Title: The Prevalence of Depression in General Hospital Inpatients: A Systematic Review and Meta-Analysis of Interview Based Studies

Authors: Jane Walker PhD\(^1\), Katy Burke MB BS \(^1\), Marta Wanat PhD\(^1\), Rebecca Fisher MBBChir\(^3\), Josephine Fielding MBChir\(^1\), Amy Mulick MSc\(^2\), Stephen Puntis DPhil\(^1\), Joseph Sharpe BA\(^1\), Michelle Degli Esposti BA\(^1\), Eli Harriss MSc\(^4\), Chris Frost MA\(^2\), Michael Sharpe MD\(^1\).

Affiliations:
1. Psychological Medicine Research, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK.
2. Department of Medical Statistics, London School of Hygiene and Tropical Medicine
3. Department of Primary Care Health Sciences, University of Oxford

Correspondence: Dr Jane Walker PhD, Psychological Medicine Research, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, OX3 7JX, UK.
Telephone: +44(0)1865 226 477, Fax: +44(0)1865 793 101
Email: jane.walker@psych.ox.ac.uk

Word count: 3514 including in-text citations (abstract 236)
ABSTRACT

Background

Comorbid depression in the medically ill is clinically important. Admission to a general hospital offers an opportunity to identify and initiate treatment for depression. However we first need to know how common depression is in general hospital inpatients. We aimed to address this question by systematically reviewing the relevant literature.

Methods

We reviewed published prevalence studies in any language which had used diagnostic interviews of general hospital inpatients and met basic methodological quality criteria. We focussed on interview-based studies in order to estimate the proportion of patients with a diagnosis of depressive illness.

Results

Of 158 relevant articles, 65 (41%) describing 60 separate studies met our inclusion criteria. The 31 studies that focussed on general medical and surgical inpatients reported prevalence estimates ranging from 5% to 34%. There was substantial, highly statistically significant, heterogeneity between studies which was not materially explained by the covariates we were able to consider. The average of the reported prevalences was 12% (95% C.I. 10% to 15%), with a 95% prediction interval of 4% to 32%. The remaining 29 studies, of a variety of specific clinical populations, are described.

Conclusions
The available evidence suggests a likely prevalence high enough to make it worthwhile screening hospital inpatients for depression and initiating treatment where appropriate.

Further, higher quality, research is needed to clarify the prevalence of depression in specific settings and to further explore the reasons for the observed heterogeneity in estimates.
Depression frequently accompanies physical illness (Vos et al., 2012). It is clinically important because it is associated with worse physical symptoms, poorer quality of life and greater functional disability (Egede, 2007, Katon, 1996). Patients with depression also spend more time in hospital, are less likely to adhere to medical treatments, and consequently incur higher healthcare costs (DiMatteo et al., 2000). Despite its importance the management of depression comorbid with physical illness is often suboptimal with low rates of detection and treatment (Balestrieri et al., 2002, Cepoiu et al., 2008, Hirschfeld et al., 1997, Katon and Sullivan, 1990, Kessler et al., 1999, Walker et al., 2014).

Admission to a general hospital therefore provides an important opportunity to improve the management of comorbid depression. Detection can be improved by incorporating systematic screening into hospital admission procedures and treatment can be initiated where appropriate (Beach et al., 2015). However, in order to plan such a management strategy we need to know how common depression is in general hospital inpatients. Surprisingly, we currently lack a clear answer to this basic question. This is largely because previous systematic reviews of the prevalence of depression have focussed on study populations with specific physical diseases such as cancer, rather than on specific clinical settings such as general hospital wards (Castro-de-Araujo et al., 2013, Craig et al., 2009, Davydow et al., 2009, Delville and McDougall, 2008, Kouwenhoven et al., 2011, Mitchell et al., 2011, Poynter et al., 2009, Shanmugasegaram et al., 2012, Singer et al., 2013, Thombs et al., 2006a, Thombs et al., 2006b, Walker et al., 2013, Wilder Schaaf et al., 2013, Wiseman et al., 2013).
We therefore aimed to conduct a systematic review of studies of the prevalence of depression in general hospital inpatients. We focussed on interview-based studies in order to estimate the proportion of patients with a definite diagnosis of depressive illness.

**METHODS**

**Search strategy and selection criteria**

We performed a systematic review of studies of the prevalence of depression in general hospital inpatients, using procedures that accorded with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher et al., 2009). We identified studies by searching EMBASE, Medline and PsycINFO (from 1974, 1946 and 1967 respectively) to December 2015. Searches were run for the combination of ‘prevalence’, ‘general hospital inpatient’ and ‘depression’ using both standardised subject terms and free text terms, including synonyms and alternative spellings. We provide full details of the searches used in the online appendix. We also manually searched the reference lists of review articles obtained through the electronic searches.

