

# Reading Between the Lines: A Pursuit of Estimating the Population Prevalence of Mental Illness Using Multiple Data Sources

The Canadian Journal of Psychiatry /  
La Revue Canadienne de Psychiatrie  
2022, Vol. 67(2) 101–103  
© The Author(s) 2021



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/07067437211016255  
TheCJP.ca | LaRCP.ca



Jordan Edwards, PhD<sup>1</sup>  and Katholiki Georgiades, PhD<sup>1</sup>

## Abstract

Population-based prevalence estimates of mental illness are foundational to health service planning, strategic resource allocation, and the development and evaluation of public mental health policy. Generating valid, reliable, and context-specific population-level estimates is of utmost importance and can be achieved by combining various data sources. This pursuit benefits from the right combination of theory, applied statistics, and the conceptualization of available data sources as a collective rather than in isolation. We believe there is a need to read between the lines as theory, methodology, and context (i.e., strengths and limitations) are what determines the meaningfulness of a combined prevalence estimate. Currently lacking is a gold standard approach to combining estimates from multiple data sources. Here, we compare and contrast various approaches to combining data and introduce an idea that leverages the strengths of pre-existing individually linked population-based survey and health administrative data sources currently available in Canada.

## Keywords

epidemiology, prevalence, mental disorders, Bayesian statistics, triangulation, evidence synthesis, mental health service planning, mental health policy

As psychiatric epidemiologists interested in producing population-based estimates of mental illness, we must be guided by the strengths and limitations of the data we use. The pursuit of estimating prevalence by combining multiple sources of evidence benefits from the right combination of theory and applied statistics. Important questions remain, however, on how best to conceptualize and combine data arising from different sources, such as survey and health administrative data, to arrive at an accurate estimate of mental illness in the population. Recently, there have been a number of suggested approaches, including both a Bayesian and now a triangulation approach from Vigo et al., which have helped to move this discussion forward.<sup>1,2</sup>

One of the greatest challenges of this work is establishing common measurement and determining whether various data sources are measuring the same underlying construct. One method for testing this is to use linked data to measure concordance across sources. Canada is positioned to tackle this challenge, as the number of provincial and national population-level data linkages has increased. Recent Canadian findings using linked population-level survey and health administrative data revealed that while both sources of data provided comparable prevalence estimates of mood

or anxiety disorders, there was high discordance between these data sources.<sup>3</sup> As such, we now know that these distinct data sources are either capturing different people with a mood or anxiety disorder or the same group of people at different stages of their mental health trajectory.<sup>3</sup> To fully capitalize on the strengths of various data sources, researchers must conceptualize available sources as a collective rather than a series of sources in isolation. The idea that the strength of one data source may be its ability to reduce bias in another is important. Combining evidence involves balancing estimate precision, which can be increased by maximizing available evidence, with estimate accuracy (i.e., validity), which involves quantifying and when possible reducing estimated biases from various data sources.

<sup>1</sup> Offord Centre for Child Studies, McMaster University, Hamilton, Ontario, Canada

## Corresponding Author:

Jordan Edwards, Offord Centre for Child Studies & Department of Psychiatry and Behavioural Neurosciences, Michael G. DeGroote School of Medicine Department, McMaster University, Hamilton, Ontario, Canada.  
Email: jordan.edwards@mcmaster.ca

While not mutually exclusive, this balance often influences methodological considerations.

In the pursuit of estimating mental illness in the population using multiple data sources, there is a need to *read between the lines* as theory, methodology, and context (i.e., strengths and limitations) are what determines the meaningfulness of a combined prevalence estimate. Here, we compare and contrast various approaches to combining data, and we introduce an idea that leverages the strengths of pre-existing individually linked population-based survey and health administrative data sources currently available in Canada.

## Methodology for Combining Estimates from Multiple Data Sources

### Triangulation Approach

Vigo and colleagues introduced an approach to combining estimates from many different data sources using meta-analysis and triangulation.<sup>1</sup> This approach offers the ability to provide timely estimates that are informed from various Canadian and international data sources, including evidence from survey and health administrative data. Furthermore, the authors introduce the idea of contextualizing estimates of mental illness by disorder class, age, sex, and presenting severity fractions that are informed by the Global Burden of Disease framework.<sup>4</sup> This approach accompanied with the proposed framework for estimating service needs based on the findings from their triangulation analysis, positions this work as a forward-thinking approach to using multiple data sources to estimate mental illness in the population.

There are a number of key assumptions that need to be met for the triangulation approach to improve the validity of an estimate. These include (1) various data sources included in the model are measuring the same underlying construct and (2) the various data sources that are being combined have differing and unrelated key sources of potential biases.<sup>5</sup> We believe these assumptions might compromise the validity of estimates drawn from a triangulation approach. First, evidence suggests that various data sources (i.e., survey data, administrative data) may be measuring different components of the same underlying construct rather than the exact same outcome.<sup>3</sup> Second, enumerating the magnitude, direction, and independency of biases between and within various data sources is a challenge. Finally, while paired with a meta-analysis, the triangulation approach maximizes available evidence, thus increasing the precision of a combined estimate, accuracy (i.e., validity) may be compromised. Combining sources that provide greater homogeneity in the assessment of the underlying construct may provide more accurate estimates.

