**ARTICLE DETAILS**

| TITLE (PROVISIONAL) | Berkson's bias in biobank sampling in a specialized mental health care setting: a comparative cross-sectional study |
|---------------------|------------------------------------------------------------------------------------------------------------------|
| AUTHORS             | Laliberté, Vincent; Giguère, Charles-Edouard; Potvin, Stephane; Lesage, Alain |

**VERSION 1 – REVIEW**

| REVIEWER             | Cecilia Björkelund University of Gothenburg Sahlgrenska Academy, Primary Health Care/School of Public Health and Community Medicine |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------|
| REVIEW RETURNED      | 20-Mar-2020                                                                                                                     |
| GENERAL COMMENTS     | Thank you for the possibility to review the manuscript" Berkson's bias in biobank sampling in a specialized mental health care setting" for BMJOpen. It’s an interesting manuscript and very relevant not only concerning the mental health field, but also concerning many other diagnose domains. Further, It’s relevant as much of the international literature draws conclusions of e.g. morbidity and mortality data received from hospital based populations and draw conclusions not valid for the whole group of depressed people most of which are seen in primary care.. I have the following comments 1. Page 7 line 15: “Depression is a common chronic disease that affects about 5% of the population yearly and 10% over the lifetime.” It’s important to specify the affection rate of women and men respectively, as there is rather a great difference both concerning yearly incidence and over lifetime. 2. The participation frequency of CCHS is not specified 3. Page 8 line 43: Clinical diagnoses in CCHS were made during the survey in the CCHS using variables derived from the CIDI. However, at page 10 line 24 you refer that it were only those individuals who were retained who reported an episode of major depression in the preceding year. This means, that it was not only those who were diagnosed with depression via CIDI who were included, but also individuals who stated depression during last year, and thus could have recovered from depression in the CCHS interview. Please comment and explain. 4. You excluded psychotic patients – but also severe depression includes psychotic phases. 5. You included only primary diagnosis of depressive disorder – please explain what this implies. |
6. Please explain the implication of the division of physical abuse into three categories: a) slapped on face, head or ears, hit or spanked; b) pushed, grabbed, shoved, something thrown at; and c) kicked, bit, punched, choked, burned, attacked.

7. Page 10: All numeral results that are presented in Table 2 do not need to be repeated in the text with figures of OR and CI.

8. Table text in Table 2 needs to be expanded and more informative.

9. Table 2: Please either present explanations on estimate, intercept etc, or exclude these rows. Further, either present p-values or ORs and CI (preferably ORs and CI).

10. In Fig 1, all "n"s should not be capital.

11. Fig 2: This is an illustrative way of showing the results, but it should be explained where the "Not depressed" part emanates from. All individuals included – both from CCHS (n=860) and Signature Bank were diagnosed with depression?

12. In Discussion, the authors should comment on how including any socioeconomic variable in the analyses could influence on the results.

---

**REVIEWER**
Herika Cristina da Silva
Federal University of Rio de Janeiro, Brazil

**REVIEW RETURNED**
12-Apr-2020

**GENERAL COMMENTS**
The study approaches a relevant subject regarding the interference of Berkson’s bias in the study of etiological pathways of depression through biobanks. However, the study doesn’t clearly present the information already existent in the literature about this theme and what are the gaps, what hasn’t been yet answered in this subject. Therefore, the relevance of this study for the literature isn’t efficiently presented. This must be presented in the introduction. And, in the discussion, the results found must be compared to the similar ones, already published in the field. And, if there are no other papers like this one, this needs to be explicit.

Below, there are some comments with suggestions for edits:

**Introduction**
1st paragraph: Are these the only available data in the psychiatry field? Is that what is most relevant and impactful in the field? I suggest presenting, concisely, more information about the psychiatry field in general and then focusing on depression. Present the most important data already available in the literature.

2nd paragraph: This paragraph can be summarized in one sentence and added to the first one, before mentioning the psychiatry field, to put Berkson’s bias into context.

