PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The prevalence of common mental health disorders in adults who are high or costly users of health care services: Protocol for a systematic review and meta-analysis
AUTHORS	Jadhakhan, Ferozkhan; Lindner, Oana; Blakemore, Amy; Guthrie, Elspeth

VERSION 1 – REVIEW

REVIEWER	Iliana Kohler University of Pennsylvania
REVIEW RETURNED	11-Jan-2019

GENERAL COMMENTS	This is a nicely developed protocol to conduct a systematic review on the prevalence of depression and anxiety among high cost users aged 18+ yrs of health care services. A major concern I have is related to the relevance of the proposed study. I urge the authors to develop and outline the study's relevance in a much better way than in the current version of the manuscript: for instance, are there any other studies that have conducted a systematic review like this? If yes, how does the proposed systematic review differ from prior studies? What is the innovation of the present study? Also, why should the readers care about this topic in general? What is the public health relevance? Why do we need this study? Some of these aspects have been mentioned in the discussion and implications of results sections of the protocol, but they are not particularly convincing and need to be better developed, and also sketched in the introduction. See also my specific comments on these parts of the protocol. Clarify the specific dates of the study: the protocol says studies from whenever published on the topic to June 2018. First, why stop in June 2018? Second, I am somewhat concerned with going very far back in the past since this may affect the quality of the data and conclusions. We know that the diagnosis of depression/anxiety has improved over time so very old studies may underestimate their prevalence and or costs associated with them among high health care users. How would this impact the proposed study and especially if the authors conduct a meta-analysis? I suggest that the authors explicitly consider the difference in costs related to managed versus unmanaged depression/anxiety.

High health care costs can also be due to the management/costs because of co-morbidity, but also because of managed depression. In other words, there are many pathways how high health costs can be accumulated and associated with the presence of depression (e.g., present co-morbidity, different types of co-morbidity that can be also associated with different levels of depression/anxiety, managed vs. not-managed depression/anxiety, etc.) and all these different pathways can results in different health care costs. It is not clear that the systematic review does consider these different pathways. I urge the authors to think more about these aspects and differentiate them in their classification and indicators.

The authors correctly include categories for depression/anxiety diagnosed via a standardized questionnaire versus clinical interview. However, there are several issues here related with the estimates of the prevalence of depression and anxiety: first, the 2 approaches can results in different estimates and this bias the results; why not focus primarily on studies that are based on standardized instruments and then as part of the analytical approach propose sensitivity analysis that consider clinical interviews?

Second, there are multiple depression/anxiety instruments that are used in different settings and that result in different estimates of the presence/prevalence of both conditions. Some of these instruments have been clinically validated, others not, some have been clinically validated only in specific settings (e.g., countries, languages, population, clinical settings versus non-clinical, etc). The proposed protocol and criteria does not outline any categories reflecting these differences and does not discuss how to deal with these issues and their implications on the analysis/results. Please address these issues.

Why not consider estimating the health care costs associated with depression/anxiety by location? It is well known that health care costs differ substantially between European countries and the US, and there are also systematic regional/country-specific differences in how the diagnosis of depression/anxiety is handled. These differences are not considered in the proposed protocol.

How is high use of health services defined? Is there a general agreement on this? Does the definition change over time and how does this impact the protocol study?

Similarly, what is the definition of high health costs? Does the definition change over time and how does this impact the protocol study?

- Aim 2, p. 5 – I am not sure how based on the proposed analytical approach the authors can determine the magnitude of health care use and costs of depression;

– p. 6, line 22 – add “among high users”

--p. 6 – criteria fir exclusion listed here are not listed later in the tables summarizing exclusion criteria; Also explain ration for excluding these settings and what is gained from focusing on general health settings

	p. 7 – the authors mention that they will include previous meta-analysis on this topic; this brings me back to what I mentioned before – how does the proposed study and meta-analysis differ from prior research? p. 7, line 2 – studies in all languages will be considered: I am not convinced that the authors will be able to do this and make all these translations. p.8, line 23 – what is "recent" review article (which time period does recent refer to)? p.6 – the comparator's criteria are not clear. These will be studies also focused on depression/anxiety? Or as stated also studies without depression/anxiety? How many groups and which exactly are compared here: why not listing the groups specifically? p.9, line 29 – how will be the qualification of the use and costs developed? Can you outline this more precisely? p. 11, lines 6-11 – I am not convinced that by doing the proposed analysis all these issues listed in the conclusions can be indeed addressed. p.4, appendix – include categories for scale/instruments used to measure depression and anxiety – confounders on p.4 to be specified --p. 3, appendix – location of study – need classification (region, type of health care facility, rural/urban, etc.)

REVIEWER	Lucinda Leung, MD, PhD Assistant Professor of Medicine, UCLA David Geffen School of Medicine Staff Physician, Division of General Internal Medicine, West Los Angeles VA Medical Center Core Investigator, VA HSR&D Center for the Study of Healthcare Innovation, Implementation, & Policy (CSHIIP) 11301 Wilshire Blvd (111G), Los Angeles 90073
REVIEW RETURNED	29-Jan-2019

GENERAL COMMENTS	Thank you for the opportunity to review. The study protocol proposes to answer an important research question regarding the prevalence and magnitude of depression/anxiety among high-utilizing adults (i.e., #patients with depression/anxiety [numerator]/ total # of high-cost patients [denominator]). The reviewer notes some areas for clarification regarding the methods and research implications. 1. Homelessness is a crucial risk factor for high utilization and should be included in the list on Page 5, line 27. 2. The reviewer has difficulty understanding the study methods proposed to answer this question (Page 6, Lines 6-23). The study cohort will include adults who are high-utilizers of health care and who "have been evaluated through standardized questionnaires or clinical interviews." Please clarify how will this be determined, since you are not planning to include individuals with psychiatric diagnostic codes (Page 6, Line 15). To the reviewer's

	knowledge, there are few health care systems from where the authors are can easily obtain PHQ-9 scores or clinical notes at the population level in order to draw valid conclusions. It would seem likely that the authors need to use diagnostic codes or procedure codes to make inferences about depression/anxiety diagnoses. 3. Please include reason for excluding specialty med-surg departments (Page 6, Line 25). Pts with depression/anxiety are known to overutilize in these departments as well, due to somatic complaints necessitating excessive specialty workup. 4. Suggested subgroup analyses are perplexing. Related to my comment in #2 above, Page 9, Line 43 highlights that subgroup analysis #1 would include “type of outcome measurement for depression/anxiety” but it remains unclear what this data source is. Please clarify if using diagnostic codes, chart review, etc. In Page 9, Line 45, please explain how patients would be assigned to a specific healthcare setting and why there would be suspected differences by healthcare setting. Are you comparing patients that ONLY get there care in acute care settings (ER, hospital) to those that get there care in primary care? How would you handle high-utilizing patients that have primary care AND ER visits, for example? 5. Authors should be cognizant of existing literature when discussing research implications. Re (Page 11, Lines 6-10) “It may also reveal an unmet need in the diagnosis and appropriate management of anxiety/depression in populations routinely seen in general health care settings, ” this is already firmly established in the reviewer’s opinion; thus, it may be a bigger contribution to “confirm and clarify the extent of unmet need…” Also, Re (Page 7, Lines 15-22) “it could suggest the type of integrated, collaborative services, or management methods…”, this is also well known and documented in the literature (Archer et al, Cochrane, 2012; Coventry et al, BMJ, 2015), so I would suggest that your findings may “further support” collaborative care model implementation.

