
The National Population Health Survey's assessment of depression risk factor associations: a simulation study assessing vulnerability to bias

S. B. Patten, MD, PhD (1,2,3)

This article has been peer reviewed.

Abstract

Background: In Canada, the major source of longitudinal information on major depression epidemiology has been the National Population Health Survey (NPHS). However, the timing of NPHS interviews may raise concerns about the quality of its estimates. Specifically, the NPHS interview assesses major depressive episodes (MDE) in the year before an interview, whereas the interviews are conducted 2 years apart. The objective of this study was to determine whether this aspect of the NPHS can be expected to introduce bias into longitudinal estimates of risk factor associations.

Methods: A simulation model was used to represent the underlying epidemiology and the expected results of a study adopting the NPHS approach to assessment of MDE. The model was used to explore the extent of the resulting distortion of estimates across a range of underlying hazard ratios.

Results: The simulations indicated that the timing and coverage of depression interviews in the NPHS would not introduce substantial bias. The model suggested that incidence would be underestimated as a result of episodes being missed, but that this would not substantially distort estimates of association.

Conclusion: The timing of interviews in the NPHS is not expected to cause biased relative risk estimates. NPHS estimates may, of course, be influenced by other sources of bias.

Keywords: *major depressive disorder, mood disorder, epidemiology, longitudinal studies, simulation, mathematical model*

Background

Improved predictions of the risk of occurrence of major depressive episodes (MDE) would help target preventive efforts and support clinical management decisions. Epidemiological data are useful for determining risk, but the literature on longitudinal studies is limited. Most psychiatric epidemiological studies have been cross-sectional and have focused

on prevalence rather than incidence. Prevalence is affected by the duration of illness and does not necessarily reflect risk.

Internationally, literature on incidence studies is beginning to emerge. Notable examples are the Netherlands Mental Health Survey and Incidence Study (NEMESIS),^{1,2} and the Dunedin Birth Cohort.³ A national source of longitudinal data in Canada is the National Population Health Survey (NPHS).⁴

The NPHS includes a diagnostic instrument for past-year MDE, the Composite International Diagnostic Interview-Short Form for Major Depression (CIDI-SFMD).⁵ However, the NPHS has certain design features that may call into question the validity of its longitudinal estimates. NPHS interviews occur every 2 years whereas the CIDI-SFMD interview covers the previous year; thus the diagnostic interview does not necessarily capture all episodes occurring between NPHS cycles. In addition, the CIDI-SFMD does not determine the timing of episodes beyond determining the presence of symptoms during the same 2-week period in the year preceding an interview.

Most studies that have used NPHS data have evaluated episode incidence as the proportion of persons without MDE in the year preceding an initial interview⁶ or in several NPHS interview cycles⁷ who then experience an MDE in the year preceding a subsequent interview. However, this cannot be precisely interpreted either as a 1-year or 2-year incidence proportion. Other studies have used proportional hazard models to evaluate incidence,⁸ but the fundamental issue of the timing of interviews remains. As a measure of annual incidence in the year preceding a follow-up interview, the CIDI-SFMD may be non-specific (since some of the episodes that have their onset earlier than the year covered by the interview may be included in the numerator of an incidence proportion); if interpreted as a measure

Author references

1. Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada
2. Department of Psychiatry, University of Calgary, Calgary, Alberta, Canada
3. Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada

Correspondence: Dr. Scott Patten, Department of Community Health Sciences, University of Calgary, 3rd Floor TRW Building, 3280 Hospital Drive NW, Calgary, AB T2N 4Z6; Tel.: (403) 220-8752/4285; Fax: (403) 270-7307; Email: patten@ucalgary.ca

of a 2-year incidence proportion, the CIDI-SFMD may be insensitive (since some of the episodes that have an onset more than 1 year prior to the second interview may resolve prior to the year preceding the interview and therefore may not be recorded).⁹ However, the actual impact of this aspect of the NPHS approach to measurement is unknown.

Had the NPHS included a reference standard measure fully assessing the course of MDE in its cohort, the impact of these design issues could be explored using real data. In the absence of such a measure, we sought to explore these issues using a simulation model designed to represent both the underlying epidemiology and the NPHS measurement strategy for MDE.

