**Preliminary Form to Add Social Determinants to CSDUL**

**Request date (2025-09-11):**

| **Researcher (name and affiliation):** | Anousheh Marouzi, University of Toronto |
| --- | --- |
| **Node Lead (name and affiliation):** | Charles Plante, Saskatchewan Health Authority |
| **Indicator or Model Name:** | Population Counts for Dissemination Areas (DAs) and Census Tracts (CTs), 2011–2021 |

**Purpose of the document**

This document includes several questions that must be answered by the researcher interested in adding indicators or models into CSDUL. These questions pretend to briefly explain the mathematical and theoretical framework of the indicator or model being incorporated. The researcher must be able to fill out every question clearly and concisely, supporting their explanation with respectable academic sources.

The document will be added to the model or indicator documentation in CSDUL-OUT and CSDUL-RDC. It must serve as a quick and straightforward introduction to the indicator or model for anyone interested and give relevant references to guide the learning process to other researchers.

**To be completed by the responsible analyst.**

**If there are questions that cannot be answered because of the nature of the indicator/model, write N/A.**

**You can support your completion using the example document located in this link:** [**Documents - Add inputs to CSDUL - 02 - Example.docx - Google Docs**](https://docs.google.com/document/d/1t4_Bh5pRtHzd8GQ3ifJWY2zjjjch8DFf/edit)

1. **Will you share the inputs through CSDUL-RDC, CSDUL-OUT, or both?** CSDUL-OUT.

1. **Explanation of the indicator/model.** 
   1. **In simple words, explain what the indicator/model to be added consists of.**

This dataset provides **annual population counts at the Dissemination Area (DA) and Census Tract (CT) levels in Canada from 2011 to 2021**. Statistics Canada only publishes counts for census years (2011, 2016, 2021), so this project fills the gap by producing **intercensal estimates** (2012–2015 and 2017–2020). Counts are harmonized across census boundaries, allowing for consistent longitudinal and neighbourhood-level analyses.

* 1. **Are there assumptions associated with the indicator/model? If there are, please briefly describe them.**

Yes. The intercensal values are estimated using **linear interpolation**, assuming population changes occur at a constant rate between census years. Areal-weighted interpolation assumes populations are uniformly distributed within geographic units.

* 1. **How is the indicator/model derived? Support your explanation with formulas when possible.**
* Census-year population counts (2011, 2016, 2021) were calculated by aggregating Dissemination Block (DB) counts from Statistics Canada’s Geographic Attribute Files.
* Intercensal estimates were derived using **linear interpolation**. For example, using this approach to interpolate between the 2011 and 2016 censuses, we calculated the population counts for intercensal periods *pt*​ where *t ∈ {*2012, 2013, 2014, 2015*}* as:

|  |
| --- |

where *p2011* and *p2016*​ are the known population counts from census years 2011 and 2016, respectively. Since DA boundaries are nested within CT boundaries, we derived CT-level population estimates by aggregating interpolated DA counts. For consistency, all interpolated values were rounded to the nearest integer.

* Geographic boundary changes were adjusted using **areal-weighted interpolation**. For example, to estimate DA population counts in 2016 based on the boundaries of the previous census, 2011, *p2016↦2011*, using the following equation:

|  |
| --- |

where *p2016↦2016* is the population of DA *j* in 2016 based on its 2016 boundaries, *ajk* is the land area of DA *j* in 2016 that overlaps with the land area of a target DA *k* in 2011, and *aj* is the total land area of DA *j* in 2016.

* We obtained CT-level population count estimates for the intercensal years by aggregating interpolated DA counts.  
  1. **What is the unit of analysis of the indicator/model? (e.g. households, persons, cities)**

Dissemination Areas (400–700 residents) and Census Tracts (2,500–7,500 residents).

* 1. **How can the indicator be integrated with other datasets?**

Population counts can be linked with other datasets at DA or CT level. It can also be linked to datasets containing postal codes, such as the administrative health database, via postal codes using Postal Code OM Conversion File Plud (PCCF+).[(Statistics Canada, 2016b)](https://paperpile.com/c/0yyVSH/Ksp8) Researchers can also aggregate DAs to align with health regions, municipalities, provinces, or CMAs.