3rd paragraph: This paragraph is related to the relevance of the paper. I suggest placing it after the paragraph of depression and childhood abuse, so that all concepts related to the study are presented before presenting the relevance of the study.

4th paragraph: I suggest that all the data about depression is presented before making the connection with childhood abuse.
as an etiological factor, presenting in general and summarized the important data about this connection.

5th paragraph: Present what the literature is lacking in this subject. What are the gaps to be filled? Then, present the objective.

Results
The results include data about individuals without depression, as displayed in Figure 2.
Who are these individuals, since the sample only contains people with depression? It has been described in the method that individuals without the diagnosis of depression had been excluded from the study.

Discussion
In the first paragraph it’s said “Depressed individuals who were hospitalized for psychiatric treatment or saw a psychiatrist represented less than one-fifth of the individuals with depression in the 2012 CCHS population. This is in line with previous Canadian and U.S. mental disorder population surveys that showed that half of respondents with depression reported not seeing any health professionals”. However, the numbers found in the present study are significantly lower than on compared studies, so they can’t be considered similar results. This could be discussed.
In the discussion, individuals without depression are also mentioned and the same questions brought up on results session apply here.
Furthermore, it’s necessary to dialogue and compare this study’s results with the ones of similar studies, already existing in the literature. In case there are none, it needed to be mentioned, as this study becomes even more important.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)’ Comments to Author:

Reviewer: 1

Reviewer Name
Cecilia Björkelund

Institution and Country
Primary Health Care, School of Public Health and Social Medicine
Institute of Medicine, University of Gothenburg
Sweden

Please state any competing interests or state ‘None declared’:
None declared

Please leave your comments for the authors below
Thank you for the possibility to review the manuscript” Berkson’s bias in biobank sampling in a specialized mental health care setting” for BMJ Open.
It’s an interesting manuscript and very relevant not only concerning the mental health field, but also concerning many other diagnose domains. Further, it’s relevant as much of the international literature draws conclusions of e.g. morbidity and mortality data received from hospital based populations and draw conclusions not valid for the whole group of depressed people most of which are seen in primary care.

Thank you very much for your appreciation of our work.

I have the following comments

1. Page 7 line 15: “Depression is a common chronic disease that affects about 5% of the population yearly and 10% over the lifetime.”
   It’s important to specify the affection rate of women and men respectively, as there is rather a great difference both concerning yearly incidence and over lifetime.

   We changed the information to reflect the lifetime prevalence for both gender, on p.5:

   “Major depressive disorder is a common chronic disease affecting 4.4 – 5.9% of Canadian men and 11.4 – 11.5 % of Canadian women in their lifetime [11,12].”

2. The participation frequency of CCHS is not specified

   The participation frequency of 2012 CCHS was added and the reference added, on p.7:

   “The response rates (combining household and person) was 68.9% for the 2012 CCHS [33].”

3. Page 8 line 43: Clinical diagnoses in CCHS were made during the survey in the CCHS using variables derived from the CIDI. However, at page 10 line 24 you refer that it were only those individuals who were retained who reported an episode of major depression in the preceding year. This means, that it was not only those who were diagnosed with depression via CIDI who were included, but also individuals who stated depression during last year, and thus could have recovered from depression in the CCHS interview. Please comment and explain.

   Thank you very much for your question.

   In both case (in the survey and in our study) we refer to the same thing: a diagnosis of depression in the preceding 12 months.
For sake of extra clarity, we emphasized the timespan in the sentence that discusses CIDI, p.8:

"The presence of these diagnoses in the last year was ascertained during the survey using variables derived from the Composite International Diagnostic Interview [34]."

4. You excluded psychotic patients – but also severe depression includes psychotic phases.

This is true. We chose to exclude psychotic patients, which included individual suffering from depression with psychotic features. There are different reasons why we chose to exclude psychosis as part of the research design of our study. First, we aim at studying common mental disorder in both specialized setting and primary health care. Psychotic disorder belongs to the “severe” mental disorder categories. Furthermore, the CCHS survey was also meant to report common mental disorders. There was only one question asked about the presence of psychosis. This would not have been sufficient to allow for meaningful comparison.

