioritized a cautious approach, we recognize that the balance achieved by these recommendations may shift in the upcoming years as we continue gaining experience disseminating gender information and as the Canadian society’s views in the matter change. A similar progressive approach was adopted years ago when disseminating information concerning same-sex couples: at first, the dissemination strategy was quite strict but was gradually relaxed over the years as the reality of same-sex couples gained wider societal acceptance and the demand for reliable information about them grew further.

References

- de Blok, C. J. M., Wiepjes, C. M., Nota, N. M., van Engelen, K., Adank, M. A., Dreijerink, K. M. A., Barbé, E., Konings, I. R. H. M., and den Heijer, M. (2019), “Breast cancer risk in transgender people receiving hormone treatment: nationwide cohort study in the Netherlands”, *The British Medical Journal*, (Clinical research ed.), 365, 11652. <https://doi.org/10.1136/bmj.11652>.

- Duggan, J., Gambino, J., Girard, C., and Marquis, M. (2018), “The Framework behind the Confidentiality Classification Tool and its Role in the Modernization”, In the *Proceedings of the Statistics Canada Symposium 2018*.
- Duncan, G., Elliot, M. and Salazar-Gonzalez, J. (2011), *Statistical Confidentiality: Principles and Practice*, Springer, New-York.
- Government of Canada (2019). [Modernizing the Government of Canada's Sex and Gender Information Practices](#)
- Green, E. and Ritchie, F. (2023), “The Present and Future of the Five Safes Framework”, *Journal of Privacy and Confidentiality*, 13(2), <https://doi.org/10.29012/jpc.831>.
- Jaffray, B. (2020), “Experiences of violence victimization and unwanted sexual behaviours among gay, lesbian, bisexual and other sexual minority people, and the transgender population, in Canada, 2018”, Statistics Canada.
- Knarr, M., Ménard, F-P., and Milan, A. (2024), “Comparing sex/gender from the T1 tax file with sex at birth and gender from the 2021 Census of Population”, Statistics Canada internal document.
- Statistics Canada (2020). [Sex at birth and gender: Technical report on changes for the 2021 Census](#) .
- Statistics Canada (2022a) [Filling the gaps: Information on gender in the 2021 Census](#).
- Statistics Canada (2022b). [Understanding who we are: Sex at birth and gender of people in Canada](#).
- Statistics Canada (2022c). [The Daily — Canada is the first country to provide census data on transgender and non-binary people](#).
- Statistics Canada (2022d). [Immigrants with a foreign degree are twice as likely to be overqualified as those with a Canadian degree](#).
- Statistics Canada (2023). [2026 Census of Population Content Consultation Results: What we heard from Canadians](#).
- Statistics Canada (2024). [Fertility in Canada, 1921 to 2022](#).
- Taormina, J.M. and Iwamoto, S.J. (2023), “Filling a gap in care: addressing obesity in transgender and gender diverse patients”, *International Journal of Obesity*, 47, pp. 761-763.
- Thomas, S. (2020), “Safe Data” – Statistics Canada’s Data Confidentiality Classification Tool. Tiered Access Workshop. Council of Professional Associations on Federal Statistics.
- UNECE (1992), [Fundamental Principles of Official Statistics | UNECE](#).
- UNECE (2023), [Synthetic Data for Official Statistics - A Starter Guide | UNECE](#).