Table of Contents

Merged PDF contents (file start pages):

2021-02-23-Data Engineering Challenges and Solutions - Dmitry Gorodnichy - D.pdf 2
2021-11-gorodnichy-little-Intern.MetholodologySymposium-submitted.pdf 13
2022-02-02-invite for seminar (reducted)-on-4 feb 2022.pdf 21
2022-02-02-invite for seminar-on-4 feb 2022.pdf 22
2022-02-04-r4gc-meetup-SpecialEdition(OneYearOfVaccineInCanada)-deck.pdf 24
2022-02-07-Email-change-of-logistics for LnL.pdf 45
2022-04-14-OpenLettertoChiefScienceAdvisorofCanada.pdf 49
comment.pdf 52
comment2.pdf 66
Data Science with Open Canada Data 'Lunch and Learn' seminars _ GCcollab.pdf 70
Dmitry Gorodnichy's Profile _ GCcollab.pdf 71
my-batch-adverse-reactions.pdf 72
PHAC 'Cases following vaccination' processed - letter being prepared _ GCcollab.pdf 73
R4GC (Use R!) Community _ GCcollab.pdf 75
SideEffectReporting-example-1.pdf 76
Today's meetup via zoom - more on working Googlesheets with R _ GCcollab.pdf 81
vitals-how-to-submit-side-effect.pdf 82
vitals.pdf 84
Welcome to ‘Open Canada Use R!’ Portal _ UseR.pdf 88

Slido Poll:

www.sli.do
(code: CBSA1)

Demo URL:

https://rCanada.shinyapps.io/demo

Data engineering challenges and solutions: demonstration of Shiny

Dmitry Gorodnichy
Chief Data Office
Data Science Division

GC Data Conference 2021, Data Literacy Fest
22-23 February 2021, Ottawa

Workshop layout

- “Round table” poll (Slido)

- Common data quality problems
- Data engineering as a solution
- Common tools for common problems
- Discussion (Slido)

Common data quality problems

- Facing same challenges turning “records” into “data”
  - Dates : ‘20210820’ vs. ‘dob 20 Aug 2021’
  - Names: ‘Dmitry Gorodnichy’ vs. ‘Dimitri Horodnytchyyi’
  - Business Names: AC, AirCanada, Air Canada Corp.
  - Geographic Names: ‘Ottawa Airport’, ‘YOW’, Ottawa International Airport

Poor data quality impedes interoperability

- Good interoperability allows various data to be linked and enriched
- Probabilistic (approximate, fuzzy) matching is used to link “noisy” data
  - All words need to be compared to each other
  - Various techniques in data linkages include: using edit metrics, look-up tables, q-grams, phonetic, heuristics, …
- However, probabilistic matching has its share of challenges as well
  - How to assign threshold?
  - How to measure quality?
  - Lost nuances?
    - E.g., Bell Canada vs. Shell Canada
- No perfect solution

Data engineering to address data quality

- Data engineers develop techniques to standardize and organize data to help address data integrity, e.g.,
  - ‘Ottawa Airport’ → ‘YOW’
  - ‘YOW’ → ‘YOW’
  - ‘Ottawa International Airport → ‘YOW’
- On average, 80% of efforts of data scientists goes to address data engineering issues

Common Tools for Common Problems

- In GoC, we are working on the same set of data engineering problems

  - Standardizing various data fields
  - Cleaning, linking and searching data

  so we can carry out analysis

- Often, data scientists end up “reinventing the wheel”
- We need to build common data engineering tools for common GoC data engineering problems

Vision for Solution

- We need a set of ‘libraries’ that are built and maintained by GoC data science community that is
  - Open
  - Free
  - Available to any data scientist who needs them
  - GCcode helps us to do that
  - R already has many libraries, supported by global community



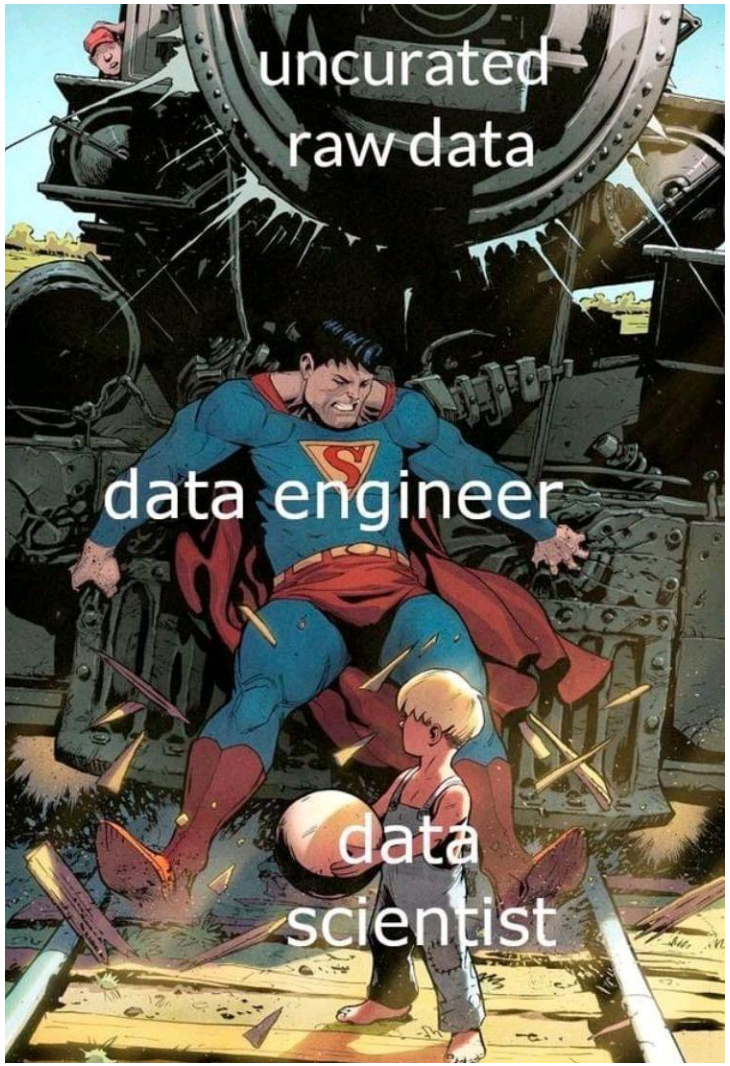
- This presentation will demo a proof of concept for feedback and discussion

Discussion

- Our works has just started. Much more ahead.
  - We build on Public Data (esp. Open Canada Data) and Public knowledge (esp. R global community)
  - We build solution (for entire GC community) and also we build Community of Practice
  - Codes and resources: https://gccode.ssc-spc.gc.ca/r4gc/
- Planned milestones:
  - rCanada Package, Testbed App, Toolkit App: 2021-2022
  - Use cases (for on-going Agency needs): Spring - Winter 2021
- We need your help!
  - curating & organizing DE challenges and public domain solutions (codes/papers)
  - curating & organizing public domain Data-sets
  - testing & benchmarking

Thank you !

Dmitry.Gorodnichy@cbsa-asfc.gc.ca



The author gratefully acknowledges the use of the following public domain material in this project:

R packages: data.table, dtplyr, lubridate, magrittr; soundex, phonetic; antiword, filehash, textreg, textreuse, stringr, stringi, textclean, syuzhet, hunspell, textshape; stringdist, recline, fuzzyjoin, RecordLinkage, fastlink; NLP, quanteda, udpipe, spacyr, tidytext; qdap, tm, lexicon, Rnewsflow, textcat; shiny, rsconnect, knitr, rmarkdown, flexdashboard, DT; cancensus; googleway, tidygeocoder; rvest, httr, xml2, jsonlite; microbenchmark; ggplot2, wordcloud.

Papers: Sariyar M / Borg A (2010). The RecordLinkage Package: Detecting Errors in Data“ The R Journal. van der Loo (2014) The stringdist package for approximate string matching. The R Journal ; L. Boytsov (2011). Indexing methods for approximate dictionary searching: comparative analyses. ACM Journal of experimental algorithmics G. Navarro (2001). A guided tour to approximate string matching. ACM Computing Surveys; Enamorado et al. (2019). ``Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records.'' American Political Science Review

Data: Statistics Canada, Post Canada, Simple Maps

Other: image by Anna Nyulund (LinkedIn post), stimulating discussions with many CBSA and GC colleagues, in particular at Friday's Data Science meet-ups.

Demo time

https://rCanada.shinyapps.io/demo

Proceedings of Statistics Canada Symposium 2021
Adopting Data Science in Official Statistics to Meet Society's Emerging Needs

Collaborative Data Science within Government of Canada: Development of R libraries for common tasks with Open Canada data

Dmitry O. Gorodnichy[1] and Patrick Little [2]

Abstract

Many Government of Canada groups are developing codes to process and visualize various kinds data, often duplicating each other's efforts, with sub-optimal efficiency and limited level of code quality reviewing. This paper informally presents a working-level approach to addressing this technical problem. The idea is to collaboratively build a common repository of code and knowledgebase for use by anyone in the public sector to perform many common data science tasks, and, in doing that, help each other to master both the data science coding skills and the industry standard collaborative practices. The paper explains why R language is used as the language of choice for collaborative data science code development. It summaries R advantages and addresses its limitations, establishes the taxonomy of discussion topics of highest interested to the GC data scientists working with R, provides an overview of used collaborative platforms, and presents the results obtained to date. Even though the code knowledgebase is developed mainly in R, it is meant to be valuable also for data scientists coding in Python and other development environments.

Key Words: Collaboration; Data science; Data Engineering; R; Open Government; Open Data; Open Science

Disclaimer: The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any agency of the Government of Canada.

1. Introduction

Data is the 'electricity' of the 21st century. It is everywhere. Everybody is using it, and one may not need special training or certification to work with data. However, in contrast to working with electricity, working with data relies on the use of data processing tools (codes), the number and complexity of which increases daily, and which are developed by other data practitioners. In other words, data science - by its nature - is the science that is fueled by collaboration of data scientists and that heavily relies on the quality of the shared data processing codes.

Globally, and in the Government of Canada (GC) in particular, data practitioners come from many different backgrounds and may not have equal level of code programming training, which hinders the development of high-quality (efficient, scalable and re-usable) codes for data science problems. This gap and the need for collaboration for data scientists, specifically those coding in R language, has been raised at the 2021 GC Data Conference Workshop on Data Engineering [1] in February of this year. This presentation describes a pan-government collaborative approach that has started following this workshop in order to address the identified need.

First (in Section 2), we review one the main challenges in data science, which is data engineering and which triggered the creation of the R4GC community of practice - the community that brings together data practitioners interested in using (or learning) the R language. The reasons why R is chosen as the language of choice for building the data science collaboration are explained in Section 3. Then (in Section 4), we describe the collaboration portals that have been set in support of the R4GC community activities and knowledgebase and overview key repositories and discussion threads that have been created there. Finally (in Section 5 and Appendix), we summarize the main outputs produced by the community to date. Whereas this paper serves as an introduction to the R4GC community efforts, the complete collection of the community knowledgebase is maintained as a shareable, free to use and collaboratively edited book (The R4GC Book) hosted at https://open-canada.github.io/r4gc. Feedback and contributions to this book are welcome, noting that only unclassified public domain knowledge is archived there.

[1]Dmitry Gorodnichy, Data Science Division, Chief Data Office, Canada Border Services Agency (Dmitry.Gorodnichy@cbsa-asfc.gc.ca)
[2]Patrick Little, Open Government Portal, Office of the Chief Information Officer, Treasury Board of Canada Secretariat (Patrick.Little@tbs-sct.gc.ca)

2. Data Engineering challenge

This effort has started from addressing the problem of data engineering, where it is understood - in analogy with the definition of software engineering by IEEE [2] - as the field of science and technology that deals with *"developing scientific and technological knowledge, methods, and experience to the design, implementation, testing, and documentation"* of data-driven systems and solutions; or - in analogy with the definition of software engineering at Google [3] - as the field that *"encompasses not just the act of writing code* [for data analysis, in our case], *but all of the tools and processes an organization uses to build and maintain that code over time"*.

Whereas globally the spectacular growth of data science tools development is overwhelmingly attributed to the collaborative open nature of the current data science code development practices, Canadian government is at the early stages of adopting open source industry standards for data coding and reporting. Major shift towards enabling and promoting such practices within the government has started recently, when Government of Canada adopted a number of policies in support of Digital, Open Science, and Open Government [4], when Shared Services of Canada deployed a number of collaboration platforms that are now available to all GC organizations, and when more GC IT organizations are embracing the use of open-source data science tools such as R and R Studio.

3. R advantages

A quick review of the presentations made at the 2021 International Methodology Symposium showed that about one third of projects presented at the symposium were done using R and RStudio. If counting the percentage of projects dealing with data visualization, including geo/spatial mapping, then this number becomes close to two thirds of all presented projects. Clearly, even though Python remains the most frequently used language for machine learning and complex data analysis tasks, R has become de-facto the language of choice in many government departments when it comes to developing interactive reports, web applications and complex visualizations, thanks to such its popular packages as "*rmarkdown*", "*ggplot*" and "*shiny*".

Other advantages of R highlighted at the Symposium include: common tidy data approach shared across all R packages; peer-reviewing and curation of packages by CRAN (Comprehensive R Archive Network), which is facilitated through the "*devtool*" R package that is specifically designed for this purpose; and RStudio-led movement for R education and deployment. All of these are in contrast to the 'wild-west' development of Python packages (which is a popular expression even among Python users), and are very important for enabling collaborative development of data science codes and knowledgebase. To summarize, Table 1 presents the Top 10, most quoted by the R4GC community, reasons to use R for data science.

Table 1: The R4GC community Top 10 reasons to use R for Data Science.

1.	Advanced graphics with *'ggplot2'* and its extensions
2.	Automated generation of reports, tutorials, and textbooks with *'rmarkdown'*
3.	Streamlined package development with '*devtools'*
4.	Development and deployment of interactive interfaces, web applications and dashboards with 'shiny'
5.	Convenient for geo/spatial computation and visualization
6.	Common tidy data design shared across packages
7.	Curated peer-tested repo of packages at CRAN (Comprehensive R Archive Network)
8.	RStudio IDE (Integrated Development Environment) on desktop and cloud (at *rstudio.cloud*)
9.	Full support and inter-operability with Python from the same IDE
10.	Global RStudio-led movement for R education and advancement (at *rstudio.com*)

Tackling the limitations of R

The main deficiency of R (in particular, when compared to Python) is that many R functions and packages are not developed to be memory- or processing time- efficient, and "by default", i.e., unless special care is taken in developing the code, most codes written in R are very slow and consume prohibitively large amount of memory, making it often impossible to run data processing codes on ordinarily laptops (with not more than 16 GB of RAM) such as those used by majority of GC employees. This is because all variables in R, including data array and data frame variables (which are naturally very large) are passed to a function by copying the entire variable content from one memory location to another, and then copying it back after the function completes its operation. In other programming languages, such as Python, C++, or Java, such operations normally are not done by passing the entire object to a function, but rather by passing only a pointer to (or memory address of) an object that needs to be processed by the function, which is

much faster and does not consume a lot of memory. This deficiency however can be practically entirely eliminated, should the R code be developed using the "efficiency-by-design" methodology that is offered by the 'data.table' R package. This is why much emphasis in the R4GC community discussions is put on encouraging the community members to use 'data.table' class by default (or always) instead of base 'data.frame' class.

Another deficiency of R is that object-oriented programming (OOP) is not native in R. There are a number of ways to implement an object in R - using S3, S4 and R6 classes, however each of them has its own limitations. This deficiency however is not critical, because fast and memory-efficient codes can still be developed with and without OOP, and, if needed, a portion of the OOP code can be also built in Python (or C++) and then used from within R using 'Rcpp' and 'reticulate' packages.

4. Overview of Collaborative Platforms

GCcode group: r4gc - https://gccode.ssc-spc.gc.ca/r4gc

GCcode is the GitLab solution that is accessible from within the GC network. As such, it allows one to view and update (pull and push) codes and documentation with a single click of button on a GC laptop from an RStudio. A step-by-step tutorial on how to do it is developed. The 'r4gc' group has been created in GCcode, where R4GC community codes, tutorials and other resources are gathered. It contains three main folders:

- */codes*. - This is where "raw" (not-reviewed, unedited) R codes contributed by GC community are uploaded. Currently, this includes codes for analyzing and visualizing PSES (Public Service Employee Survey) results, ATIP requests, COVID-19 statistics, and various codes for ease of day to day work and maintenance. Some codes are readily available to become packages, some are short code snippets taken from various blogs, question and answer portals, such as www.stackoverflow.org and www.rseek.org, and open-source textbooks.
- */gc-packages*. - This is where the work on packages being developed from the submitted "raw" codes is happening. Currently it includes repositories for building packages to process PSES results, COVID-19 data, and the utility functions package for data engineering and efficient data processing.
- */resources*. - This where the rest of knowledge-base is gathered, including the tutorials, slides, and codes presented at the community weekly 'Lunch and Learn' meetups.

GCcollab group: R4GC (Use R!) - https://gccollab.ca/groups/profile/7391537/R4GC

GCcollab allows one to participate in the discussion from within and outside GC network (for registered users). This makes it convenient for gathering information from any sources, including those that may not be available from within the GC network. A group called "R4GC" (originally called "Use R!") is created in GCcollab, where a number of discussion threads have been established to address the topics of highest interest for the R4GC community. They are listed in Table 2, divided into four main categories (parts).

Table 2: Taxonomy of R4GC community discussion topics

Part I: General discussions	*Part III: Visualization and Reporting*
1. Why R?	12. Literate programming and automated reports with 'rmarkdown'
2. Learn R: *R*ight way!	13. Data visualization with 'ggplot2' and its extension
3. Open-source textbooks	14. Geo/Spatial coding and visualization in R
4. Events and forums for R users	15. Interactive interfaces, applications and dashboards with 'shiny'
5. Using R with GC data infrastructure	16. Interactive html with 'plotly', 'Datatable', 'reactable'
6. Open source policies and guidelines	*Part IV: Machine Learning and AI*
Part II: Art of R programming	17. Entity resolution and record linking in R
7. Use R efficiently with 'data.table'!	18. Statistical tests and mixed-effects analysis in R
8. Python and R, unite!	19. Machine Learning and Modeling in R
9. From Excel to R	20. Text Analysis in R
10. Reading various kinds of data in R	21. Computer vision and Deep learning in R
11. Object oriented programming in R	22. Simulation and Optimization in R

These discussions are reviewed and updated regularly, commonly as part of weekly community meetups. A dedicated GCcollab subgroup (https://gccollab.ca/groups/about/7855030) is created for sharing minutes, notes and video-

recordings from these meetups. Additionally, a GCwiki page (https://wiki.gccollab.ca/UseR!) is also created to consolidate all discussion topics in one place and link them with other data science resources in the wiki space.

GitHub: open-canada - https://open-canada.github.io/UseR/

Inline with the GC Directive on Service and Digital [4], since most information gathered by the R4GC community is unclassified and comes from public domain, a public facing organizational account has been created on GitHub (https://github.com/open-canada) for sharing and growing the R4GC community knowledgebase. This is where public-facing community outputs are gathered, including the growing collection of Web Apps built with contributions from GC data scientists using open source tools and data (https://open-canada.github.io/Apps), as further described.

5. Outputs to date

Table 3 provides the list of tutorials and interactive applications developed by the community to date. The descriptions and screenshots of the applications are also provided in the Appendix. The updated list and more details are provided in "The R4GC Book" at https://open-canada.github.io/r4gc.

Table 3: R4GC community outputs

Tutorials:	**Interactive Shiny Apps :**
• R101: Building COVID-19 Tracker App from scratch • GCcode 101 for GC employees • R packages 101 for GC employees • Python and R unite! • Geo-spatial analysis and visualization (working with polygons, spatial intensity smoothing) • Text analysis (topics modeling, similar pages detection, string alignment and matching) • Working with Open Government Portal API within R (using 'ckanr' and 'adobeanalyticsr') • Automating R scripts with GitHub Actions	• PSES results interactive analysis and visualization • NLP analysis of ATIP requests • Automated COVID-19 results search and tracker • Geo-mapped current, historical, and predicted border wait times • Data engineering testbed demo tool for cleaning and linking noisy data records • Web page compassion tool to automate the detection of similar web pages (in gccode only) • Large data summarizer (in gccode only) • Masked name comparison tool (in gccode only)

Acknowledgements

R4GC is a collaborative effort of many people who have contributed to the development of the codes and knowledgebase that is gathered in the R4GC portals and highlighted in this paper. In particular, contributions from Jonathan Dench, Joseph Stinziano, Henry Luan, Eric Littlewood, Philippe-Israel Morin, Tony Machado, Chris Lavoie, Sylvain Paquet, Dejan Pavlic, Utku Suleymanoglu are gratefully acknowledged, as is the support from the wider international R community through stackoverflow.org portal, and conferences and webinars organized by the RStudio.
.

References

1. Dmitry Gorodnichy. "Data Engineering Challenges and Solutions: Demo of Shiny". Presentation at the 2021 GC Data Conference, Data Literacy Fest Workshop: https://youtu.be/QWv6E3e7bek.
2. "Software Engineering Body of Knowledge", IEEE Computer Society: https://www.computer.org/education/bodies-of-knowledge/software-engineering
3. "Software Engineering at Google.", O'Reilly Media: https://www.oreilly.com/library/view/software-engineering-at/9781492082781/preface01.html
4. "Directive on Service and Digital" – Appendix A: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32602.

Appendix: Web Apps developed by the R4GC community

A. PSES App: https://open-canada.github.io/Apps/pses

No other Open Canada data is of as much common interest across the government as the PSES results[3]. These data contain the information about all GC departments, their organizational structure and performance. A Shiny App prototype is developed to perform three most desired tasks one wished to do with these data (see Figure 1):

1. Vertical result tracking: results comparison across an organization, automated detection and visualization of the comparative performance within the organization - for any given PSES question.
2. Horizontal results tracking: results comparison over time - for any given unit, in comparison to the organization and Public Service averages
3. Performance summary by theme: automated generation of report cards that show the performance at each level of the organization for each of theme and in relation to the rest of the organization and Public Service average.

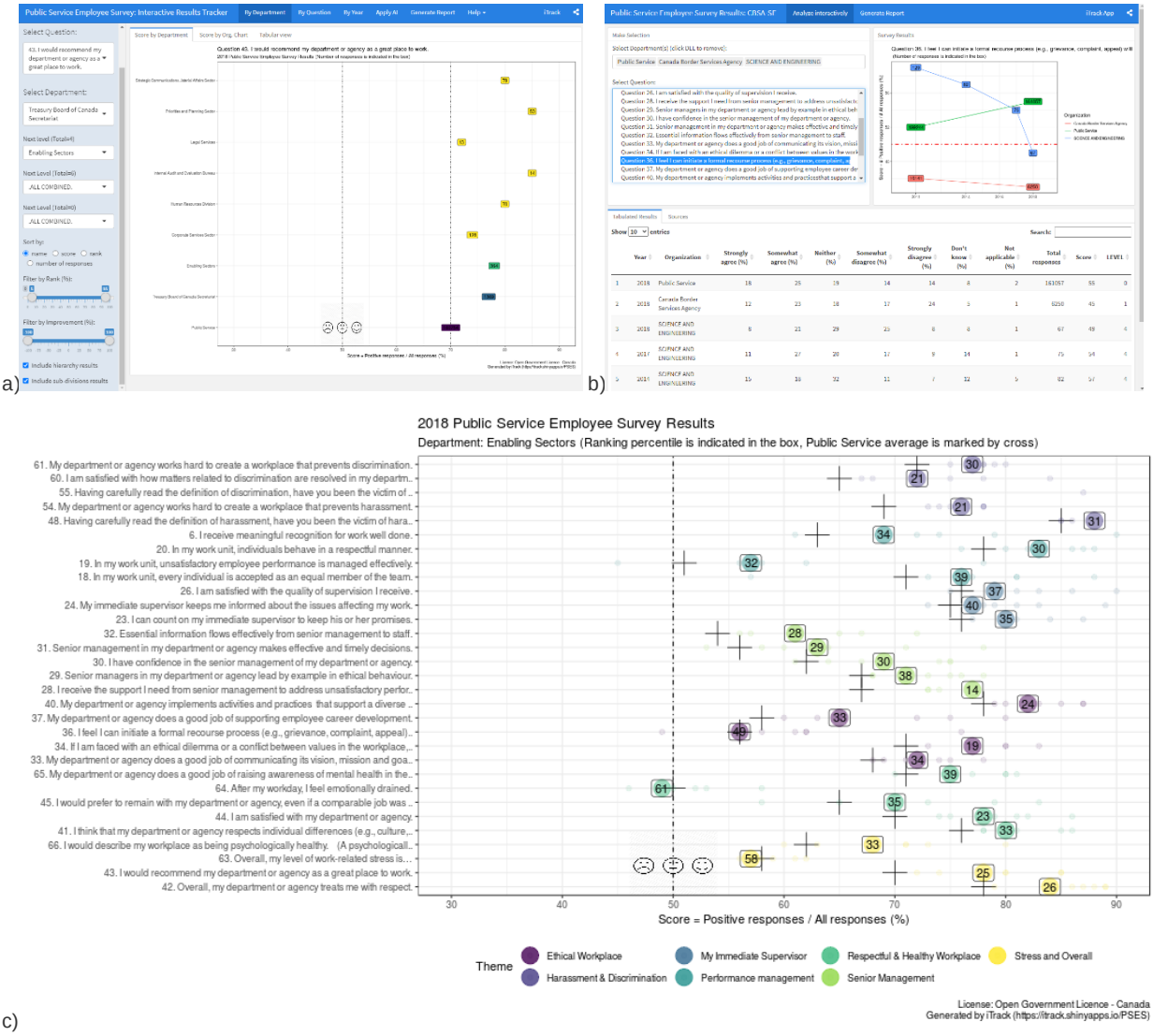


Figure 1. Key functionalities of the PSES App prototype: a) vertical results tracking - for any question over entire organization, b) horizontal results tracking - for any unit over time, and c) performance report summary - for each unit, by theme, and in comparison to the Public Service average (shown as crosses) and other units within organization (shown as small dots). The results can be displayed filtered and sorted by score, ranking percentile, org. structure, number of responses, by theme or question.

[3] https://www.canada.ca/en/treasury-board-secretariat/services/innovation/public-service-employee-survey.html

B. ATIP App: https://open-canada.github.io/Apps/atip

Another Open Canada dataset that relates to the performance of many GC departments is the ATIP requests dataset [4]. An interactive Natural Language Processing (NLP) application has been developed to enable the analysis and visualization of these requests for each participating department (see Figure 2). Its functionalities include: statistics summary, automated key-words and topic extraction using N-grams, document term matrix and Latent Dirichlet Allocation[5]. The topics can be visualized as word-clouds or as graphs that connect the related words.