We therefore added the following two sentences, on p. 9:

“We excluded psychosis and mania since our goal is to study common mental disorders in both specialized setting and primary health care. Furthermore, the CCSS survey was also primarily designed to report common mental disorders.”

5. You included only primary diagnosis of depressive disorder – please explain what this implies.

By this, we simply mean that we took individuals who received a diagnosis of depression by the ER admitting clinician. We still took them if the diagnosis of depression was thought to be a consequence of substance use or of a personality disorder. We only excluded participants who received a diagnosis of psychosis or bipolar disorder, in addition to the diagnosis of depression.

For sake of clarity, we removed the word “primary”, on p. 10 and figure 1:

“Of the 1,073 remaining, we retained only the 353 who had received a primary diagnosis of depressive disorder or mood disorder NOS. We were left with 207 after eliminating those with comorbid bipolar disorder or psychotic symptoms.”

6. Please explain the implication of the division of physical abuse into three categories: a)
slapped on face, head or ears, hit or spanked; b) pushed, grabbed, shoved, something thrown at; and c) kicked, bit, punched, choked, burned, attacked.

The main reasons why the Childhood Experience of Violence Questionnaire uses these categories is that childhood mistreatment was largely underreported “because of failure in its detection and recognition”. The solution has been to ask children directly about their specific experience of victimization in words that are easily understandable, instead of asking more generally about “abuse”. The authors of the questionnaire do not specify their rationale for dividing mistreatments in these three categories, but we follow the mode of reporting of fellow Canadian researchers and our recent paper.2,3

1. Walsh CA et al. (2008) Measurement of victimization in adolescence: development and validation of the Childhood Experiences of Violence Questionnaire. *Child Abuse and Neglect*. 32(11):1037-1057.
2. Afifi TO (2014) Child abuse and mental disorders in Canada. *CMAJ*, 186(9):E324-E332.
3. Ligier F, Giguère CE, Séguin M, Lesage A. (2019) Survey evidence of the decline in child abuse in younger Canadian cohorts. *Eur J Pediatr*. Sep;178(9):1423-1432.

7. Page 10: All numeral results that are presented in Table 2 do not need to be repeated in the text with figures of OR and CI.

The numeral results were removed from the text and now only appear in Table 2.

8. Table text in Table 2 needs to be expanded and more informative.

The title was changed to be more informative, p.24:

List of independent predictors of the variable “Any Child Abuse” from the Childhood Experience of Violence Questionnaire calculated through a regression analysis.

9. Table 2: Please either present explanations on estimate, intercept etc, or exclude these rows. Further, either present p-values or ORs and CI (preferably ORs and CI).

Estimate and Intercept were removed from the Table. We also removed p-values as you recommended.

10. In Fig 1, all “n”s should not be capital.
Figure 1 was changed accordingly.

11. **Fig 2**: This is an illustrative way of showing the results, but it should be explained where the “Not depressed” part emanates from. All individuals included – both from CCHS (n=860) and Signature Bank were diagnosed with depression?

Thank you for this comment.

It is possible to determine who exactly are the “not depressed” group, based on Figure 1. After we excluded participants according to age, we were left with 21,506 participants in the CCHS and 1,073 participants in the Signature Bank. Then, after we removed participants who did not report major depression in the past year, there were 1,162 participants left in CCHS and 353 participants left in Signature Bank. The “not depressed” corresponds to the (21,506 – 1,162 = 20,344) participants in CCHS and the (1,073 – 353 = 720) participants in Signature Bank, for a total of (20,344 + 720 = 21,064).

We added the following sentence on page 12:

“The “not depressed” group corresponds to the 20,344 (21,506 – 1,162) participants in CCHS and the 720 (1,073 – 353) participants in Signature Bank that had the appropriate age but did not have depression. They are represented in figure 2 as a reference value, but not included in the multivariate analyses.”

