The Synthesis of Prevalence and Incidence Data

Zachary Munn
Sandeep Moola
Karolina Lisy
Dagmara Riitano
Thank you for purchasing this e-book. To receive special offers and news about our latest products, sign up below.

Sign Up

Or visit LWW.com
The Synthesis of Prevalence and Incidence Data

Zachary Munn
Sandeep Moola
Karolina Lisy
Dagmara Riitano

Lippincott-Joanna Briggs Institute Synthesis Science in Healthcare Series: Book 24
The Synthesis of Prevalence and Incidence Data

The Lippincott-Joanna Briggs Institute Synthesis Science in Healthcare Series

Series Editor: Professor Alan Pearson AM

This series of concise texts is designed to provide a “toolkit” on synthesizing evidence for healthcare care decision-making and for translating evidence in action in both policy and practice. The series seeks to expand understandings of the basis of evidence-based healthcare and brings together an international range of contributors to describe, discuss, and debate critical issues in the field.

Incredible developments have occurred in the synthesis and use of evidence in healthcare over the last several years, but the science and emerging practices that underpin evidence based healthcare are often poorly understood by policy makers and health professionals. Several emerging and exciting developments have much to offer health professionals. First, new, deeper understandings of the nature of evidence and of ways to appraise and synthesize it have led to the development of more sophisticated methodologies for synthesis science. Second, the realization that the rapid increase in the availability of high quality evidence has not been matched by increases in the translation of this evidence into policy and/or clinical action has spurred on developments in the science of knowledge implementation and practice improvement.

The burgeoning publications in this area – particularly books on evidence based healthcare – can go only so in informing responsible and conscientious policy makers and healthcare practitioners. This new series Lippincott/Joanna Briggs Institute, "Synthesis Science in Healthcare," is devoted to communicating these exciting new interventions to researchers, clinicians on the frontline of practice and policy makers.

The books in this series contain step-by-step detailed discussions and practical processes for assessing, pooling, disseminating and using the best available international evidence. In all healthcare systems, the growing consensus is that evidence-based practice offers the most responsible course of action for improving health outcomes. All clinicians and health scientists want to provide the best possible care for patients, families, and communities. In this series, our aim is to close the evidence to action gap and make that possible.
About the Authors

Dr. Zachary Munn, BMedRad (NM), GradDip HlthSc, PhD, is a Senior Research Fellow in Implementation Science at the Joanna Briggs Institute, School of Translational Health Science, Faculty of Health Sciences, the University of Adelaide. He has a clinical background in medical radiation and his particular interests focus on the synthesis of research evidence and the translation of evidence into practice, and particularly in the systems and processes that can be used to achieve this. He has a keen interest in a number of research methodologies, and is involved in a range of research and teaching activities within the Joanna Briggs Institute. Additionally, he currently convenes a methodological group investigating processes for synthesizing evidence of incidence and prevalence.

Sandeep Moola, BDS, MHSM (Hons), MPhil, is a Research Fellow in Implementation Science at the Joanna Briggs Institute (JBI), School of Translational Health Science, Faculty of Health Sciences, the University of Adelaide. Sandeep has a clinical background in dentistry. His role at JBI is varied in that it involves undertaking systematic reviews and literature reviews and writing evidence summaries, systematic reviews and ‘Best Practice Information Sheets’. He also administers and manages ‘Surgical Services’ and ‘Tropical and Infectious Disease’ nodes of JBI COnNECT Plus and assists in developing online short course modules. He has contributed to various consultancy research projects funded by external organizations like Cancer Australia and NSW Health. He is part of the new methodological group investigating processes for synthesizing evidence of incidence and prevalence.

Dr. Karolina Lisy, BSc (Biotech), PhD, is a Research Fellow in Communication Science at the Joanna Briggs Institute and School of Translational Health Science at The University of Adelaide. Karolina has a background in biomedical research with a focus on cellular signaling mechanisms and cancer, and a current interest in the translation of primary research results into policy and practice change. She is currently involved in a variety of research projects within the Joanna Briggs Institute, and also has roles in teaching systematic review methodology, disseminating evidence as managing editor of the JBI Database of Systematic Reviews and Implementation Reports, and in advancing methodologies of data synthesis.