We judged studies to be relevant to the review if they met all the following criteria: (1) the study clearly aimed to estimate the prevalence of depression (i.e. studies that were designed to address a different research question but happened to include a prevalence estimate, such as clinical trials or questionnaire validation studies, were not included); (2) all study participants were adults (aged 16 or older); (3) all study participants (or a clearly defined subgroup for which there was an estimate of depression prevalence) were general hospital inpatients at the time of depression assessment; (4) the presence of depressive illness was determined using a diagnostic interview. We only included primary studies (i.e.
not reviews) for which we could obtain the full paper to allow data extraction. No language restrictions were applied.

When selecting publications to include in the review, we also applied quality criteria to the reported study methods. In order to ensure a consistent and transparent approach to this quality assessment we used a checklist based on the work of Loney et al. (Loney et al., 1998, Walker et al., 2013). We used a checklist rather than a continuous scale to ensure that all the key aspects of the study methods met basic quality criteria (Juni et al., 1999). The basic methodological standards required for inclusion were: (1) the study sample was obtained using a random or consecutive sampling method; (2) data were available for analysis on at least 70% of the eligible patients (either as reported by the authors or derived from presented data); (3) depressive illness was defined using standard diagnostic criteria: major depression from the Diagnostic and Statistical Manual of Mental Disorders (DSM), depressive episode from the International Classification of Diseases (ICD) or similar (American Psychiatric Association, 1994, 2013, World Health Organization, 1992). The first two of these criteria aimed to minimise selection bias, and the third aimed to ensure that estimates could be compared across studies.

Data collection

We screened the titles and abstracts of all articles identified by the searches to determine whether each might meet the selection criteria. We then reviewed the full text of the article, with the help of a translator where necessary, if there was any possibility that it might be relevant and would meet our quality criteria. This process (including screening of
titles and abstracts) was conducted independently by two researchers with reference to a third researcher to resolve disagreements.

Two researchers independently extracted the following data from all the articles included in the review, using a specially designed, standardised data extraction form: country in which the study took place; age, sex and clinical characteristics of participants; sample size; type of depression interview used and profession of interviewer; diagnostic criteria used to determine the presence of depressive illness; prevalence of depression in the sample (for cohort studies, we extracted the prevalence of depression at the first time point only).

**Clinical analysis**

Two researchers reviewed the data extracted on participants’ clinical characteristics in order to assess their similarity across studies. We found that there was high clinical heterogeneity, indicating that a meta-analysis of all the studies would not yield meaningful results. Whilst some studies had recruited general medical and surgical inpatients, others had recruited inpatients with very specific clinical characteristics, and therefore were of samples unrepresentative of the general hospital inpatient population. In order to deal with this clinical heterogeneity, we restricted our statistical analysis to studies of general medical and surgical inpatients. The studies of more specific patient groups are described in our results and online appendix in order to provide the reader with a comprehensive overview of the relevant literature.

**Statistical analysis**
We used forest plots to display the proportion (with exact binomial 95% confidence intervals) of participants diagnosed with depression in each study (Newcombe, 1998).

We used random-effects models to describe the prevalence of depression in general medical and surgical inpatients. 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 exactly 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 (Stroup et al., 2000). Random effects models assume that the populations investigated in each study are themselves drawn from a wider population of populations and that the underlying study-specific prevalences in these populations therefore follow a statistical distribution, rather than taking a single value.

As is common for proportions, we used the logit transformation expressing each of the prevalences as a log-odds. Accordingly our random-effects models assume that the logit transformed prevalences follow a normal distribution with a mean and standard deviation. This mean can be thought of as a “typical” prevalence, while the standard deviation quantifies the underlying between-study variability in prevalences. This variability is summarised by a 95% prediction interval, which is the interval within which 95% of underlying study-specific prevalences are predicted to lie (for a thorough discussion of this topic see Guddat et al.) (Guddat et al., 2012). As such it differs from a 95% confidence interval which quantifies the precision of the mean of the study-specific prevalences (with the mean defined after logit transformation).
We used 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 (Higgins et al., 2003). The assumption that underlying prevalences are normally distributed after logit transformation was not contradicted by our data.