Method

To address the objectives of this study, we developed a discrete event simulation model using the software Arena version 10.¹⁰ We set up the model so that the simulation clock would cover the same duration of follow-up as was available from the NPHS at that time, from 1994 to 2006. Entities in the model represented people, members of the NPHS cohort at their time of entry into the study and in a non-depressed state (at risk of incident episodes) at that time. The simulation model depicted the underlying epidemiology by representing incidence and recovery from episodes; it simulated the experience of each entity from their time of entry in 1994 until the occurrence of an episode of depression or for a maximum of 12 years of follow-up. The model did not attempt to represent mortality. There were three possible simulation paths for each entity: (1) they could

complete 624 weeks (12 years) of simulated observation without an episode; (2) they could have an episode at a time when it would have been detected, in which case the episode was recorded as an incident case and the entity was then removed from the simulation; or (3) they could have an episode associated with an onset and recovery time that would have rendered it undetectable given the timing of the NPHS measurement strategy, in which case the entity was returned to the part of the model simulating incidence. Entities following the third pathway could then follow either of the available paths, experiencing recurrences (or not) which could then be detected (or not).

The general goal of this simulation study was to develop a representation of the epidemiology and the NPHS measurement strategy so that we could explore the extent of distortion introduced by the NPHS measurement strategy (see Figure 1). A first step was to represent recovery from MDE. This was important since longer episodes occurring in a 2-year period would be more likely to persist into the second year of this interval, potentially affecting the extent of introduced bias. The description of episode duration relied initially on some NPHS estimates [depicted (a) in Figure 1], specifically an ordinal logistic regression model describing self-reported episode durations in the NPHS cohort in relation to age (which was found to be the most important determinant of episode duration).¹¹ We used an equation representing the time-dependent pattern of recovery in different age groups in the model and calibrated its parameters to the NPHS estimates. Once recovery was depicted, the model could represent MDE incidence while also

representing the possibility that the timing of onset and recovery from an episode might result in it not being identified. For example, in order to be detected at the 1996 interview, an episode had to include at least 2 weeks of sufficient depressive symptoms in the year preceding the interview. Ratios of incidence in respondents exposed or not exposed to risk factors have various strengths of association with MDE, as they would appear in the NPHS data.

The type of equation chosen to represent incidence of MDE (and recovery from MDE) was one that could depict incidence as diminishing over time spent free of depressive episodes, as is expected clinically and as has previously been observed in the NPHS data.⁹ Equation 1 was used to calculate a linear function (LF) for each entity, using an attribute (here labelled with the non-specific term “covariate”) assigned to that entity as a value of 0 or 1 and representing a risk factor exposure:

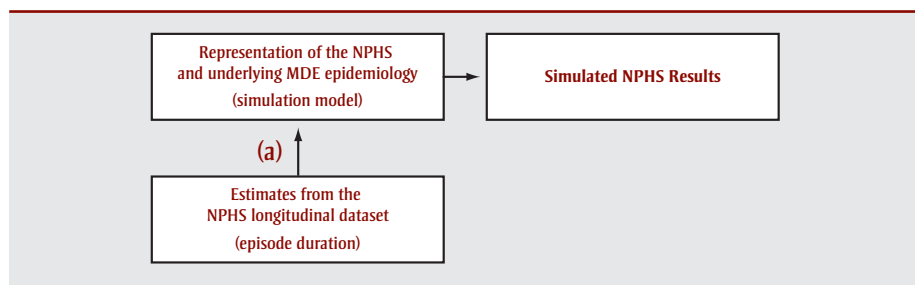
$$LF_{incidence} = \alpha + \beta_{covariate} * covariate + \beta_{\log t} * \log t$$

where t is the time in weeks. The time interval represented in the simulations were evaluated in 1-week intervals within the simulation. For the sake of simplicity, Equation 1 includes only a single covariate term, an indicator variable assuming a value of 0 or 1. Age was, however, depicted in most simulations as an attribute at five levels (12–18 years, 19–25 years, 26–45 years, 46–65 years and 66 years plus) using four separate indicator variables (such that the 12–18 years age group was the baseline category) since age is the main determinant of episode duration.¹¹ The linear function (LF) was transformed into a weekly risk (that changed with each passing week in the non-depressed state) using Equation 2:

$$\text{Weekly risk} = 1 - \exp[-\exp(LF)]$$

We used this weekly risk equation to simulate the risk of a new episode during each week of the simulation. This was represented in the simulation path by a loop, with passage around this loop corresponding to 1 week of simulation time. With each transit around the loop the time variable counted up by 1 week and the risk was recalculated. Recovery from an episode