REVIEWER	Luc Jansen Erasmus Medical Center, Rotterdam, the Netherlands
REVIEW RETURNED	07-Mar-2019

GENERAL COMMENTS	Dear Authors, Thanks for giving me the opportunity to review your study protocol I believe the aim of your study is very relevant. Indeed only a few patients can be accounted for a large proportion of healthcare use and costs. Some studies have shown that frequent healthcare users often suffer from psychiatric comorbidity. I believe this systematic review can provide some important information but I have some concerns about the search strategy and inclusion criteria. 1. It is not clear to me how you are going to define what high users of physical health care services are and who accrue high health care costs. Could you make the inclusion criteria more explicit? 2. You aim is to "determine the prevalence and magnitude of depression and anxiety in adults who are high users of physical health care services" but in the inclusion criteria you say that the focus is on studies that include patients with anxiety and/or

	depression. I believe this gives a major bias since you exclude all studies that include high users of physical health care serves without anxiety or depression. With this search strategy the answer to your research question is going to be a over estimation of the prevalence. I believe these two points need attention before you can start with this study. After updating your strategy I suggests you consult an expert to review the search strategy. This will improve the quality of your strategy and the results of your study. Thanks again for giving me the opportunity to review this study protocol

REVIEWER	Ahmed Jerome ROMAIN University of Montreal Hospital Research Centre
REVIEW RETURNED	08-Mar-2019

GENERAL COMMENTS	Thank you for giving me the possibility to review this systematic review and meta-analysis protocol. The protocol is well written in general, easy to read and the topic of interest. However, I would have some comments. In my opinion, one of the major points is to accurately define what is a high use of health care services or what do they call “high users of health care”. Even though it is easy to understand, we do not know how it is operationalized. Do they talk about healthcare costs? Number of contacts with health professionals? Hospitalization? Etc. I think the concept should be better defined as it is central in the present protocol. Also, given the importance of health care cost, how research will be harmonized? I mean, how currencies from different countries will be used and adjusted? I did not see any information about that. I also believe that the introduction needs to be extended to better acknowledge the current state of art. Why it is important to realize this review? Comments The literature review and dates should be extended. The authors mentioned they will include papers from inception to June 2018 but there is a 6-month difference with today (march 2019). So, when they will realize their review, it is ls likely that the literature will be outdated. Page 6, line 37, could you please define what is a naturalistic general health services intervention? Some specific examples would be useful. Also, could you better explain why you will exclude clinical trials? Could you assume that, in your review, clinical trials could constitute a sub-sample of studies (or sensitivity analyses)? Page 6, comparators. Could you please define what is a high level of depression/ anxiety? (line 59). Page 7, search strategy: please pay attention that the databases included in the abstract are not the same as those indicated in the study search section. Also, in this section, the authors underlined that they will focus on terms related to frequent use of health care services. However, this focus is not included in the outcomes section (page 7, lines 3-8).

	Table 1: in the eligibility criteria, the criteria are not the same as in the text. Please, modify Data analysis and synthesis section Could you be more accurate about the pooling methods (Der Simon & Laird, Maximum of likelihood, etc.) you will use if you have to realize a meta-analysis? Moreover, could you give more details about the I^2 threshold you will use? Also, some reference to support the use of these statistics would be appreciated. Also, could you provide explanations about the Egger's test and it will be used? Discussion Page 10, lines 31 – 38, the authors reported the limitations of previous research in terms of sample size, methods, instruments, evaluation, health care cost. I was wondering why the authors did not consider these limitations in subgroup analysis (if possible)? It might help for future studies/ protocols. Appendix 3, I am not sure to understand the interest of the table named baseline characteristics of patients given that authors do not want to include intervention studies. In the same table, what is "number of protocol violations"? How it is defined? Other comments Page 5, line 13, lease, define the acronym A&E.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 (Iliana Kohler)

This is a nicely developed protocol to conduct a systematic review on the prevalence of depression and anxiety among high cost users aged 18+ yrs of health care services. A major concern I have is related to the relevance of the proposed study. I urge the authors to develop and outline the study's relevance in a much better way than in the current version of the manuscript: for instance, are there any other studies that have conducted a systematic review like this? If yes, how does the proposed systematic review differ from prior studies? What is the innovation of the present study? Also, why should the readers care about this topic in general? What is the public health relevance? Why do we need this study? Some of these aspects have been mentioned in the discussion and implications of results sections of the protocol, but they are not particularly convincing and need to be better developed, and also sketched in the introduction. See also my specific comments on these parts of the protocol.

Response: We have changed the introduction to take account of the reviewer's comments, added details of a relevant systematic review, and how the proposed review will differ from existing work. We are particularly interested in the prevalence of anxiety and depression in the 'high-cost' group, as most interventions that target mental health conditions associated with multimorbidity have been designed to specifically treat depression/anxiety. It is important therefore to understand the magnitude of these particular mental health problems, as opposed to mental health problems in general, which can include schizophrenia, bipolar affective disorder, drugs and alcohol addictions, cognitive problems, other anxiety states like PTSD, panic disorder, OCD, etc. In terms of public interest, at present, co-morbid depression in multimorbidity is not managed well which leads both to poorer outcomes for patients and 3 increased health care costs, in the setting of increasing demand and rise in health care costs. The results of this review will help personalise treatment approaches for particular subgroups

of high-cost patients (pages 5-6).
2.1.Clarify the specific dates of the study: the protocol says studies from whenever published on the topic to June 2018. First, why stop in June 2018?
Response: We updated the dates for the literature search to 1st April 2019 in our manuscript and in our PROSPERO registration.

2.2.Second, I am somewhat concerned with going very far back in the past since this may affect the quality of the data and conclusions. We know that the diagnosis of depression/anxiety has improved over time so very old studies may underestimate their prevalence and or costs associated with them among high health care users. How would this impact the proposed study and especially if the authors conduct a meta-analysis?

We recognise the reviewer's point and it was something that we considered carefully when designing the protocol. Imposing an inception date limit on our search will decrease the chances of finding as many relevant studies as possible reporting on anxiety/depression and health use or costs. We recognise that there has been an improvement in the reporting of depression in physical disease by general hospital staff or primary care staff over time. However, we are only including studies that have used a standardised self-report measure or clinical interview or clinical diagnosis according to a recognised international classification system, so even older studies are unlikely to underestimate the prevalence of depression/anxiety. One of the most commonly used self-report measures of depression is the Beck Depression Inventory which was developed in 1961 but is still in widespread use today. It is of course true that both major classification systems (DSM and ICD) have slightly altered their definitions of depressive disorders and generalised anxiety disorder. However, the changes in both diagnoses have been relatively minor and are unlikely to make major differences to our findings.

2.3. I suggest that the authors explicitly consider the difference in costs related to managed versus unmanaged depression/anxiety. High health care costs can also be due to the management/costs because of co-morbidity, but also because of managed depression. In other words, there are many pathways how high health costs can be accumulated and associated with the presence of depression (e.g., present co-morbidity, different types of co-morbidity that can be also associated with different levels of 4 depression/anxiety, managed vs. not-managed depression/anxiety, etc.) and all these different pathways can results in different health care costs. It is not clear that the systematic review does consider these different pathways. I urge the authors to think more about these aspects and differentiate them in their classification and indicators.