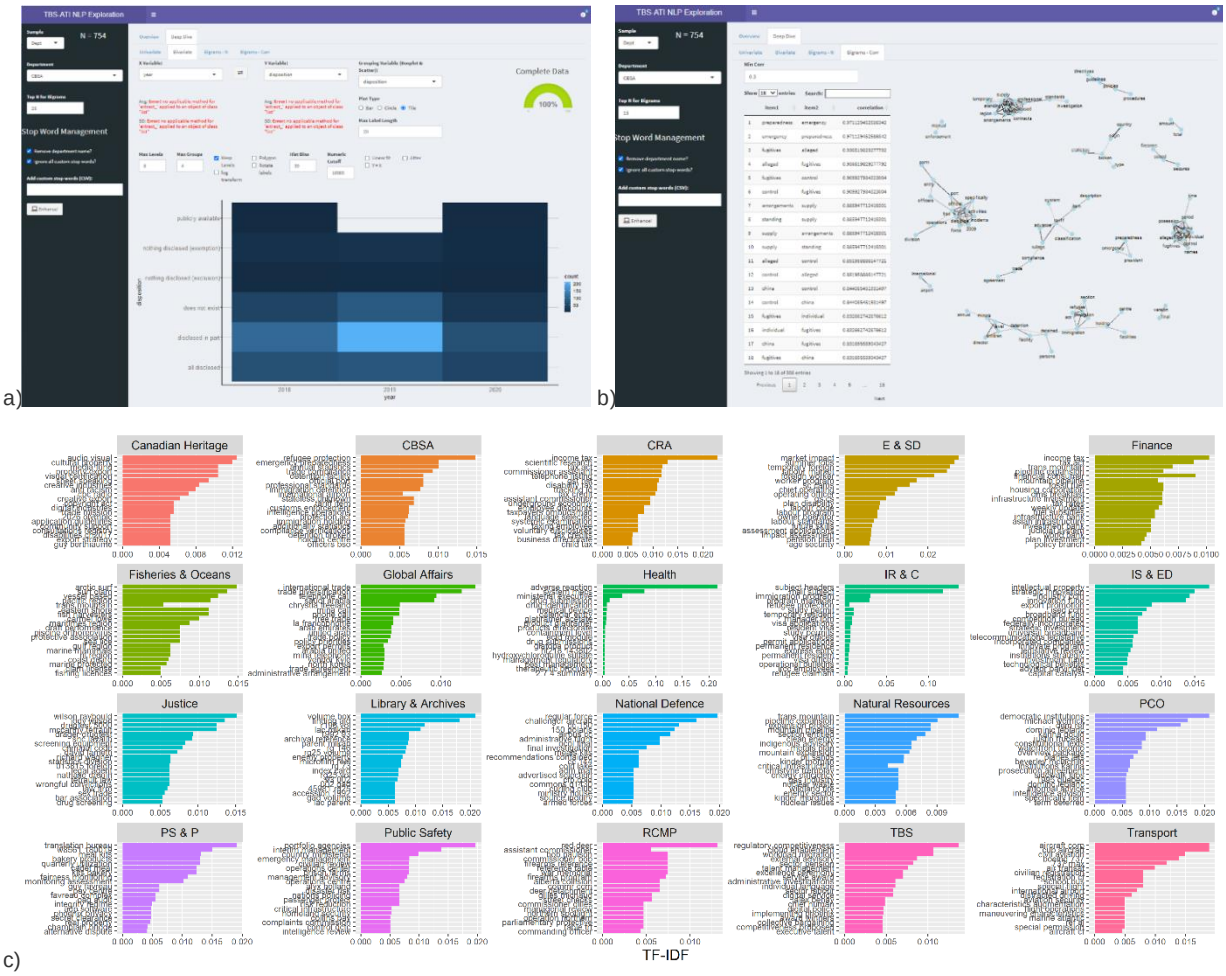


Figure 2. Key functionalities of the ATIP App: a) department specific bi-variable statics (such as dispositions by year, shown in the image), b) department specific key topics, visualized as correlated terms graphs, c) key topics for each participating department, visualized as N-gram frequency bars (such as 2-gram, or two-word combination, shown in the image).

[4] https://open.canada.ca/en/search/ati

[5] See "Text Mining with R!" by Julia Silge and David Robinson (https://www.tidytextmining.com) for definitions of the terms.

C. COVID App: https://open-canada.github.io/Apps/covid

This App was built at the beginning of the pandemic. It uses open COVID-19 data for Canada and US to allow one to summarize, search and sort the results by geographical proximity using a variety of criteria (see Figure 3). More details about it and a tutorial on how to build it are provided at https://open-canada.github.io/UseR/learn2020

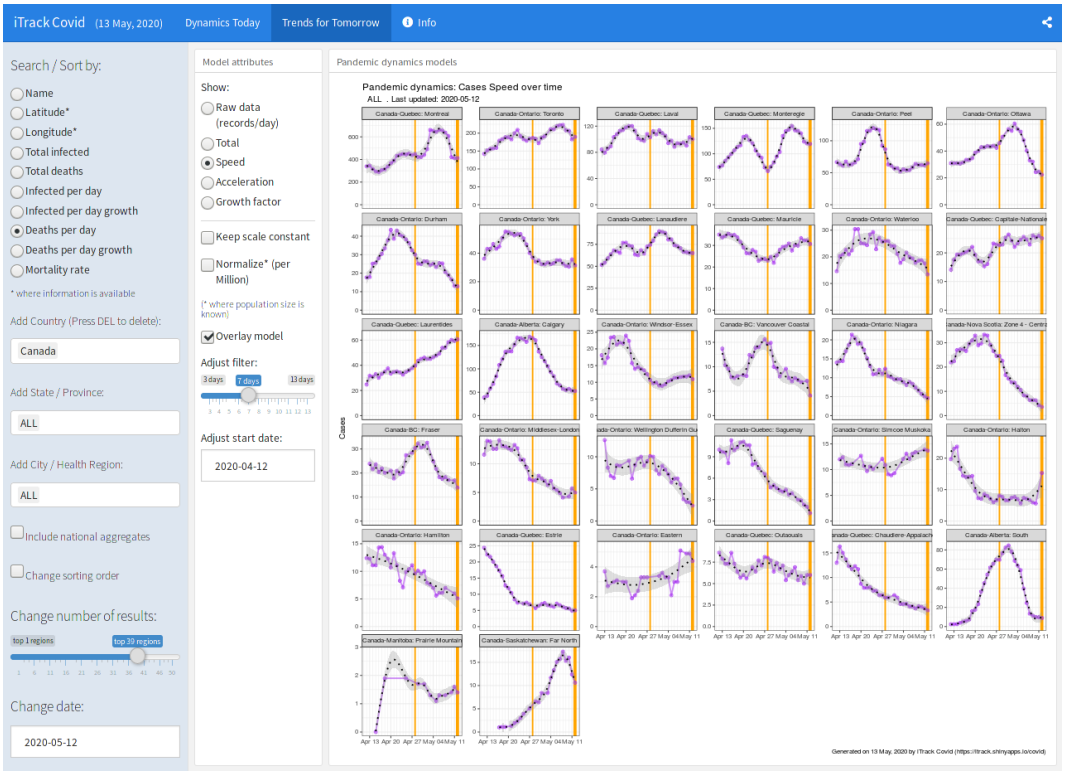


Figure 3. COVID App allows one to summarize, search and sort COVID-19 results by geographical location using a variety of criteria.

D. Border Wait Times App: https://open-canada.github.io/Apps/border

This App (shown in Figure 4) combines Open geo-spatial data with Open historical and current border wait times data to predict and visualize delays at Canadian land border crossings. The App is included (entitled iTrack-Border) in the Open Canada Apps Gallery at https://open.canada.ca/en/apps, where more information about it can be found.

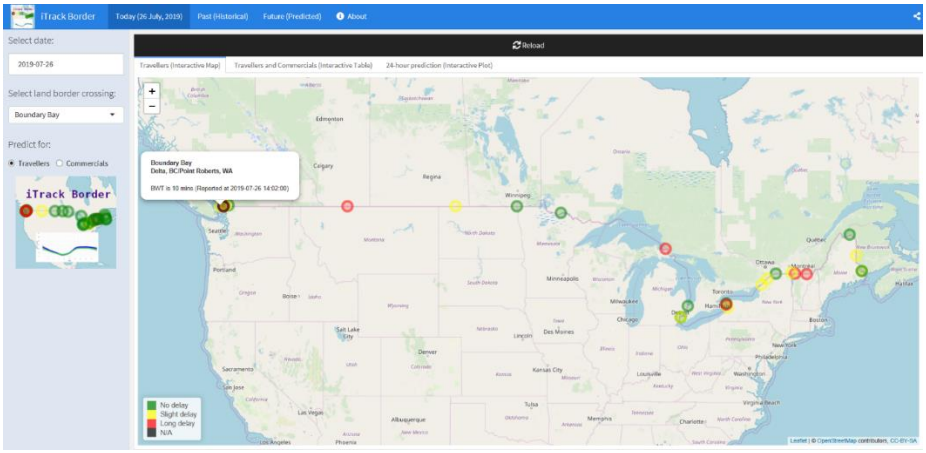


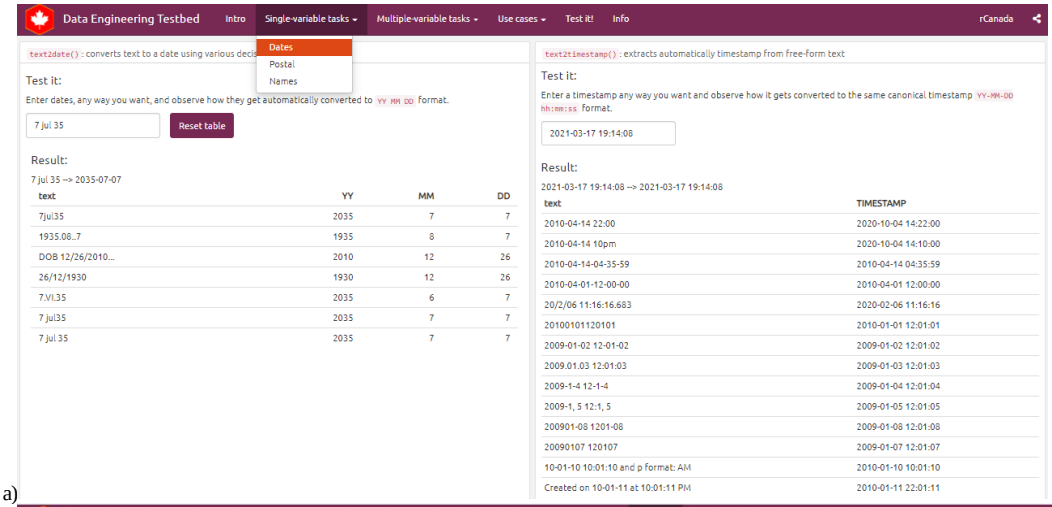
Figure 4. Border Wait Time Interactive Tracker (iTrack-Border) application allows one to visualize and predict current Border Wait Time at Canadian land border crossings.

E. Data Engineering Testbed App: https://rCanada.shinyapps.io/demo

This App was demonstrated and used for training in the "Data Engineering Challenges and Solutions: Demo of Shiny" workshop at 2021 GC Data Conference, Data Literacy Fest [1]. Its key functionalities include:

1. Single variable cleaning tasks (demo mode): automated extraction of date and timestamps from arbitrarily typed strings (as shown in Figure 5.a); rectification and correction of Canadian postal codes; fuzzy matching of names.
2. Multi-variable linking tasks (demo mode): city names deduplication; deduplication and linking of name records.
3. Use cases (live Web crawling mode): automated extraction of name variants and important dates from Web.
4. "Test it!" mode: fuzzy linking and deduplication of user-uploaded multi-variable data records, using a variety of string similarity metrics and thresholds (as shown in Figure 5.b).

a)



b)

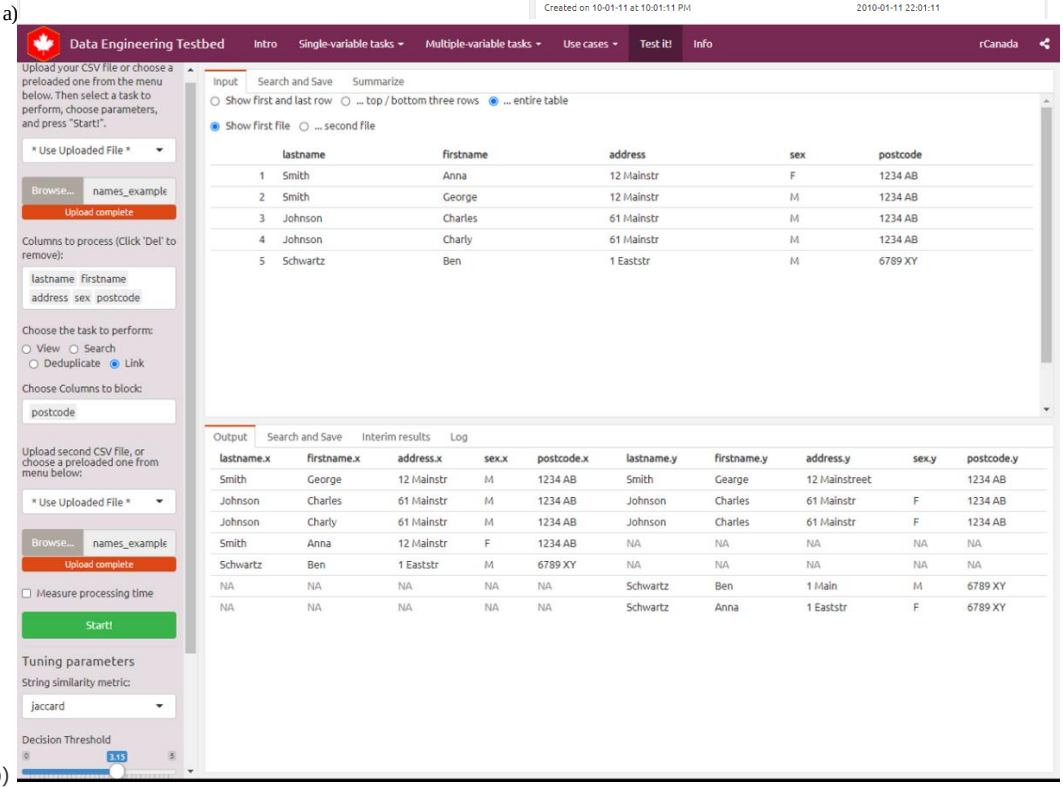


Figure 5. Data Engineering Testbed App allows one to test various data cleaning (a) and data linking (b) techniques.

566-02-48305

"Lunch and Learn" (Data Science with R) meet-up



Gorodnichy, Dmitry

Accept | Tentative | Decline | ...

Required La
Vi

ejan;
arah; **+7 others**

Thu 2022-02-03 10:48 AM

February 4, 2022 12:05 PM-01:05 PM

Every Friday from 12:05 to 12:55
Lunch and Learn (Data Science with R) - Friday's R4GC Community Meet-ups

Mark your calendars and share with others, as appropriate.
If there's anything you'd like to hear/learn about at our Meetups – please let me know!

February 4, 2022:

LUNCH AND LEARN (Special Edition): One Year since Vaccination. What have we learnt - using Open Canada Data and Data Science : GCcollab

By request of many and in coordination with SPB Engagement team, we'll be having this very Special Edition of our weekly Lunch and Learn "Data Science with R" meetup dedicated entirely to the one-year anniversary of COVID-19 vaccine in Canada.

Much is known now, which was not known a year ago (when vaccination started with high-risk population), or even in May (when vaccination was extended to general public), and even in October (when GC issued the mandatory vaccination policy for its employees).

There will no usual R coding this time, only the presentation and the analysis of the results obtained to date. Everyone is invited to come, invite others, and participate in the discussion - whether you work with data or not.

The presentation will also provide information on how to make a voluntary vaccine side effect reporting via Health Canada Portal.

Comments and feedback are sought for. Please email them directly to dmitry.gorodnichy@cbsa-asgc.gc.ca.
Slido will also be setup (code: **r4gc**) to allow everyone to provide comments/questions anonymously.

Dial-in numbers:

Microsoft Teams meeting

Click here to join the meeting

Or call in (audio only)

+1 343-804-0291,,185924081# Canada, Ottawa-Hull

Phone Conference ID: 185 924 081#

Everything else:

The R4GC Book: https://open-canada.github.io/r4gc/

Subject: “Lunch and Learn” (Data Science with R) meet-up

Start: Fri 2022-02-04 12:05 PM
End: Fri 2022-02-04 01:05 PM
Show Time As: Tentative

Recurrence: Weekly
Recurrence Pattern: every Friday from 12:00 PM to 01:30 PM

Meeting Status: Not yet responded

Organizer: Gorodnichy, Dmitry
Required Attendees: Lavoie, Christopher; Davidson, Fiona; Mojaverian, Nassim; Pavlic, Dejan;
Vincent, Nicolas; Yang, Jiangbin; Suleymanoglu, Utku; Mcculloch, Sarah; Saadawi, Hesham;
Carayanniotis, Jason; Pouliot, Frédéric; Paquet, Sylvain; Nahmias, Sherine S.; Till, Matthew; Afshar-
Zanjani, Kaveh

Every Friday from 12:05 to 12:55
Lunch and Learn (Data Science with R) - Friday’s R4GC Community Meet-ups

Mark your calendars and share with others, as appropriate.
If there’s anything you’d like to hear/learn about at our Meetups – please let me know!

February 4, 2022:

LUNCH AND LEARN (Special Edition): One Year since Vaccination. What have we learnt - using Open
Canada Data and Data Science : GCcollab

By request of many and in coordination with SPB Engagement team, we'll be having this very Special Edition of our weekly Lunch and Learn "Data Science with R" meetup
dedicated entirely to the one-year anniversary of COVID-19 vaccine in Canada.

Much is known now, which was not known a year ago (when vaccination started with high-risk population), or even in May (when vaccination was extended to general public), and even in October (when GC issued the mandatory vaccination policy for its
employees).

There will no usual R coding this time, only the presentation and the analysis of the
results obtained to date. Everyone is invited to come, invite others, and participate in the
discussion - whether you work with data or not.

The presentation will also provide information on how to make a voluntary vaccine side
effect reporting via Health Canada Portal.
Comments and feedback are sought for. Please email them directly to dmitry.gorodnichy@cbsa-asgc.gc.ca.
Slido will also be setup (code: r4gc) to allow everyone to provide comments/questions
anonymously.
Dial-in numbers:
Microsoft Teams meeting
Click here to join the meeting
Or call in (audio only)
+1 343-804-0291,,185924081# Canada, Ottawa-Hull
Phone Conference ID: 185 924 081#
Everything else:
The R4GC Book: https://open-canada.github.io/r4gc/

One Year since Vaccination. What we have learnt - using Open Canada Data and Data Science

Dr. Dmitry O. Gorodnichy

Ottawa, 4 February 2022
R4GC Community (Lunch and Learn) Meetup
open-canada.github.io/r4gc

Disclaimer

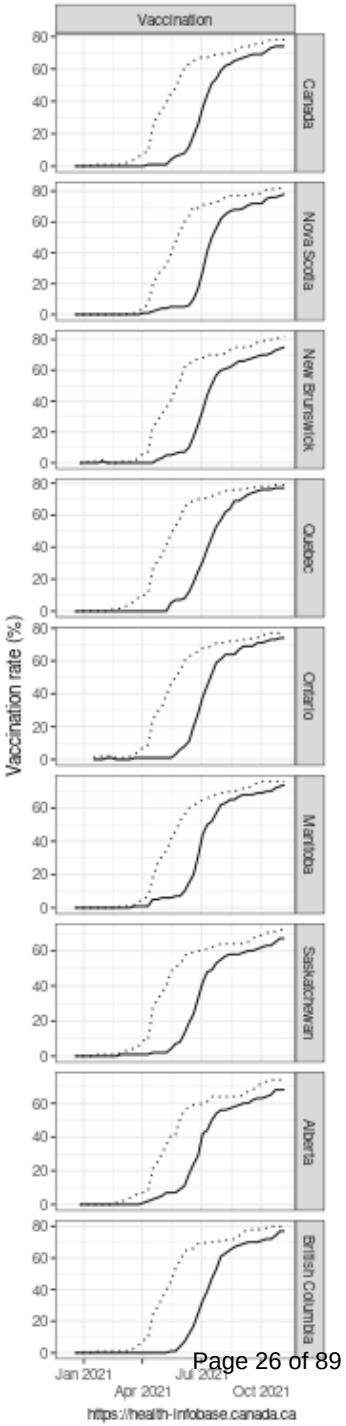
The views expressed in this publication are those of the authors.

They do not purport to reflect the views of any Canadian Agency, University, or Department.

The authoritative source for COVID-19 information is Canada.ca/coronavirus

Milestones

- December 14, 2020: First shot administered
- January 1, 2021: First Canadian fully immunized (2 shots)
- March 1, 2021: Full vaccination rate reached 1%
- May 1, 2021: Full vaccination rate reached 3%
- November 1, 2021: Full vaccination rate reached 75%



Lets put it in perspective



Diversity of Canadians:

- Affected by mental health (20%)
- Smokers (15%)
- Vegetarians (8%)
- Data scientists (~0.1%)
- Ice water dippers (~0.01%)



Google: “CBC Gorodnichy”

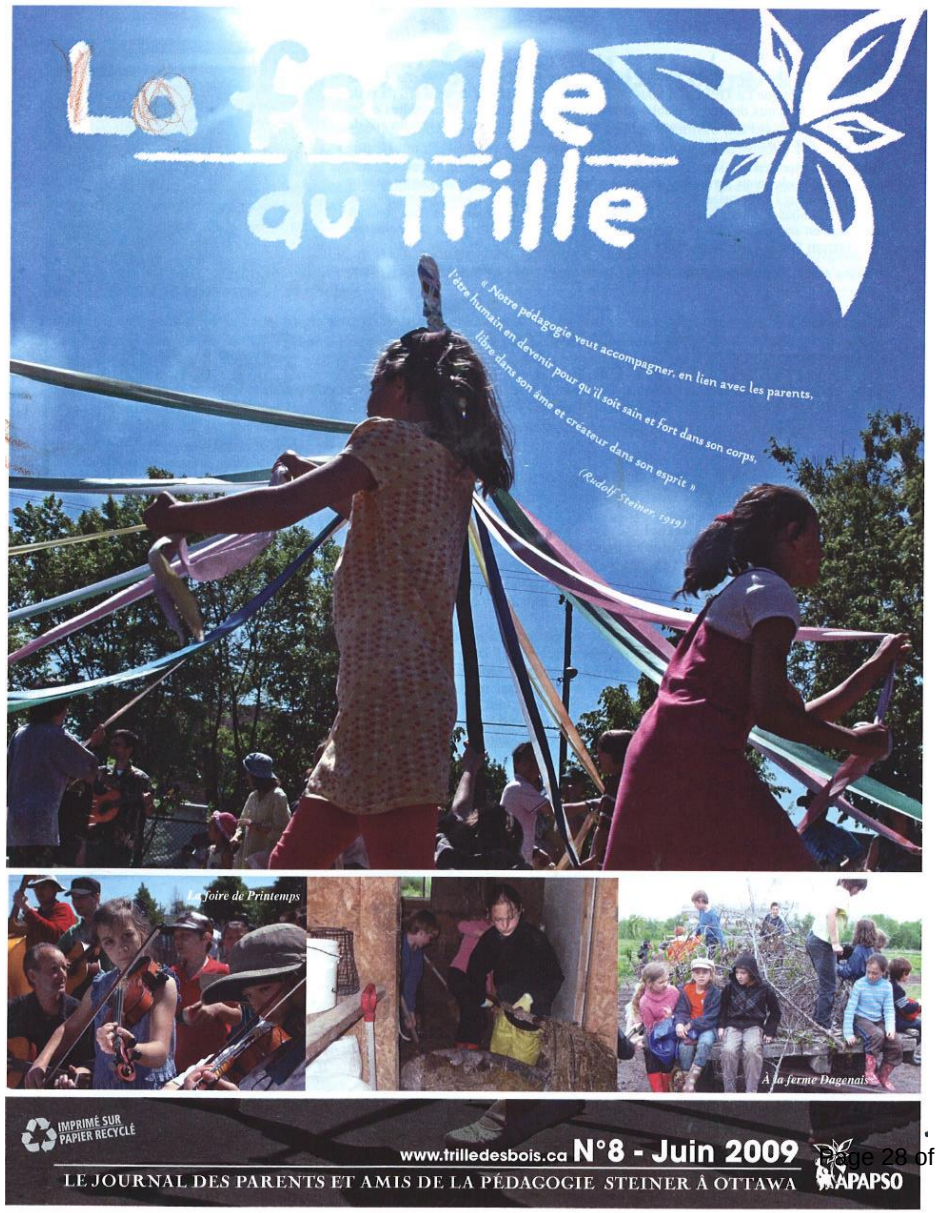


Luigina Fabrianesi, left, and Alexandra Gorodnichy, right, chill in the water. (Giacomo Panico/CBC)

Another perspective (generational)

Children are most affected

Why are we vaccinating children against COVID-19?, Toxicology Reports, Volume 8, 2021, Pages 1665-1684, ISSN 2214-7500 (14 September 2021). https://doi.org/10.1016/j.toxrep.2021.08.010



Another perspective (technological)

All claims made by Vendors need to be validated – by third party

- That's what we do

Example: AVATAR

- "Clear safeguards needed around technology planned for border checkpoints", CBC, May 2021

- "Design and Evaluation of Biometric-enabled Interview Assisting Traveller Screening Technology" (D. Gorodnichy), DRDC-RDDC-2018-C223



https://www.cbc.ca/news/opinion/opinion-technology-border-canada-1.6005907 - Google: "CBC AVATAR CBSA"
https://cradpdf.drdc-rddc.gc.ca/PDFS/unc331/p808530_A1b.pdf - Google: "DRDC Gorodnichy Traveller Screening"

Policy on Scientific Integrity

https://science.gc.ca/eic/site/063.nsf/eng/h_97643.html

- 2017: Govn't creates Chief Science Advisor post (Dr. Mona Nemer - uOttawa)
- 2018: Introduced Scientific Integrity Policy*
- Government has become (more) Open, (more) pro-science
- Many Open Canada databases are now available
- A lot with data related to COVID-19, vaccines and vital statistics (deaths).

So why not to use these Open Canada Datasets
to practice Data Science / Visualization skills,
and help Canadians at the same time?

That's what we do: https://open-canada.github.io/Apps

Policy on Scientific Integrity (cntd)

https://science.gc.ca/eic/site/063.nsf/eng/h_97643.html

"GC recognizes two complementary approaches to fostering a culture of scientific integrity:

- One focuses on instilling the virtues that underlie responsible conduct in research, science and related activities (s. 7.2.1).

- A second focuses on the procedure for bringing allegations of breaches forward, the investigation of these allegations, and the consequences of a finding that a breach has occurred (s. 7.2.2)"

Official sources

[1] ‘Cases following vaccination’, COVID-19 Daily Epidemiology Update, Public Health Agency of Canada, https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html

[2] Hospitalizations by vaccination status, COVID-19 (coronavirus) in Ontario, https://covid-19.ontario.ca/data/hospitalizations#hospitalizationsByVaccinationStatus, Weekly epidemiology report (PDF) available on the Government of Canada’s COVID-19 data trends page

[3] ‘Reported side effects following COVID-19 vaccination in Canada’, Canadian COVID-19 vaccination safety report, Public Health Agency of Canada, https://health-infobase.canada.ca/covid-19/vaccine-safety/

[4] Statistics Canada, Canadian Vital Statistics - Death database . Provisional weekly death counts, by selected grouped causes of death. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310081001 (Dashboard: https://open-canada.github.io/App/vitals)

[5] Recalls and safety alerts, Government of Canada: https://recalls-rappels.canada.ca/

Big Questions

- Should COVID-19 vaccines be mandatory? For everyone? Even now?
- What do we know now - compared to year ago?
- How really efficient and safe are they - based on data *after* the start of mass vaccination?
- How do they compare to natural immunity?
- What is the cost?
- Can Canadians have unbiased, safe, respectful discussion about it all?