FIGURE 1
Schematic diagram depicting the simulation approach used in the study



Abbreviations: MDE, major depressive episode; NPHS, National Population Health Survey.

was simulated using a similar approach. In order to explore the relationship of β coefficients from Equation 1 to analogous NPHS estimates, it was necessary to identify a value for α and $\beta_{\log t}$, also from Equation 1. This we did by fitting a grouped time proportional hazard model for 7029 NPHS respondents with complete data collection across all of the relevant NPHS cycles. We chose this subset because an important MDE risk factor, family history of MDE, was only evaluated in the 2004 NPHS. Table 1 compares the characteristics of this cohort with those of the entire NPHS cohort. The main difference was that the $n = 7029$ cohort was younger, probably due to attrition of elderly respondents from the original sample (e.g. due to death or institutionalization) over time. Table 2 shows the estimated hazard ratios from the proportional hazard model. The α value (-7.435) and $\beta_{\log t}$ (-0.128) parameters from Equation 1 were calibrated, using an automated procedure in Arena¹⁰ called OptQuest to identify parameter values that allowed the model output to resemble the NPHS estimates.

In reality, the NPHS interviews do not take place at a single point in time, but rather occur over the course of several months. For the sake of simplicity, this aspect of the data collection was not represented in the simulation model. Instead, each entity was generated at week 0 (baseline) and the subsequent interviews were represented as occurring at 104 weeks, 208 weeks, 312 weeks, 416 weeks, 520 weeks and 624 weeks. Tracking variables recorded the times of simulated onset and resolution of depressive episodes for each entity and tracked weeks elapsed in the simulation for each entity. The detection of episodes in the NPHS was assumed to have occurred if an entity had 2 or more weeks in the depressed state during the 52 weeks before a simulated assessment time. In this way, the model was able to represent the “under-detection” of episodes expected to occur in the NPHS and also the “over-detection” whereby episodes from previous years persisted into the year before an interview. The experience of a large number of entities was simulated each time a simulation was run, allowing evaluations of frequency distributions.

TABLE 1
Characteristics of the NPHS sample at baseline and of the respondents eligible for inclusion in the analysis

	NPHS sample at baseline (N = 13 175), %	Eligible sample (n = 7029), (%)
Sex		
Male	48.4	46.4
Female	51.6	53.6
Age (years)		
12–18	11.8	11.9
19–25	10.2	8.8
26–45	40.6	44.8
46–65	24.5	26.8
66+	13.0	7.6
Marital status		
Married/common-law	59.8	65.3
Single	28.0	25.3
Widowed/separated/divorced	12.2	9.4
Education		
High school graduation or less	47.5	42.4
Some post-secondary or higher	52.5	57.6
Income^a		
Lowest	17.5	12.6
Low/mid/high	82.5	87.4
Injuries in the past 12 months^b		
Yes	17.1	17.4
No	82.9	82.6
Chronic condition		
Yes	49.9	48.0
No	50.1	52.0
Smoking status		
Current	28.3	25.3
Former/never	71.7	74.7
Childhood stress or trauma		
Yes	47.7	48.0
No	52.3	52.0
Stress^c		
Yes	28.5	27.8
No	71.5	72.2
Mastery^d		
Low	25.5	23.5
Not low	74.5	76.5
Self-esteem		
Low	34.6	32.8
Not low	65.4	67.2
Benzodiazepine use in the past 2 days		
Yes	2.0	1.4
No	98.0	98.6
Pain		
Moderate/severe	10.8	9.7
Mild/no	89.2	90.3

Abbreviation: NPHS, National Population Health Survey.

^a The lowest income group corresponds to an income of less than \$15,000 for a household of 1–2 persons, \$20,000 for a family of 3–4 persons and less than \$30,000 for 5 or more persons.