Response: This is an important point that we had not considered, and we thank the reviewer for pointing this out. We will consider the importance of managed versus unmanaged depression/anxiety, if there are sufficient publications which enable us to do this. We have added the criteria of managed/un-managed common mental health problems and comorbidities in our data extraction sheet in Appendix 2, under Patient characteristics. We have also made it clearer that we will record the number and type of co-morbidities, if specified by the study (see Appendix 2, page 7). We also made this point clearer in the data extraction section of the manuscript (bottom of page 12/top of page 13):

It will include: year and country of study, type of health care system, criteria used to define high use or high costs, method used to record depression/anxiety (self-report measure validated or non-validated, clinical interview), prevalence of depression and anxiety, health care use, and costs and associated ranges, the methods used to evaluate these, health care settings (e.g. primary, secondary or ED, or total health care), the odds ratios of use of health care by depressed/anxious patients compared with non-depressed/anxious patients, patient

characteristics (e.g. co-morbidities, whether anxiety/depression is managed).
2.4.The authors correctly include categories for depression/anxiety diagnosed via a standardized questionnaire versus clinical interview. However, there are several issues here related with the estimates of the prevalence of depression and anxiety: first, the 2 approaches can results in different estimates and this bias the results; why not focus primarily on studies that are based on standardized instruments and then as part of the analytical approach propose sensitivity analysis that consider clinical interviews? Second, there are multiple depression/anxiety instruments that are used in different settings and that result in different estimates of the presence/prevalence of both conditions. Some of these instruments have been clinically validated, others not, some have been clinically validated only in specific settings (e.g., countries, languages, population, clinical settings versus non-clinical, etc). The proposed protocol and criteria does not outline any categories reflecting these differences and does not discuss how to deal with these issues and their implications on the analysis/results. Please 5 address these issues. Why not consider estimating the health care costs associated with depression/anxiety by location? It is well known that health care costs differ substantially between European countries and the US, and there are also systematic regional/country-specific differences in how the diagnosis of depression/anxiety is handled. These differences are not considered in the proposed protocol.

Response: We agree that estimates of depression and anxiety can vary depending upon the type of measure used. Although, there is variability between self-report measures, the greatest differences lie between use of standardised clinical interviews and clinical interviews (both using diagnostic criteria) and self-report measures. The latter often reporting much higher rates reflecting caseness, rather than diagnosis. The reviewer is right to imply that the number of studies that have used interviews, will be much smaller than those employing self-report measures, as interviews are very labour intensive. If we are able to conduct a meta-analysis, we will focus on studies that have included self-report measures and include those using interviews in a sensitivity analysis. If we are unable to conduct a meta-analysis, we will report the studies using self-report measures and interviews separately.

In relation to validation of self-report measures, in the scoping that we have done so far, most measures that have been used to report depression/anxiety in general medical populations have been validated for this setting or have subsequently been validated. Usually the threshold for caseness is slightly increased when measures are used in patients with physical co-morbidities. We will check whether measures have been validated for use in general medical settings, and report accordingly with a narrative synthesis or with a sensitivity analysis if we are able to carry out a meta-analysis (page 15).

Sensitivity analyses will be pursued at minimum on high/low quality studies, on the use of unvalidated standardised questionnaires, and use of structured clinical interviews.

The point about country of study and type of health care system is important and these are now clearly included in the data extraction form and will be considered in the narrative review and any potential meta-analysis. We will consider country and type of health care system as potential subgroups.

2.5.How is high use of health services defined? Is there a general agreement on this? Does the definition change over time and how does this impact the protocol study? Similarly, what is the definition of high health costs? Does the definition change over time and how does this impact the protocol study?
Response: There are no agreed definitions of either high use or high health care costs. Most

studies of high health care costs have studied the top percentiles of patients, but percentiles vary. A recent systematic review included all studies of ‘high-cost’ patients due to the lack of definition but reported that most studies included populations belonging to the top-1%, top5%, top-10% and top-20% of all patients (Wammes JJG, van der Wees PJ, Tanke MAC, et al. Systematic review of high-cost patients’ characteristics and healthcare utilisation. BMJ Open 2018;8:e023113. doi:10.1136/bmjopen-2018-023113). We will therefore include cost studies with patients in the top 1st, 5th, 10th, and 20th percentiles. In relation to use of health care, we will include studies that have used a clearly defined threshold for high use according to the setting. For primary care this is either the top 10% of consulters (Luciano J V, Fernández A, Pinto-Meza A, et al. Frequent attendance in primary care: Comparison and implications of different definitions. Br J Gen Pract Published Online First: 2010. doi:10.3399/bjgp10X483139) or 10 or more attendances per year (Vedsted P, Christensen MB. Frequent attenders in general practice care: A literature review with special reference to methodological considerations. Public Health Published Online First: 2005. doi:10.1016/j.puhe.2004.03.007). For ED it is 4 or more attendances per annum (Locker TE, Baston S, Mason SM, et al. Defining frequent use of an urban emergency department. Emerg Med J Published Online First: 2007. doi:10.1136/emj.2006.043844). (Bottom of page 7 and top of page 8).

The definitions in certain areas have changed over time. For example, in relation to health care use, frequent attendance in primary care was defined as an arbitrary number or numbers which varied from study to study, but more recent work has refined this threshold and it is clear that using a percentage (i.e. 10%) threshold is better than a hard number, but to use a threshold alone, would exclude most studies. In ED the thresholds have again changed and vary from study to study. The Locker et al study, however, is now recognised as the best way of calculating frequent attendance and its threshold would mean most studies on frequent attendance at ED would be included in the review.

2.6.Aim 2, p. 5 – I am not sure how based on the proposed analytical approach the authors can determine the magnitude of health care use and costs of depression;

Response: We offered additional clarification to this aspect within the data analysis section (page 13): We are not attempting to pool or calculate costs or health use across studies. We will only be able to determine the magnitude of health care use associated with depression/anxiety in relation to studies that have specifically calculated or estimated these. This will be studies where high health care use/costs are compared between patients with anxiety/depression versus patients without anxiety/depression. Odds ratios and 95% confidence intervals will be extracted from studies presenting the number of health care contacts (e.g. ED attendances or GP contacts or number of hospital admissions) by subjects with and without depression.
2.7.p. 6, line 22 – add “among high users”

Response: We have added this (page 7)
…as the aim is to estimate the prevalence of anxiety/depression among high users of general health care...
2.8.p. 6 – criteria for exclusion listed here are not listed later in the tables summarizing exclusion criteria; Also explain ration for excluding these settings and what is gained from focusing on general health settings

Response: The inclusion and exclusion criteria in Appendix 2 (page 4) have been amended to maintain consistency with descriptions in the methods and design section. We have also added the rationale for focusing on specific general health care settings (page 7; last full

paragraph):
We have focused on general hospital and primary care services, to ensure the review is relevant to as wide a population as possible. There is great variability in the way costs, healthcare and depression/anxiety have been recorded in the literature. To add studies on individual disease conditions or specialities would considerably inflate the variability within the population of this review.
2.9.p 7 – the authors mention that they will include previous meta-analysis on this topic; this brings me back to what I mentioned before – how does the proposed study and meta-analysis differ from prior research?