Dagmara Riitano, BA, GradDip(Psych), BHlthSc(Hons), is a Research Officer in the Synthesis Science program at The Joanna Briggs Institute, School of Translational Health Science, Faculty of Health Sciences, at The University of Adelaide. She has a background in providing high-level research support to a range of government and non-government research projects and consultancies across the health sciences. This has included playing a key role in the development and co-ordination of a series of deliberative stakeholder engagements to in-form implementation approaches to healthcare disinvestment as part of the NHMRC funded ASTUTE health study.
Introduction

The amount of literature in the health and social fields has increased at an exponential rate over the last 30 years and each year hundreds of thousands of new articles are being published in social and health journals. This increase in research output has been paralleled with a growing focus on the need for healthcare decisions, policies and funding to be based on the best available evidence, with consideration of patient preferences, clinical expertise and available resources. This need to base health and social care policy on evidence from research is now well-accepted internationally and is seen as the ideal way to practice healthcare. However, in real world settings this is not always the case.

There are many barriers that inhibit the uptake of research evidence into practice, one of which is the difficulty for the practicing professional to keep up to date with the expanding body of literature. The systematic review of evidence has been proposed and is well now accepted as the best method to summarize the literature relating to a certain social or healthcare topic. The systematic review is a type of research design that provides a reliable summary of the literature to assist health professionals to keep up to date. Key features of a systematic review include the creation of an a priori protocol, clear inclusion criteria, a structured and systematic search process, critical appraisal of studies, and a formal process of data extraction followed by methods to synthesize, or combine, this data. In this way, systematic reviews extend beyond the subjective, narrative reporting characteristics of a traditional literature review to provide a comprehensive, rigorous and transparent synthesis of the literature on a certain topic.

Historically reviews have focused on the synthesis of evidence of effects, particularly trying to establish the effectiveness of various treatments on social and health outcomes. However, decisions made in social and health care require more information than can be provided by the simple question “does this work?” As systematic review methodology has evolved so have the types of evidence that have been synthesized using this approach. There now exists methods and guidance for conducting reviews of various forms of evidence, including qualitative research, cost data, diagnostics, prognostics, harms and risk.

Whilst the steps included in the systematic review process are mirrored across the various types of evidence, there are important considerations that need to be taken into account when conducting a systematic review pertaining to the type of research to be synthesized. There are established methods for conducting meta-analyses of randomized controlled trials (RCTs) and some observational study designs. However, no clear guidance currently exists on synthesizing frequency data from incidence and prevalence estimates. This book seeks to fill this gap by outlining methods and guidance for an emerging type of systematic review, that of prevalence and incidence data. Prevalence and incidence data systematic reviews are becoming more important as policy makers realize the usefulness of syntheses of this type of information. Synthesis of this type of information has the potential to better inform social and healthcare professionals, policy makers and consumers in decisions made relating to a range of healthcare decisions, but particularly regarding the burden of healthcare both now and in to the future. This book provides comprehensive guidance to authors wishing to conduct systematic reviews on prevalence and incidence data.
Chapter 1: Prevalence and Incidence Data

The accurate measurement of disease among populations, whether at a local, national, or global level, is of critical importance for governments, policy makers, health professionals and the general population to inform the development, delivery and use of health services. Accurate information regarding measures of disease can assist in planning management of disease services (by ensuring resources are available to cope with the burden of disease), set priorities regarding public health initiatives, and evaluate changes and trends in diseases over time.

There are a number of measurements of disease, including determining the proportion of a population who have a certain disease at a specific point in time, or within a given time frame (the prevalence) and how often a disease occurs within a given time frame (the incidence). Questions of disease prevalence and incidence are often asked by researchers and the findings of their research provide crucial data for policy makers. Both of these measures enable health researchers to quantify disease amongst populations (Webb, Bain, & Pirozzo, 2005).

