

Representation in Brain Imaging Research: A Quebec demographic overview

DOI: [10.55458/neurolibre.00035](https://doi.org/10.55458/neurolibre.00035)

Interactive Version

- [Living Preprint](#)

Code

- [Technical Screening](#)
- [Submitted Repository](#)

Reproducibility Assets

- [Repository](#)
- [Dataset](#)
- [Living Preprint](#)
- [Container](#)

Moderator: [Agah Karakuzu](#)

Screeener(s):

- [@mathieuboudreau](#)

Submitted: 29 May 2025

Published: 08 June 2025

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Tudor Sintu¹, Olujide Oyeniran⁵, Udunna Anazodo², Benjamin De Leener^{1,3,4}, and Eva Alonso-Ortiz^{1,3}

1 NeuroPoly Lab, Polytechnique Montreal, Montreal, Quebec, Canada. 2 Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada. 3 CHU Sainte-Justine, Montreal, Quebec, Canada. 4 Department of Computer Engineering and Software Engineering, Polytechnique Montreal, Montreal, Quebec, Canada. 5 Medical Artificial Intelligence Laboratory, Crestview Radiology, Lagos, Nigeria

Disclaimer



This pdf is intended for content registration purposes only! For full access and interactive reading of this publication, please visit [the live version](#).

Introduction

The purpose of this reproducible preprint is to explore demographic data gathered from human MRI and/or PET brain imaging studies conducted in Quebec, Canada from 1992 to 2023. We would like to visualize trends and relationships between various demographic factors, as well as assess whether there are issues in participant selection and in the reporting of demographic data.

Why do we focus on demographics? Much of our understanding of human health is built on scientific studies that often focus on specific aspects of disease, treatment, or prevention. While these targeted investigations collectively inform broader medical knowledge and public health practices, their impact can be limited if participant samples are not representative of the broader population. For example, if a landmark study on Alzheimer's disease primarily includes male participants but draws general conclusions about disease progression, it risks creating a misleading narrative, one that might differ if the study included more female participants or a balanced cohort.

What is the potential problem with reporting? Standards for reporting demographic data vary across publications and institutions, and many reputable journals may not prioritize details like sex, race and ethnicity, or age. When this information is omitted, readers may struggle to determine which populations the study's conclusions accurately apply to.

The content of this reproducible preprint In this preprint, you will find figures illustrating the demographics of Quebec MRI and PET brain imaging study participants, some of which contain interactive elements, and brief descriptions of what has been done to create those figures and why. A discussion and interpretation of results can be found in the full publication.

Study selection

We used Medline, Embase, and Google Scholar databases, retrieved all records involving MRI and/or PET brain imaging in Quebec between 1992 and 2023, and screened those results.

The Sankey diagram below shows the screening process in detail. Users can hover over the connections between blocks for information regarding screening criteria.

Figure 1

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

We identified 1,549 studies that were conducted in Quebec using MRI, PET, or both imaging modalities to study the healthy or diseased brain. These studies took place in different parts of Quebec, and focused on a wide variety of conditions. A breakdown of how studies were categorized is shown in the following treemap.

Users can click on each box in the figure to expand the category, and each individual study can be expanded to see the title and a DOI link to the article. Users can click at the top of the current box to zoom back out.

The first level of distinction is the imaging technique. The second level is the region(s) of Quebec in which the studies take place. The final level is a categorization of study populations among the following: healthy population (studies of healthy brain connectivity and cognition, testing of imaging techniques in healthy populations, etc.), epilepsy, Parkinson's disease, Alzheimer's disease, mental and behavioral disorders (excluding dementia), multiple sclerosis, cerebrovascular disease, cancer and tumors, newborns and infants, concussion and injury, vision disorders, and pain. Categories also exist for populations concerning multiple of the above categories, labelled "Multiple Categories", and for studies concerning any other patient population which has not been categorized here, labelled "Other".

Figure 2

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

Regional breakdown of studies

Studies were conducted within different administrative regions of the province of Quebec. The following map shows the regional distribution of studies that were conducted with MRI and/or PET. Users can zoom and drag the map. The dropdown menu can be used to change the imaging modality, and each bubble can be hovered over to see the corresponding region's name and number of associated studies.

Figure 3

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

Basic demographic breakdown

We aimed to visualize how imaging studies are distributed across age groups in order to understand which populations are most represented, and why certain age groups are commonly

studied. To do this, we categorized studies based on the average age of the participants and display the results as a histogram.