Response: We are not intending to include any prior meta-analyses in the proposed review. We are not aware of any meta-analytic studies of relevance. Prior systematic reviews and other reviews of relevance will be used to help search the literature. We hope we have addressed the main point the reviewer makes about previous studies in our response to the reviewer's first comment and in the expanded introduction to the manuscript (page 5 and 6).
2.10.p 7, line 2 – studies in all languages will be considered: I am not convinced that the authors will be able to do this and make all these translations.

Response: We included more details regarding language translation possibilities in our collaborating universities (page 10):
We will include studies in all languages; translations will be pursued either by co-authors or by international colleagues/students at the Universities of Birmingham, Leeds, and Manchester.
We understand the reviewer's scepticism, but across three large universities (Birmingham, Manchester and Leeds) we are confident we will be able to obtain translations of all relevant papers.
2.11.p8, line 23 – what is "recent" review article (which time period does recent refer to)?

Response: We rephrased this (page 9), under heading Search Strategy
We will hand-search reference lists of relevant reviews/meta-analyses.
2.12.p6 – the comparator's criteria are not clear. These will be studies also focused on depression/anxiety? Or as stated also studies without depression/anxiety? How many groups and which exactly are compared here: why not listing the groups specifically?

Response: We have clarified our comparators (bottom of page 8/top of page 9):
We will include studies where anxiety/depression is described in groups of patients considered 'high/frequent users' and/or 'high cost users' versus non-high cost and non-high users of health care services. We will include studies where high health care use/costs are compared between patients with anxiety/depression versus study patients without anxiety/depression.
2.13.p.9, line 29 – how will be the qualification of the use and costs developed? Can you outline this more precisely?

Response: Please see the response to 2.5.
2.14.p. 11, lines 6-11 – I am not convinced that by doing the proposed analysis all these issues listed in the conclusions can be indeed addressed.

Response: We have revised our discussion and limited the points we make to be directly relevant to the two main aims of the review (page 16):
Our review will build upon the recent systematic review by Wammes and colleagues [1] that described the characteristics of high-cost patients and found that a high prevalence of high cost patients had associated mental health disorders. This review will specifically focus upon depression/anxiety and include both studies of cost and health care use. It will also provide

information about the prevalence of depression/anxiety in different health care settings, including primary care and ED. There is trade-off between diagnostic accuracy versus size of study. Our results will complement those of Wammes and colleagues [1], and increase our understanding of the role of depression/anxiety in driving health care use and costs.
2.15.p.4, appendix – include categories for scale/instruments used to measure depression and anxiety

Response: We have included these in Appendix 2 under data extraction template for full-text articles- section outcome measurement (page 6):
- Specify instrument and range (e.g. PHQ-9, GAD-7, SCID, etc.)
- Is it self-report
- Is it standardised?
- Is it validated to context?
- Is it a standardised clinical interview
- Is it a clinical interview with diagnosis according to recognised diagnostic system

2.16.confounders on p.4 to be specified

Response: We have specified this in Appendix 2, under data extraction template for full-text articles- section covariates/confounders (page 6):

Please consider patient demographics (e.g. age, ethnicity, gender, homelessness) and clinical factors (e.g. anxiety/depression management, physical co-morbidities)" (page 6)

2.17.Location of study – need classification (region, type of health care facility, rural/urban, etc.)

Response: This classification has now been included in Appendix 2 (page 6):
Location of study (please specify country, type of health care system, healthcare setting (general hospital, primary care, ED/all settings)

3. Reviewer 2 (Lucinda Leung)
3.1.Homelessness is a crucial risk factor for high utilization and should be included in the list on Page 5, line 27.

Response: Whilst we recognise that homelessness is a risk factor for high utilization, it is not a mental health disorder or problem, although clearly it impacts upon mental health. We do not think it should be added to a list of mental health problems. We do think it is an important point raised by the reviewer, particularly in relation to high attendance at ED departments and will add homelessness to our list of confounders, in Appendix 2 under patient
examined the characteristics of high cost patients using such methods and identified the proportion of high-cost patients with mental and behavioural disorders [1]. We are seeking to specifically identify patients with depression/anxiety and hence, will include studies that have involved a standardised self-report measure, or clinical interview. There are many studies of which we are aware that have done so on different populations of high users (e.g. frequent attenders in primary care, frequent attenders at ED). We recognise that several studies of health care systems may not be included in this review but we believe that the broad question regarding the relationship between cost and overall mental health has been answered by Wammes et al., 2018. We are seeking a more nuanced answer in relation to depression/anxiety.
3.3. Please include reason for excluding specialty med-surg departments (Page 6, Line 25). Pts with depression/anxiety are known to overutilize in these departments as well, due

to somatic complaints necessitating excessive specialty workup.
Response: We have focused on general hospital and primary care services, to ensure the review is relevant to as wide a population as possible. There is great variability in the way costs, health care and depression/anxiety have been recorded in the literature. This is a major challenge in relation to this review. To add studies on individual disease conditions or medical specialties like surgery would considerably inflate the variability within the population of this review. We recognise that patients with anxiety/depression are associated with high use in many medical specialties and disease conditions.

3.4.Suggested subgroup analyses are perplexing. Related to my comment in #2 above, Page 9, Line 43 highlights that subgroup analysis #1 would include “type of outcome measurement for depression/anxiety” but it remains unclear what this data source is. Please clarify if using diagnostic codes, chart review, etc. In Page 9, Line 45, please explain how patients would be assigned to a specific healthcare setting and why there would be suspected differences by healthcare setting. Are you comparing patients that ONLY get there care in acute care settings (ER, hospital) to those that get there care in primary care? How would you handle high-utilizing patients that have primary care AND ER visits, for example?

Response: The analyses will be partly guided by the area of health care focused upon in the papers. We expect to find studies that: only focus on ED (i.e. studies on frequent attendance at ED), only focus on primary care, only focus on secondary care, and focus on the whole health care system. We are not comparing patients that only get their care in particular settings, we are instead including studies that have focused on a particular health care setting. The studies will form subgroups according to their different settings, the most likely being: studies that include all healthcare; primary care studies; studies on ED; and studies focused on hospital admissions.

3.5.Authors should be cognizant of existing literature when discussing research implications. Re (Page 11, Lines 6-10) “It may also reveal an unmet need in the diagnosis and appropriate management of anxiety/depression in populations routinely seen in general health care settings, ” this is already firmly established in the reviewer’s opinion; thus, it may be a bigger contribution to “confirm and clarify the extent of unmet need…” Also, Re (Page 7, Lines 15-22) “it could suggest the type of integrated, collaborative services, or management methods…”, this is also well known and documented in the literature (Archer et al, Cochrane, 2012; Coventry et al, BMJ, 2015), so I would suggest that your findings may “further support” collaborative care model implementation.

Response: We have made the suggested changes to the implications section (page 16):
The results of this systematic review will provide an estimate of the prevalence of common mental health disorders in high users of health care services, while also providing an estimate of the magnitude of use associated with depression/anxiety. It will enable treatments, such as the collaborative care model, that have already been developed for the treatment of depression/anxiety in the physically ill, to be evaluated in high-cost patients with co-morbid depression/anxiety resulting in a more personalised approach to both treatment and policy.