^b Affirmative responses to the question: “In the past 12 months did you have any injuries serious enough to limit your normal activities?”

^c Upper quartile scores on a scale containing up to 16 questions concerned with ongoing sources of stress.

^d Mastery is the extent to which individuals believe that their life-chances are under their control. This was assessed in the NPHS on a scale of 7 questions.

TABLE 2
Estimated hazard ratios from the NPHS

Variable	Hazard ratios estimated directly from the NPHS data
Female	1.5
Age group, years ^a	
19–25	1.0
26–45	0.8
46–65	0.5
66+	0.3
Injury	1.3
Chronic condition ^b	1.3
Current smoking	1.3
Childhood stress ^c	1.4
Stress ^d	1.5
Mastery ^e	1.3
Family history of depression	1.6
Pain ^f	1.8
Benzodiazepine use	1.8

Abbreviations: MDE, major depressive episodes; NPHS, National Population Health Survey.

^a Baseline category is the 12–18 year age group.

^b One or more reported conditions.

^c Any one or more reported childhood traumas.

^d Upper quartile on a stress scale. The scale contained up to 16 questions concerned with ongoing sources of stress.

^e Lower quartile on a mastery scale. Mastery is the extent to which individuals believe that their life-chances are under their control and was assessed in the NPHS on a scale consisting of 7 questions.

^f Assessed using items from a scale associated with the Health Utility Index.

every day during that same 2-week period is required by the scoring algorithm, consistent with the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV).¹² In data analyses for this study (as in most other NPHS-based depression studies), the instrument was scored at the 90% predictive cut-off point.⁵ This cut-off requires endorsement of 5 of 9 specified depressive symptoms during the same 2-week period, a standard that is also broadly consistent with DSM-IV criteria. (While the approach taken by the CIDI-SFMD is consistent with DSM-IV, it should be noted that the instrument was developed using DSM-III-R data collected in the National Comorbidity Survey.⁵)

Results

Figure 2 shows durations of simulated episode for the youngest (12–18 years) and oldest (66+ years) age groups. These had the shortest and longest episodes, respectively. The curves for other age groups fell between these two. The shape of all the age-specific curves were broadly consistent with other international estimates.¹³ As described above, these curves were then included in subsequent simulations as a representation of the recovery pattern and as a means of assessing the likely impact of the NPHS measurement strategy. Figures 3 and 4 show simulated predictions (200 000 simulated entities for each data point) of what the NPHS would be expected to identify as the relative risk given underlying hazard ratios of 1 to 5, according to the model. The logarithms of these various hazard ratios were used as β coefficients for incidence in producing these simulations, as in Equation 1. Two age groups are represented: one that includes those who were aged 12 to 18 years in 1994 (see Figure 3) and one that includes those who were 66 years plus in 1994 (see Figure 4). The hazard ratios in both age groups correspond very closely. The grey line on the figures represents equivalence of the two sets of hazard ratios. How we approach measuring MDE in the NPHS appears to make a negligible difference to the relative risk estimates arising from the approach.

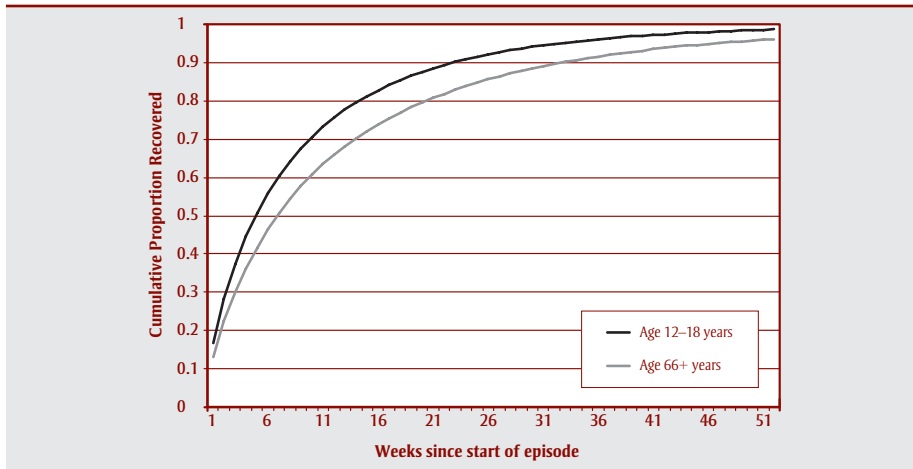
Whereas our simulations indicate that hazard ratio estimates for MDE risk factors are not likely to be biased substantially