Open Canada Data and Data Science (with R) allows to do that

What did Chief Science Advisor say (1)

https://science.gc.ca/eic/site/063.nsf/eng/h_98229.html

Scientific Considerations for Using COVID-19 Vaccination Certificates

Report of the Chief Science Advisor of Canada, March 31, 2021

Scientific uncertainties: Given the short time since the COVID-19 vaccines have become available, it is not surprising that many scientific uncertainties persist and are the subject of intense ongoing studies.

Ethical and Social Considerations: It should also be noted that while the anticipation of more freedom may be an incentive for some to get vaccinated, vaccine acceptance could decrease in others if there was a sense of coercion tied to using vaccination certificates.

Legal Considerations: The processes by which vaccine certificates are issued and controlled need to be fraud proof

Conclusion: all levels of government could work to develop a framework that Maximizes consistent post-vaccine monitoring,

What did Chief Science Advisor say (2)

https://science.gc.ca/eic/site/063.nsf/eng/h_98291.html

COVID-19 vaccine-associated myocarditis/pericarditis
Report of the Chief Science Advisor of Canada, July 16, 2021

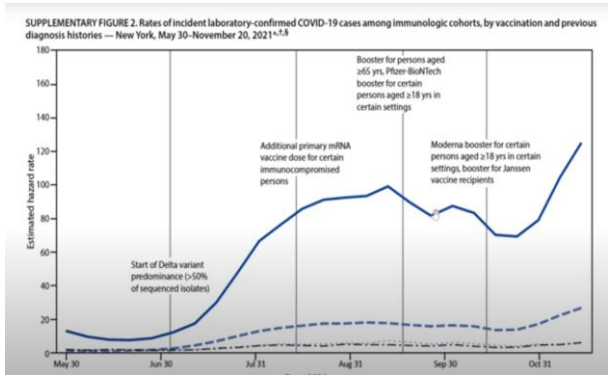
Priority actions moving forward: The emerging issue of vaccine-associated heart disease requires attention on two important levels: addressing data and knowledge gaps on the one hand, and promoting awareness and clinical care on the other

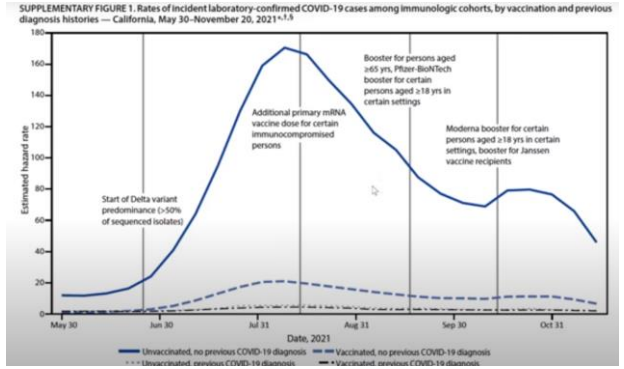
Nothing on the topic of vaccine since then…

So, What did other countries' governments say?

Stats: vaccines vs. natural immunity

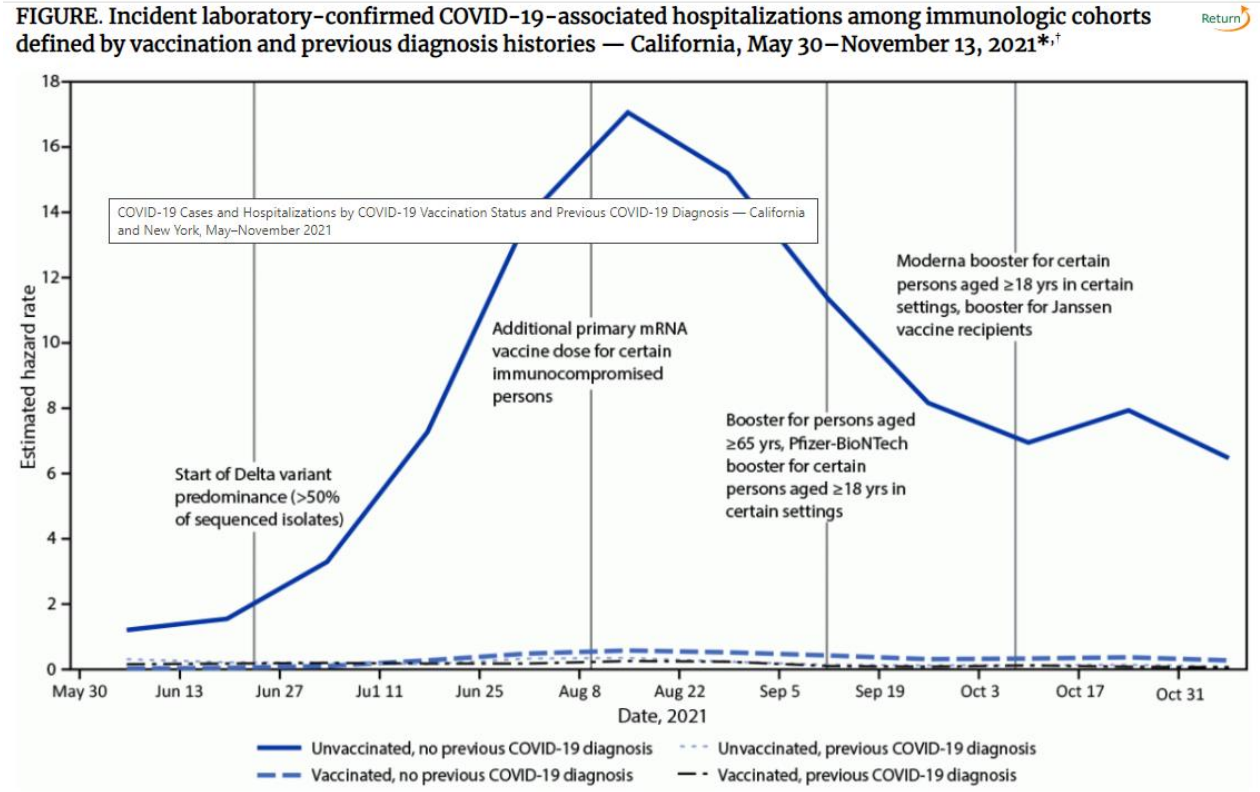
https://www.cdc.gov/mmwr/volumes/71/wr/mm7104e1.htm





Based on data from California and New York from May to November 2021:

- Vaccination (upper dash line) significantly reduced risk of hospitalization (solid line) (x10-x20)
- However, natural immunity (lower dash line) reduces this risk even more



* The SARS-CoV-2 Delta variant exceeded 50% of sequences in U.S. Department of Health and Human Services Region 9 (containing California) during the week of June 26. https://covid.cdc.gov/covid-data-tracker/#variant-proportions

† Estimated hazard rate is laboratory-confirmed COVID-19-associated hospitalizations per 100,000 person-days visualized at midpoint of each reporting interval.

Based on these new data, the main question becomes:

What is less costly, safer, and to whom: having a vaccine or become infected with COVID ?

Stats: Google “Canada, covid”

Population: 38.246.108

Total (since the start of pandemic)

- COVID Cases: 3.1 M
- COVID Deaths: 34 K ~= 1% Cases
  - Most are comorbidities (>90% based on UKG and USG data)
  - Most in people 65+ (>90% based on GC, UKG, USG data)

- COVID Deaths for people <65 with no health problems < 0.01% Cases

Lets put it perspective

Within the same period

- All deaths :
- Cancer:
- Heart problems:
- Homicides:
- Suicides:
- COVID:
- COVID w/o comorbidities*:
- COVID w/o comorbidities for people <65 *:

To compute these, use Vitals App Calculator:
https://open-canada.github.io/Apps/vitals (click on Statistics Tab)

* needs to be estimated using other data

Numbers, numbers, numbers…

The way Stats are reported tells …
… about the (embedded) bias of Stats reporters

Related terms:

- political bias
- vendor bias
- confirmation bias
- group bias
- propaganda

Reverse-engineering Algorithmic Bias from COVID data

What is Algorithm?

Predefined sets of commands which produces desired output from input: Deterministically or Stochastically (in Probabilities)

Algorithms for computers = AI

Algorithms for humans = Policies

By examining the input (raw data) and the output produced by the algorithm (posted stats), one can uncover the algorithm and the bias embedded in it. This is called “reverse-engineering”.

Case study 1: Vaccine efficacy reporting

Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine, C4591001 Clinical Trial Group, N Engl J Med 2020; 383:2603-2615, https://www.nejm.org/doi/full/10.1056/nejmoa2034577

21,720 with BNT162b2 - 8 cases
21,728 with placebo - 162 cases

Difference:
(162 - 8) / 21,720 = 0.7%

which is reported as 95%

Table 1: COVID-19 vaccines currently in use in different countries

Platform	Name	Number of Doses	% Efficacy**
mRNA:	Moderna*	2	~94
	Pfizer-BioNTech*	2	~95
Viral Vectors:	AstraZeneca & Serum Institute of India COVID-19*	2	~76
	Janssen*	1	~66
	CanSino	1	~65
	Gamaleya ('Sputnik')	2	~92
Inactivated Virus:	Sinovac	2	~50
	Sinopharm (Bejing)	2	~79
	Sinopharm (Wuhan)	2	~72
	Bharat biotech	2	~81
Protein/VLP:	Vektor Institute/EpiVacCorona	2	N/A

** Vaccines approved for emergency use in Canada.*

"COVID-19 vaccine efficacy and effectiveness—the elephant (not) in the room", The Lancet, 7, E279-E280, https://doi.org/10.1016/S2666-5247(21)00069-0
Principles of Epidemiology in Public Health Practice: https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section6.html

Case Study 2: Vaccinated vs. Non-vaccinated deaths

https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html

Recall: There were no fully-vaccinated before January, <1% before March, <3% before May, and … most* (50-75%), deaths happened exactly then.

Try it yourself:
- Open link above, write down numbers in red boxes and compare them a week later!
- Or, compare these numbers in past reports since November (archived here).

Observe:
These numbers monotonically change (one down, one up) every week - according to math behind them…

*computed using Vitals App Calculator

Table 2. Characteristics and severe outcomes associated unvaccinated, partially vaccinated and fully vaccinated confirmed cases reported to PHAC, as of December 18, 2021

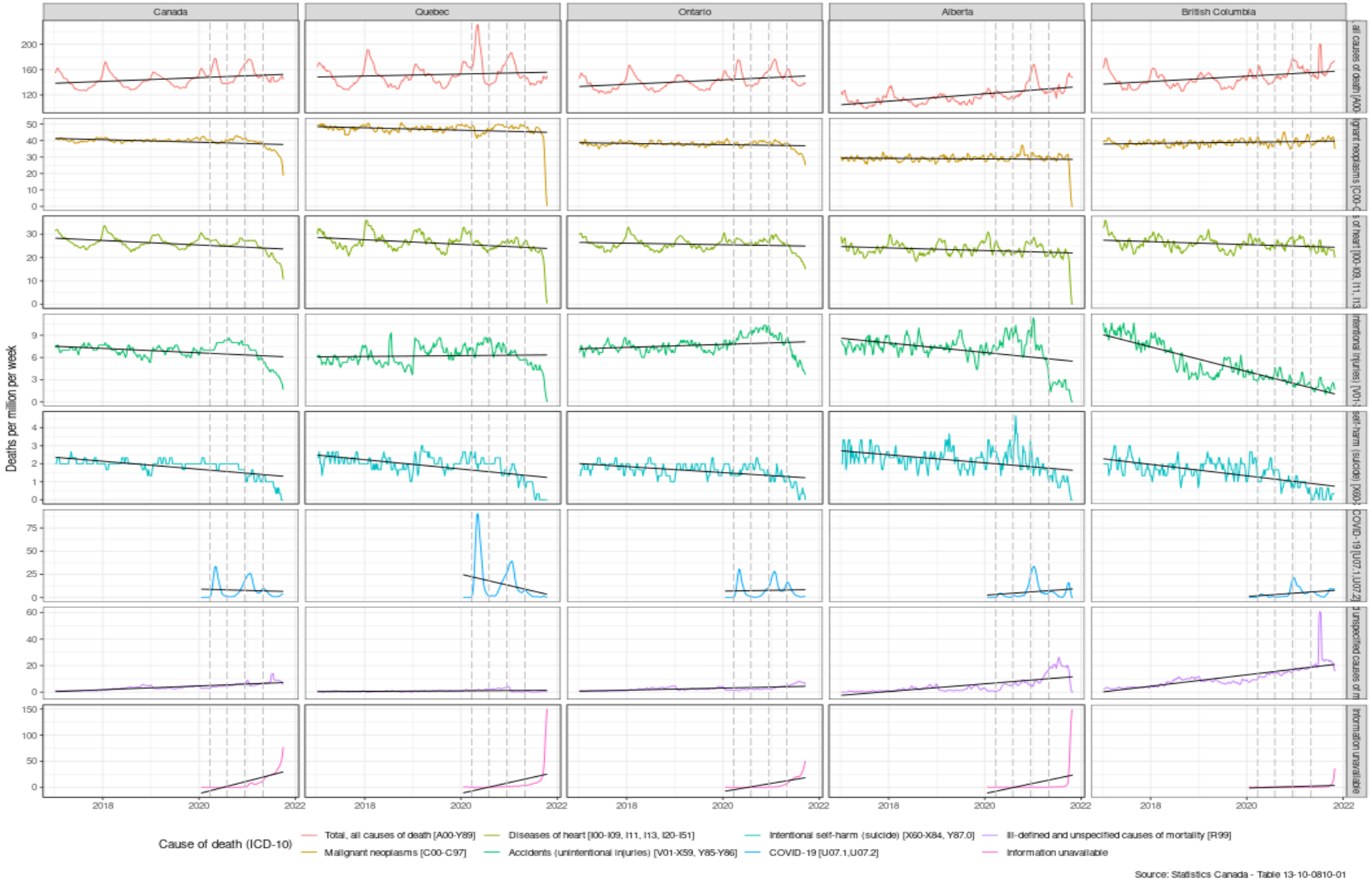
		Unvaccinated (n=727,925)	Cases not yet protected (n=43,471)	Partially vaccinated (n=53,171)	Fully vaccinated (n=122,843)	Total cases (n=947,410)
Gender*	Male	371,897 (78.5%)	20,959 (4.4%)	24,980 (5.3%)	56,197 (11.9%)	474,033 (100%)
	Female	354,365 (75.2%)	22,454 (4.8%)	28,113 (6.0%)	66,344 (14.1%)	471,276 (100%)
Hospitalizations		40,788 (79.5%)	3,062 (6.0%)	3,374 (6.6%)	4,099 (8.0%)	51,323 (100%)
Deaths		8,013 (75.6%)	759 (7.2%)	744 (7.0%)	1,077 (10.2%)	10,593 (100%)

Source: Detailed case information received by PHAC from provinces and territories, since December 14, 2020

As of today, February 4, 2022 : Total cases (n=1,458,433)
Fully vaccinated 2,032 (16.8%), +3% since last week, +8% since November

Case Study 3: ‘COVID’ deaths reporting

Go to https://open-canada.github.io/Apps/vitals



Latest News:

- Denmark (where full vaccination rate is 81%) just lifted all COVID restrictions:

https://www.telegraph.co.uk/global-health/science-and-disease/land-covid-now-no-worse-cold/, 1 February 2022. *"The land where Covid is now no worse than a cold"*

- *A Literature Review and Meta-Analysis of the Effects of Lockdowns on COVID-19 Mortality, Studies in Applied Economics,* The Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise (January 2022)

https://sites.krieger.jhu.edu/iae/files/2022/01/A-Literature-Review-and-Meta-Analysis-of-the-Effects-of-Lockdowns-on-COVID-19-Mortality.pdf

Gorodnichy, Dmitry (Ph.D. / Dr.)

From: Luan, Henry
Sent: February 7, 2022 04:21 PM
To: Gorodnichy, Dmitry; Morin,Philippe-Israel (ECCC); Littlewood,Eric (ECCC); Little, Patrick
Subject: RE: change of logistics for L&L

Hi Dimitry,

I made some comments in yellow. ☺

Best,

Henry

From: Gorodnichy, Dmitry <Dmitry.Gorodnichy@cbsa-asfc.gc.ca>
Sent: February 7, 2022 3:38 PM
To: Luan, Henry <Henry.Luan@cra-arc.gc.ca>; Morin,Philippe-Israel (ECCC) <Philippe-Israel.Morin@ec.gc.ca>; Littlewood,Eric (ECCC) <Eric.Littlewood@ec.gc.ca>; Little, Patrick <Patrick.Little@tbs-sct.gc.ca>
Subject: RE: change of logistics for L&L

Henry, you will need to create a new MS Team link – to be the host of it. (Mine will not work)

What I can do meanwhile – is also create a Zoom link (accessible from anywhere) and we can dedicate next few meetups to development of the Open Canada/Ontario dashboard
I can show, what has been done so far. – unless, Henry you want to show your stuff I think this is great. I prefer the zoom idea for now (people can access from web browsers without software installation). I think the Open Canada/Ontario dashboard is more relevant for everyone.

I do have some demo codes for building a simple pre-trained image classifier (with RShiny+ Javascript). I could show it too. I think it would be easier if we work on the dashboard for the next two weeks; I could do a formal presentation after Eric and Philippe come back.

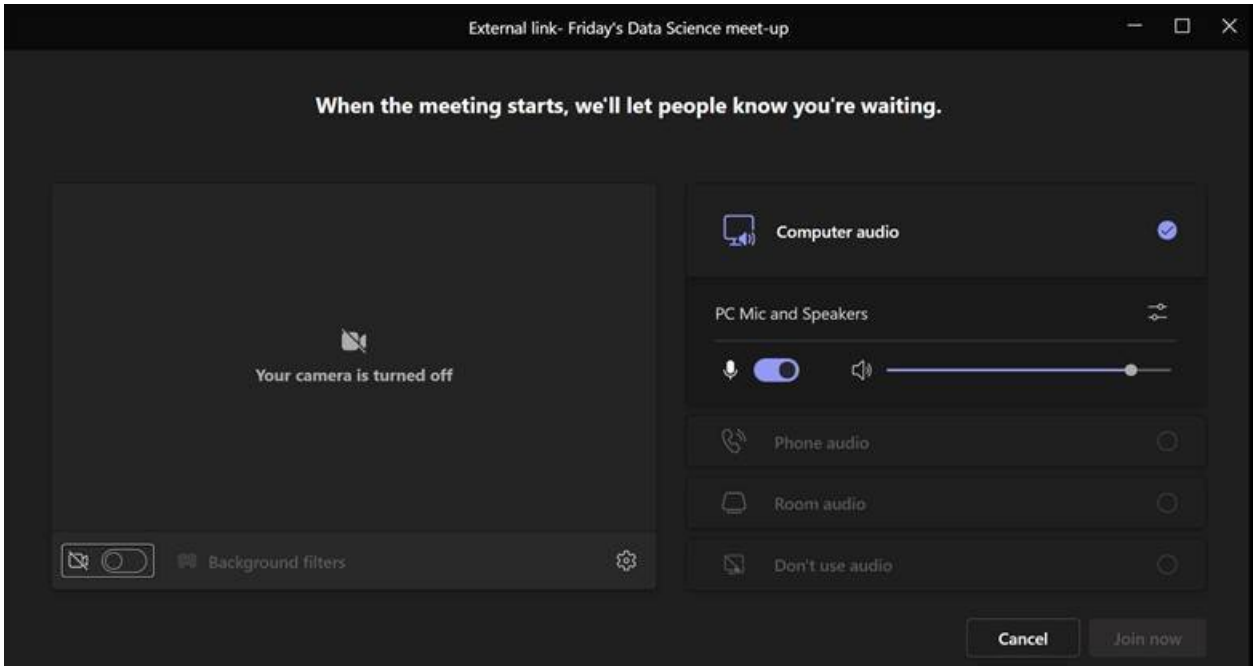
Let me know

Dmitry

From: Luan, Henry <Henry.Luan@cra-arc.gc.ca>
Sent: February 7, 2022 12:13 PM
To: Gorodnichy, Dmitry <Dmitry.Gorodnichy@cbsa-asfc.gc.ca>; Morin,Philippe-Israel (ECCC) <Philippe-Israel.Morin@ec.gc.ca>; Littlewood,Eric (ECCC) <Eric.Littlewood@ec.gc.ca>; Little, Patrick <Patrick.Little@tbs-sct.gc.ca>
Subject: RE: change of logistics for L&L

Hi everyone,

It looks like my MSTeam won't let me host the meeting (see the attachment). When I tried to join the meeting, I had to wait for the organizer to approve. Any idea how I can resolve this issue?



Also, there seems to be no agenda for this week/upcoming weeks? Dimitry, would you still be able to add them?

Best,

Henry

-----Original Message-----
From: Gorodnichy, Dmitry <Dmitry.Gorodnichy@cbsa-asfc.gc.ca>
Sent: February 7, 2022 12:02 PM
To: Morin,Philippe-Israel (ECCC) <Philippe-Israel.Morin@ec.gc.ca>; Littlewood,Eric (ECCC) <Eric.Littlewood@ec.gc.ca>; Luan, Henry <Henry.Luan@cra-arc.gc.ca>; Little, Patrick <Patrick.Little@tbs-sct.gc.ca>
Subject: RE: change of logistics for L&L

Sure. Thanks for letting us know

Dmitry

-----Original Message-----
From: Morin,Philippe-Israel (ECCC) <Philippe-Israel.Morin@ec.gc.ca>
Sent: February 7, 2022 11:14 AM
To: Gorodnichy, Dmitry <Dmitry.Gorodnichy@cbsa-asfc.gc.ca>; Littlewood,Eric (ECCC) <Eric.Littlewood@ec.gc.ca>; Luan, Henry <Henry.Luan@cra-arc.gc.ca>; Little, Patrick <Patrick.Little@tbs-sct.gc.ca>
Cc: dmitry.gorodnichy@gmail.com
Subject: RE: change of logistics for L&L

Hello Dmitry,

Thank you for your e-mail.

Due shifting priorities with my schedule, I will not be able to host/attend the next two L&L meetings. I will also need to pause my contribution to the R+python+Git project and other related projects.

Apologies for the short notice.

Regards,

Philippe

From: Gorodnichy, Dmitry <Dmitry.Gorodnichy@cbsa-asfc.gc.ca>
Sent: Monday, February 7, 2022 10:15 AM
To: Morin,Philippe-Israel (ECCC) <Philippe-Israel.Morin@ec.gc.ca>; Littlewood,Eric (ECCC) <Eric.Littlewood@ec.gc.ca>; Luan, Henry <Henry.Luan@cra-arc.gc.ca>; Little, Patrick <Patrick.Little@tbs-sct.gc.ca>
Cc: dmitry.gorodnichy@gmail.com
Subject: change of logistics for L&L

Good morning Philippe, Eric, Henry, Patrick

I decided to take a leave from GC for time being (starting tomorrow) – to help with the crisis happening now in Ottawa/Canada

This should not affect the course of our weekly R4GC meetups or efforts in building an R4GC knowledgebase (book), with all your great contributions.
In fact , I hope for the opposite – that it will give me more time to dedicate to it.

But some logistic will change. -

I will be accessing gccollab from home and I will not be able to host the MS Team (I can join only as guest)

Hence, I wonder if any of you would like to host it ? –

Perhaps Philippe, you , for the next two Fridays (with your python+R tutorials)? and Henry, you – the next two, after that with your ones?

Also, I wonder if you are ok / support moving your codes to github.com/open-canada/ ?

Eric, your latest semantic text matching, and Philippe your new codes, and Henry your, old and new ones?

Please let me know

You can reach me on gmail.

To learn more - you can follow me on Facebook and Twitter :) I just posted my ever first Twitter post, yahoo!.. check it out – all links are on my page: https://gorodnichy.github.io <https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgorodnichy.github.io%2F&data=04%7C01%7CPhilippe-Israel.Morin%40ec.gc.ca%7Cc101360fc2cd46f8231708d9ea4c9b9b%7C740c5fd36e8b41769cc9454dbe4e62c4%7C0%7C0%7C637798437015186972%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000&sdata=43k8lfzV4OILbEiH9vZlIFnwuzZHuOJ4sonEusFzAhs%3D&reserved=0>

Dmitry

PS . Found out about this just yesterday: https://ottawa.ctvnews.ca/video?clipId=2308517 <https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fottawa.ctvnews.ca%2Fvideo%3FclipId%3D2308517&data=04%7C01%7CPhilippe-Israel.Morin%40ec.gc.ca%7Cc101360fc2cd46f8231708d9ea4c9b9b%7C740c5fd36e8b41769cc9454dbe4e62c4%7C0%7C0%7C637798437015186972%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000&sdata=M8OYpR%2F8giB8RPUTZs0oRHUvCOaOwBKjQQ%2B29wgRkWw%3D&reserved=0> … (no wonder they support demonstrators)

This is what every agency/employer should have done, and we would not be having all this mess…

OPEN LETTER
The copy of this letter is posted at www.IVIM.ca

April 14, 2022

Dr. Mona Nemer
Chief Science Advisor of Canada
science@canada.ca

Subject: Request to produce a report on COVID-19 Vaccine Efficacy and Safety that would be based on post-marketing vaccine data that have become available after July 2021 in Canada

Dear Dr. Nemer

I applaud your past efforts in establishing the Model Policy on Scientific Integrity and directives for Open Science in Canada. Establishing these policies and directives opened the way for conducting transparent and responsible science in Canada. It also made it possible for all Canadians to access the data related to COVID-19 vaccines and pandemic directly from the official Government of Canada sources and conduct analysis of COVID-19 Vaccine efficiency and safety directly using the data provided in these sources, the main two of which are listed below:

[PHAC-1] 'COVID-19 Daily Epidemiology Update', Public Health Agency of Canada:
https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html

[PHAC-2] Reported side effects following COVID-19 vaccination in Canada, Public Health Agency of Canada:
https://health-infobase.canada.ca/covid-19/vaccine-safety/

I equally applaud your past efforts in raising public awareness in matters related to COVID-19 vaccines efficacy and safety through your reports published last year in March and July.

Specifically, on March 31, 2021, just before COVID-19 vaccination of the general public commenced, you published *"Scientific Considerations for Using COVID-19 Vaccination Certificates"* report that outlined a number of concerns and recommendations related to the uncertainties associated with using new COVID-19 vaccines. Some of these are quoted below:

- Scientific uncertainties: *Given the short time since the COVID-19 vaccines have become available, it is not surprising that many scientific uncertainties persist and are the subject of intense ongoing studies*.

- Ethical and Social Considerations: *It should also be noted that while the anticipation of more freedom may be an incentive for some to get vaccinated, vaccine acceptance could decrease in others if there was a sense of coercion tied to using vaccination certificates.*
- Legal Considerations: *The processes by which vaccine certificates are issued and controlled need to be fraud proof.*
- Conclusion: *[All levels of government could work to develop a framework that] Maximizes consistent post-vaccine monitoring.*

In your other report that was issued on July 16, 2021, in the light of the COVID-19 vaccine side effects that were just uncovered, when the vaccination rates started to pick up (but still being low - less than 10%): "*COVID-19 vaccine-associated myocarditis/pericarditis"*, you have further emphasized the following:

- Priority actions moving forward: *The emerging issue of vaccine associated heart disease requires attention on two important levels: addressing data and knowledge gaps on the one hand, and promoting awareness and clinical care on the other.*

It could have been reasonably expected that - following on your own recommendations of "addressing data and knowledge gaps" and "promoting awareness" - there would be more follow-up reports produced by your office after July 2021, as it was only after July 2021 when post-marketing data, such as 'Cases following vaccination' statistics, 'Variation by Age' statistics, and 'COVID-19 Vaccine Severe Adverse Reactions' statistics, have been published in the above listed PHAC sources. However, as of today April 2022 there are still no new follow-up reports produced by your Office on this very important subject.