4. Reviewer 3 (Luc Jansen)
4.1.It is not clear to me how you are going to define what high users of physical health care services are and who accrue high health care costs. Could you make the inclusion criteria more explicit?

Response: We have clarified this point which was also raised by Reviewer 1.
There are no agreed definitions of either high use or high health care costs. Most studies of high health care costs have studied the top percentiles of patients, but percentiles vary. A recent systematic review included all studies of ‘high-cost’ patients due to the lack of definition but reported that most studies included populations belonging to the top-1%, top-5%, top-10% and top-20% of all patients [1]. We will therefore include cost studies with patients in the top 1st 5th, 10th, and 20th percentiles. In relation to use of health care, we will include studies that have used a clearly defined threshold for high use according to the setting. For primary care this is either the top 10% of consulters [2] or 10 or more attendances per year [3]. For ED it is 4 or more attendances per annum [4].

4.2.You aim is to "determine the prevalence and magnitude of depression and anxiety in adults who are high users of physical health care services" but in the inclusion criteria you say that the focus is on studies that include patients with anxiety and/or depression. I believe this gives a major bias since you exclude all studies that include high users of physical health care serves without anxiety or depression. With this search strategy the answer to your research question is going to be a over estimation of the prevalence. I believe these two points need attention before you can start with this study. After updating your strategy I suggests you consult an expert to review the search strategy. This will improve the quality of your strategy and the results of your study.

Response: We apologise for confusion. Our intention is not to focus on studies that include patients with anxiety/depression, but studies that have used a measure of anxiety/depression in order to determine the prevalence of anxiety/depression in a high cost population. The inclusion criteria now state that we will review studies that include measures that assess anxiety/depression. There would be no way of estimating the prevalence of depression/anxiety unless it is measured.

In Appendix 2, page 4, our eligibility criteria require that any study included will report the prevalence of anxiety/depression in high cost/high use patients.

Our search strategy has also been reviewed by two experts, in the Universities of Manchester and Birmingham. This is now included in the manuscript (page 10):

The strategy was developed in collaboration with experts in these fields and experienced librarians in the Universities of Birmingham and Manchester, to ensure it yields appropriate studies.

And in the Acknowledgements (page 18):
The authors would like to thank Rosalind McNally (Outreach Librarian in the Research and Innovation Department & Knowledge Service in the Greater Manchester Mental Health NHS Foundation Trust) and Anita Phul (Librarian at the Barberry, National Centre for Mental
14
Health, Birmingham and Solihull Mental Health NHS Foundation Trust, Birmingham) for their critical review of the search strategy.
5. Reviewer 4 (Ahmed Jerome Romain)
5.1.In my opinion, one of the major points is to accurately define what is a high use of

health care services or what do they call “high users of health care”. Even though it is easy to understand, we do not know how it is operationalized. Do they talk about healthcare costs? Number of contacts with health professionals? Hospitalization? Etc.
I think the concept should be better defined as it is central in the present protocol.
Also, given the importance of health care cost, how research will be harmonized? I mean, how currencies from different countries will be used and adjusted? I did not see any information about that.

Response: We have responded to the first point raised by the reviewer in 4.1. We hope we have answered the point to the reviewer’s satisfaction. In relation to the second point, we intend to analyse the data according to subgroups including ‘country’ and type of health care system. We are not intending to pool costs or healthcare use. Both vary widely across and within countries. We are intending to compare the prevalence of depression/anxiety and the ratio of use of care by depressed/anxious patients versus non-depressed anxious patients.

5.2.I also believe that the introduction needs to be extended to better acknowledge the current state of art. Why it is important to realize this review?

Response: We have re-written parts of the introduction to address the reviewer’s point, which was also raised by reviewer 1. We have reviewed recent work and more clearly delineated the reasons for focusing on depression/anxiety (page 6):

Specific interventions for treating depression and anxiety in people with co-morbid physical health problems have shown promising results [5,6] but have not been targeted at high-cost patients with co-morbid depression/anxiety.
Improved recognition of the association of depression and anxiety with high health care use and costs will enable treatments that have already been developed to treat depression/anxiety in physical disease, to be evaluated in this high need/high cost group. Although, there has been a general call for better integration of physical and mental health services, the treatment and management of co-morbid depression/anxiety in chronic physical disease remains poorly managed [7].

5.3.The literature review and dates should be extended. The authors mentioned they will include papers from inception to June 2018 but there is a 6-month difference with today (march 2019). So, when they will realize their review, it is ls likely that the literature will be outdated.

Response: We have amended the time range for our literature search to 1st April 2019, as also requested by Reviewer 1.

5.4.Page 6, line 37, could you please define what is a naturalistic general health services intervention? Some specific examples would be useful. Also, could you better explain why you will exclude clinical trials? Could you assume that, in your review, clinical trials could constitute a sub-sample of studies (or sensitivity analyses)?

Response: We were referring to implementation studies involving changes in service delivery or introductions of integrated care pathways. We clarified the information offered in the interventions section and explain why interventions involving health service implementation will be considered, but not randomized controlled trials (page 8):
We will not include randomised controlled trials, due to their selective nature. We will include cohort studies of naturalistic changes in health service delivery e.g. implementation of a new integrated care pathway across a geographical region, where external validity is likely to be

high.
We considered the inclusion of RCTs in the review but decided against this, due to the restrictive nature of entry criteria into trials and the usual large differences between those entered into a study and the base population.
5.5.Page 6, comparators. Could you please define what is a high level of depression/ anxiety? (line 59).

Response: We apologise for this loose terminology. For studies that have used self-report measures, we will use caseness to determine the presence of anxiety/depression (ie. above a recognised threshold). For studies that have used an interview, we will use the presence of a clinical diagnosis of either a depressive disorder or a diagnosed of generalised anxiety disorder.

5.6.Page 7, search strategy: please pay attention that the databases included in the abstract are not the same as those indicated in the study search section. Also, in this section, the authors underlined that they will focus on terms related to frequent use of health care services. However, this focus is not included in the outcomes section (page 7, lines 3-8).

Response: We have checked the Abstract and Main text for consistency. We have corrected the omission in the outcomes section (page 9):
MEDLINE, PsychINFO, EMBASE, CINAHL, PROSPERO, Cochrane Library
The primary outcome is the prevalence of anxiety/depression in high/frequent and/or costly users of general health care services

5.7.Table 1: in the eligibility criteria, the criteria are not the same as in the text. Please, modify

Response: Thank you. We have changed the phrasing to ensure consistency between text and Table 1.

5.8.Data analysis and synthesis section
Could you be more accurate about the pooling methods (Der Simon & Laird, Maximum of likelihood, etc.) you will use if you have to realize a meta-analysis? Moreover, could you give more details about the I^2 threshold you will use? Also, some reference to support the use of these statistics would be appreciated. Also, could you provide explanations about the Egger’s test and it will be used?

Response: We have offered more details in the data analysis section (pages 14-15):
We will use random-effects models to describe the prevalence of depression/anxiety high use or high cost populations. This is because it is implausible that the underlying study-specific prevalence of depression (i.e. the prevalence that would be observed were a study of infinite size) is the same for each study. Prevalence is likely to vary from study to study according to factors, both measured and unmeasured, that differ between them [8].