Prevalence

Point prevalence

The prevalence of a disease indicates the number of people in a population that have the disease at a given point in time, and is defined as the number of people who have a disease at a point in time divided by the number of people in the population at that time. This is also known as point prevalence as it provides a picture of what is occurring at a certain point of time. The point prevalence is calculated using the following formula (Webb et al., 2005):

\[
\text{Prevalence} = \frac{\text{Number of people with disease in the population at a specific point in time}}{\text{Total number of people in the population at that time}}
\]

This is often presented as a proportion or percentage, or as the number cases per unit of a population. As point prevalence provides a snapshot in time of a specific disease or condition in a population, it is often used to measure chronic diseases or conditions. For example, the prevalence of HIV/AIDS in Australia in 2011 was 0.155%, or 115 out of every 100 000 people in Australia (The Kirby Institute, 2012). Please note that prevalence here does not have to refer only to prevalence of disease; prevalence of other variables such as a symptom, event, or risk factor may also be measured.

Point prevalence data should be interpreted carefully as the duration of a disease will have an impact on its prevalence in a population at any specific point in time. Diseases that are incurable, take a long time to cure, or have a long survival time may have a higher prevalence in a population at a given point in time than diseases that can be either cured quickly, or that quickly lead to death. For example, a disease with a long survival time will affect more people in a population at a single point in time compared with a disease that rapidly causes the death of affected individuals. Likewise, a disease that takes a long time to treat and cure will have a
higher point prevalence compared with a disease that is cured rapidly. To continue with the earlier example of HIV, improvements in HIV therapies that enable affected individuals to live longer with their condition actually increase the prevalence of HIV at a given point in time by delaying deaths of HIV positive people.

**Period Prevalence**
Period prevalence is similar to point prevalence, except that it assesses the proportion of a population that had a disease at any time within a specified period of time. Rather than asking the number of people who currently have a disease, it asks “how many people had the disease in a defined period of time?” As an example, period prevalence could be used to describe the number of people who had influenza in a one year period. For an illness of short duration, period prevalence is more useful than point prevalence as it captures the total number of cases that occurred over the time period of interest, including cases that were cured or otherwise resolved over the specified time period.

**Lifetime prevalence**
A third measure of disease prevalence is lifetime prevalence, and this measures the proportion of individuals in a population that had a disease at some point over the course of their lifetimes. Lifetime prevalence is therefore not affected by the time taken to cure a disease or time taken for affected individuals to die from the disease, and is most often used to measure the prevalence of relatively rare diseases or conditions.

**Incidence**
The incidence of a disease indicates how many new cases of a disease occur during a defined period of time. Incidence data is obtained through tracking of an initially disease-free cohort and determining the number of those who developed the disease in that cohort over a period of time. There are two main types of incidence used in epidemiological studies: cumulative incidence (sometimes called incidence proportion, attack rate or risk) and person-time incidence rate (also called incidence density). Both cumulative incidence and person-time incidence rate are both often referred to simply as incidence rate or incidence, which can sometimes make understanding incidence measures more difficult than prevalence. Incidence is not affected by disease duration; it simply indicates the number of new occurrences of a disease in a certain period of time. The incidence of a disease in a one-year period, for example, would be found using the equation (Webb, Bain, & Pirozzo, 2005):

\[
\text{Incidence} = \frac{\text{Number of people who develop disease in one year}}{\text{Average number of people in the population in the same year}}
\]

For example, the number of new HIV diagnoses, or the incidence of HIV, in Australia in 2010 was 1,031 total cases. This means that during 2010, 0.0055% of the population was newly diagnosed with HIV, or in other words, approximately 5.5/100,000 people were newly diagnosed (Australian Bureau of Statistics, 2012).

**Cumulative Incidence**
Cumulative incidence measures the total number of new cases of a disease in a defined population over a specified period of time; it gives the number of people who were at risk of developing a certain disease and who did develop the disease, in a certain time period. Cumulative incidence can be presented as either a proportion or percentage, or number of cases out of the population. This can be calculated using the following equation:
Cumulative incidence = \textbf{Number of people who develop disease in a specified time period} \\
\text{Size of population initially at risk of developing disease at the start} \\
\text{of that time period}

The population at risk should exclude people who cannot get the disease. For example, if the disease of interest is cervical cancer, the population at risk would not include men. Also excluded should be people who already have the disease, and people who cannot contract the disease, either through prior exposure or immunization.

To measure cumulative incidence, one assumes that the population remains constant over the observed period of time. In these so-called closed populations, all of those who are included in the population group at the beginning of the time period are also included in the count at the end of the time period. Closed populations do not necessarily refer to settings or locations; a closed population may also be a group of people who share a common characteristic, such as people who shared an exposure or who are all taking a certain drug.