Hence - in order to maintain the highest level of trust and confidence of the Canadian public in the decisions of our government - on behalf of all concerned and affected Canadians, I would like to request that your Office produce a new follow-up report on this very important subject - namely, on the subject of COVID-19 vaccines efficacy and safety and related scientific, ethical, and legal considerations that would be based on the new post-marketing data that have become available after July 2021.

I hope that your Office could produce such a report as soon as possible, as your highest most urgent priority, knowing that decisions related to the continuing use of vaccines and, in particular, on their use with children as well as the mandatory use of these vaccines as a condition for employment are being discussed right now by our government.

Evidently, the Mandatory Vaccination Policy that was introduced last year was developed without the knowledge of the post marketing data that have become available after July 2021. It is of the primordial importance therefore - for the well-being of all Canadians - that the new revision of this policy, which is currently being prepared, be based on the new post-marketing data that has become available since July 2021. This should include the data with new Vaccine 'Cases following vaccination' statistics, new 'Variation by Age' statistics, new 'COVID-19 Vaccine Severe Adverse Reactions' statistics that were published by the Government of Canada in the mentioned above PHAC sources since July 2021. It may also include the new data that were recently made available by vendor manufacturers and other governments.

To assist your Office in analyzing post-marketing data that have become available from August 2021, the collective of Data Scientists from several Universities and Government departments (most of whom prefer to remain anonymous) developed a number of data analysis tools, which, we hope, could be of use for your post-marketing analysis of vaccine safety and efficacy in Canada. These tools are made available, with source codes and all cached official data at www.IVIM.ca, and include: All Data Spreadsheet and Dashboard, Algorithmic Bias Analysis, Death Tracking App and several White Papers, all of which are based on Official Government of Canada data that have become available after July 2021.

I'd like to thank you again for your all contributions to the openness and transparency of our government and all policies that our government makes using scientifically-validated evidence. If there is anything I could help with in your data analysis work, I would be very glad to assist. Please let me know.

Yours sincerely

Dr. Dmitry Gorodnichy, Ph.D.
Senior Data Scientist of the Government of Canada
Adjunct professor with University of Ottawa
Partnership Group for Science and Engineering of the Royal Academy of Sciences of Canada
"Leader of Tomorrow" Nominee

Some observations from Canadian Vital Statistics Death (CVSD) Database related to the "Elephant in the Room"[1]

Dmitry Gorodnichy, Rafal Kulik, Stan Matwin[2]

Disclaimer: The views expressed in this publication are those of the authors. They do not purport to reflect the views of any Canadian Agency, University, or Department.

Canada provides open access to their Canadian Vital Statistics - Death (CVSD) data [1] that show weekly mortality rates categorized by the cause of death defined by the "International Statistical Classification of Diseases and Related Health Problems" Revision 10 (ICD-10). These data are used to calculate basic indicators on deaths of residents of Canada as well as in the calculation of statistics, such as cause-specific death rates and life expectancy. Recently, these data have been used to analyze the effects of the COVID-19 pandemic and related public health measures on non-natural manners of death in Ontario [2]. In this paper, it is used to conduct temporal and regional cross-factor analysis of all reported and unreported causes of death in Canada since pandemic and the start of mass vaccination. For this purpose, an interactive Open Source Web Application (App) is developed [3], using which a number of (un)expected observations related to the pandemic and vaccination are made. These are described below (Figure, Appendix). 'ME' indicates Mean weekly mortality rate (number of deaths per million per week).

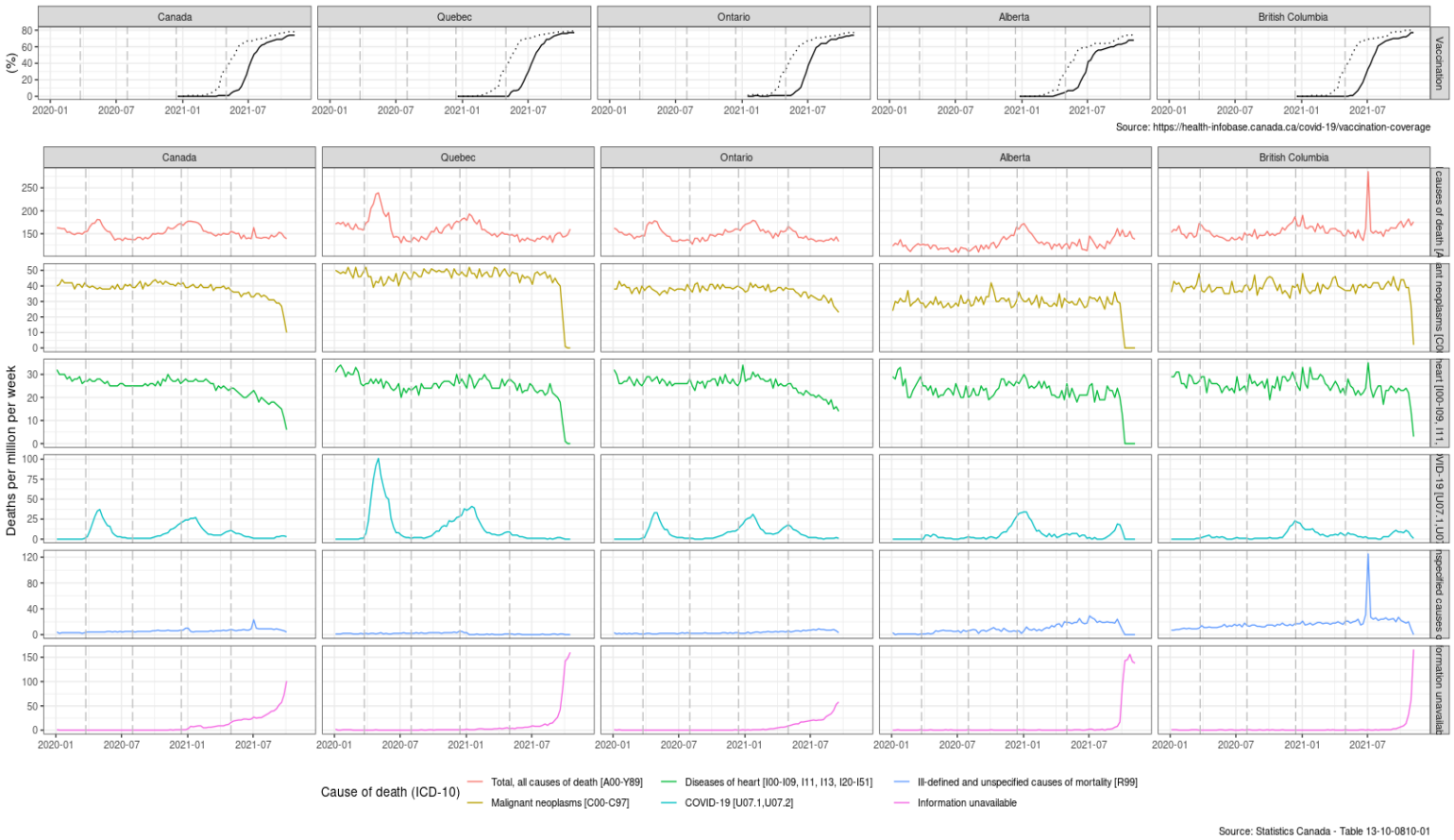


Figure 1: Vaccination and weekly death rates by cause in Canada and during the pandemic and vaccination.
Coloured time-series lines show CVSD statistics at the national level and in four most populated provinces for the total number of deaths and six death cause categories from January 2019 to October 2022. In the top graph, solid lines show full vaccination rate, dash lines show the rate of vaccinated with at least one dose. Vertical lines mark four distinct periods observed: early pandemic pre-vaccination (varying policies across provinces), late pandemic pre-vaccination phase (consistent policies across policies), early vaccination (vaccination rate 0-3%), mass vaccination (vaccination rate: 3-75%). The use of different scales for each cause is noted. Varying dynamics of COVID deaths across the provinces in the early pandemic period and the rapid increase of Ill-defined and unspecified causes [R99] and the number of unreported causes [Information unavailable] in the mass vaccination period are observed.

[1] Submission (Unsolicited Comment) to The Lancet journal. Submitted: 2022/01/16. Decision ("it would be better placed elsewhere"): 2022/01/20. Suggested citation: *"Some observations from Canadian Vital Statistics Death (CVSD) Database related to the 'Elephant in the Room'" (D.Gorodnichy, R.Kulik, S.Matwin). Submission to The Lancet Journal. 2022/01/16, https://open-canada.github.io/vitals.*

[2] Dmitry Gorodnichy and Rafal Kulik (Corresponding author, e-mail: rafal.kulik@uottawa.ca) are with the Department of Mathematics and Statistics, University of Ottawa. Stan Matwin is with the Faculty of Computer Science, Institute for Big Data Analytics, Dalhousie University, Halifax, and the Institute of Computer science, Polish Academy of Sciences, Warsaw.

At national level, it is observed that, despite the large number of COVID case numbers that have been reported since the declaration of the pandemic, the total number of deaths has remained stable (ME=152.2). Furthermore, in contrast to the historically observed trends, one observes a slight decrease in the total number of deaths and consistent decrease in practically all (except COVID) death causes since the start of the pandemic (negative correlation with date). The number of COVID deaths (ME=7.7) is considerably lower than that of leading causes, which remain to be Malignant neoplasms [C00-C97] (ME=38) and Diseases of heart [I00-I09, I11, I13, I20-I51] (ME=24.7), and other reported causes combined (ME=64.3). The increase in reporting of COVID deaths correlates with the decrease in reporting of other causes of deaths.

Examining the death statistics during the pandemic closer, one can further identify four distinct periods, for each of which the following observations are made.

Period 1 (Early pandemic pre-vaccination phase) is characterized by varying policies between East (Ontario and Quebec) and West (Alberta and British Columbia) provinces [4] and includes data from the "first wave", starting from 14 March 2020 (first COVID-19 death in Canada) ending 30 July 2020 (end of the "first wave"). It is observed that in this period East provinces have significantly lower COVID death rate compared to that in West provinces (ME 1.9 and 1.8 in Alberta and British Columbia vs. 33.2 and 10.3 in Quebec and Ontario). Looking at the death toll only, one could state that West provinces did not experience the first COVID "wave". The number of Ill-defined and unspecified causes [R99] is slightly higher than but comparable to the historical data (ME=4.2 compared to 2.2). The “Information unavailable” category (NA) is introduced in this period, however no records under this category are yet reported.

Period 2 (Later pandemic pre-vaccination phase) starts 1 August 2020 and ends on 13 December 2020 (last day before the commencement of vaccination). It is characterized by much more consistent (same) health measures across provinces, and includes second wave, which, in contrast to the "first one", is now equally observed (as a spike of COVID deaths) in all main provinces (ME=6.5 and 4.4 in Alberta and British Columbia vs. 10.8 and 3.9 in Quebec and Ontario). The rate of [R99] deaths grows (ME=5.7). The rate of NA records is insignificant (ME=0.2).

Period 3 (Early high-risk population vaccination phase) is characterized by the small percentage of vaccinated people, includes post-Christmas third wave and runs from 14 December 2020 (when the first vaccination shot was administered in Canada) to 1 May 2021 (when vaccination rate reaches 3%). In this period, the rate of [R99] deaths continued to grow (ME=6.0). The rate of deaths with NA started to grow more rapidly (ME=7).

Period 4 (Later general population vaccination phase) corresponds to the period when the majority population of the country has become vaccinated, including low-risk people and children, with full vaccination rate growing rapidly from 3% in 1 May 2021 to 74% on 11 October (the last record in the CVSD database at the time of writing this communication). The main difference of this period from previous periods is seen in a significant jump in the rate of NA records, which started to grow exponentially in all provinces (ME=35 at national level). In observing this, it is noted that the provinces, as known from the description of CVSD database [1], collect and report statistics independently from each other.

To summarize, CVSD statistics show that since the start of mass vaccination, unreported cases have become (and remain to be) the most dominant portion of all deaths. There can be two main explanations for this. First, reporting of valid death causes has become significantly more difficult since the beginning of mass vaccination and one can wait until the reporting of valid death causes is resumed. Second, the data could in fact reflect the new normal in the Canadian death causes reporting, with the majority of deaths now being left unexplained. In either case, it is important to maintain a trustful Vital Statistics database, open for research and analysis. It is especially important during the mass vaccination campaign. Indeed, the presented above observations from the Canadian Vital Statistics Database, coupled with the data presented in other reports presenting statistics related to deaths due to COVID and/vs due to COVID vaccines [5-10] explain the challenges in keeping public trust in the vaccination as well as in other COVID-19 prevention measures, keeping of which is known to be an important task during the pandemic [11]. The recent Angus Reid poll [12] has also indicated that "health concerns" (besides "personal freedom") is now the most common argument against vaccination. In this regard, we applaud the efforts of the Government of Canada in keeping its death by cause statistics open and regularly updating it. We also encourage other data scientists and health practitioners around the world to take advantage of the free data visualization tools and open source codes, such as the ones provided with the Web App used here [3], to build similar comparative death causes trackers for other countries and regions.

References

1. Statistics Canada, Canadian Vital Statistics - Death database (CVSD), https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3233 .
(Provisional weekly death counts, by selected grouped causes of death. Table: 13-10-0810-01.
URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310081001)

2. Jennifer M. Dmetrichuk, Jeffrey Rosenthal, Julia Mann, Mackenzie Cullip,Richard A. Wells, Retrospective study of non-natural manners of death in Ontario: Effects of the COVID-19 pandemic and related public health measures, The Lancet Regional Health Americas, (7), 100130, March 2021.
https://www.thelancet.com/journals/lanam/article/PIIS2667-193X(21)00126-5/fulltext

3. Open Canada Vital Statistics Tracker.
URL: https://open-canada.github.io/Apps/vitals (Mirror: https://o-canada.shinyapps.io/vitals/).
Code Source: https://github.com/open-canada/vitals

4. An Inconsistent Canadian Provincial and Territorial Response During the Early COVID-19 Pandemic
Amelie Cyr, Prosanta Mondal and Gregory Hansen,
Frontiers in Public Health, 27 September 2021 https://doi.org/10.3389/fpubh.2021.708903

5. 'Cases following vaccination',
COVID-19 Daily Epidemiology Update, Public Health Agency of Canada,
https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html

6. 'Reported side effects following COVID-19 vaccination in Canada',
Canadian COVID-19 vaccination safety report, Public Health Agency of Canada,
https://health-infobase.canada.ca/covid-19/vaccine-safety/

7. 'COVID-19 vaccine efficacy and effectiveness—the elephant (not) in the room', Piero Olliaro, Els Torreele, Michel Vaillant,
The Lancet, VOLUME 2, ISSUE 7, E279-E280, JULY 01, 2021: https://doi.org/10.1016/S2666-5247(21)00069-0

8. Thacker P D. Covid-19: Researcher blows the whistle on data integrity issues in Pfizer's vaccine trial BMJ 2021; 375 :n2635 doi:10.1136/bmj.n2635. https://www.bmj.com/content/375/bmj.n2635

9. COVID-19: stigmatising the unvaccinated is not justified.
Günter Kampf, The Lancet, VOLUME 398, ISSUE 10314, P1871, NOVEMBER 20, 2021
https://doi.org/10.1016/S0140-6736(21)02243-1

10. Ronald N. Kostoff, Daniela Calina, Darja Kanduc, Michael B. Briggs, Panayiotis Vlachoyiannopoulos, Andrey A. Svistunov, Aristidis Tsatsakis, Why are we vaccinating children against COVID-19?, Toxicology Reports, Volume 8, 2021, Pages 1665-1684, ISSN 2214-7500, https://doi.org/10.1016/j.toxrep.2021.08.010.
(https://www.sciencedirect.com/science/article/pii/S221475002100161X)

11. Vickery, Jamie and Atkinson, Paul and Lin, Leesa and Rubin, Olivier and Upshur, Ross and Yeoh, EK and Boyer, Chris and Errett, Nicole, Challenges to Evidence-Informed Decision-Making in the Context of Pandemics: The Case of COVID-19. Available at SSRN: https://ssrn.com/abstract=3971551 .

12. Dwindling group of unvaccinated cite 'personal freedom' and 'health concerns' as main reasons for dodging the jab, Angus Reid poll, https://angusreid.org/canada-unvaccinated-freedom-reasons/

Appendix

This Appendix contains:

- The list of ICD-10 death categories reported in the CVSD database
- Table 1: The national weekly deaths rates per million for leading causes of death in four analyzed periods.
- Table 2: Death due to COVID-19 in Quebec, Ontario, Alberta and British Columbia in Period 1 (2020-03-14 - 2020-08-01) vs. Period 2 (2020-08-01 --2020-12-14)
- Additional images: Time-series, regression lines and correlation plots - at National level, four largest populated and four lower populated provinces
- Analysis of other official Government of Canada data (not included in original submission)
- Note related to the estimate of COVID only deaths from all COVID-reported deaths (not included in original submission)
- Additional reference from the vaccine manufacturer (not included in original submission)

ICD-10 death categories reported in the CVSD database

Statistics Canada uses "International Statistical Classification of Diseases and Related Health Problems" Revision 10 (ICD-10), which reports death causes using the following categories:

[1] Total, all causes of death [A00-Y89]
[2] Malignant neoplasms [C00-C97]
[3] Diabetes mellitus [E10-E14]
[4] Alzheimer's disease [G30]
[5] Diseases of heart [I00-I09, I11, I13, I20-I51]
[6] Cerebrovascular diseases [I60-I69]
[7] Influenza and pneumonia [J09-J18]
[8] Chronic lower respiratory diseases [J40-J47]
[9] Nephritis, nephrotic syndrome and nephrosis [N00-N07, N17-N19, N25-N27]
[10] Accidents (unintentional injuries) [V01-X59, Y85-Y86]
[11] Intentional self-harm (suicide) [X60-X84, Y87.0]
[12] COVID-19 [U07.1,U07.2] (since January 2020)
[13] COVID-19, virus identified [U07.1] (since January 2020)
[14] COVID-19, virus not identified [U07.2] (since January 2020)
[15] Ill-defined and unspecified causes of mortality [R99]
[16] Information unavailable (since January 2020)
[17] All other causes of death

Table 1: The weekly deaths rates per million for leading causes in four analyzed periods in Canada.

The numbers in the table show weekly deaths rate means and dynamics (the slope of the regression). It is noted that in some cases linear modeling may not be appropriate to model the data.

Cause	Period 1	Period 2	Period 3	Period 4	Historical
All causes	153.8, -1.7	150.8, 1.8	159.5, -2.1	138.7, -0.8	139.4, 10.4
Malignant neoplasms	39.2, 0	41.3, 0.1	39.2,-0.1	29.8, -0.9	39.9, 1.3
Diseases of heart	26.2, -0.1	26.3, 0.2	25.8,-0.3	17.6, -0.7	26.4, 2.5
Influenza and pneumonia	2.8, -0.2	2, 0.1	2.1, 0	1.1, 0	3.6, 1.8
Accidents	7.7, 0.1	8.1, 0	6.5, -0.1	3.5, -0.2	6.9, 0.7
Suicides	2, 0	1.9, 0	1.1, 0	0.5, -0.1	2, 0.3
COVID-19	12.1, -0.7	6, 1	13.9,-1.3	3, -0.4	NA
Ill-defined & unspecified	4.2, 0.1	5.7, 0.1	6.2, 0	8.3, -0.1	2.2, 1.4
Information unavailable	0, 0	0.2, 0	7.2, 0.4	34.8,3	NA

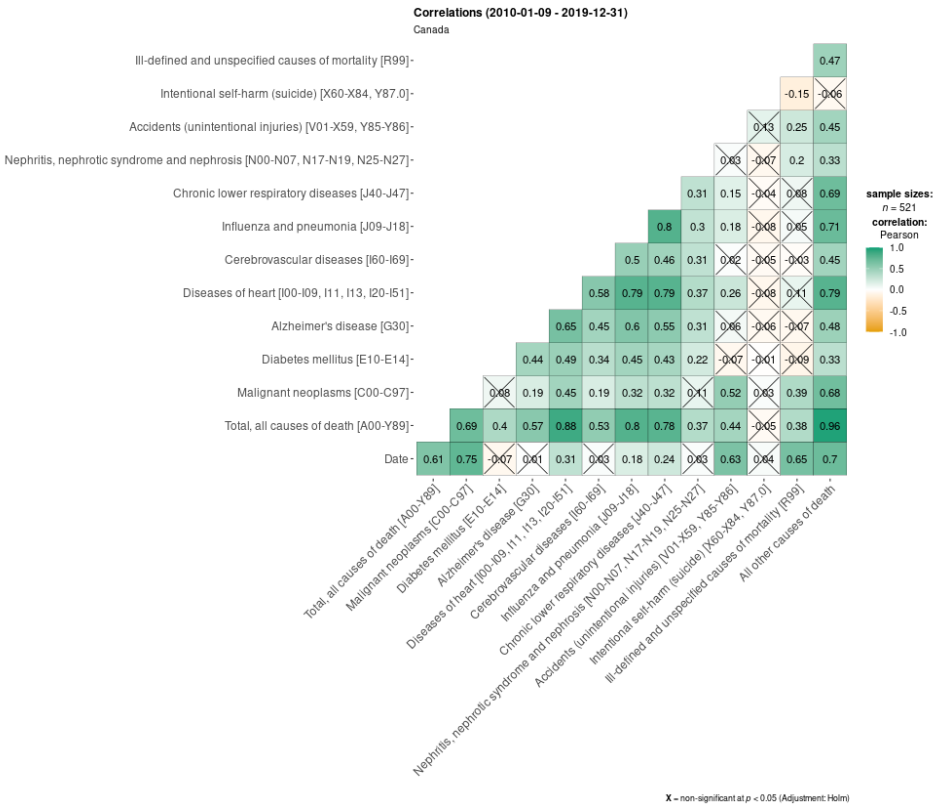
Table 2: Weekly rates of deaths due to COVID-19 in Quebec, Ontario, Alberta and British Columbia Period 1 (2020-03-14 - 2020-08-01) vs. Period 2 (2020-08-01 - 2020-12-14)

Cause	Period 1 (QE, ON, AB, BC)	Period 2 (QE, ON, AB, BC)
COVID-19 [U07.1,U07.2]	33.2, 10.3, 1.9, 1.8	10.8, 3.9, 6.5, 4.4

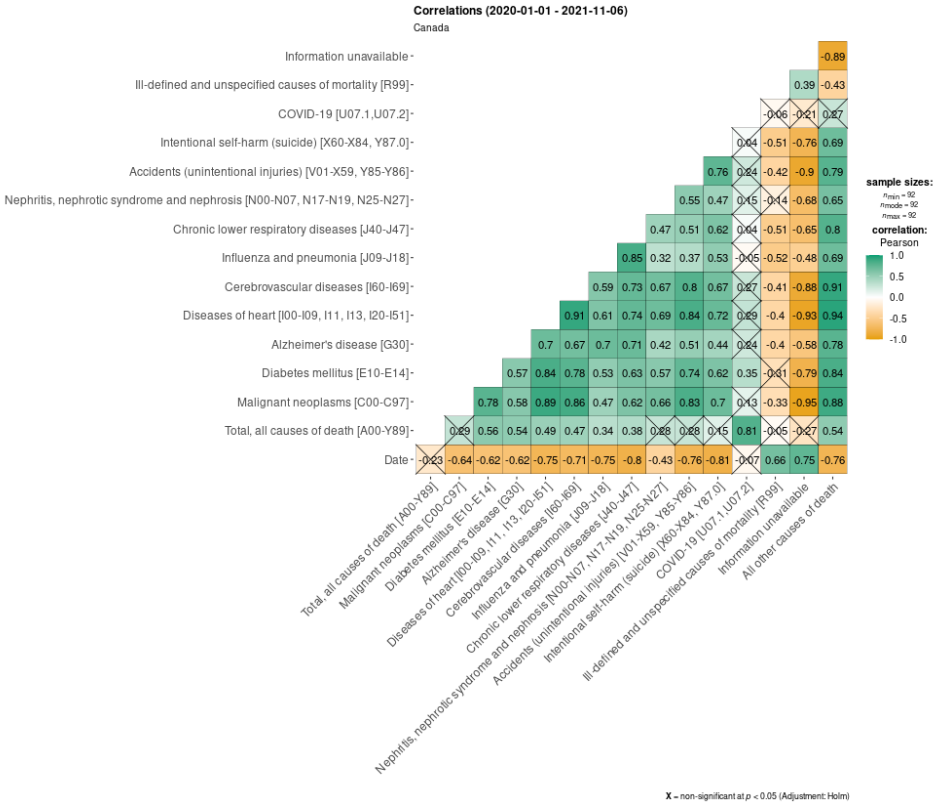
Death causes at national level: cross-correlations and correlations with time (Date)

Green colour indicates positive correlation, orange colour indicates negative correlation. Statistically non-significant correlations (at p<0.05%) are crossed over. Change in the sign of the correlation with Date (last row) is observed between pre-pandemic and pandemic periods.