The model was initially verified by assessing expected outputs associated with various input values. For example, the frequency of missed episodes was evaluated in relation to various episode durations. When the recovery rate from episodes was represented as being very high, for example, by entering a large value for the α coefficient in the recovery equation (see Equation 1), the proportion of entities with undetected episodes became approximately 50%; when the recovery probability was very low (e.g. a large negative value for the same coefficient), the proportion of entities with undetected episodes became 0. To calibrate the model, we created output variables representing the sum of squared differences between NPHS episode duration estimates and used the simulated output and OptQuest to identify parameter values that allowed the model output to resemble the NPHS estimates. We also identified values for the recovery equation that produced simulated episode duration frequencies resembling those predicted by the ordinal logistic regression model describing age-specific episode durations in the NPHS.¹¹

These approximate representations of episode duration were then used to explore the relationship between NPHS estimates of hazard ratios for risk factors and a series of hypothetical hazard ratios crossing a range of relevant values. The logarithm of this set of hypothetical hazard ratios was entered into the model as β coefficients in Equation 1, where the β coefficients are log hazard ratios. The output from the model (reflecting the simulated onset, resolution and measurement of episodes), in the form of risk ratios for the first 104-week risk interval, was then compared to these hazard ratios to see how closely they agreed.

The NPHS used the CIDI-SFMD to assess MDE. This is a brief, fully structured interview designed to identify probable past-year episodes. The CIDI-SFMD interview is designed for use by non-clinician interviewers and is scored with a predictive probability algorithm based on the number of symptom-based criteria fulfilled during a 2-week period in the preceding year. Either depressed mood or loss of interest or pleasure, most of the time and nearly

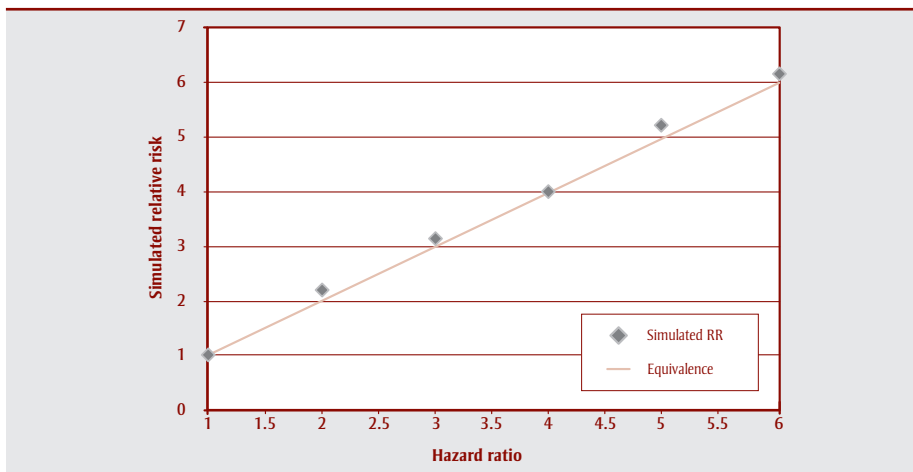
FIGURE 2
Pattern of recovery from MDE in the NPHS, in 2 age groups



Abbreviations: MDE, major depressive episode; NPHS, National Population Health Survey.