\textbf{Person-time Incidence Rate (Incidence Density)}

To calculate incidence in an open population, where over time individuals flow in and out of the population at risk, it is necessary to count the amount of time in total that each person in the population spent at risk of developing the disease. Examples of open populations include populations of whole countries, where the population will change as people immigrate to and from the country, or as people are born into the population and also as people die, or the patients in a particular hospital over a given period of time who have varying lengths of stay. Rather than counting the number of new cases per the number of people in the population, the person-time incidence rate measures the number of disease cases per person-time (often person-years) spent at risk of developing the disease. To put it another way, the person-time incidence rate is the number of new disease cases over the sum of the length of time each individual was at risk in the population.

The person-time incidence rate of a disease can be found using the equation:

\text{Person-time incidence} = \frac{\text{Number of people who develop disease in a specified time period}}{\text{Person-time when people were at risk of developing a disease}}

For example, if measuring the incidence of prostate cancer in 1000 men over 10 years, and 80 men developed prostate cancer over that 10-year period, this can be presented as 80 per 10000 (1000x10) or 8/1000 person years.

\textbf{Relationship between prevalence and incidence}

There is a relationship between prevalence and incidence. As described above, the prevalence of a disease in a population is dependent on the duration of the disease, and it is also dependent on the incidence of the disease. For example, two diseases can have a similar incidence, but if the duration of disease is longer for one of them (perhaps because it is a chronic or longer lasting condition compared to an acute condition), than the prevalence will be much higher in the condition with the longer duration (Webb et al., 2005). To give another example, an incurable disease may have a high incidence in a given year, and then the incidence may decrease the following year due to the introduction of a new preventive measure. In this scenario, the prevalence will remain the same or even slightly increase, even if the incidence has dramatically decreased. This is because even though the number of new...
The Synthesis of Prevalence and Incidence Data

cases has dropped, the disease is incurable meaning that the number of people who have the disease the following year will remain high, only to decrease as people with the disease gradually die and leave the population. It is clear to see that here, the prevalence of the disease is not indicative of the success of the prevention initiative. Due to the effect of disease duration on prevalence, prevalence data must be interpreted cautiously, particularly when seeking correlations between disease prevalence and other variables.

Use of Observational Studies in Healthcare

Now that the measures of prevalence and incidence of disease have been discussed, it is important to detail the kinds of studies that are useful in obtaining this kind of data. Whilst RCTs are the best study design to answer questions of the effectiveness of interventions due to their ability to determine causality, they are not ideally suited to provide data of rates and patterns of disease occurrence. Having said that, certain prevalence or incidence data can be gleaned from RCTs, as discussed further below.

To address issues regarding prevalence and incidence, epidemiological study designs, such as those classified under the term observational and descriptive studies, are required. Observational studies do not involve manipulation on the part of the researcher. These studies rely on the natural or ecological events of exposures and disease, where the researcher simply observes certain characteristics of the sample population as they occur naturally, and records the relevant data. Observational studies can therefore be distinguished from experimental or quasi-experimental studies such as RCTs and controlled clinical trials, where there is manipulation of the independent variable (or the intervention) by the researcher. Observational studies have a number of advantages over experimental study designs and are particularly valuable in instances where conduct of an RCT is unethical. It is certainly unethical to conduct an RCT to investigate the effects of a variable that is thought to be harmful, such as the effect of alcohol use during pregnancy for example, or the effect of asbestos exposure. Questions such as these can only be addressed using observational studies, where an exposure, behavior or event occurs and the researcher observes participants over time to investigate any outcomes.

Observational studies address questions such as: how many people have the disease? Who is getting the disease? Where is the disease occurring? This kind of information is particularly valuable for governments when making decisions regarding health policy and planning. Furthermore, observational studies can often be used to infer correlations between two variables, for example, between a variable such as an exposure, risk factor or protective factor, and a disease outcome. Data from observational studies can therefore be useful in enabling the formation of hypotheses regarding risk or preventive factors in disease development and progression. It is important to note that these studies are not able to determine causality; rather they are able only to infer correlations or relationships between variables.