Pre-pandemic:

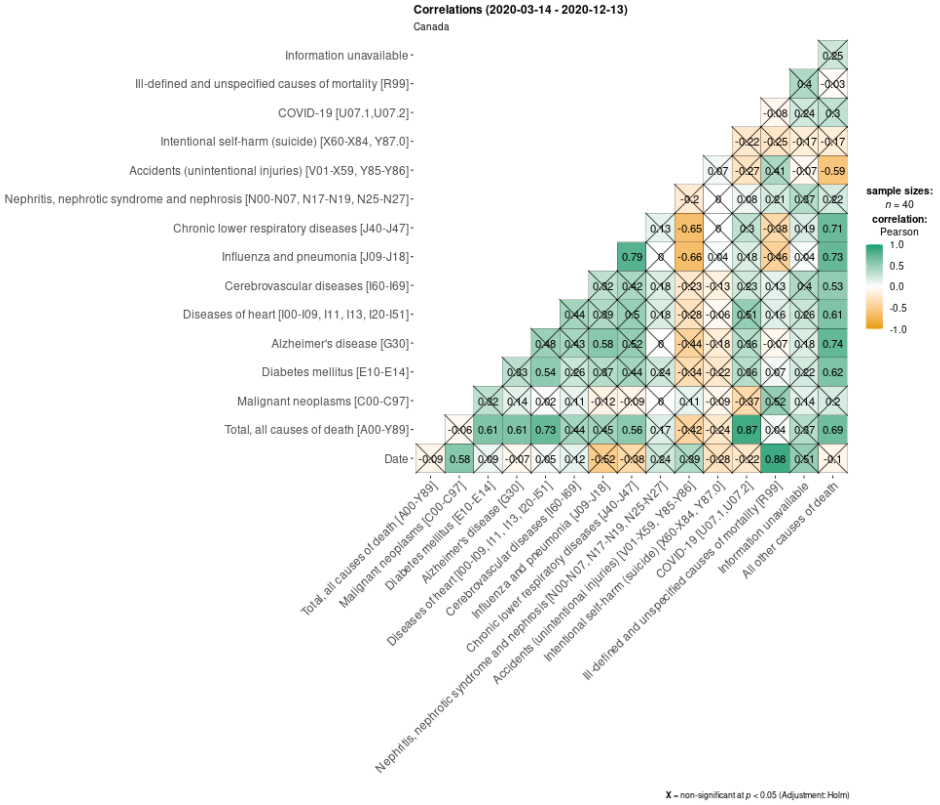


Entire Pandemic:

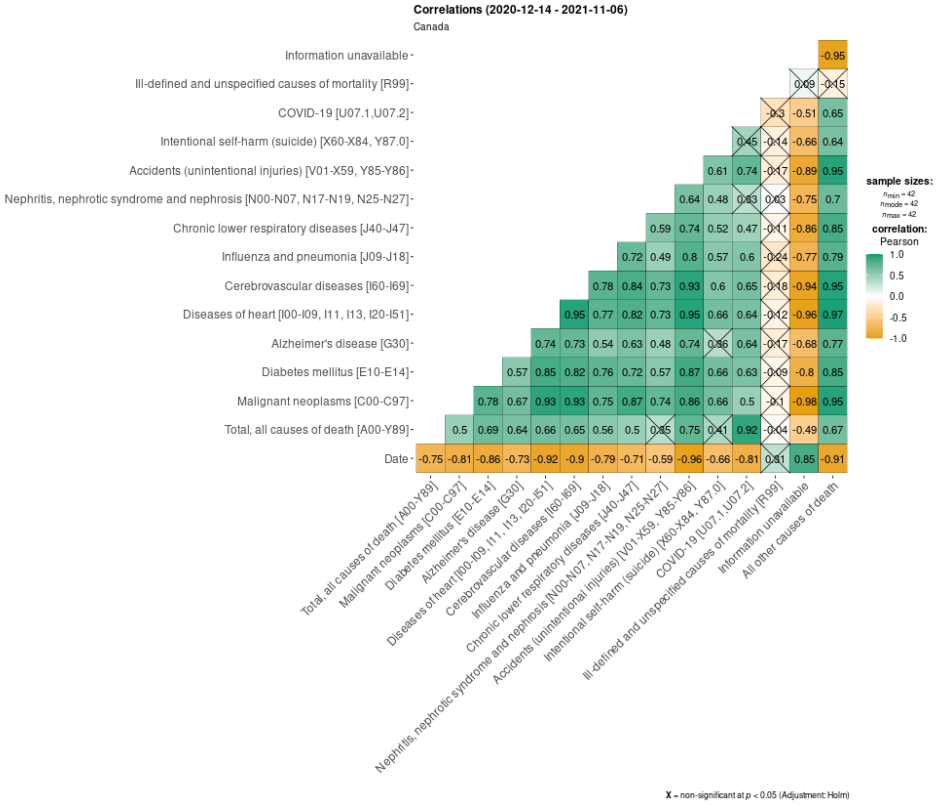


566-02-48305

Pandemic pre-vaccination:



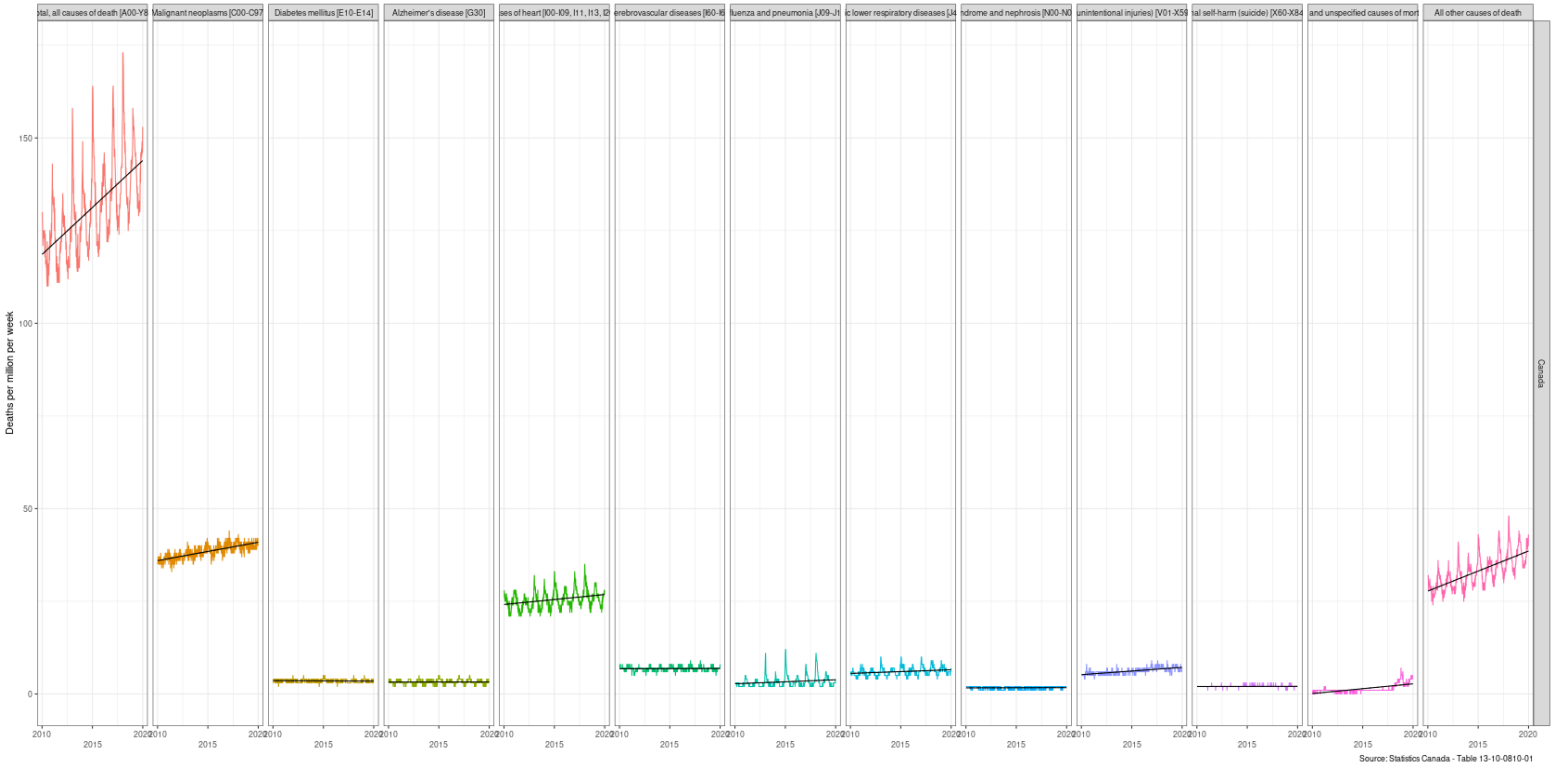
Pandemic since vaccination:



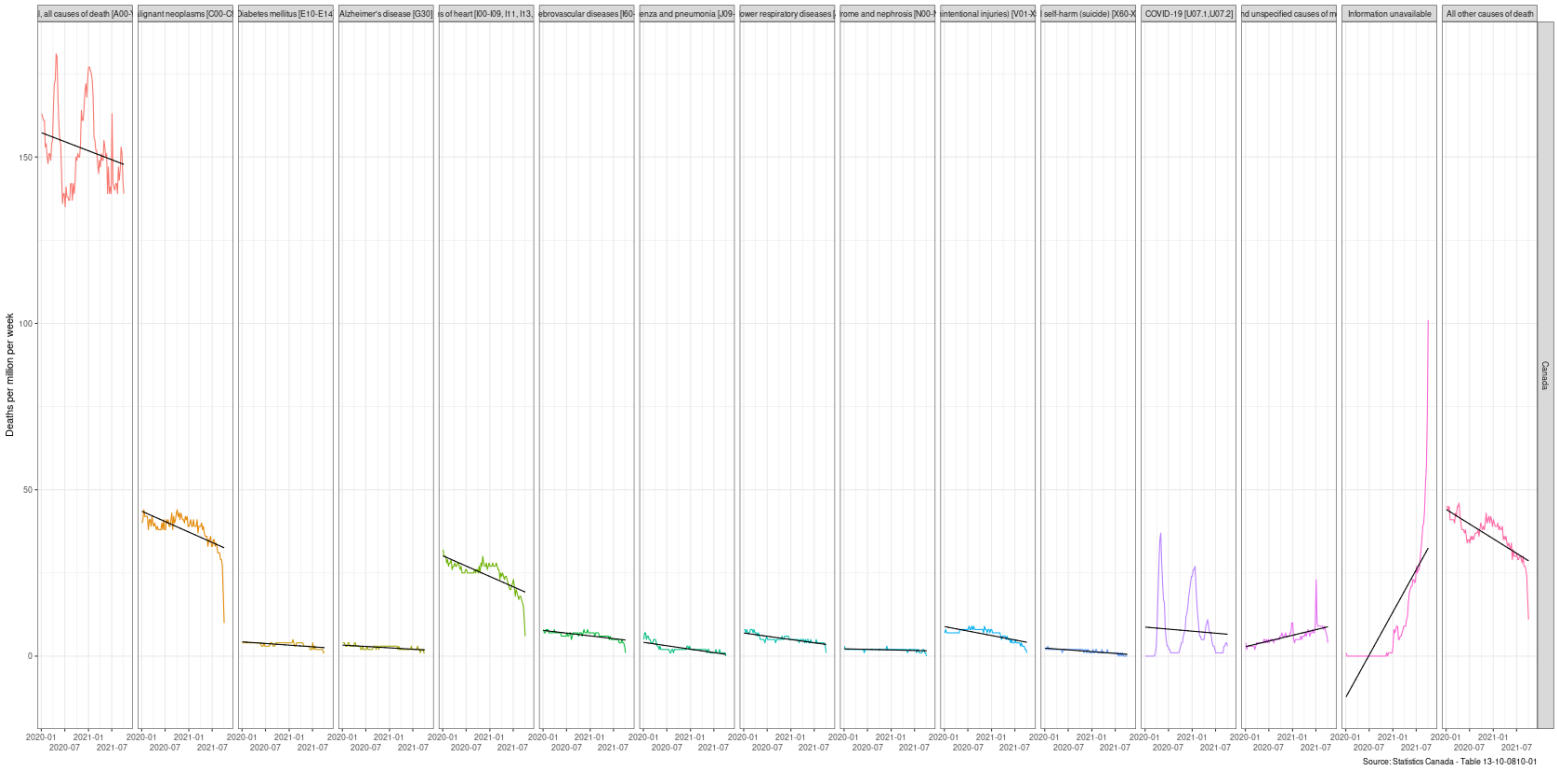
Deaths by cause - at national level: time-series and linear regression (trendlines)

In the following we provide additional images that show the time-series (dynamics and changes) in each death cause category, before and during the pandemic and vaccination. First we show time-series at the National level - for all death cause categories. Then we show time-series for four largest populated provinces and four lower population provinces - for six most frequent death cause categories. The change in dynamics (regression slope) before and during the pandemic is observed.

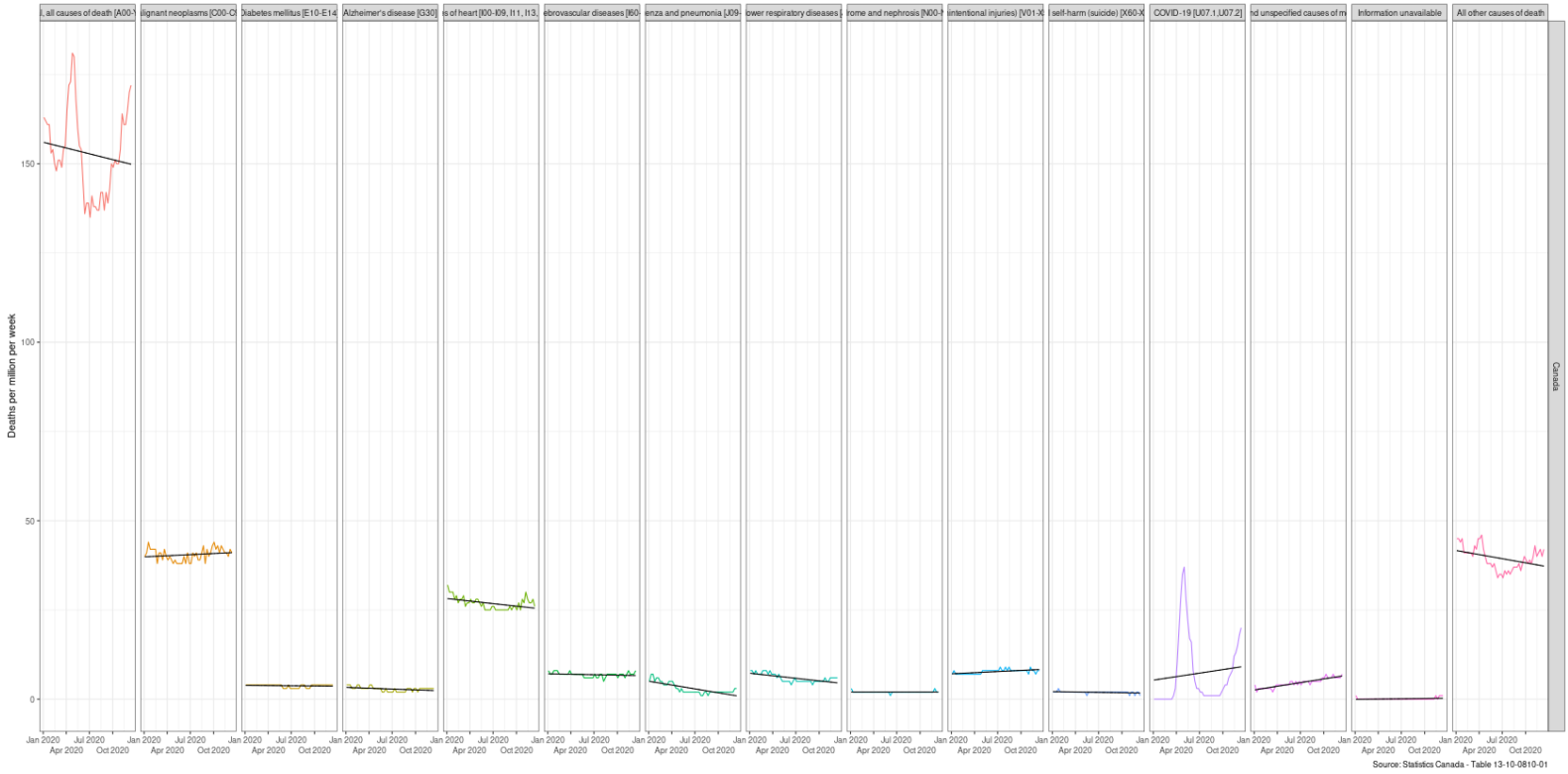
Historical pre-pandemic (January 2010 - December 2019):



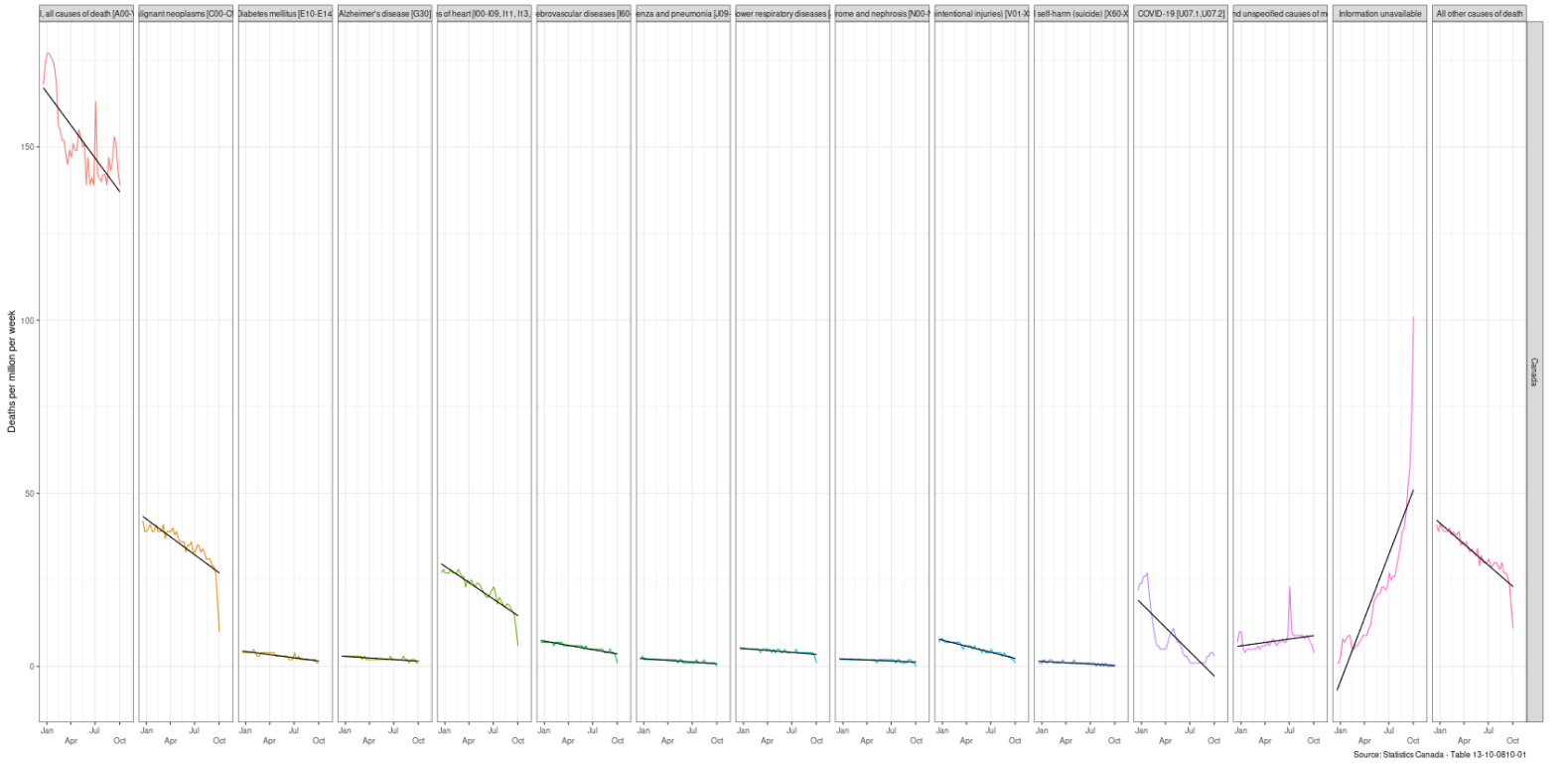
During entire pandemic (1 January 2020 - 6 November 2021):



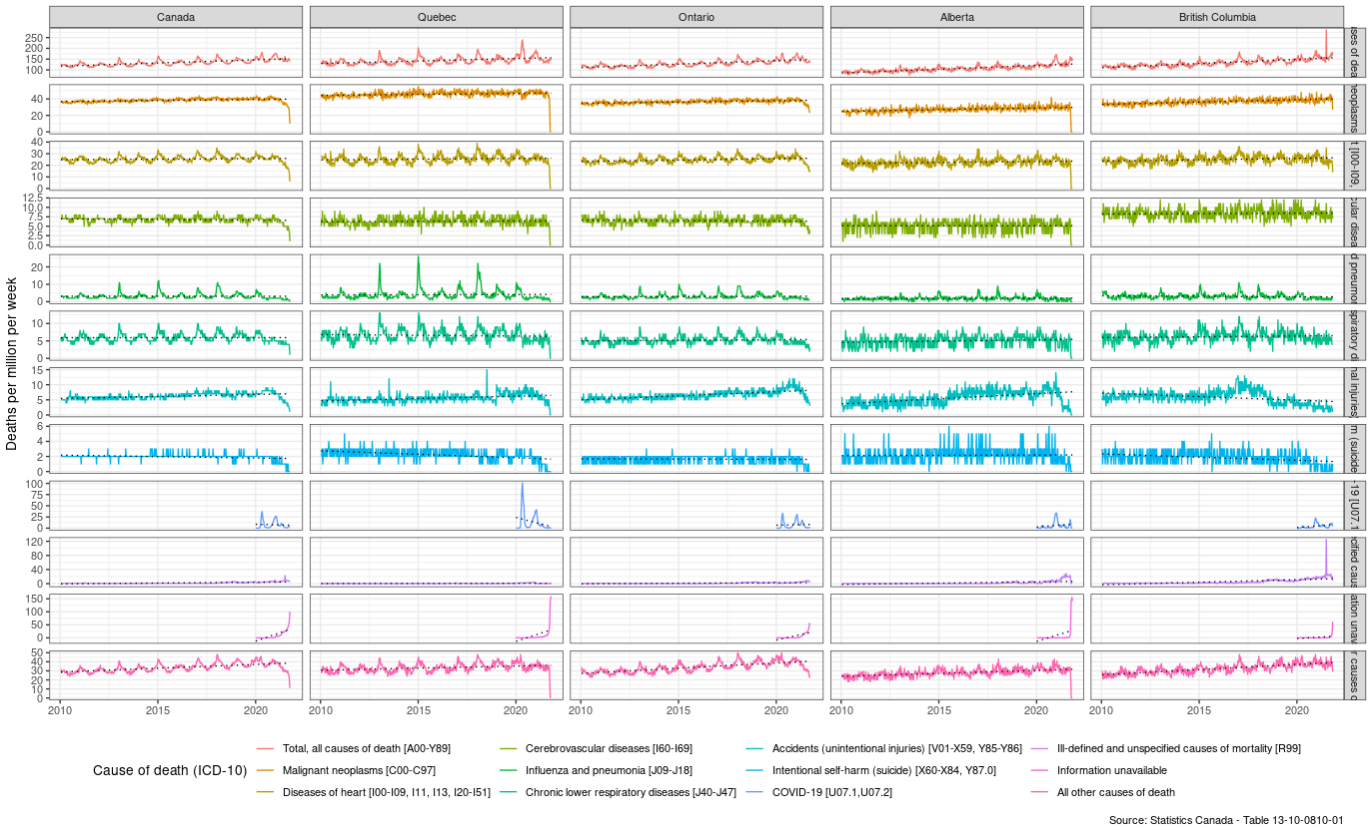
Pandemic pre-vaccination:



Since start of vaccination:

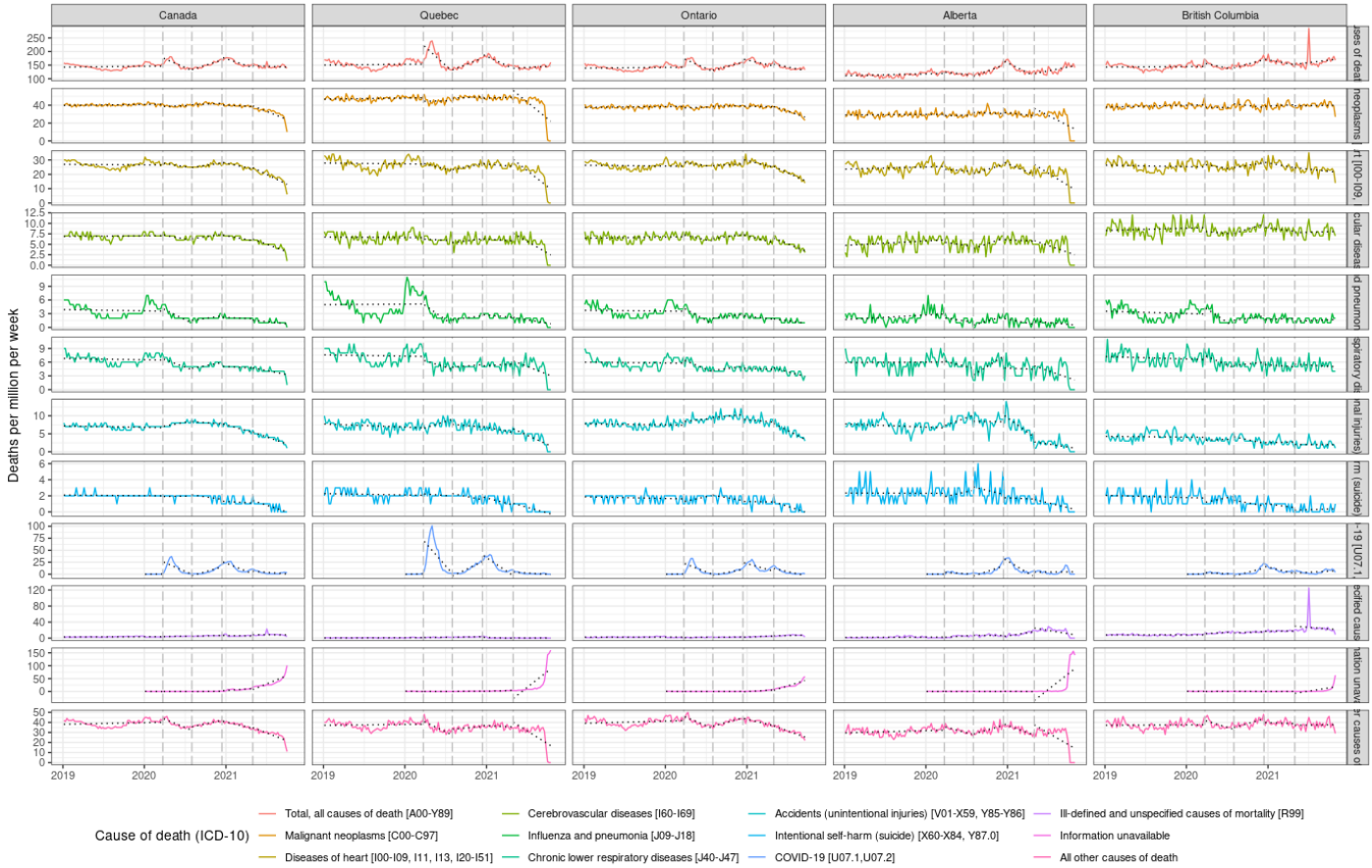


Deaths by cause - at National level and four largest provinces: 2010-2021



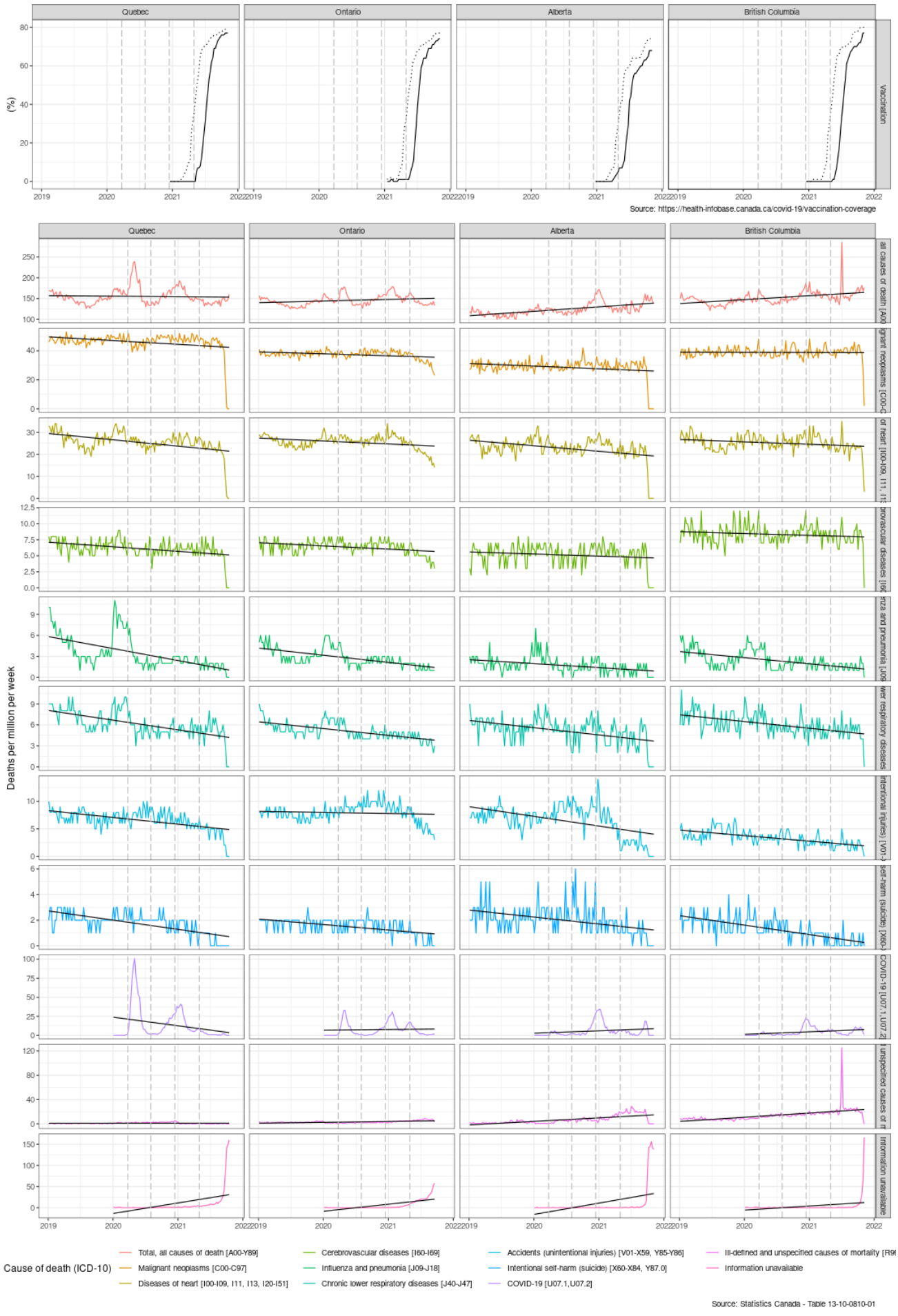
Source: Statistics Canada - Table 13-10-0810-01