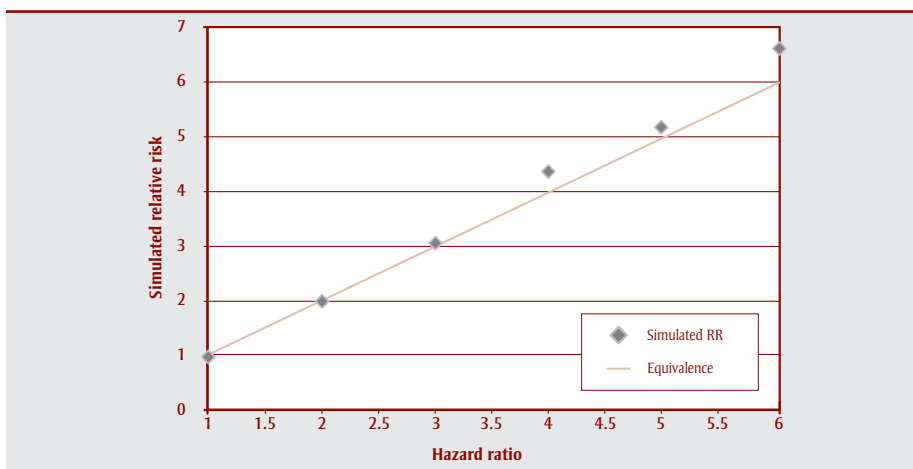
Note: The curves for the 19–25, 26–45 and 46–65 year age groups fall between those for the 12–18 and 66 plus year age groups. They are not shown to prevent clutter.

FIGURE 3
Simulated NPHS relative risks across a range of plausible hazard ratio values, 12–18 years



Abbreviations: NPHS, National Population Health Survey; RR, relative risk.

FIGURE 4
Simulated NPHS relative risks across a range of plausible hazard ratio values, 66+ years



Abbreviations: NPHS, National Population Health Survey; RR, relative risk.

by the timing of the NPHS interviews, incidence estimates are subject to bias. At strengths of association in the most relevant range (hazard ratios of 1 to 2), approximately 40% of entities with at least one MDE would not have been detected during follow-up, according to the simulations. Approximately 15% of the cases that would be regarded as incident (not depressed at one cycle, depressed at the next cycle) would have had their onset in the first year of the 2-year interval and could be considered false positives if the intention is to estimate a 1-year incidence proportion.

Conclusion

Psychiatric epidemiology is a fairly young discipline. The first of the current generation of studies (those employing standardized diagnostic procedures in representative samples) occurred in the 1980s.¹⁴ To date, the literature has been largely descriptive and mostly cross-sectional. As a result, many estimates of prevalence are available, although these estimates have not been as consistent as might have been hoped.¹⁵ A comprehensive understanding of the epidemiology of this condition will depend on longitudinal data clarifying the association of MDE incidence with various potential determinants, and associations between those determinants and the prognosis of MDE. Unfortunately, with the exception of a few international studies,^{1,16} longitudinal data are scarce. In Canada, for example, the NPHS has been the major source of information on incidence⁷ and of associations between longitudinal risk factors.⁶ Unfortunately, aspects of the NPHS study design may cast some doubt on the validity of these estimates. In this sense, the results of this simulation study are encouraging because the simulations reported here do not suggest that hazard ratio estimates from the longitudinal NPHS data are likely to be substantially biased by the design features of the study.

Limitations and strengths

A notable limitation of our study involves the way in which incidence was depicted in the model. Rather than the incidence of depressive disorders, it was necessary to simulate MDE incidence. Some of those

considered “at risk” of MDE by virtue of not having had MDE at the baseline interview may actually have had major depressive disorders in the past and their incident episodes may have been recurrences of those disorders. The pattern of declining incidence over time represented by Equation 1 may be due partially to the gradual removal of those at highest risk from the population as they experience episodes.

Additional factors may, of course, affect the validity of estimates arising from the NPHS. The CIDI-SFMD is a brief diagnostic interview that does not include all of the detailed symptoms covered by the full CIDI. The CIDI-SFMD does not include probes for carefully delineating the duration and severity of each symptom or for distinguishing between organic and non-organic etiology of symptoms.^{17,18} Misclassification bias arising from measurement error associated with the CIDI-SFMD (unrelated to the timing of its administration) may also distort estimates arising from the NPHS. Another important issue is that of attrition over time due to factors such as loss to follow-up, mortality and institutionalization; if such attrition depends on the outcome (MDE) in a way that differs with respect to risk factor exposures, bias may result. The estimates used in this study arose only from respondents with complete data collection, which is pertinent to the question of the validity and generalizability of the estimates. As the NPHS estimates used in the project arose from a subset of the longitudinal cohort (those providing complete data over seven cycles) the results may not be generalizable to the population. The simulation methods used here were intended to address one specific concern about the NPHS—the timing of its interviews in relation to its measurement of depression. The results can reassure us on this specific point.