* 1. **What are the boundaries of the indicator/model?**

Counts range from **0 to several thousand,** depending on DA/CT size.

* 1. **If you want to add a model to CSDUL, is this associated with a hypothesis? If yes, please describe their:**
     1. **Null hypothesis**
     2. **Alternative hypothesis**
     3. **The implications of rejecting the null hypothesis**

N/A

* 1. **What is the interpretation of the values of the indicator/model?**

The values represent the estimated population of each DA or CT for a given year. They can be used as denominators for calculating rates (e.g., mortality, hospitalization, prevalence of conditions) or to track demographic changes at the local level.

* 1. **Based on the literature and your experience working with this indicator/model, is it possible to identify weaknesses in its calculations or assumptions? To facilitate your answer, you can focus on:**

1. **Potential biases**
2. **Overestimation**
3. **Underestimation**
4. **Omitted variables**
5. **Endogeneity**
6. **Datasets’ problems**

* Linear interpolation may under- or over-estimate changes in rapidly growing or declining areas.
* Areal-weighted methods assume uniform distribution within DAs, which may not reflect reality.
* Random rounding in census data introduces a minor error.
* Intercensal counts are estimates, not official Statistics Canada values.

1. **Does the indicator/model have other mathematical or computational versions (not syntax) to build it? (provide references)**

Yes. Alternatives include **population-weighted areal interpolation** and **dasymetric mapping techniques** (used in the Canadian Longitudinal Census Tract Database, CLTD).[(Allen & Taylor, 2018)](https://paperpile.com/c/0yyVSH/YXHR)

* 1. **Why are you building the indicator/model as you propose? Are there advantages compared to other versions?**

This method is simple, transparent, reproducible, and scalable. It requires no GIS expertise and aligns with Statistics Canada’s own intercensal estimation methods at higher geographic levels.[(Statistics Canada, 2016a; Weden et al., 2015)](https://paperpile.com/c/0yyVSH/W04Kw+F7mP3)

1. **Do you see potential improvements for the indicator/model? This could involve using other datasets, refining calculations, or modifying assumptions, among others.**

* Incorporate population-weighted or dasymetric methods for greater accuracy.
* Automate updating as new census data is released.

**What inputs are to be added to CSDUL? Write “X”**

| **Publicly available on the Statscan website** | Raw or intermediate datasets required to create the indicator/model. |
| --- | --- |
| **X** | Codes that create the indicator/model (be sure that your code is clear enough to be replicated in the future for yourself or any other researcher). |
| **X** | Documentation that explains step by step the entire process that builds the indicator or model. |
| **X** | Results, which consist of the list of variables, indicators, or model results. |
| **X** | Support files. They can be papers, chapter books, codes, etc. |

**References**

[Allen, J., & Taylor, Z. (2018). A new tool for neighbourhood change research: The Canadian Longitudinal Census Tract Database, 1971–2016: Canadian Longitudinal Tract Database. *The Canadian Geographer. Geographe Canadien*, *62*(4), 575–588.](http://paperpile.com/b/0yyVSH/YXHR)

[Statistics Canada. (2016a). *Population and Family Estimation Methods at Statistics Canada*. Statistics Canada.](http://paperpile.com/b/0yyVSH/W04Kw)

[Statistics Canada. (2016b). *Postal CodeOM Conversion File Plus (PCCF+) Version 6C, Reference Guide. August 2015 Postal Codes* (Version 6c). Statistics Canada.](http://paperpile.com/b/0yyVSH/Ksp8) <https://mdl.library.utoronto.ca/sites/default/files/mdldata/open/canada/national/statcan/postalcodes/pccfplus/2011/2015aug/82-F0086-XDB-2016v6c-eng.pdf>

[Weden, M. M., Peterson, C. E., Miles, J. N., & Shih, R. A. (2015). Evaluating linearly interpolated intercensal estimates of demographic and socioeconomic characteristics of U.s. counties and census tracts 2001-2009. *Population Research and Policy Review*, *34*(4), 541–559.](http://paperpile.com/b/0yyVSH/F7mP3)