Deaths by cause - at National level and four largest provinces: zoomed on 2019-2021

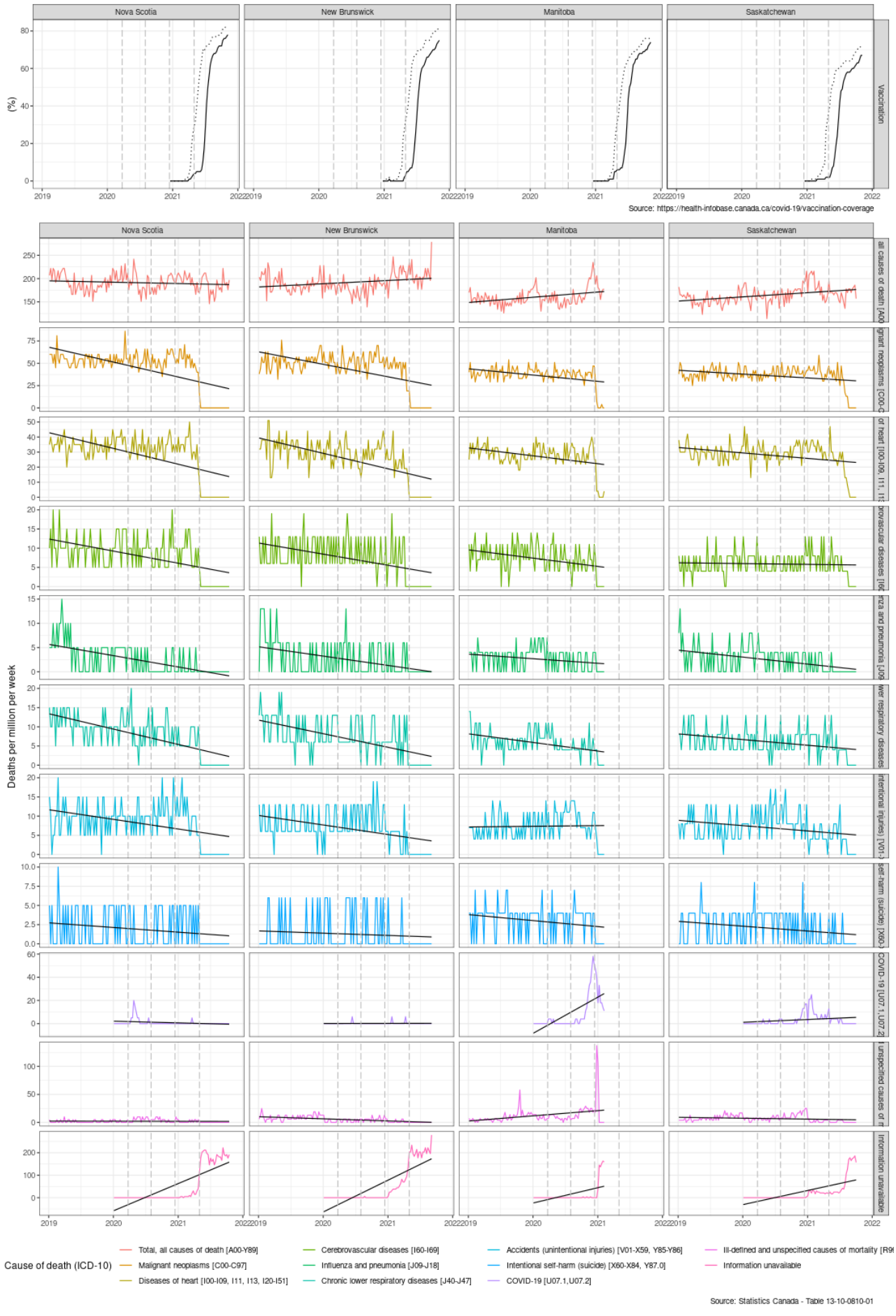


Source: Statistics Canada - Table 13-10-0810-01

Deaths by cause and vaccination rate at four largest provinces



Deaths by cause and vaccination rate at four populated provinces



Analysis of the official Government of Canada data related to the vaccine efficacy and safety
(Last updated: 2202/01/22)

The information below is not part of the original submission to The Lancet journal and is included below for the completeness of the presentation.

Source 1:
'Cases following vaccination', COVID-19 Daily Epidemiology Update, Public Health Agency of Canada, https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html

Quote: *"For cases from December 14, 2020 up until November 27, 2021, fully vaccinated cases were 68% less likely to die as a result of their illness, compared to unvaccinated cases (Table 3).*

Observation: Using the App [3], one can find that in the considered period (December 14, 2020 – November 27, 2021) over 75% (10100 / 13075) of all deaths occurred prior to May 1, 2021, when the majority of population (over 97%) was unvaccinated. Recomputing these numbers for the period *after May 1* (when the proportion of vaccinated was approximately equal to the proportion of unvaccinated) results in the opposite conclusion – the majority (75% or more) of people who died of COVID-19 between May and November were fully vaccinated.

Source 2:

Hospitalizations by vaccination status, COVID-19 (coronavirus) in Ontario, https://covid-19.ontario.ca/data/hospitalizations#hospitalizationsByVaccinationStatus,

Weekly epidemiology report (PDF) available on the Government of Canada's COVID-19 data trends page, https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19/epidemiological-economic-research-data.html

Quote: "*Incidence among fully vaccinated cases has surpassed those among unvaccinated since mid-December 2021."*

Observation: Incidents among fully vaccinated cases continues to grow at a rate higher than of unvaccinated.

Source 3:

'Reported side effects following COVID-19 vaccination in Canada', Canadian COVID-19 vaccination safety report, Public Health Agency of Canada, https://health-infobase.canada.ca/covid-19/vaccine-safety/

Quotes: *"Table 1. Count of reported adverse events of special interest up to and including October 22, 2021 (n = 3,793). This includes: 166 Auto-immune diseases, 1,111 heart diseases (including 64 Cardiac arrests and failures, 77 Heart Attacks, 970 Myocarditis1/Pericarditis), 947 thrombosis and blood clots, 37 Acute kidney injury, 25 Liver injury,781 Nerves and central nervous system failures (including 594 Bell's Palsy), 487 Anaphylaxis, and 208 deaths (some of which are still under investigation)." "Cumulative serious reports: 6124 (January–October), 5038 (May-October)."*

Observation: The above numbers (5038) are almost twice as high as the number of *COVID-reported deaths* in the same period (May-October), which is only 2975 (or 8.1 deaths per week per million), and much higher (by one order of magnitude) than the number of deaths from *COVID alone* (estimated using UK data as 13.6 % * 2975 = 404 - see note below), and even much more higher (by two orders of magnitude) when compared to the number of deaths *from COVID alone for population younger than 64 years* (estimated using UK data as 22% *13.6 % * 2975 = 89 - see note below).

It is noted that most people who received vaccine since May are less than 64 years of age, meaning that they have a very small chance of dying from COVID alone only (22% *13.6 % * 8.1 = 0.24 deaths per week per million, which is less than from

any other cause reported in CVSD database), while having a much higher chance of dying or having serious health damage from the vaccine.

It is also noted that the above number of adverse events (5038) is quoted from the reports for up to November 6 2021 (to allow comparison with death statistics which are reported with two months delay). Since that time, the number of adverse events from vaccination, including deaths, continues to grow at the rate faster than that of vaccination and faster than that of COVID-reported deaths, having reached 7378 at the time of the last update of this comment.

Source 4:

Recalls and safety alerts, Government of Canada: https://recalls-rappels.canada.ca/

Quote: *"Health Canada updates Pfizer-BioNTech and Moderna COVID-19 vaccine labels to include information on myocarditis and pericarditis"* and *"…to reflect very rare reports of Bell's Palsy"*

Observation: The listed three complications count for less than third of reported serious complications, and less 7% of all reported adverse events. A several months lag between when severe events are reported and when the vaccine recall alerts are posted is observed.

Source 5:

Statistics Canada, Canadian Vital Statistics -Death database. Provisional weekly death counts, by selected grouped causes of death. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310081001 (Dashboard:https://open-canada.github.io/vitals)

Observation: Since May (i.e. since the beginning of mass vaccination, when the rate of vaccinations started to rapidly increase from 3% in May to 75% in October) the rate of unreported and Ill-defined / unspecified causes of mortality has also started to rapidly increase in all provinces. Additionally, it is observed that, since the introduction of the 'COVID death' category in 2020, the reporting of all non-COVID causes deaths started to decline, even for those causes that historically have a trend to go up (such as cancer, heart diseases, accidents). This indicated that the COVID death cause category likely consumed some other causes of death. This agrees with the analysis from UK (See note below) that shows that only 13.6% of COVID-reported deaths in UK were COVID only deaths.

Note related to the estimate of 'COVID only deaths' from all 'COVID-reported deaths'

In the above calculations of *COVID only* deaths from all *COVID-reported deaths* the following statistics from UK is used:

In 2020 and first three quarters of 2021 there were in total 127,704 deaths reported in UK as due to COVID [a]. Of these, 13.6% (17,371) were due to COVID alone [b]. From the same source [b], one also finds that 78% of all COVID alone deaths occurred in people with 64+ of age, and that the average age of death from covid alone is 82.5 years, which is approximately equal to the average life expectancy in the UK.

Sources:

[a] Deaths in United Kingdom, https://coronavirus.data.gov.uk/details/deaths

[b] Deaths from COVID-19 with no other underlying causes, FOI Ref: FOI/2021/3240, Release date: 16 December 2021, https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/deathsfromcovid19withnootherunderlyingcauses

Additional reference from the vaccine manufacturer

The following reference to the report is not part of the original submission to The Lancet Journal, provided for completeness.

- CUMULATIVE ANALYSIS OF POST-AUTHORIZATION ADVERSE EVENT REPORTS OF PF-07302048 (BNT162B2) RECEIVED THROUGH 28-FEB-2021, Prepared by Pfizer, Released to Public Health and Medical Professionals for Transparency (https://phmpt.org/) under the court order on 17 November 2021.
  Direct link: https://phmpt.org/wp-content/uploads/2021/11/5.3.6-postmarketing-experience.pdf

This report provides data that was collected, but not disclosed prior to 17 November 2021, by the vaccine manufacturer on real world adverse events that occurred in the first 2.5 months after Emergency Use Authorization. It lists: over 1,200 deaths; over 25,000 nervous system adverse events; and Anaphylaxis and Vaccine-Associated Enhanced Disease (listed under “Safety concerns”).

Additional observations related to the "Elephant in the Room" – from Government of Canada COVID-19 daily epidemiology updates

Dmitry Gorodnichy, Stan Matwin, Anonymous 1, Anonymous 2, Anonymous 3, Anonymous 4 [1]

Disclaimer: **The views expressed in this publication are those of the authors. They do not purport to reflect the views of any Canadian Agency, University, or Department.**

This comment follows up on previous observations [1] made in relation to the COVID-19 vaccine and death statistics since the start of the general public vaccination in Canada in Spring 2021 that were made by analyzing Open Canada Vital Statistics Death (CVSD) database [2,3]. Here we make additional observations related to the same "Elephant in the Room" (safety and efficacy of COVID-19 accines). This time the observations are made from another Open Canada data source: "COVID-19 epidemiology updates" that are published by the Public Health Agency of Canada (PHAC) [4], the cached copies of which can be retrieved using "WayBackMachine" Internet Archive [5] and archived on GitHub [6].

Main observation: algorithmic bias detected

The slide below summarizes the main observation. - The way statistics is reported in "COVID-19 epidemiology updates" [2] exposes the algorithmic bias that is introduced in calculating the statistics related to the percentage of unvaccinated COVID-19 deaths vs. that of vaccinated COVID-19 deaths (pages 20-22, Figure 5, Tables 2-3 in PDF versions of the report).

Reverse-engineering algorithmic bias. Case Study: Vaccinated vs. Non-vaccinated deaths

https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html

Recall: There were no fully-vaccinated before January, <1% before March, <3% before May, and … most* (50-75%), deaths happened exactly then.

Try it yourself:
- Open link above, write down numbers in red boxes and compare them a week later!
- Or, compare these numbers in past reports since November (archived here).

Observe:
These numbers monotonically change (one down, one up) every week - according to math behind them…

*computed using Vitals App Calculator

Table 2. Characteristics and severe outcomes associated unvaccinated, partially vaccinated and fully vaccinated confirmed cases reported to PHAC, as of December 18, 2021

		Unvaccinated (n=727,925)	Cases not yet protected (n=43,471)	Partially vaccinated (n=53,171)	Fully vaccinated (n=122,843)	Total cases (n=947,410)
Gender*	Male	371,897 (78.5%)	20,959 (4.4%)	24,980 (5.3%)	56,197 (11.9%)	474,033 (100%)
	Female	354,365 (75.2%)	22,454 (4.8%)	28,113 (6.0%)	66,344 (14.1%)	471,276 (100%)
Hospitalizations		40,788 (79.5%)	3,062 (6.0%)	3,374 (6.6%)	4,099 (8.0%)	51,323 (100%)
Deaths		8,013 (75.6%)	759 (7.2%)	744 (7.0%)	1,077 (10.2%)	10,593 (100%)

Source: Detailed case information received by PHAC from provinces and territories, since December 14, 2020

As of today, February 4, 2022 : Total cases (n=1,458,433)
Fully vaccinated 2,032 (16.8%), +3% since last week, +8% since November

While this comment was being written, the proportion of Fully vaccinated deaths (the number in the right red box) has increased within a week from 2032 (16.8%) reported on Feb. 4, 2022, to 2490 (19.5%) reported on Feb.10, 2022.

[1] Dmitry Gorodnichy is a Senior Data Scientist with the Government of Canada, Adjunct Professor at the Department of Mathematics and Statistics, University of Ottawa, and Royal Academy of Sciences of Canada's Partnership Group for Science and Engineering "Leader of Tomorrow" nominee. Stan Matwin is Full Professor and Director of the Institute for Big Data Analytics, Dalhousie University, Halifax, Emeritus Distinguished Professor of Computer Science at the University of Ottawa, and Professor in the Institute of Computer Science of the Polish Academy of Sciences in Warsaw. Anonymous 1-4 are Data Scientists with the Government of Canada and Academia who wished to remain anonymous. Corrections and comments to this paper should be emailed to dmitry@gorodnichy.ca.
Suggested reference: *D.Gorodnichy et al. "Additional observations related to the 'Elephant in the Room' – from Government of Canada COVID-19 daily epidemiology updates. Time to act!". 2022/02/10, https://ivi-m.github.io/vitals/comment2.pdf*

As shown in Appendix and as can be validated using the Open Canada Vital Statistics (Deaths) Tracker App (https://open-canada.github.io/Apps/vitals), the majority (over 75%) of all deaths that happened in the analyzed period (which is from December 14, 2020, when the first vaccine dose was administered, to the date when the report is published) happened right after Christmas 2020, when there were practically no (statistically speaking, insignificant number of) fully-vaccinated people. More specifically:

> There were 0 fully-vaccinated as of January 4, less than 1% as of February 18, less than 2 % as of April 6, less than 3% as of April 30, and less than 10% as of June 10, 2021, after which the full-vaccination quickly reached 78% by November, and staying the same approximately the same (+/- 2%) since then.

In other words, whether deliberately or undeliberately, a significant positive bias (or "advantage") for fully-vaccinated has been embedded directly in the way the proportion of fully-vaccinated deaths has been computed and reported to the public. Having access to the past archived "COVID-19 epidemiology updates" however allows one to reverse-engineer and remove this bias from computation.

Removing algorithmic bias from reported data statistics

To remove algorithmic bias from the reported weekly (or monthly) data statistics, it is simply sufficient to subtract the numbers (of vaccinated and unvaccinated COVID-19 deaths) reported in two consecutive reports. The difference between those numbers will be the true observed number of deaths observed between the dates of the two reports, which can then be used to compute the true (bias-removed) proportions of unvaccinated COVID deaths vs. vaccinated COVID deaths in the considered period.

The following example illustrates how to use the described technique to compute the true ratio of fully-vaccinated cases to unvaccinated cases for any given time period (e.g., for the period between December 4,2021 and January 15, 2022).

Consider two published reports one for January 15, 2022 and the other for December 04, 2021 (see below)

Characteristics and severe outcomes associated unvaccinated, partially vaccinated and fully vaccinated confirmed cases reported to PHAC, as of January 15, 2022

Status	Cases
Unvaccinated	771,095
Cases not yet protected	44,494
Partially vaccinated	61,209
Fully vaccinated	581,635

Characteristics and severe outcomes associated unvaccinated, partially vaccinated and fully vaccinated confirmed cases reported to PHAC, as of December 04, 2021

Status	Cases
Unvaccinated	709,123
Cases not yet protected	43,114
Partially vaccinated	52,116
Fully vaccinated	88,742

By subtracting the numbers in two tables, one obtains that between December 04, 2021 and January 15, 2022, there were 492,893 cases among fully vaccinated, 61,971 cases among the unvaccinated, and 9.093 cases among partially vaccinated.

That is, the number of COVID-19 cases in the considered period (December 04, 2021 and January 15, 2022) among fully-vaccinated was almost eight times larger than among unvaccinated and partially-vaccinated (87% vs. 11% and 2%).

Taking into account the number of fully vaccinated people at that time was only five times larger than the number of unvaccinated and partially vaccinated people (77% vs. 18% and 5%, according to [6]), leads to the conclusion that, in the considered period, fully-vaccinated people were *more likely* to be infected with COVID-19 than unvaccinated people.

The analysis still continues, as the latest reported data are still being processed. In the meanwhile, some conclusions and recommendations can already be made.

DRAFT FOR CONSULTATION

Conclusions:

The data from the Government of Canada's "COVID-19 epidemiology updates" do not support the claims about COVID-19 vaccines, as having over high (over 90%) efficacy [8], as claimed originally a year ago. In contrast, they provide evidence that COVID-19 vaccines, which as known, were approved under Emergency Use Authorization and did not go through the same lengthy testing protocol as all previous vaccines, may have a negative impact on a person's immunity towards the COVID-19 infection. Whether this is true only for a particular variant of the infection, or only for a particular portion of the Canadian population needs to be further investigated.

Recommendations:

- Start computing and reporting Cases and Deaths following vaccination statistics in 'COVID-19 Daily Epidemiology Updates' [4] without the embedded bias, as done elsewhere [9-10], i.e. using weekly cases totals, instead reporting totals numbers since December 14, 2020.

- Report empirically measured COVID-19 vaccines efficacy and track it over time based on the data that has become available since October 2021 and using the unbiased data reporting metrics discussed in this paper.

References

[1] "Some observations from Canadian Vital Statistics Death (CVSD) Database related to the 'Elephant in the Room'" (D.Gorodnichy, R.Kulik, S.Matwin). Submission to The Lancet Journal. 2022/01/06, https://open-canada.github.io/vitals/comment.pdf

[2] Statistics Canada, Canadian Vital Statistics - Death database (CVSD), https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3233 .
Provisional weekly death counts, by selected grouped causes of death. Table: 13-10-0810-01. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310081001)

[3] Open Canada Vital Statistics Tracker. URL: https://open-canada.github.io/Apps/vitals (Mirror: https://o-canada.shinyapps.io/vitals/). Code Source: https://github.com/open-canada/vitals

[4] 'Cases following vaccination', COVID-19 Daily Epidemiology Update, Public Health Agency of Canada, https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html#a9

[5] "WayBackMachine" Internet Archive: https://web.archive.org

[6] Archive of COVID-19 Daily Epidemiology Update from Public Health Agency of Canada: https://github.com/IVI-M/vv/tree/main/docs/COVID-19%20epidemiology%20updates

[7] https://health-infobase.canada.ca/covid-19/vaccination-coverage/

[8] Scientific Considerations for Using COVID-19 Vaccination Certificates, Report of the Chief Science Advisor of Canada, March 31, 2021, https://science.gc.ca/eic/site/063.nsf/eng/h_98229.html

[9] 'Reported side effects following COVID-19 vaccination in Canada', Canadian COVID-19 vaccination safety report, Public Health Agency of Canada, https://health-infobase.canada.ca/covid-19/vaccine-safety/

[10] Hospitalizations by vaccination status, COVID-19 (coronavirus) in Ontario, https://covid-19.ontario.ca/data/hospitalizations#hospitalizationsByVaccinationStatus

DRAFT FOR CONSULTATION

Appendix:

Analysis of Canadian Vital Statistics Death Statistics using the Vitals App: https://open-canada.github.io/Apps/vitals

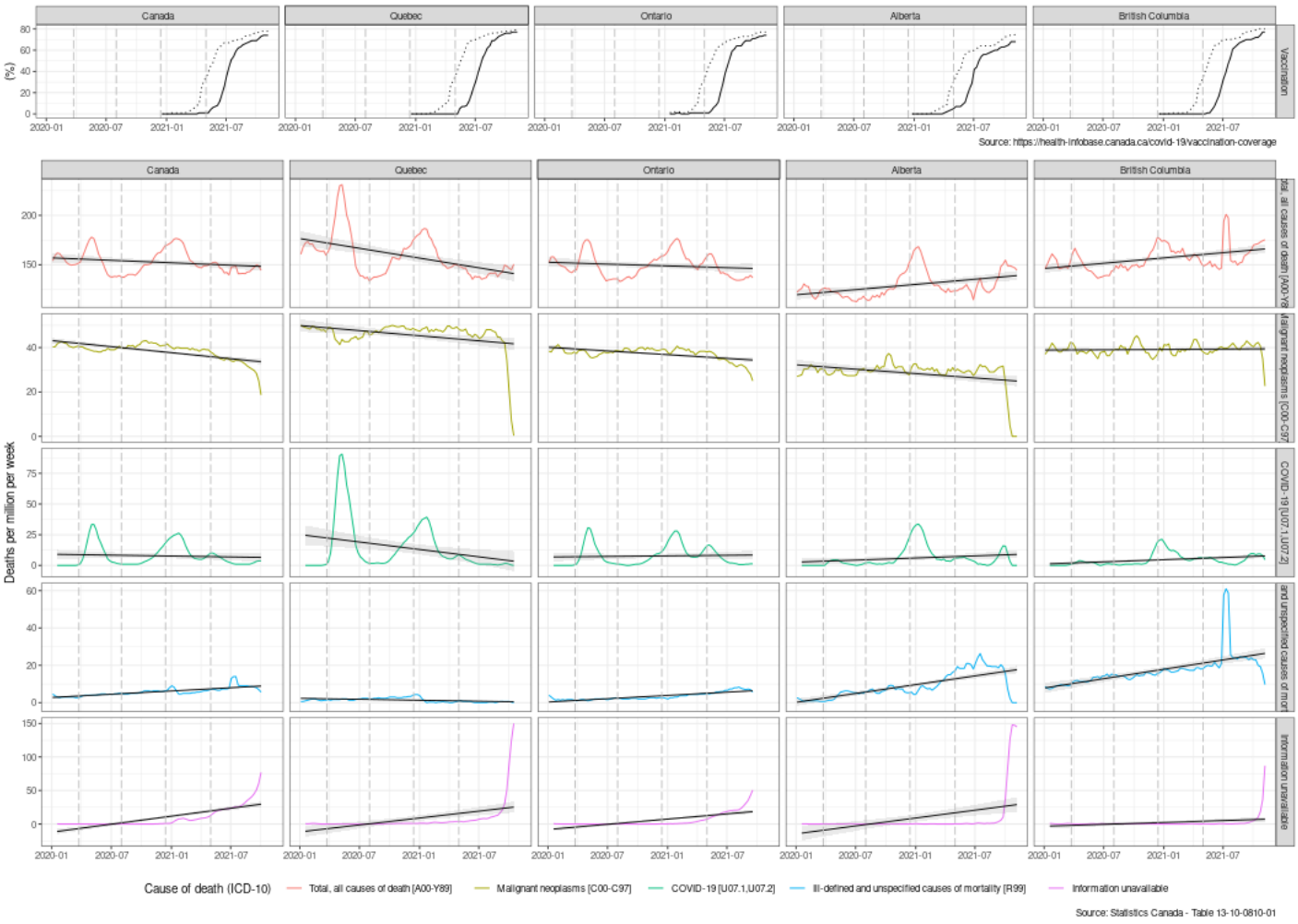


Figure 1: Vaccination and weekly death rates by cause in Canada and during the pandemic and vaccination.

Coloured time-series lines show CVSD statistics (deaths per million per week) at the national level and in four most populated provinces for the total number of deaths, cancer (the leading cause of death), “COVID deaths”, as well for two unexplained causes (Ill-defined and unspecified causes [R99] and the number of unreported causes [Information unavailable]). In the top graph, solid lines show full vaccination rate, dash lines show the rate of vaccination with at least one dose. Vertical lines mark four distinct periods observed: early pandemic pre-vaccination (varying policies across provinces), late pandemic pre-vaccination phase (consistent policies across policies), early (high-risk population) vaccination, late (general public) vaccination). The use of different scales for each cause category is noted.