Observational Study Designs

Observational study designs include prospective and retrospective cohort studies, case-control studies, cross-sectional studies, case series, and case reports, and can be broken down into the broad categories of analytical studies and descriptive studies (Figure 1). Generally, descriptive studies describe the occurrence/presence of an outcome only, whereas analytical studies describe the relationship between a variable and an outcome. Some observational studies may report both analytic and descriptive data particularly in the case of certain case-studies and cross-sectional studies. Due to the nature of observational study designs, they are more at risk of confounding factors and different sources of bias that are unavoidable, which
will be discussed further below. Despite this, observational studies are essential in answering questions of prevalence and incidence.

**Figure 1: Sources of Prevalence and Incidence Data According to Study Type**

<table>
<thead>
<tr>
<th>Sources of Prevalence and Incidence Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive</strong></td>
</tr>
<tr>
<td>- Case studies</td>
</tr>
<tr>
<td>- Case series</td>
</tr>
<tr>
<td>- Population Surveys</td>
</tr>
<tr>
<td>- Population Censuses</td>
</tr>
<tr>
<td><strong>Analytical</strong></td>
</tr>
<tr>
<td>- Cohort studies</td>
</tr>
<tr>
<td>- Case-control studies</td>
</tr>
<tr>
<td>- Cross-sectional studies</td>
</tr>
<tr>
<td><strong>Experimental</strong></td>
</tr>
<tr>
<td>- Randomized controlled trials</td>
</tr>
<tr>
<td>- Pseudo-randomized trials</td>
</tr>
</tbody>
</table>

**Descriptive studies**

Descriptive studies aim to collect certain information about a given individual or group and can be used to provide information on the distribution of disease. Examples of descriptive study designs include case reports and case series. In healthcare, these types of studies are typically used to describe the occurrence of disease or a risk factor. Descriptive studies have value and powerful examples from history illustrate their place in the study design hierarchy. Other descriptive data sources include population censuses, which collect data on whole populations, and survey data.

**Case series**

Case reports describe the details of individual patients, and are often used to report unusual or unexpected occurrences of a disease, or a unique finding. These types of reports are not particularly useful in providing prevalence or incidence data, however, when a group of linked cases are presented as a case series, they can be informative for rare or emerging diseases. For example, a case series detailing 41 cases of Kaposi’s Sarcoma among homosexual men in California and New York provided the first evidence of what would later be found to be HIV infection (Centers for Disease Control, 1981).

**Analytical studies**

Analytical studies seek to make comparisons between different subgroups of the population and ask why a disease occurs in a subset of people or has a particular distribution. For example, a study may involve two groups; one group may have a disease and the other group may not, or one group may have experienced a certain event or been exposed to something,
and another group have not. By assessing relationships between certain exposures, or indeed any variable such as behavior, risk factors, or characteristics, and a disease, researchers can form hypotheses regarding why disease occurs.

**Cohort studies**

Cohort studies are longitudinal studies that are typically used to analyze relationships between exposures and disease by comparing the outcomes between two groups over time, where individuals in one group were exposed to a common event or share a certain risk factor, and the other group does not. Sampling in cohort studies is based on the presence or absence of a certain exposure, and participants are followed over time to observe development of any disease outcomes. A prospective cohort study begins prior to or following an exposure, and participants are followed forward through time to observe any outcomes that may occur. A retrospective cohort study begins after outcomes are already known. A researcher sifts through patient records and groups patients according to exposures, and identifies any differences in outcomes.

**Case-control studies**

Case-control studies, on the other hand, select participants based on presence of disease or a specific condition, and look for prior exposures or possible risk factors that may have led to the disease or condition developing. In this study design, those with the disease (cases) are matched with comparable individuals who do not have the disease (controls), and both groups are studied to determine if any differences in characteristics or past exposures exist.

As the study population in case-control designs is based on individuals already having a disease, rates of prevalence or incidence of disease cannot be derived from these studies.

**Cross sectional studies**

Cross sectional studies are used to describe characteristics of a population at a given point in time, and as such provide a single snapshot of disease and other variables at one point in time. Prevalence studies describe the health of a population. Cross sectional studies, or surveys of a ‘cross section’ of society at a point in time, are the study designs primarily used to determine the prevalence of a disease.