12. In Discussion, the authors should comment on how including any socioeconomic variable in the analyses could influence on the results.

In many countries, coming from a lower socioeconomic background implies lower access to care. However, this is not necessarily true in Canada, given the universal health coverage. Previous work showed that individuals coming from lower economic background had higher rates of mood disorders, psychiatric consultation and emergency visits.\(^1\) We did not expect that including socioeconomic variable would influence the results, but we added this as a limitation to our study, p.15:

“Finally, our data did not allow us to take into consideration socioeconomic variables in our analysis. However, we think that such variables would not have influenced our results, since the universal coverage in Canada provides access to care to everyone. Previous work showed that individuals coming from lower economic background have higher rates of mood disorders, but also higher rates of services utilization (psychiatric consultation and emergency visits) [46]”
The study approaches a relevant subject regarding the interference of Berkson's bias in the study of etiological pathways of depression through biobanks. However, the study doesn't clearly present the information already existent in the literature about this theme and what are the gaps, what hasn't been yet answered in this subject. Therefore, the relevance of this study for the literature isn't efficiently presented. This must be presented in the introduction. And, in the discussion, the results found must be compared to the similar ones, already published in the field. And, if there are no other papers like this one, this needs to be explicit.

Below, there are some comments with suggestions for edits:

Introduction

- 1st paragraph: Are these the only available data in the psychiatry field? Is that what is most relevant and impactful in the field? I suggest presenting, concisely, more information about the psychiatry field in general and then focusing on depression. Present the most important data already available in the literature.

Thank you for your comment. We considerably reworked on the first paragraph of the introduction, that now discusses three other articles that examine the impact of the Berkson's bias in psychiatry, p.5:

Berkson's bias or "paradox" is a selection bias whereby a factor associated with a study's sampling framework gives rise to an etiological association with the dependent variable of interest [1]. Joseph Berkson, who first pointed out this bias, identified the role of hospital sampling in the association between two conditions: cholecystitis and diabetes [1]. Berkson's bias was later revealed in other associations, including between respiratory disease and "disease of the bones and organs of movement" [2], multiple myelomas and the elderly [3], and bladder cancer and smoking [4], as well as in the comorbidity of migraine and hypertension [5]. In the field of psychiatry, the Berkson's bias was shown to play a role the associations of positive and negative symptoms that we assume to constitute schizophrenia, since each of them are associated independently with mental health care use [6]. The same would be true of the association between manic and depressive symptoms in bipolar disorder [7]. Considering this bias may therefore provide support to consider the value of a dimension-specific approach to ascertain the aetiology of psychosis and mood dysregulation, based on a "continuum hypothesis" [8]. The Berkson's bias also helped researchers show that depressed
patients treated in specialized care settings are more strongly associated with parental loss and separation, compared with patients treated in primary care settings [9,10].

- 2nd paragraph: This paragraph can be summarized in one sentence and added to the first one, before mentioning the psychiatry field, to put Berkson’s bias into context.

This paragraph was summarized and included in the first paragraph (see above).

- 3rd paragraph: This paragraph is related to the relevance of the paper. I suggest placing it after the paragraph of depression and childhood abuse, so that all concepts related to the study are presented before presenting the relevance of the study.

Done

- 4th paragraph: I suggest that all the data about depression is presented before making the connection with childhood abuse as an etiological factor, presenting in general and summarized the important data about this connection.

Done

- 5th paragraph: Present what the literature is lacking in this subject. What are the gaps to be filled? Then, present the objective.

Thank you for this comment. We added a sentence that clearly summarize what the literature is lacking and what gaps we are planning to fill, p.6:

“To our knowledge, although the presence of a Berkson’s bias in biobanks was considered by different researchers [27–29], it was never tested empirically in any field of medicine, including psychiatry.”

Results
The results include data about individuals without depression, as displayed in Figure 2.

- Who are these individuals, since the sample only contains people with depression? It has been described in the method that individuals without the diagnosis of depression had been excluded from the study.