The following observations from the figure are made (see [1] for more details):

- total number of deaths over the entire period is comparable to (and in some provinces lower than) their five-year average;
- small number of COVID deaths, compared to the number of reported COVID cases, and leading causes of death;
- absence of “first wave” in Alberta and British Columbia, in contrast to Quebec and Ontario;
- decrease in all non-COVID causes, since the start of the pandemic, including those that historically go up;
- increased number of unexplained deaths since the start of pandemic and especially since the start of mass vaccination;
- increased delay in reporting deaths statistics

DRAFT

Home (https://gccollab.ca/) ➜ Groups (https://gccollab.ca/groups/all)



Data Science with Open Canada Data 'Lunch and Learn' seminars

(https://gccollab.ca/ajax/view/thewire_tools/reshare?reshare_guid=7855030) (https://gccollab.ca/action/cp_notify/unsubscribe?guid=7855030&__elgg_ts=1768411131&__elgg_token=xkNfJYSDmF4hcJwDD8GnJw)

Owner: Dmitry Gorodnichy (https://gccollab.ca/profile/Dmitry.Gorodnichy)

+ Invite (https://gccollab.ca/groups/invite/7855030)

Group members: 111 (https://gccollab.ca/groups/members/7855030)

Activity (https://gccollab.ca/groups/profile/7855030/data-science-with-open-canada-data-lunch-and-learn-seminars)

Discussion (https://gccollab.ca/discussion/owner/7855030)

Files (https://gccollab.ca/file/group/7855030/all)

Blog (https://gccollab.ca/blog/group/7855030/all) More ()

Description

NOTE, AS OF APRIL 2022, THESE SEMINARS ARE NO LONGER CONDUCTED

Building advanced Data Science skills, while also helping each other and community - one meeting at a time!

Every Friday Lunch (12:00 -13:00) we have Lunch and Learn seminars where we practice and discuss various data manipulation and data visualization skills in R.

We've been having these virtual seminars since the beginning of pandemic in Spring 2020, when we used COVID-19 cases data for practicing.

Since the end of 2021, we use the new data provided by Public Health Agency of Canada and Statistics Canada to analyze and visualize various statistics related vital statistics and COVID-19.

While mastering the R skills, we also get many interesting insights from the data we visualize, including those that are not discussed anywhere in main stream media. All codes are made public at https://github.com/open-canada/vitals (https://github.com/open-canada/vitals). The insights are posted at https://open-canada.github.io/vitals/ (https://open-canada.github.io/vitals/).

No registration is required to join the meeting. Just join us via MS Team at link provided.

To participate in internal discussion, you need to join this sub-group (https://gccollab.ca/groups/edit/7855030): https://gccollab.ca/groups/about/7855030 (https://gccollab.ca/groups/about/7855030)

Additional information:

- GCcollab: R4GC (Use R!) community (https://gccollab.ca/groups/profile/7391537/r (https://gccollab.ca/groups/profile/7391537/enuse-rfruse-r)4gc (https://gccollab.ca/groups/profile/7391537/r4gc))
- The R4GC Book: https://open-canada.github.io/r4gc/ (https://open-canada.github.io/r4gc/)
- R4GC Web Apps: https://open-canada.github.io/Apps/ (https://open-canada.github.io/Apps/)
  - PSES Results: https://itrack.shinyapps.io/ (https://itrack.shinyapps.io/covid/us.Rmd)pses
  - Covid tracker App: https://itrack.shinyapps.io/covid/us.Rmd (https://itrack.shinyapps.io/covid/us.Rmd).
  - StatsCan and PHAC vital statistics related to covid: https://o-canada.shinyapps.io/vitals/ (https://o-canada.shinyapps.io/vitals/)

Add widgets

Group discussion (https://gccollab.ca/discussion/owner/7855030)

Dmitry Gorodnichy

Edit



Research Data Scientist

Canada Border Services Agency
dmitry.gorodnichy@cbsa-asfc.gc.ca (mailto:dmitry.gorodnichy@cbsa-asfc.gc.ca)
https://www.gorodnichy.ca (https://www.gorodnichy.ca)

(https://gccollab.ca/Avatar/edit/Dmitry.Gorodnichy) (http://www.facebook.com/dmitry.gorodnichy) (https://github.com/gorodnichy)
(http://ca.linkedin.com/in/dmitry.gorodnichy) (http://www.youtube.com/dmitry.gorodnichy)

Profile Files Blog Wire posts Bookmarks Polls Photos (https://gccollab.ca/photos/siteimagesowner/55030) More ()

About Me

Edit

Senior Research Data Scientist with Computer Science and Applied Math background, 30+ years of experience innovating, coding, developing , leading projects and groups, supporting clients, 25 years of which within Government of Canada, including five years in bilingual Manager position.

Author of several inventions, technologies and technical recommendations, including to International Standard Organization (ISO) and internal clients, that were implemented in the field. Adjunct Professor at Univ. of Ottawa (CompSci & Math), Univ. of Dalhousie (CompSci), Univ. de Quebec (École de technologie supérieure).

I've been the chair for the Video Technology for National Security (VT4NS) Government of Canada Community of Practice and currently lead the R4GC Community of Practice (https://gccollab.ca/groups/profile/7391537/r4gc-use-r-community-utilisez-r-r4gc-communaut) here on GCCollab.

I create environments where initiative, creativity, team spirit, and data challenges excite both the management and their sub-ordinates. Love bringing people - especially from diverse backgrounds - together, and come up with new ideas to better our communities and the environment, the latest of which are GC "Lunch and Learn R" meet-ups (https://gccollab.ca/groups/about/7855030).

Professional skills include extensive experience with Biometrics, Computer Vision and Identity Analytics technologies, Data Science and Research methods with focus on design and evaluation of complex systems, collaborative programming and advanced data analysis and visualization.

Play and compose music (http://music.gorodnichy.ca) in between the lines.
LinkedIn: https://www.linkedin.com/in/dmitry-gorodnichy (https://www.linkedin.com/in/dmitry-gorodnichy/)

Education

Edit

University of Alberta
Ph.D. - Computing Science (Artificial intelligence)
August, 1995 - July, 2000

Moscow Insitute of Physics and Technology
M.Sc (Cum Laude) - Applied Mathematics (Neural networks)
September, 1988 - June, 1994

Work Experience

Edit

Research Data Scientist
Canada Border Services Agency, Chief Data Office Directorate, Data Science Division
June 2020 - Present

- Providing scientific and technical advice related to Data Science, including Artificial Intelligence and Machine Learning ; - Developing DS / AI / ML tools for Agency needs (data engineering toolkit, traffic prediction, border wait time optimization, sensor data recognition, automated stress/lie detection, biometrics); - General statistical & mathematical support for the agency (sampling, questionnaires, confidence intervals, multivariate / mixed-effect analysis, R programming, etc.)

Senior Research Scientist, Acting Manager - Video Surveillance and Biometrics
Canada Border Services Agency, Science and Engineering Directorate
June 2008 - June 2020

- Providing scientific and technical advice related to the use of video analytics and biometric technologies by the agency; - Supervising a team of research scientists and research engineers (including coop students and casuals); - Seeking out and developing proposals for external funding in the area of video analytics and biometrics in support of agency's mandate.

Research Officer
National Research Council of Canada - Institute of Information Technologies
December 2000 - January 2008

- Developing video recognition technologies for security and health applications. - Supervising a team of junior researchers (including Ph.D and M.Sc. students)

Skills

Edit

R 0
Machine learning 0
data engineering 0
biometrics 0

Opt In

Edit

Career Opportunities

Casual Work Seeker No
Casual Work Opportunity Creator No
Student Integration Seeker No
Student Integration Opportunity Creator No
Interchange Canada Seeker Yes
Interchange Canada Opportunity Creator No

Developmental Opportunities

Mentee No
Mentor Yes
Collaboration Seeker Yes
Collaboration Opportunity Creator Yes

https://knollfrank.github.io/HowBadIsMyBatch/HowBadIsMyBatch.html?batchCode=FA8721

How Bad is My Batch?

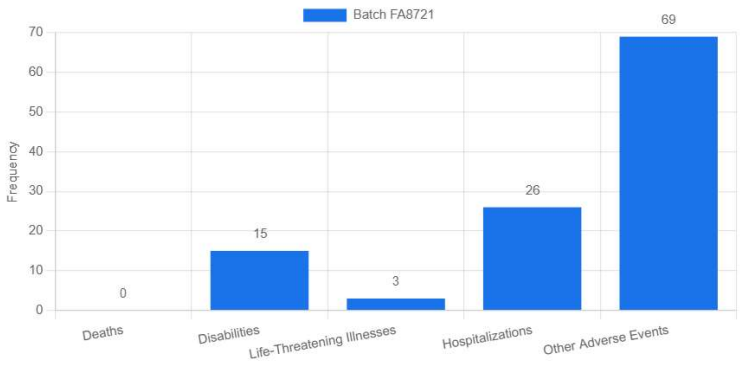
566-02-48305

By Craig Paardekooper

Check out your batch code: FA8721

Batch FA8721 (PFIZER\BIONTECH)

Outcomes of Adverse Events



Frequencies of Reported Symptoms

Show 10 entries Search: Enter Symptom

Symptom	Frequency
Chest pain	24
Fatigue	22
Myocarditis	22
Dyspnoea	20
SARS-CoV-2 test	18
Pericarditis	17
Headache	14
Nausea	13
Pain	12
Arthralgia	11

Home (https://gccollab.ca/) ➜ Discussion (https://gccollab.ca/discussion/all)
➜ Data Science with Open Canada Data 'Lunch and Learn' seminars (https://gccollab.ca/discussion/owner/7855030)
➜ PHAC 'Cases following vaccination' processed - letter being prepared.

Data Science with Open Canada Data 'Lunch and Learn' seminars

(https://gccollab.ca/ajax/view/thewire_tools/reshare?reshare_guid=7855030) (https://gccollab.ca/action/cp_notify/unsubscribe?guid=7855030&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)

Owner: Dmitry Gorodnichy (https://gccollab.ca/profile/Dmitry.Gorodnichy)

Group members: 111 (https://gccollab.ca/groups/members/7855030) + Invite (https://gccollab.ca/groups/invite/7855030)

Activity (https://gccollab.ca/groups/profile/7855030/data-science-with-open-canada-data-lunch-and-learn-seminars)

Discussion (https://gccollab.ca/discussion/owner/7855030) Files (https://gccollab.ca/file/group/7855030/all)

Blog (https://gccollab.ca/blog/group/7855030/all) More ()

PHAC 'Cases following vaccination' processed - letter being prepared.

Following on the discussion we had last month, I have now processed all historical PHAC 'Cases following vaccination' reports from its COVID-19 Daily Epidemiology Updates. The results are in table below. Briefly, they can be summarised as follows.

> Because PHAC computes statistics counting Deaths from the first day of Vaccination (December 14, 2020),
> when there was negligible percent of Fully vaccinated, the results reported by PHAC are skewed significantly (by order of magnitude) in favour of Fully vaccinated.
>
> The skew was most noticeable at early stages of reporting these statistics (July - September),
> when instead of the observed 10%-30% deaths among Fully vaccinated, PHAC reported only 1%-3% of them.

Spreadsheet with all raw data and calculations are here: google sheet (https://docs.google.com/spreadsheets/d/1lSO5jJwTkMNGl9kk6Y5jeGORGS11KzEE5yTxBGrGYKw/edit?usp=sharing), xlsx (https://github.com/IVI-M/vitals/tree/main/docs/COVID-19%20epidemiology%20updates/Table2-log-calculations.xlsx), pdf (http://github.com/IVI-M/vitals/tree/main/docs/COVID-19%20epidemiology%20updates/Table2-log-calculations.pdf). Feedback / comments are very welcome.

A letter in this regard is also being prepared to the Chief Science Advisor of Canada. If you'd like to join in preparing and/or signing this letter, please contact me directly at dmitry.gorodnichy@gmail.com (your identity will be protected)

Please join me tomorrow for Zoom (at numbers below) - to discuss anything related to the above, or anything else, inc. invasion of Russia to Ukraine...

Last updated: 22 February 2022

Source: Cases following vaccination, COVID-19 Daily Epidemiology Update, Public Health Agency of Canada

URL: https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html#9 (Table 2). Archived reports: web.archive.org

NB1: PHAC started reporting of 'Cases by vaccination stutus' statistics **on 2021/07/30**, showing data for up to **2021/07/10**

NB2: PHAC computes percentages counting the cases **from 2020/12/14** (the day when vaccination started and when there were no fully vacci

		Fully vaccinated TO TOTAL (%)			Fully vaccinated TO TOTAL (%)		
		Reported by PHAC counting cases from December 2020			Recomputed as weekly average cases from July 2021		
For week ending	Full vaccination rate	Cases	Hospitalizations	Deaths	Cases	Hospitalizations	Deaths
2021-05-01	2.9						
2021-06-05	8.0						
2021-06-26	27.0						
2021-07-10	**40.0**	0.5%	0.69%	1.19%			
2021-07-17	49.9	0.5%	0.73%	1.24%	16%	16%	9%
2021-08-07	61.5	0.8%	0.97%	1.43%	13%	11%	14%
2021-08-14	63.4	1.1%	1.10%	1.62%	17%	10%	20%
2021-09-04	67.0	2.6%	2.19%	3.01%	15%	11%	16%
2021-09-11	68.0	3.3%	2.66%	3.74%	27%	19%	35%
2021-10-02	71.0	5.4%	4.18%	5.75%	33%	21%	30%
2021-10-30	74.0	7.3%	5.64%	7.95%	42%	28%	35%
2021-11-27	76.0	9.3%	7.07%	9.50%	55%	26%	36%
2021-12-04	76.0	9.9%	7.37%	9.76%	62%	35%	38%
2021-12-18	76.5	13.0%	7.99%	10.17%	63%	37%	35%
2022-01-15	77.5	39.9%	17.09%	16.84%	90%	67%	65%
2022-01-22	77.9	38.8%	20.06%	19.50%	31%	62%	66%
2022-01-30	78.69	40.4%	22.59%	21.64%	80%	68%	65%
2022-02-06	79.36						

When: 25 February 2022, noon-12:45

Join Zoom Meeting
https://us04web.zoom.us/j/72260868830?pwd=dgsDcmFPb82kRI8I-8wkhiKdv8IPuX.1 (https://us04web.zoom.us/j/72260868830?pwd=dgsDcmFPb82kRI8I-8wkhiKdv8IPuX.1)

Meeting ID: 722 6086 8830
Passcode: HqwR90

(https://gccollab.ca/profile/Dmitry.Gorodnichy) Started by Dmitry Gorodnichy (https://gccollab.ca/profile/Dmitry.Gorodnichy) 2022-02-24 17:29

(https://gccollab.ca/bookmarks/add/55030?address=https%3A%2F%2Fgccollab.ca%2Fdiscussion%2Fview%2F11701600%2Fphac-cases-following-vaccination-processed-letter-being-prepared&title=PHAC 'Cases following vaccination' processed - letter being prepared.&title2=)
(https://gccollab.ca/action/discussion/toggle_status?guid=11701600&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)
(https://gccollab.ca/discussion/edit/11701600)
(https://gccollab.ca/action/discussion/delete?guid=11701600&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)
(https://gccollab.ca/ajax/view/thewire_tools/reshare?reshare_guid=11701600)
(https://gccollab.ca/action/likes/add?guid=11701600&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)
(https://gccollab.ca/action/cp_notify/unsubscribe?guid=11701600&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)

Replies

(https://gccollab.ca/profile/zhuoyuwang-3)

By Zhuoyu Wang (https://gccollab.ca/profile/zhuoyuwang-3) 2022-02-24 20:27
(https://gccollab.ca/discussion_reply/edit/11709079)
(https://gccollab.ca/action/discussion_reply/delete?guid=11709079&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)
(https://gccollab.ca/action/likes/delete?guid=11709079&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)
2 likes

There are obvious errors in your calculation. For example, for the weekly cases in week 2022-01-31:

Cases						
Unvaccinated	Cases not yet protected	Partially vaccinated	Fully vaccinated	Fully vaccinated with an additional dose	Total Cases	Fully vaccinated TO TOTAL (%)
8,975	382	3,046	48,330	87,090	60,733	**80%**

The total cases in the cell (60,733) doesn't match with the sum of the other numbers. 87,090 > 60,733.


(https://gccollab.ca/profile/Dmitry.Gorodnichy)

By Dmitry Gorodnichy (https://gccollab.ca/profile/Dmitry.Gorodnichy) 2022-03-04 11:41
(https://gccollab.ca/discussion_reply/edit/11810537)
(https://gccollab.ca/action/discussion_reply/delete?guid=11810537&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)
(https://gccollab.ca/action/likes/add?guid=11810537&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)

Much appreciated the review of it.

Sorry, I should have made it note there clarifying that

'Fully vaccinated with an additional dose' is included in 'Fully vaccinated'

That is , Total is computed by summing up four columns only:

Unvaccinated Cases not yet protected Partially vaccinated Fully vaccinated

So all numbers are correct.

We can discuss it today at meetup

Also, I have rearranged sheets in the spreadsheet - so they are easier now for easier processing

Have a look:

(https://docs.google.com/spreadsheets/d/1lSO5jJwTkMNGl9kk6Y5jeGORGS11KzEE5yTxBGrGYI usp=sharing)https://docs.google.com/spreadsheets/d/1lSO5jJwTkMNGl9kk6Y5jeGORGS11KzEE usp=sharing (https://docs.google.com/spreadsheets/d/1lSO5jJwTkMNGl9kk6Y5jeGORGS11KzEE5yTxBGrGYI usp=sharing)

I also added - as per other requests - 'Severe outcomes / Cases (%)' analysis.

It does show reduction of risk of Severe outcomes for Vaxed (over Unvaxed) by factor of to ~2-3.

But is that really what they call "significantly protected"?..

I have also added the comparison of C-19 deaths vs. C-19 vaccine serious events

which show that the number of -19 vaccine serious events is of the same order of scale as C-19 deaths... Is that "safe vaccine"

and The worst news comes when you split this comparison by age groups -

Then you'll see that for people <65 (esp. <19), the number of serious events is actually more (and much more) t

And ... there's a delay in reporting adverse events by about three months... So the number of events is even hig


(https://gccollab.ca/profile/zhuoyuwang-3)

By Zhuoyu Wang (https://gccollab.ca/profile/zhuoyuwang-3) 2022-03-04 14:59
(https://gccollab.ca/discussion_reply/edit/11810994)
(https://gccollab.ca/action/discussion_reply/delete?guid=11810994&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)
(https://gccollab.ca/action/likes/add?guid=11810994&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)

Fully vaccinated with an additional dose' (87,090) is included in 'Fully vaccinated (48,330) ???


(https://gccollab.ca/profile/Dmitry.Gorodnichy)

By Dmitry Gorodnichy (https://gccollab.ca/profile/Dmitry.Gorodnichy) 2022-03-04 16:37
(https://gccollab.ca/discussion_reply/edit/11815896)
(https://gccollab.ca/action/discussion_reply/delete?guid=11815896&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)
(https://gccollab.ca/action/likes/add?guid=11815896&__elgg_ts=1768408120&__elgg_token=jaHf2nUlTHQmMLyJlCQZaQ)

Actually this column (Fully vaxed w. 1 dose) should be ignored altogether not to confuse anyone.

It can be used ONLY after at least the second week after they started reporting 'Fully vaccinated with an adc dose' - because for "weekly numbers" you need to substract one week from previous week.

so what you see (87,090) is the difference between (0 and 99531) / 8 * 7. Makes sense?

0
99,531

You should be able to see the formula yourself -from link below - can you ?

Spreadsheet with raw data and calculation: Google spreadsheet (https://docs.google.com/spreadsheets/d/1lSO5jJwTkMNGl9kk6Y5jeGORGS11KzEE5yTxBGrGYI usp=sharing)

H20
=('Reported (from Dec 2020)'!H20-'Reported (from Dec 2020)'!H19)/('Reported (from Dec 2020)'!$B20-'Repor (from Dec 2020)'!$B19)*7

Reply to this

Embed content

Home (https://gccollab.ca/) ➜ Groups (https://gccollab.ca/groups/all)



R4GC (Use R!) Community

(https://gccollab.ca/ajax/view/thewire_tools/reshare?reshare_guid=7391537)
Owner: Dmitry Gorodnichy (https://gccollab.ca/profile/Dmitry.Gorodnichy)
(https://gccollab.ca/action/cp_notify/unsubscribe?guid=7391537&__elgg_ts=1768410347&__elgg_token=CLA4eoRdOFjX3NOsymBNcQ)
Group members: 364 (https://gccollab.ca/groups/members/7391537)
+ Invite (https://gccollab.ca/groups/invite/7391537)

Activity (https://gccollab.ca/groups/profile/7391537/r4gc-use-r-community-utilisez-r-r4gc-communaut)

Discussion (https://gccollab.ca/discussion/owner/7391537)

Files (https://gccollab.ca/file/group/7391537/all)

Blog (https://gccollab.ca/blog/group/7391537/all) More ()

Ce contenu est aussi disponible en Français

Description

This GCcollab group is created to bring together data scientists and data lovers within the Government of Canada. We discuss data problems that are of interest to our community and tackle them using the R language. Everyone is welcome to join, whether you are an advanced R user, or just curious about data-related issues. For those who already code in R, or interested in learning R, we gather and curate resources and knowledgebase related to the use of **R** within the Government of Canada.

Background:

R is one of the fastest growing programming languages and environment for data science, visualization and processing. **R** integrates well with Python and other free and commercial Data Science tools. It can also do something that other tools cannot and is very well supported by growing international community (https://stackoverflow.com/questions/tagged/r). It also comes with **RStudio** - a free Integrated Development Environment (IDE) that is now supported by most

GC Agencies and Departments, and that is becoming one of the main tools for data related problems worldwide. For Microsoft data users, anything you do with Excel, Access, Power BI, you can also do with R and **RStudio.**

The idea to create this group came after the #GCData2021 (https://www.csps-efpc.gc.ca/events/data-conference2021) Data Literacy Fest workshop on Data Engineering Challenges and Solutions: Demonstration of Shiny (https://gccollab.ca/groups/activity/7071541). The highest voted question there was : "*How can I get more help for our members to enhance their knowledge and how can I "spread the word" and raise more awareness regard to this tool?* This group is the answer to this question.

Additional information:

- Public facing page: https://open-canada.github.io/UseR/ (https://open-canada.github.io/UseR/)
- Subgroup: Lunch and Learn R: https://gccollab.ca/groups/profile/7855030/enfridays-lunch-and-learn-r (https://gccollab.ca/groups/profile/7855030/enfridays-lunch-and-learn-r-meet-upsfr)
- GCcode group - r4gc: https://gccode.ssc-spc.gc.ca/r4gc/gc-packages (https://gccode.ssc-spc.gc.ca/r4gc/gc-packages)
- GitHub (Apps repository): https://github.com/open-canada/Apps (https://github.com/open-canada/Apps)
- GCwiki: https://wiki.gccollab.ca/UseR! (https://wiki.gccollab.ca/UseR!)

For questions and feedback please contact:

- Dmitry Gorodnichy: Dmitry@Gorodnichy.Ca (mailto:Dmitry.Gorodnichy@cbsa-asfc.gc.ca)

Add widgets

Tag cloud

(https://gccollab.ca/ajax/view/widget_manager/widgets/settings?guid=8340239&show_access=)
(https://gccollab.ca/action/widgets/delete?widget_guid=8340239&context=7391537&__elgg_ts=1768410346&__elgg_token=miq-63ztvWyPbiXnoNGJzg)

Geosptial (https://gccollab.ca/search?q=Geosptial&search_type=tags), knowledgebase (https://gccollab.ca/search?q=knowledgebase&search_type=tags), gcwiki

Government of Canada Gouvernement du Canada

Report a side effect

Download PDF Form

* Mandatory Field

*Complete all mandatory fields, marked by a *, and provide as much detail as possible for the remaining fields.*

Specific field instructions are included in the Instructions section above.

A. Report and Reporter Information

***** Type of report

Initial ◉ Follow-up ○

Health Canada Reference No. (for follow-up reports only)

Reference number for a previously submitted initial report

Reporter File No.

***** Reporter First Name

***** Reporter Last Name

***** Telephone

A telephone number, a mailing address or an email address must be provided.

613

Ext.

***** Address

A telephone number, a mailing address or an email address must be provided.

City

Ottawa

Province / Territory

Ontario

***** Postal Code

A telephone number, a mailing address or an email address must be provided.

* Email Address

A telephone number, a mailing address or an email address must be provided.

[redacted]

Organization (if applicable)

Select one that best describes you

Consumer or other non health professional

B. About the person who had the side effect

* *Patient ID, age, or sex must be provided*

* Patient ID (for health care providers)

Patient identifier for follow-up purposes (e.g., patient initials, patient record number). Please do not provide patient's full name.

lv

* Age at the time of the side effect

[redacted]

Select Age Unit

* Sex

Male

Height

cm 170 ft in

Weight

kg 70 lb oz

Known medical conditions and relevant lifestyle factors

e.g. liver and/or kidney impairment, diabetes, current pregnancy, tobacco, cannabis or alcohol use, recreational drug use, etc.

[redacted]. Healthy life style.

Known Allergies

e.g. food, drugs, environmental, etc.; provide details

none

C. Information on the Side Effect

Reason for seriousness

(More than one can be selected)

- Caused/prolonged in-patient hospitalization
- Disability
- Birth defect
- Needed medical attention

Explain

- Life-threatening
- Death

Date of death (if known)

Partial dates are acceptable

YYYY-MM-DD

☑ None of the above (non-serious report)

Did the person recover from the side effect?

- ○ Recovered
- ○ Recovering
- ○ Not recovered
- ◉ Recovered with residual medical complications
- ○ Died
- ○ Unknown

Side effect start date (YYYY-MM-DD)

Partial dates are acceptable

2021-05-15

Side effect end date (YYYY-MM-DD)

Partial dates are acceptable

2021-12-01

* Describe the side effect (timelines, treatment, etc.)

Shingles (HERPES ZOSTER) - painful rushes all around torso started 1.5 months after first dose of Moderna (22 Mar 2021), and became worse after the second dose of Moderna (12 Jul 2021). Painful rashes were not going away for over six months, medication was not helping much (7 pills a day made feeling weak with little improvement), finally gone 2021, the pain however continues to present day; never entirely recovered, debilated for the entire summer and most of fall, was not able to do favourite outdoor activites such as biking and swimming, which was normally doing always in past years, overall weakness and pain after the Shingles has started, till present.

D. Suspect Product

Product Information

* *DIN#/NPN# or Product name must be provided*

* DIN #/NPN #

MODERNA COVID-19 m

* Product Name

MODERNA COVID-19 m

Strength

Amount of active ingredient per single dosage form of the drug

Strength unit

Select | Strength unit other

Dosage form

e.g. tablet, powder, liquid

Manufacturer name

MODERNA COVID-19 m

Lot #

3001176

Expiry date

Partial dates are acceptable

YYYY-MM-DD

Therapy information at the time of side effect:

Product start date (YYYY-MM-DD)

Partial dates are acceptable

YYYY-MM-DD

Product end date (YYYY-MM-DD)

Partial dates are acceptable

YYYY-MM-DD

Dose

Quantity of product taken at a time (e.g. 50 mg or 5 tablets)

Frequency

How often the product is taken (e.g. twice daily)

How was the product taken?