For example, cross sectional studies can be used to describe the prevalence of a disease and a risk factor at a specific point in time and can be carried out using a survey. Cross sectional studies have been used to examine the relationship between diseases (or other health-related characteristics) and other variables of interest as they exist in a defined population at one particular time (Last, 1995). Data can still be used to infer relationships between a disease and other variables, however as the data is gathered simultaneously, chronological sequences of exposures and outcomes cannot be determined.

These types of studies collect information from sample populations that are selected regardless of exposure or disease status; they seek to describe the prevalence of exposure or disease in the population as a whole. Therefore, for this study design it is crucial that the sample obtained for the study represents the population as a whole, and random sampling methods should be used. Rare diseases can be problematic as they may require studies with particularly large sample sizes.

**Randomized-Controlled Trials**

Lastly, data from RCTs can also be used to derive prevalence data to answer certain questions. For example, if one was interested in ascertaining the prevalence of a certain
symptom or event related to a condition, for example fatigue in type 2 diabetics, using baseline data from an RCT with a sample of patients with type 2 diabetes, where the prevalence of dizziness is measured at baseline, would be acceptable. However, systematic reviewer's need to carefully check the inclusion criteria when considering including experimental studies, as they are often quite strict and may not be reflective of the target population. As the conduct of the subsequent trial would not be of interest to a question of prevalence, it is important to note that a critical appraisal tool designed for RCTs would not be appropriate to use, and a tool to assess studies reporting prevalence data is more appropriate.

Chapter 2: Systematic Reviews and Evidence-Based Healthcare

Evidence-Based Healthcare

Evidence-based health care takes place when decisions that affect the care of patients are taken with due weight accorded to all valid, relevant information (Hicks, 1997). Sackett, Rosenberg, Gray, Haynes, and Richardson (1996) define evidence-based medicine as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research” (p.71).

Pearson, Wiechula, Court, and Lockwood (2005) state that evidence-based practice (EBP) is clinical decision-making that considers the best available evidence; the context in which the care is delivered; client preference; and the professional judgment of the health professional. Guyatt, Drummond, Meade, and Cook (2008) suggest that:

EBP is clinical practice in which patient management decisions are consistent with the principles of evidence-based health care. This means that decisions will be, first of all, consistent with the best evidence about the benefits and downsides of the alternative management strategies. Second, decisions will be consistent with the values and preferences of the individual patient. (p.783)

Policymaking is evidence-based when practice policies (e.g. use of resources by clinicians), service policies (e.g. resource allocation, pattern of services), and governance policies (e.g. organizational and financial structures) are based on research evidence of benefit or cost-benefit (Guyatt et al., 2008).

Systematic Reviews

Systematic reviews are a method to synthesize the existent evidence on a certain topic. A systematic review extends beyond the subjective, narrative reporting characteristics of a traditional literature review by employing procedures to rigorously extract data from studies that have been included following assessment of their quality, and to synthesize, or combine, that data where appropriate. The systematic review of the literature on a particular condition, intervention or issue is seen as core to defining reliable evidence for practice. Systematic reviews aim to provide comprehensive and unbiased summaries of the evidence on a single topic bringing together multiple individual studies in a single document. As part of the systematic review process, individual research studies are subjected to critical appraisal. Even when research evidence is limited or non-existent, systematic reviews summarize the best available evidence on a specific topic providing the best evidence for clinical decision-making as well as identifying future research needs.
The process of conducting a systematic review is a scientific exercise, and as the results will influence health care decisions, it is required to have the same rigor expected of all research. The quality of a review, and its recommendations, depends on the extent to which scientific review methods are followed to minimize the risk of error and bias. The explicit and rigorous methods of the process distinguish systematic reviews from traditional reviews of the literature.

A systematic review is essentially an analysis of all of the available literature (i.e. evidence) involving the following steps (Pearson et al., 2005) (Figure 2):

- The development of a rigorous proposal or protocol.
- Stating the questions or hypotheses that will be pursued in the review.
- Identifying the criteria that will be used to select the literature.
- Detailing a strategy that will be used to identify all relevant literature within an agreed time frame.
- Establishing how the quality of each study/paper will be assessed or critically appraised and any exclusion criteria based on quality considerations.
- Detailing how data will be extracted from the primary research or text.
- Setting out a plan of how the data extracted will be synthesized.