On the right side of the plot, in gray, are studies that did not report participant age or only provided limited information, such as an age range. Using the dropdown menu, users can choose whether the histogram displays the number of studies with participants with an average age falling within a given age range or the total number of participants.

Both counting methods share a key limitation: grouping studies by average age masks the actual distribution of ages within each study, which can lead to a misleading picture of participant demographics. However, because most studies do not provide detailed age breakdowns, this visualization offers the best possible insight given the available data.

Figure 4

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

We also assessed whether studies reported participant sex, and whether male and female participants were appropriately represented. The pie chart shows the distribution of male, female, and unreported sex across the included studies.

Figure 5

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

Relation between age and sex

Our initial approach to visualizing age and sex demographics was to use a population pyramid, where the number of studies reporting male and female participants with an average age within a given age bracket are displayed as opposing horizontal bars. While this allows for some insight into male and female participant age distributions, it relies only on the reported mean ages for each sex, and neglects within-study age distributions. This simplification can obscure important differences in how age is distributed across participants.

Figure 6

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

To explore the relationship between age and sex representation in a different way, we created a relational scatter plot. Each marker represents an individual study, with the x-axis indicating the study participants' average age and the y-axis indicating the proportion of female to male participants. Marker color indicates the number of participants per study, and a slider allows users to filter studies based on sample size. This is particularly useful for reducing noise from small studies, which often have less balanced sex distributions. Hovering over each marker reveals the study's sex ratio, mean age, and sample size.

However, like the population pyramid, this scatter plot still relies solely on reported mean age values. It does not resolve the broader issue of missing or insufficient age distribution data,

which limits our ability to perform deeper demographic analyses.

Figure 7

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

Ethnicity

This treemap shows the reported ethnicity of participants. Each box represents a category, with its size proportional to the number of participants. Boxes can be clicked on to expand and explore subcategories. Each category displays both the participant count and its percentage relative to the total participant pool. For subcategories within the “Reported Ethnicity” group, percentages are also calculated relative to the total number of participants with reported ethnicity.

Figure 8

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

Trends in participant sex by year of publication

Based on observed age distributions, it appears that many healthy participant studies likely recruit from student university volunteer populations. We wanted to explore whether this, or other factors, might influence the sex distribution of participants over time, given the significant improvements in gender representation within scientific fields in recent decades. Additionally it is valuable to visualize the total number of participants included in imaging studies over time. The stacked bar chart shows the number of male and female participants by year of publication, with the full height of each bar representing the total participants studied in that year. The red line overlaid on the graph indicates the percentage of female participants in studies each year.

Figure 9

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

Age distribution by region

Given that the majority of MRI and PET studies are conducted in Montreal, it is important to examine whether location influences participant age distribution. In contrast, fewer studies originate from other regions, such as Sherbrooke and Quebec City. These sites may be more selective in recruitment or more focused on specific populations, rather than broadly sampling healthy volunteers.

The raincloud plot below shows age distribution across regions. Each region is represented as a box plot showing the median (centerline), first and third quartiles (box edges), and the minimum and maximum data points at the edge of the whiskers (known as fences), excluding

any outliers. Scatter plots below each boxplot show individual studies as black dots. The overlaid half-violin plots indicate the density of studies across the average age range. Hovering over each distribution reveals summary statistics.

Figure 10

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

Age distribution by study population category

We are also interested in how the average participant age varies across different study population categories, especially those likely to show skewed distributions, such as Alzheimer's disease. The overall age distribution shown in Figure 4 can be understood as the sum of these individual population-specific distributions. Disaggregating them provides more insight into how participants are selected for different study types.

The following raincloud plot presents these distributions by study population category, using the same format as Figure 10. Each box plots shows the median (center line), first and third quartiles (box edges), and the minimum and maximum data points at the edge of the whiskers (known as fences), excluding any outliers. Scatter plots below each box plot represent individual studies as black dots. The overlaid half-violin plots visualize the density of studies across average participant age for each category. Hovering over each distribution reveals summary statistics.

Figure 11

Figure placeholder

Please see [the living preprint](#) to interact with this figure.

Trends in demographic reporting

Given our findings on the lack of demographic reporting, we wanted to examine whether this trend has changed over time. This final visualization tracks the reporting of demographic information over time. A bar plot in the background represents the total number of participants per year, while three line plots show the percentage of studies reporting each demographic variable over time.

Figure 12

Figure placeholder

Please see [the living preprint](#) to interact with this figure.