We will use the inverse variance method of DerSimonian and Laird to estimate between-study heterogeneity in underlying depression prevalence and the I-squared measure which represents the proportion of total variance attributable to this heterogeneity [9,10]. The I-squared measure gives the percentage of variability in the effect estimate that is due to heterogeneity rather than to chance. Suggested thresholds for the interpretations of the I-squared measure are as follows: less than 40% indicates there is no problem with heterogeneity, 30% to 60% indicates moderate 17 problems, 60% to 90% a substantial problem, and 75% and over a considerable problem [39]. We will use the threshold of less than 40%.

Egger’s statistics with 95% confidence intervals and associated funnel plot will depict potential publication or small bias related to our main outcome summaries and/or within subgroups [11]. Egger’s test is based on the Galbraith plot which is a plot of difference over standard error against one over standard error. Egger suggests that a regression of study difference over standard error on 1/standard error be undertaken to test the null hypothesis that the intercept is equal to zero. If Egger’s test is significant ($p<.05$), it means that the funnel plot is asymmetric and that smaller studies with smaller precision show larger effects sizes, suggesting bias.

5.9.Page 10, lines 31 – 38, the authors reported the limitations of previous research in terms of sample size, methods, instruments, evaluation, health care cost. I was wondering why the authors did not consider these limitations in subgroup analysis (if possible)? It might help for future studies/ protocols.

Response: Following the reviewers’ comments we have added details of most of these aspects in our subgroup analyses which are detailed on page 14.
For both outcomes, where possible subgroups will be explored based on potential differences related to: 1) country, 2) type of healthcare system, 3) medical settings (e.g. primary care, ED, inpatients, outpatients, etc), 4) metrics used to evaluate health use/costs (e.g. attendances, hospital admissions, admissions, scheduled/unscheduled, etc).

5.10.Appendix 3, I am not sure to understand the interest of the table named baseline characteristics of patients given that authors do not want to include intervention studies. In the same table, what is “number of protocol violations”? How it is defined?

Response: We note the reviewer’s point. We are using this instrument in an adapted form, hence some details which are not needed have been removed (e.g. protocol violations) and others renamed (i.e. Patient characteristics).
5.11.Page 5, line 13, lease, define the acronym A&E.

Response: We have removed the acronym A&E and referred instead to emergency department (ED).

Reference:

1 Wammes JJG, van der Wees PJ, Tanke MAC, et al. Systematic review of high-cost patients’ characteristics and healthcare utilisation. BMJ Open 2018;8:e023113. doi:10.1136/bmjopen-2018-023113
2 Luciano J V., Fernández A, Pinto-Meza A, et al. Frequent attendance in primary care: Comparison and implications of different definitions. Br J Gen Pract Published Online First: 2010. doi:10.3399/bjgp10X483139
3 Vedsted P, Christensen MB. Frequent attenders in general practice care: A literature review with special reference to methodological considerations. Public Health Published Online First: 2005. doi:10.1016/j.puhe.2004.03.007
4 Locker TE, Baston S, Mason SM, et al. Defining frequent use of an urban emergency department. Emerg Med J Published Online First: 2007. doi:10.1136/emj.2006.043844
5 Camacho EM, Ntais D, Coventry P, et al. Long-term cost-effectiveness of collaborative care (vs usual care) for people with depression and comorbid diabetes or cardiovascular disease: A Markov model informed by the COINCIDE randomised controlled trial. BMJ Open Published Online First: 2016. doi:10.1136/bmjopen-2016-012514

6 Coventry PA, Hudson JL, Kontopantelis E, et al. Characteristics of effective collaborative care for treatment of depression: A systematic review and metaregression of 74 randomised controlled trials. PLoS One Published Online First: 2014. doi:10.1371/journal.pone.0108114

7 Department of Health and Social Care. No Health Without Mental Health. 2011. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/138253/dh_124058.pdf

8 Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. J Am Med Assoc Published Online First: 2000. doi:10.1001/jama.283.15.2008

9 DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials Published Online First: 1986. doi:10.1016/0197-2456(86)90046-2

10 Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Cochrane Collab Published Online First: 2011. doi:10.1002/9780470712184.ch4

11 Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. Br Med J Published Online First: 1997. doi:10.1136/bmj.315.7109.629

VERSION 2 – REVIEW

REVIEWER	Lucinda Leung UCLA David Geffen School of Medicine, VA Greater Los Angeles Healthcare System
REVIEW RETURNED	06-May-2019

GENERAL COMMENTS	The authors have made substantial changes to improve manuscript readability. The study rationale is more clear. As I understand it, the authors are attempting to build on findings from Wammes et al, 2018 to explore the exact prevalence of depression/anxiety in different health care settings through examination of standardized tools/interviews. I am satisfied with their responses to my specific queries. However, I encourage them to more strongly emphasize what sets them apart from Wammes et al, 2018 and other work -- they are the first (I think) to look at depression/anxiety as evaluated by standardized questionnaire or clinical interview, as opposed to others whom have focused on available diagnostic codes. This is a compelling research motivation that needs to be stated more strongly upfront in the Intro (both Abstract and Text). In its current iteration, that point is not clear until the Methods and with review of Reviewer Responses.

REVIEWER	Luc Jansen Erasmus Medical Center, Netherlands
REVIEW RETURNED	01-May-2019

GENERAL COMMENTS	Thanks for your convincing responses to the raised questions, I am looking forward to see the manuscript.

REVIEWER	Ahmed Jerome ROMAIN Université de Montréal
REVIEW RETURNED	07-May-2019

GENERAL COMMENTS	Thank for this revised version of this systematic review protocol. The authors should be commended for the efforts they put in this revised version and the manuscript is clearly improved. Nevertheless, I would still some comments regarding the manuscript. Based on different sections in the manuscript, to what conditions a meta-analysis will be conducted or not? I mean, in the strengths and limitations methods, it is indicated that a meta-analysis might not be feasible given the expected level of heterogeneity. I clearly understand this assumption but what criteria will be used to realize a meta-analysis? From how many studies included for the authors will consider this analysis, and why? This information is important in terms of methods. Also, in the methods section, sorry if I missed it, but I did not see what will happen to studies without accurate information about depression and anxiety? I mean, what about a study reporting only “mental disorders” without specifications on whether it is depression, anxiety or something else? Will they be excluded? Will you try to have contact with the authors to have more accurate information? Strengths and limitations In this section, given your study is a protocol, I was wondering whether it might not be better to use to future rather than the present (e.g., this systematic review will include…). Introduction Page 7, lines 21-22, it is written that studies conducted in primary, secondary care and emergency department will be included but on line 45, it is reported that you will focus on general hospital and primary care services. Could you please explain the difference between these services in terms of methodology in your review? Methods Page 8, lines 30-32, it is reported that odds ratio will be calculated. But do the authors talk about a meta-analysis of odds ratio? Data analysis Page 14, line 41, you will use the threshold of “less than 40%” to what? This information is not provided and is important as using this value in terms of methodology is not something current. Page 15, lines 7-9, what number of studies is considered as enough? Appendix 2 Page 4, why “funding” will be considered as an information to collect?