We investigated potential sources of the heterogeneity that we observed between the studies’ prevalence estimates (that is, the large amount of between-study variability compared with the total variability) by considering some of the known differences between the studies. To this end, we inspected scatter plots of depression prevalence against year of study publication, sample size, average (or other available measure for central tendency) participant age, and percentage of female participants. We used forest plots to compare depression prevalence in studies grouped by use of DSM major depression versus other diagnostic criteria for depression, and national income of the country where the study took place (we used income groupings because the studies had been done in too many different countries to group by country) (The World Bank, 2015). Where evidence of an association with depression prevalence was apparent, we performed a mixed-effects meta-regression and present its I-squared statistic, odds ratio and p-value for the association (Thompson and Higgins, 2002). We did not present funnel plots for bias assessment because in the presence of heterogeneity, there is no reason to expect a funnel shape (Terrin et al., 2005). Statistical analysis was performed in R v3.2.2 using the “meta” package v3.8-0 (R Core Team, 2015, Schwarzer et al., 2014). Graphs were produced in R and Stata v14 (StataCorp, College Station, TX, USA).
RESULTS


These studies had been conducted in 29 countries (see appendix for map) and had included a total of 12,540 participants (median sample size 109, range 27 to 993).
A variety of interviews and associated diagnostic criteria were used. The most commonly used interview (16 studies) was the Structured Clinical Interview for DSM-IV (SCID) and the most commonly used diagnostic criteria (used in 47 studies) were those for DSM major depression (American Psychiatric Association, 1994, First et al., 1996). The majority of studies (40) had employed a psychiatrist or clinical psychologist to conduct the diagnostic interviews.

The study sample was of general medical or surgical inpatients (or both) in 31 of the studies (median sample size 215, range 65 to 993, with a total of 9,305 participants; see Table 1) (Abiodun and Ogunremi, 1990, Arolt and Driessen, 1996, Arolt et al., 1997, Feldman et al., 1987, Fenton et al., 1994, Hosaka et al., 1999, Jenkins et al., 1994, Kathol and Wenzel, 1992, Kayhan et al., 2013, Kigamwa, 1991, Koenig et al., 1997, Koenig et al., 1991, Koenig et al., 1993, Kok et al., 1992, Kok et al., 1995, Koroglu and Tural, 2010, Kumar et al., 2011, Lazaro et al., 1991, Lazaro et al., 1995, Linka et al., 1999, Linka et al., 2000, Marchesi et al., 2004, Moayedoddin et al., 2013, Nair and Pillay, 1997, Pakriev et al., 2009, Seltzer, 1989, Sharma et al., 2002, Silverstone, 1996, Soeiro et al., 2008, Thalassinos et al., 1992, Topitz et al., 2015, Uwakwe, 2000, Wancata et al., 2000, Yan et al., 2013, Zhong et al., 2010). These studies reported prevalence estimates for depression that ranged from 5% to 34% (see Figure 3). The high heterogeneity observed between study findings (I-squared 90%) indicated that no single estimate was sufficient to describe the prevalence of depression in general medical and/or surgical inpatients. Our random-effects model assumed that the underlying study-specific prevalences followed a normal distribution (on the log-odds scale).
The mean of this distribution corresponded to a prevalence of 12% (95% C.I. 10% to 15%) with 95% of all populations predicted to have an underlying depression prevalence between 4% and 32% (the prediction interval).

[Table 1 and Figure 3 about here]

In our investigations of potential sources of this observed heterogeneity visual inspection of the scatter and forest plots suggested that percentage of female participants, study sample size, the income band of the country in which the study was done, and the diagnostic criteria used (but not year of study publication or average participant age), may all be associated with the observed prevalence of depression. We therefore tested the association of these variables with depression prevalence. We found that when expressed as an odds, studies with a higher percentage of female participants reported a lower prevalence of depression (OR 0.82 per 10 percentage points increase in female participants, 95% C.I. 0.71 to 0.95, p=0.007). Studies with larger sample sizes reported lower prevalences (OR 0.82 per doubling in size, 95% C.I. 0.68 to 0.99, p=0.043). There were also non-significant associations with national income in the country in which the study was done (p=0.292), and the diagnostic criteria used for depression (p=0.154). Notably, in all our investigations the residual heterogeneity remained high (all I-squared >88%, see appendix for scatter plots and forest plots) meaning that a very high proportion of the heterogeneity remained unaccounted for by the variables we considered.