Thank you for this comment, more information about the “not depressed” group will indeed add clarity to the paper.
Figure 1 makes it possible to determine who exactly are included in the “not depressed” group. After we excluded participants according to age, we were left with 21,506 participants in the CCHS and 1,073 participants in the Signature Bank. Then, after we removed participants who did not report major depression in the past year, there were 1,162 participants left in CCHS and 353 participants left in Signature Bank. The “not depressed” corresponds to the (21,506 – 1,162 = 20,344) participants in CCHS and the (1,073 – 353 = 720) participants in Signature Bank, for a total of (20,344 + 720 = 21,064).

We added the following sentence, on page 12:

“The “not depressed” group corresponds to the 20,344 (21,506 – 1,162) participants in CCHS and the 720 (1,073 – 353) participants in Signature Bank that had the appropriate age but did not have depression. They are represented in figure 2 as a reference value, but not included in the multivariate analyses”

Discussion

- In the first paragraph it’s said “Depressed individuals who were hospitalized for psychiatric treatment or saw a psychiatrist represented less than one-fifth of the individuals with depression in the 2012 CCHS population. This is in line with previous Canadian and U.S. mental disorder population surveys that showed that half of respondents with depression reported not seeing any health professionals”. However, the numbers found in the present study are significantly lower than on compared studies, so they can’t be considered similar results. This could be discussed.

Thank you for pointing our attention at this area of unclarity in our manuscript. In fact, we meant to say that the results are similar to what is found in the literature: the majority of depressed individual do not see a psychiatrist.

In CCHS, less than one-fifth saw a psychiatrist. In the study we refer too, half of the respondent did not see any health professional. Therefore, it is fair to think that it is only a minority of individuals who saw a psychiatrist.

To remove the confusion, we now mention the following, p.13:

“This is in line with previous work that shows that only a minority of individuals with depression is seen by a psychiatrist [14][39]."
- In the discussion, individuals without depression are also mentioned and the same questions brought up on results session apply here.

Yes, see response above regarding individuals without depression.

Furthermore, it’s necessary to dialogue and compare this study’s results with the ones of similar studies, already existing in the literature. In case there are none, it needed to be mentioned, as this study becomes even more important.

We clarified this point the following way, p.14:

“This is the first study that provides empirical evidence of the presence of a Berkson’s bias in a mental disorders biobank or any similarly constructed biobank where patients are recruited in specialist health care settings. The presence of this bias has implications for the etiological links established using these biobanks for common chronic diseases (e.g., diabetes, hypertension, depression) generally treated in primary care.”

**VERSION 2 – REVIEW**

| REVIEWER | Cecilia Björkelund  
|          | Primary Health Care, School of Public Health and Community Medicine, Institute of Medicine, University of Gothenburg, Sweden |
| REVIEW RETURNED | 14-May-2020 |

**GENERAL COMMENTS**  
I think the authors have answered my questions in a good way. My only concern is the ethics declaration - declaration of ethics approval is lacking - only participant consent is declared.

| REVIEWER | Herika Silva  
|          | Federal University of Rio de Janeiro (Universidade Federal do Rio de Janeiro - UFRJ), Brazil |
| REVIEW RETURNED | 25-May-2020 |

**GENERAL COMMENTS**  
The authors took into account the notes made by the reviewers, but there are still some points to be worked on. Below, there are some comments with suggestions.

The new title contains two colons. I suggest removing the “original research” part.

Introduction  
1st paragraph:  
I suggest removing some of the associations related to the Berkson bias and mentioning just two, it is enough and avoids getting too long.  
Ex: Joseph Berkson, who first pointed out this bias, identified the role of hospital sampling in the association between two
conditions: cholecystitis and diabetes [1]. Later, it was revealed in other associations, such as... (cite just two).

Then, when begins the part of psychiatry, I suggest removing this phrase "Considering this bias may therefore provide support to consider the value of a dimension-specific approach to ascertain the aetiology of psychosis and mood dysregulation, based on a “continuum hypothesis” [8]." And connect the other remaining phrases, continuing with "Also, the Berkson’s bias helped researchers show that patients WITH DEPRESSION treated in specialized care settings ..." Change depressed patients by patients with depression can help with the connection between paragraphs if the second paragraph begins with “Depression is...”.