VERSION 2 – AUTHOR RESPONSE

Reviewer 2. Lucinda Leung
1. The authors have made substantial changes to improve manuscript readability. The study rationale is more clear.

Reply: We thank the reviewer for this comment. We reply to their questions below.

2. As I understand it, the authors are attempting to build on findings from Wammes et al, 2018 to explore the exact prevalence of depression/anxiety in different health care settings through examination of standardized tools/interviews. I am satisfied with their responses to my specific queries. However, I encourage them to more strongly emphasize what sets them apart from Wammes et al, 2018 and other work -- they are the first (I think) to look at depression/anxiety as evaluated by standardized questionnaire or clinical interview, as opposed to others whom have focused on available diagnostic codes. This is a compelling research motivation that needs to be stated more strongly upfront in the Intro (both Abstract and Text). In its current iteration, that point is not clear until the Methods and with review of Reviewer Responses.

Reply: We thank the reviewer for pointing that out. We now emphasize this point more strongly in the Abstract (page 3):

This is the first systematic review to focus on patients who are over the age of 18 whose degree of anxiety/depression has been evaluated with a standardised or by a clinical interview generating a diagnosis according to international diagnostic criteria.

(…)

We will estimate the prevalence of anxiety/depression in these populations, and the magnitude of use associated with anxiety/depression across various general physical health care settings.

In Strengths and limitations (page 4):

It will focus upon studies that have specifically recorded the presence of depression and/or anxiety in the high cost/high use population studied, using standardized questionnaires or clinical interviews leading to a clinical diagnosis.

And Introduction (page 5-6):

A recent systematic review of the general characteristics of high-cost patients found a high prevalence of multiple chronic conditions amongst this patient population [1]. Mental health problems were also common but varied according to the health care system. In US Medicaid, the prevalence of mental illness ranged from 30-75%, whereas in US Medicare, the prevalence was between 10-25%. One of the main findings of the review was that high-cost patients were more likely to have a mental health disorder. There were, however, no details as to the nature of mental health problems experienced by these high-cost patients, as data were grouped under a broad category of mental and behavioral disorders. This review will focus on patients with depression and anxiety disorders, as they are the most common form of mental disorder. We will focus upon studies where depression and anxiety are identified through standardized questionnaires or by clinical interviews leading to a clinical diagnosis. Our review will provide information about the prevalence of depression/anxiety in both high- and low-income countries and in different general physical health care settings, namely primary, secondary care, and ED.

Reviewer 3. Luc Jansen
Thanks for your convincing responses to the raised questions, I am looking forward to see the manuscript.
Reply: We would like to thank the reviewer for their comments.

Reviewer 4. Ahmed Jerome ROMAIN
1. Thank for this revised version of this systematic review protocol. The authors should be commended for the efforts they put in this revised version and the manuscript is clearly improved.
Reply: We thank the reviewer for their comments. We provide further details to their questions below.

2. Nevertheless, I would still some comments regarding the manuscript. Based on different sections in the manuscript, to what conditions a meta-analysis will be conducted or not? I mean, in the strengths and limitations methods, it is indicated that a meta-analysis might not be feasible given the expected level of heterogeneity. I clearly understand this assumption but what criteria will be used to realize a meta-analysis? From how many studies included for the authors will consider this analysis, and why? This information is important in terms of methods.

Reply: We thank the reviewer for pointing out that some clarification is needed. Our criteria for summarising the findings in a meta-analysis and the minimum number of studies is now discussed in more depth. We acknowledge that a meta-analysis can be pursued with at minimum 2 studies, but this approach has limitations. The studies in this review are highly likely to include a range of diverse settings including country, type of health care system and location (e.g. primary, secondary care, and ED), as alluded in our proposed subgroup analyses. Studies will only be combined quantitatively if this is (1) clinically meaningful and (2) effect estimates can be combined directly or through transformations. Risk of bias and sources of heterogeneity will be discussed for all quantitative or qualitative summaries. Several paragraphs have been modified to reflect this:

Data analysis and synthesis section (pages 13-15)

Where enough studies are available for quantitative summaries (minimum two studies [2]) we will offer weighted estimates of prevalence within relevant subgroups related to populations, comparators, study designs, measurement types, and geographical regions.

Data analysis will result in quantitative and narrative summaries, as appropriate, based on current recommendations for the pooling of observational studies [3,4]. Whereas there is some published guidance on the number of studies necessary to ensure the power of the effect size estimates when pooling interventional studies [5,6], there are no similar clear, agreed guidelines on the number of studies necessary for an appropriately powered meta-analysis of observational studies. We will offer a quantitative summary for any number of studies (2>) if combining their outcomes is clinically meaningful, if they report the same effect metrics, or transformations are possible [2–4]. We will comment on these pooled results in light of clinical practice and research significance and potential statistical issues that may decrease the generalizability of the effect estimates (e.g. high level of heterogeneity, potential sources of bias). For both outcomes, subgroups will be explored quantitatively and narratively, as appropriate, and depending on the type of effect estimates available, based on potential differences related to: 1) country, 2) type of healthcare system, 3) medical settings (e.g. primary, secondary care, ED, inpatients, outpatients, etc.), 4) metrics used to evaluate health use/costs (e.g. attendances, hospital admissions, etc.). For instance, we expect to find studies that may only focus on frequent attendance at ED, primary care outpatient visits, number of bed days in secondary care, or more generic attendance metrics across either of these health care settings. We

will account for such differences in reporting, but we are not planning to compare outcomes across settings, just to record and estimate the magnitude of use/cost in each of these contexts.
3. Also, in the methods section, sorry if I missed it, but I did not see what will happen to studies without accurate information about depression and anxiety? I mean, what about a study reporting only "mental disorders" without specifications on whether it is depression, anxiety or something else? Will they be excluded? Will you try to have contact with the authors to have more accurate information?

Reply: We thank the reviewer for pointing this out. We included further details earlier in the protocol to clarify this aspect (page 9)

Studies will be excluded if they do not meet our criteria for the assessment of anxiety or depression. A review concerning general mental health disorders has already been undertaken by Wammes et al. [1].
4. Strengths and limitations
In this section, given your study is a protocol, I was wondering whether it might not be better to use to future rather than the present (e.g., this systematic review will include…).

Reply: Thank you for pointing out the use of present tense. We have now amended this in the Strengths and limitations section (page 4)

1. This systematic review will include both studies of high health care use and/or high health care costs.
2. It will include studies undertaken in general physical health care settings - primary, secondary care, and emergency departments.
3. It will focus upon studies that have specifically recorded the presence of depression and/or anxiety in the high cost/high use population studied, using standardized questionnaires or clinical interviews leading to a clinical diagnosis.
4. We will provide a narrative summary of findings with sources of variation and bias based on a comprehensive data extraction framework, with relevant subgroup analyses and interpretations based upon: country, type of health care system, location of study (primary, secondary care, emergency department, or total health care), and way of recording depression/anxiety.
5. A meta-analysis may not be feasible given a likely high level of heterogeneity in outcome definitions and measurements.

5. Introduction
Page 7, lines 21-22, it is written that studies conducted in primary, secondary care and emergency department will be included but on line 45, it is reported that you will focus on general hospital and primary care services. Could you please explain the difference between these services in terms of methodology in your review?