Figure 2: The Systematic Review Process (Pearson, Robertson-Malt, & Rittenmeyer, 2011)
Systematic Reviews of Prevalence and Incidence Data

The systematic review of prevalence and incidence data is important in the description of the geographical distribution of a variable, variation between subgroups (such as gender), and informing health care planning and resource allocation. Pooling of such data is necessary to monitor trends in disease burden and emergence and to contribute to the design of further etiological studies. Systematic reviews are of particular relevance where individual studies are limited by small sample sizes. The systematic review of studies to answer questions of prevalence and incidence data still follow the same basic principles of systematic review of other types of data. A protocol must be written for the conduct of the systematic review, comprehensive searching must be performed and critical appraisal of retrieved studies must be carried out.

There are many reasons to do a review of these types of data, including the following:

- they assist in answering questions of global disease burden,
- help to measure global design burden (incidence data can be used to determine disability adjusted life years),
- in cases where it is not practical to do a large global survey,
- questions larger than a national scale,
- cumulative meta-analysis can show changes and trends over time, and can highlight emerging or decreasing diseases,
- assists policy makers and funding models,
- informs geographical distributions and comparisons of subgroups,
- inform healthcare professionals of diseases and symptoms of disease,
- can compare prevalence between groups, and
- inform further research priorities.

The methods outlined in this book relate to systematic reviews of prevalence and incidence data. The data that is to be synthesized is therefore proportional data; that is, proportions (often percentages) of a population experiencing the particular disease or condition. However, the methods can be applied more broadly than this and do not necessarily need to focus on a disease. There are examples of reviews that have been conducted addressing issues such as the prevalence and incidence of medication errors (Miller, Robinson, Lubomski, Rinke, & Pronovost, 2007), claustrophobic reactions in magnetic resonance imaging (Munn, Moola, Lisy, Riitano, & Murphy, 2014), barriers to adherence with treatment (Mills et al., 2006), and electronic health record adoption (Rao et al., 2008).
Chapter 3: Critical Appraisal of Prevalence and Incidence Data

Why critically appraise?

A main principle of evidence-based healthcare is to apply the findings of well conducted research studies into practice to ensure more efficient and effective healthcare. A key component of this process is analyzing and appraising the research evidence to determine its methodological quality (Straus, Richardson, Glasziou, & Haynes, 2005). It is of the utmost importance for the clinician attempting to implement EBP to understand the strength of evidence of a piece of research by following a process of critical appraisal to assess the methodological quality of research (Jack, 1997; Greenhalgh, 1996).

The main object of critical appraisal is to assess the methodological quality of a study and to determine the extent to which a study has addressed the possibility of bias in its design, conduct and analysis. If a study has not excluded the possibility of bias then its results should be viewed as questionable and could well be invalid. Therefore, part of the systematic review process is to evaluate how well the potential for bias has been excluded from a study, with the aim of only including high quality studies in the resulting systematic review.

A systematic review aims to synthesize the best available evidence; therefore the review should aim to include the highest quality of evidence possible. All papers selected for inclusion in the systematic review (that is – those that meet the inclusion criteria described in the protocol) need to be subjected to rigorous appraisal by two critical appraisers. The purpose of appraisal is to include only those studies that are of high quality and to exclude those of poor quality. Given that the systematic review aims to summarize the best available evidence through pooling the results of sufficiently similar studies where possible, it is important to note that the pooling of poor quality evidence may lead to outcomes that are less than desirable for patients.

Critical appraisal is probably the most difficult component of the systematic review and a good understanding of research design is required.

The major aim of critical appraisal of quantitative evidence is to establish the validity of the evidence. Validity refers to the soundness of the evidence; in other words it is about the degree to which we can accept the evidence as trustworthy and believable. The validity of quantitative studies refers to the degree to which possible bias has been limited. Bias refers to any influence that may distort the results of a study. Relying on the results of studies with variable validity and pooling these results may lead to a conclusion that is incorrect.