### Bayesian Approach

An alternative approach to triangulation, which is currently being used in the global burden of disease study,<sup>6</sup> and to forecast mental health service needs in the UK,<sup>7</sup> is a Bayesian analytical framework. One of the major strengths of this methodology is its ability to model an outcome using not only the prevalence estimates from various sources, but prior information which can be integrated to enhance our confidence in estimates drawn from various sources. This is important when combining data sources which measure different components of the same outcome. In these cases, a Bayesian model can integrate prior information that characterize the relationship between various data sources (i.e., their concordance, observed or estimated) as well as prior information that can help guide our confidence in both measures compared to a gold standard measure. This was shown in a recent Bayesian analysis conducted in Ontario.<sup>2</sup> While this approach may be used to combine estimates with varying potential biases, one of the strengths of this approach is its flexibility in using prior information to adjust the model to reduce known biases. This has important implications for estimating mental illness among Canada's diverse population, where various measures may provide more valid estimates among certain sub-populations. We believe there are two major strengths of the Bayesian approach. First, is its ability to produce combined estimates that are informed from prior information, which allow researchers to leverage the strengths of various data sources while incorporating and adjusting for potential biases to enhance the accuracy of pooled estimates. Second, it has the distinct advantage of forward uncertainty propagation through its pairing with simulation-based methodologies.<sup>8</sup> This allows for the estimation of uncertainty surrounding individual prior estimates and combined estimates. This not only provides insight into our confidence in a combined estimate but also into the quality of the evidence that is available to inform the model. Furthermore, Bayesian modelling can be used to measure mental illness over space, time, and by sub-population.<sup>9</sup> As such, we believe this approach is well suited for measuring mental illness at the population level.

One of the challenges of using a Bayesian analysis is the time required to create an initial analytical infrastructure. Once established, however, new data can be added in an iterative fashion, which would provide an opportunity for timely dissemination. A Bayesian approach may also require extensive collaboration to inform the model when prior information is unavailable. As such, this approach presents the opportunity to integrate the values and perspectives of a diverse range of researchers and knowledge users.

### Measuring Mental Illness in the Population

While the technique used to combine data is important, we believe developing our theory of measurement is equally important. The field has room for development in this area.

For example, there may be benefits to moving away from focusing on measuring one single overall prevalence to a more contextualized series of prevalence estimates, including estimates of population need, service use, and unmet need. This would provide opportunities to inform service planning, and evaluate service targeting for various mental disorders. To that end, we believe individually linked population-based survey and health administrative data can support these efforts. Specifically, by leveraging (1) the ability of surveys to estimate community need and community mental health service contacts; (2) administrative data's ability to estimate physician-based service contacts; and (3) the linkage of both data sources to measure the discordance between indicators of mental health needs and service use to estimate unmet need in the population. Fully capitalizing on currently available data linkages in Canada may reveal new ways of measuring mental illness in the population, which may better serve policy makers.

### Reading Between the Lines

The production of valid, reliable, and context-specific population-based estimates of mental illness are the foundation of health service planning and strategic resource allocation. These estimates are used to inform and evaluate equitable public mental health policy initiatives and the performance of our mental health care system. We believe in the idea of combining various data sources for improving our ability to generate reliable estimates of mental illness, including estimates of population need, service use, and unmet need. Navigating the data landscape and identifying, quantifying, and leveraging the strengths of various data sources (i.e., and adjusting for their limitations), in a way that improves these estimates is our challenge. We believe a Bayesian approach will allow us to *read between the lines* and fully capitalize on the strengths of various data sources available in Canada. Nevertheless, there is currently no gold standard approach to combining estimates from multiple data sources. As such, we must work to compare and contrast various approaches, including the triangulation approach introduced by Vigo and colleagues, which has moved this field forward.<sup>1</sup>


### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Jordan Edwards  <https://orcid.org/0000-0002-1420-3795>

### References

1. Vigo D, Jones W, Dove N, et al. Estimating the prevalence of mental and substance use disorders: a systematic approach to triangulating available data to inform health systems planning. *Can J Psychiatry*. 2022;67(2):105-114.
2. Edwards J, Pananos AD, Thind A, Stranges S, Chiu M, Anderson KK. A Bayesian approach to estimating the population prevalence of mood and anxiety disorders using multiple measures. *Epidemiol Psychiatr Sci*. 2021;30:e4.
3. Edwards J, Thind A, Stranges S, Chiu M, Anderson KK. Concordance between health administrative data and survey structured interview diagnoses for mood and anxiety disorders in Ontario, Canada. *Acta Psychiatr Scand*. 2019;141(4):385-395.
4. Burstein R, Fleming T, Haagsma J, Salomon JA, Vos T, Murray CJL. Estimating distributions of health state severity for the global burden of disease study. *Popul Health Metr*. 2015;13:31.
5. Lawlor DA, Tilling K, Smith GD. Triangulation in aetiological epidemiology. *Int J Epidemiol*. 2016;45(6):1866-1886.
6. James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet*. 2018;392(10159):1789-1858.
7. McDonald K, Ding T, Ker H, et al. Using epidemiological evidence to forecast population need for early treatment programmes in mental health: a generalisable Bayesian prediction methodology applied to and validated for first-episode psychosis in England. *Br J Psychiatry* [Internet]. 2021; <https://psyarxiv.com/bvcgu/> Accessed: 03/15/2021.
8. Lee SH, Chen W. A comparative study of uncertainty propagation methods for black-box-type problems. *Struct Multidiscip Optim*. 2009;37(3):239.
9. Alexander M, Polimis K, Zagheni E. Combining social media and survey data to nowcast migrant stocks in the United States. *arXiv*. 2020;1-36.