In addition to the 31 studies of general medical and/or surgical patients, we identified 29 studies (median sample size 72, range 27 to 502, with a total of 3,235 participants) of
inpatients who were in a variety of specialist units (such as endocrinology or haematology)
or had very specific clinical characteristics (such as a diagnosis of systemic sclerosis)
These studies reported prevalence estimates ranging from 2% to 56%. They are described in
detail in the online appendix.
This is the first systematic review of studies of the prevalence of depression in general hospital inpatients. The 60 studies that we found had been conducted in 29 countries and included a total of 12,540 participants. They reported a wide range of prevalence estimates. We reduced the clinical heterogeneity by focusing on the 31 studies of general medical and/or surgical inpatients. However even in these studies the estimated prevalence ranged from 3% to 34%. There was also a high degree of heterogeneity, indicating that even ‘general medical and/or surgical inpatients’ cannot be considered as a single population, but rather as a number of different populations, each with a different prevalence of depression. These populations had a median prevalence of depression of 12% and we can predict that 95% of them have a prevalence between 4% and 32%. This median prevalence of 12% is more than twice that in the general population, for which international studies suggest an average 12-month prevalence of approximately 5% (Kessler and Bromet, 2013).

Our analyses were unable to adequately explain the observed heterogeneity in prevalence estimates. The only variables that we found to be significantly associated with prevalence of depression were sample size and the proportion of female patients in the study samples. As these explained only a trivial amount of the heterogeneity and the latter association was not in the expected direction (a higher proportion of female patients was associated with a lower prevalence of depression), we judge this finding to be of questionable importance. Our inability to explain the observed heterogeneity indicates that it resulted from variables we were unable to investigate as they had not been consistently reported in the publications we reviewed. These unreported variables might be at the population, healthcare system, patient or methodological level. At the population level, it is likely that
national and local prevalence of depression in the general population varies. At the healthcare system level, hospital type (e.g. university, community), funding systems, admission pathways and medical staffing vary substantially. At the patient level, it is likely that the characteristics of patients admitted to general hospitals, and specifically to general medical and surgical (as opposed to sub-speciality) wards varies. Methodologically, there are likely to be unreported variations in how the studies were done. These include how the patients were sampled (e.g. who was excluded), how the diagnosis of depression was made (e.g. the details of how the diagnostic interviews were conducted, whether physical symptoms were counted toward the diagnosis of depression or not and exactly how the diagnostic criteria were applied) and when the assessment was done during the period of hospitalisation (e.g. soon after admission or later in the stay).

This review has strengths which include: (a) The use of clearly defined inclusion criteria for papers to minimise selection bias; (b) the focus on studies where the diagnosis of depression was made by interview; (c) the exclusion of studies with major design flaws (Moher et al., 2009, Stroup et al., 2000). It also has limitations which include: (a) a reliance on the published reports to assess studies’ relevance and quality, which may potentially have led to us excluding studies that were in fact well conducted, but poorly reported; (b) our inability to investigate all potential sources of heterogeneity because of the limited potentially relevant data reported in the publications we reviewed.

Given the importance of the question we addressed in this review, we found a remarkably small literature, much of which was published some time ago. Furthermore, our quality assessments indicated that much of that literature was of poor quality. Common
shortcomings were poor sampling strategies and the use of unclear case definitions for depressive disorder. Even the methodologically better studies selected for inclusion in this review were mostly small in size (median sample size 109) by epidemiological standards. There is consequently a clear and pressing need for better quality studies of the prevalence of depression in medical inpatients. If these are to inform service planning these should aim both to determine the prevalence of depression in specific settings (such as National Health Service hospitals in the UK) and to clarify the determinants of the substantial apparent variations in prevalence noted in this review. Suggestions for the design of future studies are given in Table 2.

Despite the limitations of the available evidence we can reasonably conclude that depression is sufficiently common in medical inpatients to make planning for its systematic management worthwhile. This management should include systematic identification of depression during hospital admissions, monitoring of depression once identified (during the hospital admission and thereafter, to determine whether it resolves post-discharge) and the initiation of treatment when it is persistent (Kathol and Wenzel, 1992, Mayou et al., 1988). Few hospitals currently have such systems. The approach we have tested for depression management in cancer patients provides a potential model for how we might improve depression care in all medical settings (Walker and Sharpe, 2014).
FINANCIAL SUPPORT

This work was supported by the Oxford Academic Health Science Network; Sir Michael Sobell House Hospice, Oxford; and the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

CONFLICTS OF INTEREST

None.
REFERENCES


R Core Team (2015). R: A language and environment for statistical computing.: Vienna, Austria.


World Health Organization (1992). *International Classification of Mental and Behavioural Disorders (ICD-10)*.