Reply: Thank you for this comment. We should have been clearer that the term 'general hospital' can refer to both acute inpatient settings and ED. In the first section of our Methods/Population we now more clearly define what we mean by 'general physical health care services', namely any health care services offered across primary, secondary care, and emergency departments that do not include specific medical specialties (page 7).

We include studies conducted in general rather than specialist physical health services, namely primary, secondary care, and ED, across all health care systems.

We amended what was initially line 45 to only refer to 'general physical health care services' as they had been defined above (pages 7-8)

We have focused on general hospital, ED and primary care services to ensure the review is relevant to as wide a population as possible. There is great variability in the way costs, health care use, and depression/anxiety have been recorded in the literature. To add studies on individual disease conditions or specialities would considerably inflate the variability within the population of this review.

For studies of high-cost patients, we will include studies that have defined high cost patients as being in the top 1st, 5th, 10th and 20th percentiles of the patient population [1]. For studies involving high use of health care, we will include studies that have either used similar percentiles to describe high use (i.e. 1st, 5th, 10th or 20th) or have used a recognised definition of high or frequent use for the particular health care services. For ED, we will use the definition of 4 or more attendances per annum [7]. For primary care, we will use the definition of 10 or more attendances per year [8] or the top 10% of consulters [9].

6. Methods
Page 8, lines 30-32, it is reported that odds ratio will be calculated. But do the authors talk about a meta-analysis of odds ratio?
Reply: We thank the reviewer for pointing this out. We are aware of potential pitfalls in the estimation of effect sizes based on odds ratio data and recent methods to overcome these. We discuss and offer references to this on page 8

The review will include studies reporting costs and health care use. However, resource use and costs are sensitive to variability both within and between countries, due to aspects such as local prices or aspects of service organization and delivery. This may limit the generalizability and transferability of estimates of cost and health care across settings. We will not attempt to combine costs or health use in the analyses across studies. The prevalence of depression or anxiety will be compared across studies. To determine the magnitude of health care use associated with depression/anxiety in high-use/high-cost patients, we will estimate the health care used by depressed and non-depressed individuals. If sufficient studies report similar effect measures (e.g. odds ratios, relative risk, incidence rate ratios) of the frequency of health care use in these patients [5], they will be combined in a meta-analysis, consistent with current recommendations [3,4,10]. Studies reporting different effect measures will not be combined, unless they can be transformed [3,4].

And on page 14

Pooled prevalence estimates with 95% confidence intervals will be calculated using SPSS version 25 (IBM Corp, Armonk, NY, USA); where possible and warranted, estimate transformations and quantitative summaries will be pursued using R [10].

This will be studies where high health care use/costs are compared between patients with anxiety/depression versus patients without anxiety/depression. Outcome metrics (including odds ratios, relative risk etc.) and 95% confidence intervals will be extracted from studies presenting the number of health care contacts (e.g. ED attendances or GP contacts or number of hospital admissions) by subjects with and without depression.

7. Data analysis
Page 14, line 41, you will use the threshold of "less than 40%" to what? This information is not provided and is important as using this value in terms of methodology is not something current.

Reply: We acknowledge that the suggested thresholds for the interpretation of the I-squared statistics are a rough guide and percentage of heterogeneity should also be interpreted in light of sources of bias, associated 95% confidence interval, the magnitude/direction of the effects, and their clinical significance. We include more details on page 15-16:

We will use the inverse variance method of DerSimonian and Laird to estimate between-study heterogeneity in underlying depression prevalence and the I-squared measure with associated 95% confidence intervals, which represents the proportion of total variance attributable to this heterogeneity [11,12]. The I-squared measure gives the percentage of variability in the effect estimate that is due to heterogeneity rather than to chance. A rough guide to the interpretation of the I-squared measure suggests that I-square < 40% indicates low to no problems with heterogeneity, 30% to 60% indicates moderate problems, 60% to 90% indicates significant problems, whereas an I-squared of 75% or more suggests considerable problems[6]. If I-squared is less than 40% we will consider the estimated effect to have a low degree of heterogeneity, but this will also be interpreted in light of the magnitude, direction of the effect, and its 95% confidence interval, sources of bias, and clinical significance [2,3,6,11].

8. Page 15, lines 7-9, what number of studies is considered as enough?
Reply: We hope we replied to the reviewer's satisfaction in the Reply to the query above.

9. Appendix 2 - Page 4, why "funding" will be considered as an information to collect?

Reply: There is evidence that the funding source can be associated with reporting bias. Several quality assessment and risk of bias tools include 'funding', as good practice. We have now included a reference and explanation on why we record this aspect

(page 12)

Risk of bias (including reporting bias) will be evaluated commensurate with recent recommendations for the narrative interpretation of variation in observational studies [3,4] and the recommendations of the Cochrane Collaboration [11,13].

(page 13)
We will also record the presence and source of bias, including funding, given its potential association with reporting bias [11,13].

Reference:

1 Wammes JJG, van der Wees PJ, Tanke MAC, et al. Systematic review of high-cost patients' characteristics and healthcare utilisation. BMJ Open 2018;8:e023113. doi:10.1136/bmjopen-2018-023113

2 Ioannidis JPA, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. BMJ Published Online First: 2008. doi:10.1136/bmj.a117

3 Dekkers OM, Vandenbroucke JP, Cevallos M, et al. COSMOS-E: Guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. PLoS Med Published Online First: 2019. doi:10.1371/journal.pmed.1002742

4 Mueller M, D'Addario M, Egger M, et al. Methods to systematically review and meta-analyse observational studies: A systematic scoping review of recommendations. BMC Med. Res. Methodol. 2018. doi:10.1186/s12874-018-0495-9

5 Valentine JC, Pigott TD, Rothstein HR. How Many Studies Do You Need?: A Primer on Statistical Power for Meta-Analysis. J Educ Behav Stat Published Online First: 2009. doi:10.3102/1076998609346961

6 Higgins JP, Thompson SG, Deeks JJ AD. Measuring inconsistency in meta-analyses. BMJ Published Online First: 2003. doi:10.1136/bmj.327.7414.557

7 Locker TE, Baston S, Mason SM, et al. Defining frequent use of an urban emergency department. Emerg Med J Published Online First: 2007. doi:10.1136/emj.2006.043844

8 Vedsted P, Christensen MB. Frequent attenders in general practice care: A literature review with special reference to methodological considerations. Public Health Published Online First: 2005. doi:10.1016/j.puhe.2004.03.007

9 Luciano J V., Fernández A, Pinto-Meza A, et al. Frequent attendance in primary care: Comparison and implications of different definitions. Br J Gen Pract Published Online First: 2010. doi:10.3399/bjgp10X483139

10 Chang B, Hoaglin DC. Meta-analysis of odds ratios: Current good practices. Med Care Published Online First: 2017. doi:10.1097/mlr.0000000000000696

11 Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Cochrane Collab Published Online First: 2011. doi:10.1002/9780470712184.ch4

12 DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials Published Online First: 1986. doi:10.1016/0197-2456(86)90046-2

13 Bero LA. Why the Cochrane Risk of Bias Tool Should Include Funding Source as a Standard Item. In: Cochrane Database of Systematic Reviews. 2014. doi:10.1002/14651858.ed000075