Intramuscular Other

What was the product prescribed/taken for?

† Did you also report to the manufacturer?

Yes ○ No ◉

Date reported to manufacturer

Partial dates are acceptable

YYYY-MM-DD

Manufacturer reference number (if known)

What action was taken?

- ○ Drug withdrawn
- ○ Dose reduced
- ○ Dose increased
- ○ Dose not changed
- ○ Unknown
- ○ Not applicable

If the product was stopped did the side effect stop?

- ○ Yes
- ○ No
- ○ Unknown
- ○ N/A

If the product was restarted, did the side effect return?

- ○ Yes
- ○ No
- ○ Unknown
- ○ N/A

✚ Additional Suspected Health Product(s)

E. Other health product(s)

List all known health product(s), other than the suspect product(s), taken when the side effect occurred, excluding treatment. If known, please include additional information related to the product(s) (e.g. length of use, timelines, etc.).

F. Additional Information

Use this section to include details that did not fit in the previous sections' structured boxes and that you feel would contribute to the assessment of the side effect. This section can also include additional suspected health product(s) information.

Related test/laboratory results

Before submitting your report, please review the information you provided.

Submit

Date modified:
2021-07-14

Government of Canada activities and initiatives

#YourBudget2018 – Advancement



Advancing our shared values

#YourBudget2018 – Reconciliation



Advancing reconciliation with Indigenous Peoples

#YourBudget2018 – Progress



Home (https://gccollab.ca/) ➔ Discussion (https://gccollab.ca/discussion/all)
➔ Data Science with Open Canada Data 'Lunch and Learn' seminars (https://gccollab.ca/discussion/owner/7855030)
➔ Today's meetup via zoom - more on working Googlesheets with R

Data Science with Open Canada Data 'Lunch and Learn' seminars

(https://gccollab.ca/ajax/view/thewire_tools/reshare?reshare_guid=7855030)

Owner: Dmitry Gorodnichy (https://gccollab.ca/profile/Dmitry.Gorodnichy) (https://gccollab.ca/action/cp_notify/unsubscribe?guid=7855030&__elgg_ts=1768408564&__elgg_token=zjK-nV8woFp-j1qLiDoi0g)

Group members: 111 (https://gccollab.ca/groups/members/7855030)

+ Invite (https://gccollab.ca/groups/invite/7855030)

Activity (https://gccollab.ca/groups/profile/7855030/data-science-with-open-canada-data-lunch-and-learn-seminars)

Discussion (https://gccollab.ca/discussion/owner/7855030) Files (https://gccollab.ca/file/group/7855030/all)

Blog (https://gccollab.ca/blog/group/7855030/all) More ()

Today's meetup via zoom - more on working Googlesheets with R

Sorry for last week's technical problems - should be fixed now.- If you get kicked of meeting - you should be able now to automatically rejoin it, by clicking the same zoom link again

Join Zoom Meeting
(https://us04web.zoom.us/j/72260868830?pwd=dgsDcmFPb82kRI8l-8wkhiKdv8lPuX.1)https://us04web.zoom.us/j/72260868830?pwd=dgsDcmFPb82kRI8l-8wkhiKdv8lPuX.1 (https://us04web.zoom.us/j/72260868830?pwd=dgsDcmFPb82kRI8l-8wkhiKdv8lPuX.1)
Meeting ID: 722 6086 8830
Passcode: HqwR90

I have prepared a number of GoogleSheets - of much interesting information for working with:

1. Severe C-19 outcomes by age (Source: PHAC-1)
2. Side-effects following vaccination by age (Source: PHAC-2)
3. Cases following vaccination (Source: PHAC-1)

So we can continue the analysis of data related to Safety and Efficacy of Vaccine.
Today, we can look into results by various age groups

Data:

1. https://docs.google.com/spreadsheets/d/1BIoCKilBY5UJ0_4Q6eiz_51TSn_NImkSrJeoIzNm5pY/edit?usp=sharing (https://docs.google.com/spreadsheets/d/1BIoCKilBY5UJ0_4Q6eiz_51TSn_NImkSrJeoIzNm5pY/edit?usp=sharing)
2. https://docs.google.com/spreadsheets/d/1UilgWgnvkoPLIaxdBcz3XSPTymOVDKUp-pH7pv-HyhY/edit?usp=sharing (https://docs.google.com/spreadsheets/d/1UilgWgnvkoPLIaxdBcz3XSPTymOVDKUp-pH7pv-HyhY/edit?usp=sharing)
3. https://docs.google.com/spreadsheets/d/1p27Qsk-O-vzr6aGqyAVUODyRL97wh9PIxMchpJSWFn0/edit?usp=sharing (https://docs.google.com/spreadsheets/d/1p27Qsk-O-vzr6aGqyAVUODyRL97wh9PIxMchpJSWFn0/edit?usp=sharing)

(live version:
(https://docs.google.com/spreadsheets/d/1lSO5jJwTkMNGl9kk6Y5jeGORGS11KzEE5yTxBGrGYKw/edit#gid=0)https://docs.google.com/spreadsheets/d/1
(https://docs.google.com/spreadsheets/d/1lSO5jJwTkMNGl9kk6Y5jeGORGS11KzEE5yTxBGrGYKw/edit#gid=0))

(https://gccollab.ca/profile/Dmitry.Gorodnichy) Started by Dmitry Gorodnichy (https://gccollab.ca/profile/Dmitry.Gorodnichy) 2022-03-04 11:52

(https://gccollab.ca/bookmarks/add/55030?address=https%3A%2F%2Fgccollab.ca%2Fdiscussion%2Fview%2F11810540%2Ftodays-meetup-via-zoom-more-on-working-googlesheets-with-r&title=Today's meetup via zoom - more on working Googlesheets with R&title2=)
(https://gccollab.ca/action/discussion/toggle_status?guid=11810540&__elgg_ts=1768408564&__elgg_token=zjK-nV8woFp-j1qLiDoi0g)
(https://gccollab.ca/discussion/edit/11810540)
(https://gccollab.ca/action/discussion/delete?guid=11810540&__elgg_ts=1768408564&__elgg_token=zjK-nV8woFp-j1qLiDoi0g)
(https://gccollab.ca/ajax/view/thewire_tools/reshare?reshare_guid=11810540)
(https://gccollab.ca/action/likes/add?guid=11810540&__elgg_ts=1768408564&__elgg_token=zjK-nV8woFp-j1qLiDoi0g)
(https://gccollab.ca/action/cp_notify/unsubscribe?guid=11810540&__elgg_ts=1768408564&__elgg_token=zjK-nV8woFp-j1qLiDoi0g)

Replies

(https://gccollab.ca/profile/Dmitry.Gorodnichy)

By Dmitry Gorodnichy (https://gccollab.ca/profile/Dmitry.Gorodnichy) 2022-03-04 12:05
(https://gccollab.ca/discussion_reply/edit/11810626)
(https://gccollab.ca/action/discussion_reply/delete?guid=11810626&__elgg_ts=1768408564&__elgg_token=zjK-nV8woFp-j1qLiDoi0g)
(https://gccollab.ca/action/likes/add?guid=11810626&__elgg_ts=1768408564&__elgg_token=zjK-nV8woFp-j1qLiDoi0g)

R code to read data:

Reading https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html#a9 (https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html#a9) (Table 2)

```{r}
source("00_common.R")
library(googlesheets4)
gs4_deauth()
# gs4_auth()

strSheet <- "https://docs.google.com/spreadsheets/d/1p27Qsk-O-vzr6aGqyAVUODyRL97wh9PIxMchpJSWFn0/edit?usp=sharing" (https://docs.google.com/spreadsheets/d/1p27Qsk-O-vzr6aGqyAVUODyRL97wh9PIxMchpJSWFn0/edit?usp=sharing")
col_type = paste0("DD", paste0 (rep("d", which(letters == "x") - 2),collapse = ""))
dtRaw <- read_sheet(strSheet,
                    sheet=2, range="A:X",
                    col_types = col_type,
                    col_names=F
                    ) %>% setDT
```

Reading https://health-infobase.canada.ca/covid-19/vaccine-safety/ (https://health-infobase.canada.ca/covid-19/vaccine-safety/)

```{r}
strSheetPHAC2 <- "https://docs.google.com/spreadsheets/d/1BIoCKilBY5UJ0_4Q6eiz_51TSn_NImkSrJeoIzNm5pY/edit?usp=sharing" (https://docs.google.com/spreadsheets/d/1BIoCKilBY5UJ0_4Q6eiz_51TSn_NImkSrJeoIzNm5pY/edit?usp=sharing")
dtEvents<- read_sheet(strSheetPHAC2,
                      # col_types = col_type,
                      # col_names=F,
                      sheet=2 ) %>% setDT
```

vitals

Official Canada Vital (Deaths) Statistics, analyzed using Data Science

View On GitHub

This project is maintained by open-canada

Reporting COVID-19 vaccines side effects

Background

Reporting of serious (or life threatening) side effects of vaccines is mandatory in Canada. However, reporting of non-serious (or not life threatening) side effects in Canada is voluntary. Many doctors may not have time to do it, nor they provide guidance to their patients on how they can do it.

For this reason many non-serious side effects remain unreported. Many doctors will not do it, and consumers do not know how to do it. In fact, many Canadians may not even know that they can do report vaccine side-effect themselves, i.e. without a medical doctor. There are no instructions on the Health Canada portal on how to do it. The link to the Health Canada Voluntary Side Effect Reporting portal may also be rather difficult to find.

This page is created to address this gap and to help Canadians to report any vaccine (or any drug) side-effect, as a consumer. Health Canada encourages everyone to do so. – It is thanks to such efforts that vaccines (and drugs) can be improved for the benefits of all.

Once you know how to do it, it should not take more than fifteen minutes to do it. This is how you do it. An Example is provided here.

Submitting Voluntary Side Effect Report via Health Canada portal

Step 1: Go to the Health Canada Voluntary Side Effect Reporting portal : https://hpr-rps.hres.ca/side-effects-reporting-form.php?form=voluntary

Step 2: Submit electronic form there as shown in this Example.

- There are three places there that you need to know how to fill out:
  - in Section B: Patient ID write your initials;
  - in Section D: Product DIN and Product Name - paste the vaccine names (and batch/lot number, if available), as printed on your receipt.
- Make sure you provide a valid email address, which will be used to send you to further instructions on how ontinue side-effect report process.
- If you still have problems submitting the form (i.e., clicking on button “SUBMIT” does not lead you anywhere),
  call Heath Canada Voluntary Reporting support line: Tel. 866-234-2345, if you leave a message. They should respond within 24 hours. If they did not, please call again.

Step 3: A few (2-5) days later you should receive email from *Canada Vigilance (HC/SC) canadavigilance@hc-sc.gc.ca* (with Subject: “Correspondence from the Marketed Health Products Directorate - Adverse Reaction Number 0009xxxxx), which will provide you with the
Adverse Reaction Report (AER) number and which will ask you to submit additional details though their portal.

If you don’t receive email from them within 5 business days, please check you Spam folder and, if it is not there, please call them at Tel. 866-234-2345 and leave a message.

As they note in their email: *“Posting of reports in the online database remains 3 months behind the date of reporting. This delay accounts for the necessary activities involved in report processing”*

Additional resources:

The following additional resources are provided to help you prepare your side-effect report. There you can check if your symptoms are already reported in national adverse event reporting systems, and if so, how common they are, for which type of vaccine, and which vaccine batch (a.k.a. lot, which is the number printed on your vaccination receipt)

- Recalls and safety alerts, Government of Canada https://recalls-rappels.canada.ca/: for [Moderna], for [Pfizer], for [Astrazeneca])
- 'Reported side effects following COVID-19 vaccination in Canada', Canadian COVID-19 vaccination safety report, Public Health Agency of Canada, https://health-infobase.canada.ca/covid-19/vaccine-safety/
- US Government Vaccine Adverse Event Reports (VAERS): https://vaers.hhs.gov. You can also check if your vaccine batch (lot) is among the "bad ones", according to VAERS data, here,
- Comprehensive world-wide COVID-19 vaccine reports repository
- CUMULATIVE ANALYSIS OF POST-AUTHORIZATION ADVERSE EVENT REPORTS OF PF-07302048 (BNT162B2) RECEIVED THROUGH 28-FEB-2021. Prepared by Pfizer, Released to Public Health and Medical Professionals for Transparency (https://phmpt.org) under the court order on 17 November 2021: https://phmpt.org/wp-content/uploads/2021/11/5.3.6-postmarketing-experience.pdf

The authoritative source for COVID-19 information is Canada.ca/coronavirus.

Hosted on
GitHub Pages
using the Dinky theme

vitals

Official Canada Vital (Deaths) Statistics, analyzed using Data Science

View On GitHub

This project is maintained by open-canada

Analysis of Official Data Related to COVID-19 Vaccine Efficacy and Safety

open-canada.github.io/vitals

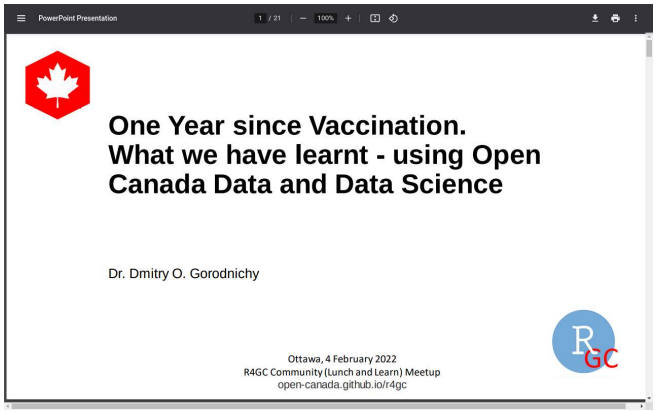
Disclaimer: The authoritative source for COVID-19 information is Canada.ca/coronavirus. The results and views presented here are those by the authors and are not to be misconstrued as representing views of any Canadian Agency, University, or Department. They were developed and gathered by the R4GC Data Science community as part of the community training and socializing, mostly on authors' own time and using authors' own resources. All codes behind the analysis and supporting documents are free, open and available in GitHub. Comments and corrections are welcome.

Talking about "The Elephant in the Room".

The following presentation was given to the Government of Canada employees, who expressed interest in learning about the latest Government of Canada data related to COVID-19 vaccination. A considerable portion of this presentation was about how to create a safe and respectful environment within Government of Canada where Government of Canada employees can discuss this topic ("The Elephant in the Room"), which is already causing much of polarization within Canadian public, including among the Government of Canada employees, and which has led to the largest in the Canadian history

demonstrations of protests of Canadians against the government. government, demanding it to revoke its "Mandatory Vaccination Policies" The solution to this problem was seen in keeping the discussion focused entirely on Canada's own sources and open data, much of which – thanks to the Open Goverment and Scientific Integrity policies introduced by the Liberal Government several years ago – have now become available, including statistics of the number of deaths by cause and by vaccination status, which can be analyzed and effectively visualized using contemporary open source data science tools such as R and RStudio.

- "One Year since Vaccination. What we have learnt - using Open Canada Data and Data Science", Special Edition presentation for the R4GC Community 'Lunch and Learn' Meetup by the R4GC Data Science community lead, Senior Data Scientist, Dr. Dmitry Gorodnichy (4 February 2022).



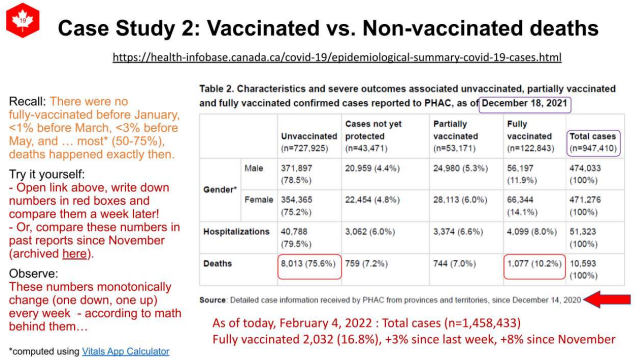
White Papers and Tools:

- Guide on how to report COVID-19 vaccine side-effects in Canada as a consumer
- Interactive *Open Canada Vitals Statitistcs (Deaths) Tracker* App

- "Some observations from Canadian Vital Statistics Death (CVSD) Database related to the 'Elephant in the Room'", Comment submission to The Lancet Journal (6 January 2022)

566-02-48305

- "Additional observations related to the 'Elephant in the Room' – from the Government of Canada COVID-19 daily epidemiology updates. Time to act!", White paper (10 January 2022)



- "Statistical analysis of official data sources related to vaccine efficacy and safety (NEW EVIDENCE SINCE OCTOBER 2021)"

Sources and Findings

Last Updated: 2022-02-10

[gc1] "Cases* following vaccination, COVID-19 Daily Epidemiology Update"*, Public Health Agency of Canada, https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html#a9 (archived reports)

NB: The statistics reported in this source are computed using death data since December 2020, when there were no fully vaccinated Canadians yet, but when many cases happened (More exactly: There were 0 fully-vaccinated in December, < 1% for most of January, <3% until May). This is known as *algorithmic bias*, which skews results favourably towards fully-vaccinated cases, and which can be removed using technique described in this paper. The numbers below are quoted directrly from the source.

COVID-19 deaths by Vaccination Status

- *on February 10, 2022 for cases from December 14, 2020 up until January 15, 2022 (n=1,670,682)*
  unvaccinated 68.1%% vs. fully vaccinated **19.5%** (+2.7% increase since last week)
- *on February 04, 2022 for cases from December 14, 2020 up until January 15, 2022 (n=1,458,433)*
  unvaccinated 70.3% vs. fully vaccinated **16.8%** (+3% increase since last week)
- *on January 21, 2022 for cases from December 14, 2020 up until January 08, 2022 (n=1,341,192)*:
  unvaccinated 72.8% vs. fully vaccinated **13.8%** (+4.3% increase since December 14)
- *on December 15, 2021 for cases from December 14, 2020 up until November 27, 2021 (n=882,988)*:
  unvaccinated 76.1% vs. fully vaccinated **9.5%**

[gc2a] Hospitalizations by vaccination status, COVID-19 (coronavirus) in Ontario, https://covid-19.ontario.ca/data/hospitalizations#hospitalizationsByVaccinationStatus

[gc2b] For more information on cases following vaccination, please see the Weekly epidemiology report (PDF) available on the Government of Canada's COVID-19 data trends page

- *"Incidence [of COVID-19 cases] among fully vaccinated cases has surpassed those among unvaccinated since mid-December 2021"*.
- In January 2022, close to 3/4 of hospitalized COVID patients in Ontario were fully vaccinated

- Total (since Jan 2021) "Serious": * 7,678. Weekly: 118
- Total (since Jan 2021) "Not serious": 28,486.

- Between *May 2021 and Oct 2021* "Serious" side-effects: **5200** (increased from 5038 posted in last month report)
- Of these: 1,111 heart diseases (including 64 Cardiac arrests, 77 Heart Attacks, 970 Myocarditis1/Pericarditis), 947 thrombosis, 781 central nervous system failures

566-02-48305

(including 594 Bell's Palsy), 487 Anaphylaxis, 166 Auto-immune diseases, 37 Acute kidney injury, 25 Liver injury, and 208 deaths (some under investigation).

[gc4] Recalls and safety alerts, Government of Canada: https://recalls-rappels.canada.ca/ (for [Moderna], [Pfizer], [Astrazeneca]):

- Risk of Facial paralysis/Bell's Palsy. Posted: 2021-08-06
- Risk of Myocarditis and pericarditis. Posted: 2021-06-30
- Risk of Capillary Leak Syndrome. Posted: 2021-06-29
- Risk of Thrombosis with Thrombocytopenia. Posted: 2021-03-24

[gc5] Statistics Canada, Canadian Vital Statistics - Death database . Provisional weekly death counts, by selected grouped causes of death. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310081001 (Dashboard: https://open-canada.github.io/vitals)

- From December 14, 2020 to April 30, 2021: 13075
- From May 1, 2021 to October 31, 2021*: **2975**
- Of these, deaths of COVID-19 without comorbidities: **404** (recomputed using [ukg1])
- Of these, deaths of COVID-19 without comorbidities for <65 : **89** (recomputed using [ukg1])

[gcA] **Vaccination rates** (Source: canada.ca):

- Dec 2020 - Apr 2021: High-risk population vaccination (rate: 0 – 2%)
- May 2021 - Oct 2021: General public vaccination (rate: **2% – 75%**)

[gcB] **COVID-19 cases vs. COVID-19 deaths** (source: google1):

- Canada population: 38.246.108
- Total COVID Cases (since March 2020): 3.150.000
- Total COVID Deaths (since March 2020): 35.00 or 1.1% Cases
- Of these, Most are comorbidities (>90% based on UKG and USG data)

- Of these, Most are in people 65+ (>90% based on GC, UKG, USG data)
- COVID Deaths for people younger than 65 with no health problems: < 0.01% Cases

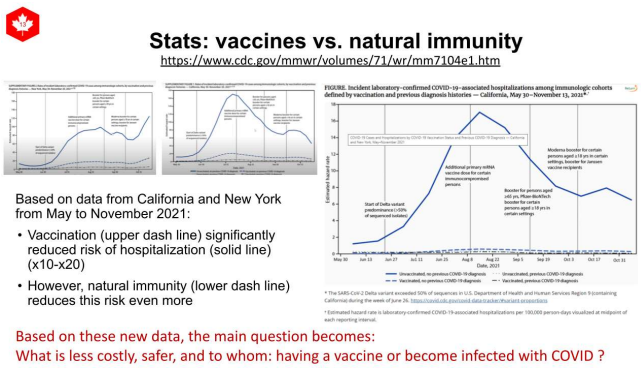
Data from other governments:

[ukg1] **COVID deaths vs. COVID alone deaths** (Source: gov.uk - Explanation):

- Of all reported in UK COVID deaths, only 13.6% are due to COVID alone.
- 78% of all COVID alone deaths occurred in people with 65+ of age.
- Average age of death from COVID is 82.5 years, which is statistically equal to the average life expectancy in UK.

[usg1] **Natural immunity vs. vaccine-induced immunity** (Source: cdc.gov - Explanation):

- COVID vaccine reduces the risk of hospitalization considerably (by x10-x20)
- However, natural immunity reduces the risk of hospitalization even more (by x20-x50), while the chance of having severe complication from COVID infection for people younger than 65 and in good health is very low (<0.01% Cases based on [ukg1]) and for children under 12 even lower (<0.0001% Cases based on [ukg2]).



Corrections/Comments: [open-canada-data-analysis @ ivim . ca]

Hosted on

GitHub Pages

using the Dinky theme

566-02-48305

./ UseR

Public facing R4GC community portal (Open Canada portal for R users)

View on GitHub

Welcome to 'Open Canada Use R!' Portal

GitHub GCCollab GCcode Resources Apps The R4GC Book R101.

Pages for people interested in learning and using R in the Government of Canada and elsewhere.

News:

2021-11-14

>> The first draft of The R4GC Book released at https://open-canada.github.io/r4gc)

2021-10-29

>> A presentation was made about the work of our R4GC community at 2021 International Methodology Symposium. The slide decks of this presentation are here: in English and en francais)

2021-09-10

>> On 8 October, we are starting a new series of Lunch and Learn's weekly meetups focused on Text Analysis with R. For more details please see GCCollab Events pages.

2021-08-04

>> A contribution is made to the 2021 International Methodology Symposium to describe the collaborative work done by the R4GC community.

>> Please join us on Friday 29th October 2021 at the Symposium for Session 9A -- Making Official Statistics More Open (13:00- 14:15) to learn about "Development of R libraries for common tasks with open Canada data".

2021-06-25

>> We finished our Season 2 - The codes and video recordings are available on GCcollab.
>> This portal goes now into the dormant state. Please visit GCCollab: R4GC Community for all information related to our future "Lunch and Learn" meet-ups and related knowledgebase.
>> The GC wiki page is created, which provides links to R4GC community discussion topics and resources: https://wiki.gccollab.ca/UseR!

2021-05-10:

>> The Apps that we have developped are now also hosted at: https://open-canada.github.io/Apps!

2021-04-30:

>> Based on the poll answers, our weekly Lunch and Learn R Friday meet-ups moved from 11:30-12:15 to 12:05-12:55. See GCCollab Group for all details.

2021-04-23:

>> We now have a new Use R! portal hosted here at open-canada GitHub account! This portal is the continuation of the efforts started last year at R Ottawa, now supported by a larger community of government employees from around the country.

2021-04-16:

>> We have started a new series of hands-on "build from scratch" tutorials - focused on developing skills for building your own R packages. The objective is to build a variety of useful packages for common use by all, using the codes contributed by the community.
>> GC employees and partners, please visit GCCollab Group for more details, notes and recordings from these meetups.

2021-03-26:

>> The deck and video-recording from the seminar “Data Engineering with R, R Markdown, Shiny and algorithms” (by Dmitry Gorodnichy) organized by Statistics Canada for the Applied machine learning for text analysis community of practice are now posted at https://gccollab.ca/event_calendar/view/7556628

2021-02-27:

>> GC employees, head to https://gccode.ssc-spc.gc.ca/r4gc for more resources and codes related to R coding and usage at GC! Join us for informal ‘Lunch and Learn’s meet-ups every Friday at 11:30. - See GCcollab for more information.

2021-02-21:

>> Data Engineering Demo is ready for the #GCData2021 Conference Data Literacy workshop: https://rCanada.shinyapps.io/demo/. See workshop GCcollab page for the associated deck and video-recording of the workshop

2021-01-06:

>> Join us at the Data Engineering workshop on 22 February, organized within the Fifth Annual Data Conference (#GCData2021 / #DonnéesGC2021). Hosted by CBSA, “the workshop will overview the challenges and solutions related to data engineering. You will be able to experience some of data engineering challenges first-hand, and see how these challenges can be solved using modern data science tools such as R and RStudio. The work on developing an R package and Shiny application to assist with data cleaning and linking will be presented.”

2020-12-10:

>> Following the announcement at the Data Science Network for the Federal Public Service (DSNFPS) newsletter, we received multiple requests for another R training, like the one we did last Summer. Why not? To enroll and more details - see R101.

2020-06-05:

>> We finished the first season, where we learnt how to write an interactive dashboard that visualizes the COVID datain US. - You can see the final result here! All codes and resourcea are posted here. Have great summer!

2020-05-03:

>> YouTube channel created: R101 “Learning R together” (Building Data Science Apps from scratch). If you missed a session, you can now watch it there. Session transcripts are also posted in R101.
>> Ideas for homework are added in Resources

2020-04-28:

>> We’ve heard you: You want to get all the power and fun of R, but you don’t know where to start – Starting this week, we’ll show you how you can start learning and using R in absolutely no time and even without any software installed on your computer! All you’ll need is to login to https://rstudio.cloud and we’ll guide you through after that. We will be developing a COVID App from scratch - together! Sessions will be recorded.

2020-04-01:

>> Invitation to Lunches with Data Challenges meet-ups is extended from CBSA to all GoC. See: GCCollab*
>> The focus of discussion becomes Covid-19 and what we (R enthusiasts and learners) can do about it.

2020-03-18

>> Our informal Wednesday’s “Morning Coffees with Data Challenges” meetings, which we’ve been having in CBSA since Summer 2019, go online, becoming “Lunches”.

Contact:

Dmitry Gorodnichy (dmitry.gorodnichy @ gmail.com or dmitry.gorodnichy @ cbsa-asfc.gc.ca)