Acknowledgements

The author is a Senior Health Scholar with the Alberta Heritage Foundation for Medical Research. Dr. Scott Patten has no conflicts of interest related to this project. He has received consulting fees and has served on a Data Safety Monitoring Board for Cipher Pharmaceuticals, and has also received

consulting and speaking fees from Servier Canada Inc. and Lundbeck Canada. This project was funded by a grant from the Canadian Institutes of Health Research. This analysis is based on data collected by Statistics Canada. However, the results do not reflect the opinions of Statistics Canada.

References

1. Bijl RV, van Zessen G, Ravelli A, de Rijk C, Langendoen Y. The Netherlands Mental Health Survey and Incidence Study (NEMESIS): objectives and design. *Soc Psychiatry Psychiatr Epidemiol.* 1998;33:581-6.
2. Bijl RV, De Graaf GR, Ravelli A, Smit F, Vollebergh WA; Netherlands Mental Health Survey and Incidence Study. Gender and age-specific first incidence of DSM-III-R psychiatric disorders in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol.* 2002 Aug;37(8):372-9.
3. Moffitt TE, Caspi A, Taylor A, Kokaua J, Milne BJ, Polanczyk G, et al. How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. *Psychol Med.* 2010 Jun;40(6):899-909.
4. Swain L, Catlin G, Beaudet MP. The National Population Health Survey—its longitudinal nature. *Health Reports.* 1999;10(4):69-82.
5. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen HU. The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF). *Int J Methods Psychiatr Res.* 1998;7:171-85.
6. Beaudet MP. Psychological health—depression. *Health Rep.* 1999;11:63-75.
7. Wang J, Williams J, Lavorato D, Schmitz N, Dewa C, Patten SB. The incidence of major depression in Canada: the National Population Health Survey. *J Affect Disord.* 2010;123:158-63.
8. Patten SB, Williams JV, Lavorato DH, Eliasziw M. Allergies and major depression: a longitudinal community study. *Biopsychosoc Med.* 2009;3:3.
9. Patten SB, Lee RC. Refining estimates of major depression incidence and episode duration in Canada using a Monte Carlo Markov model. *Med Decis Making.* 2004;24:351-8.
10. Arena simulation software. Version 10. Sewickley (PA): Rockwell Software Inc.; 2006 [accessed 2011 Sep]. Available from: <http://www.arenasimulation.com>
11. Patten SB, Wang JL, Williams JV, Lavorato DH, Khaled SM, Bulloch AG. Predictors of the longitudinal course of major depression in a Canadian population sample. *Can J Psychiatry.* 2010;55(10):669-76.
12. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th rev. ed. (DSM-IV-TR). Washington: American Psychiatric Association; 2000.
13. Vos T, Haby MM, Berendregt JJ, Kruijshaar M, Corry J, Andrews G. The burden of major depression avoidable by longer-term treatment strategies. *Arch Gen Psychiatry.* 2004;61:1097-103.
14. Regier DA, Myers JK, Kramer M, Robins LN, Blazer DG, Hough RL, et al. The NIMH Epidemiologic Catchment Area program. Historical context, major objectives, and study population characteristics. *Arch Gen Psychiatry.* 1984;41:934-41.
15. Narrow WE, Rae DS, Robins LN, Regier DA. Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Arch Gen Psychiatry.* 2002;59:115-23.
16. Eaton WW, Anthony JC, Gallo J, Cai G, Tien A, Romanoski A, et al. Natural history of Diagnostic Interview Schedule/DSM-IV major depression. The Baltimore Epidemiological Catchment Area follow-up. *Arch Gen Psychiatry.* 1997;54:993-9.
17. Patten SB, Brandon-Christie J, Devji J, Sedmak B. Performance of the Composite International Diagnostic Interview Short Form for Major Depression in a community sample. *Chronic Dis Can.* 2000;21:68-72.
18. Patten SB, Wang JL, Beck CA, Maxwell CJ. Measurement issues related to the evaluation and monitoring of major depression prevalence in Canada. *Chronic Dis Can.* 2005;26:100-6.