Results
The results still include data about individuals without depression, as displayed in Figure 2.
If individuals without depression were excluded from the study, they cannot be considered in the tables, figures, and results, unless it is to show the sample selection flowchart. Otherwise, they are participating in the study sample and the method must be corrected.

| VERSION 2 – AUTHOR RESPONSE |
|----------------------------|
| Reviewer(s)’ Comments to Author: Reviewer: 1 |
| Reviewer Name: Cecilia Björkelund |
| Institution and Country: Primary Health Care, School of Public Health and Community Medicine, Institute of Medicine, University of Gothenburg, Sweden |
| Please state any competing interests or state 'None declared': None declared |
| Please leave your comments for the authors below |
| I think the authors have answered my questions in a good way. My only concern is the ethics declaration - declaration of ethics approval is lacking - only participant consent is declared. |

Thank you for your comment. On page 10 of the manuscript, we wrote: “The study was approved by the research ethics board of the Centre intégré universitaire de santé et de services sociaux de l’Est-de-l’Île-de-Montréal (Project no. 2017-809).” We also added this sentence in the “Ethical Standards” section.
Name Herika Silva

Institution and Country
Federal University of Rio de Janeiro (Universidade Federal do Rio de Janeiro - UFRJ), Brazil

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
The authors took into account the notes made by the reviewers, but there are still some points to be worked on.
Below, there are some comments with suggestions.

The new title contains two colons. I suggest removing the “original research” part.

"Original Research” was removed.

Introduction 1st paragraph:
I suggest removing some of the associations related to the Berkson bias and mentioning just two, it is enough and avoids getting too long.
Ex: Joseph Berkson, who first pointed out this bias, identified the role of hospital sampling in the association between two conditions: cholecystitis and diabetes [1]. Later, it was revealed in other associations, such as... (cite just two).

Thank you, we did the modifications as suggested, p. 5:

Later, Berkson's bias was revealed in other associations, such as between respiratory disease and "disease of the bones and organs of movement" [2], [3], as well as between bladder cancer and smoking [4][5].

Then, when begins the part of psychiatry, I suggest removing this phrase “Considering this bias may therefore provide support to consider the value of a dimension-specific approach to ascertain the aetiology of psychosis and mood dysregulation, based on a “continuum hypothesis” [8].” And connect the other remaining phrases, continuing with “Also, the Berkson’s bias helped researchers show that patients WITH DEPRESSION treated in specialized care settings ...” Change depressed patients by patients with depression can help with the connection between paragraphs if the second paragraph begins with “Depression is...”.

Thank you for these suggesting those reformulations. Changes were made accordingly, p. 5:
Also, the Berkson’s bias helped researchers show that patients with depression treated in specialized care settings are more strongly associated with parental loss and separation, compared with patients treated in primary care settings [9,10].

Depression is a common chronic disease affecting 4.4 – 5.9% of Canadian men and 11.4 – 11.5 % of Canadian women in their lifetime [11,12].

Results

The results still include data about individuals without depression, as displayed in Figure 2. If individuals without depression were excluded from the study, they cannot be considered in the tables, figures, and results, unless it is to show the sample selection flowchart. Otherwise, they are participating in the study sample and the method must be corrected.

In figure 2, individuals without depression from the Canadian community survey were indeed added to show the sample selection and reference group values. We added the following sentence in the method section, p.10:

To reference the prevalence of child abuse depending on mental health services use for people with depression, we also computed the prevalence of child abuse among people without depression in the CCHS, from all participants who were under 20 years and over 80 years, and reported it in Figure 2 only.

And we changed 95% CI prevalence rates for each type of childhood trauma by hierarchical use of mental health services for groups of people with depression (for Canadian Community Health Survey and Signature groups combined); and, for reference, individuals without depression from the Canadian Community Health Survey.