Methodological quality is assessed by critical appraisal using validated tools/ checklists. There are a variety of checklists and tools available to assess the validity of studies that aim to identify sources of bias. Many checklists have been developed that can be used by appraisers and different checklists are used for different research designs.
Critical Appraisal of Studies Reporting Prevalence

Systematic reviews often use critical appraisal checklists that are linked to the study design used. For example, there may be separate checklists used to appraise RCTs, cohort studies, cross-sectional studies and so on. To address issues regarding prevalence and incidence, epidemiological studies, such as those classified under the term observational and descriptive studies are required. These designs address questions such as: how many people have a disease? Who is getting the disease? Where is the disease occurring? However, prevalence data can also come from experimental studies, such as RCTs as discussed in the background section of this book. However, critical appraisal tools directed at assessing the risk of bias of RCTs are aimed at assessing biases related to causal effects and hence are not appropriate when the interest of the review is the prevalence of a condition. For example, when collecting data regarding the prevalence of dizziness amongst diabetics the baseline characteristics as reported in a randomized controlled trial could be of use to the review. However, as the conduct of the subsequent trial would not be of interest to a question of prevalence, it is important to note that a critical appraisal tool designed for RCTs would not be appropriate to use. For example, criteria regarding the use of an intention to treat analysis as often seen in critical appraisal checklists for RCTs are not a true quality indicator for questions of prevalence.

Due to this a new tool assessing validity and quality indicators specific to issues of prevalence that can be used across study designs has been developed. This checklist assesses important issues when assessing validity of prevalence data. The criteria ensure the following:

- a representative sample,
- appropriate recruitment,
- an adequate sample size,
- appropriate description and reporting of study subjects and setting,
- data coverage of the identified sample is adequate,
- the condition was measured reliably and objectively,
- appropriate statistical analysis, and
- confounding factors/subgroups/differences are identified and accounted for.

Prevalence Critical Appraisal Instrument

The 10 criteria used to assess the methodological quality of studies reporting prevalence data and an explanation are described below. These questions can be answered either with a yes, no, unclear, or not applicable.

Was the sample representative of the target population?

This question relies upon knowledge of the broader characteristics of the population of interest. If the study is of women with breast cancer, knowledge of at least the characteristics, demographics, and medical history is needed. The term ‘target population’ should not be taken to infer every individual from everywhere or with similar disease or exposure characteristics. Instead, give consideration to specific population characteristics in the study, including age range, gender, morbidities, medications, and other potentially influential factors. For example, a sample may not be representative of the target population if a certain group has been used (such as those working for one organization, or one profession) and the results then inferred to the target population (i.e. working adults). Or, if including data from baseline groups of an RCT, if there were strict inclusion criteria for the trial.
Were study participants recruited in an appropriate way?

Recruitment is the calling or advertising strategy for gaining interest in the study, and is not the same as sampling. Studies may report random sampling from a population, and the methods section should report how sampling was performed. What source of data were study participants recruited from? Was the sampling frame appropriate? For example, census data is a good example of appropriate recruitment, as a good census will identify everybody. Was everybody included who should have been included? Were any groups of persons excluded? Was the whole population of interest surveyed? If not, then was random sampling from a defined subset of the population employed? Was stratified random sampling with eligibility criteria used to ensure the sample was representative of the population that the researchers were generalizing to?

Was the sample size adequate?

An adequate sample size is important to ensure good precision of the final estimate. Ideally we are looking for evidence that the authors conducted a sample size calculation to determine an adequate sample size. This will estimate how many subjects are needed to produce a reliable estimate of the measure(s) of interest. For conditions with a low prevalence, there will need to be a larger sample size. Also consider sample sizes for subgroup (or characteristics) analyses, and whether these are appropriate. Sometimes, the study will be large enough (as in large national surveys) whereby a sample size calculation is not required. In these cases, sample size can be considered adequate.

When there is no sample size calculation and it is not a large national survey, the reviewer’s may consider conducting their own sample size analysis using the following formula (Daniel, 1999; Naing, Winn, & Rusli, 2006):

\[ n = \frac{Z^2P(1-P)}{d^2} \]

Where:

- \( n \) = sample size
- \( Z \) = Z statistic for a level of confidence
- \( P \) = Expected prevalence or proportion (in proportion of one; if 20%, \( P = 0.2 \))
- \( d \) = precision (in proportion of one; if 5%, \( d=0.05 \))

Were the study subjects and the setting described in detail?

Certain diseases or conditions vary in prevalence across different geographic regions and populations (e.g., women vs. men, socio-demographic variables between countries). Has the study sample been described in sufficient detail that other researchers can determine if it is comparable to the population of interest to them?

Is the data analysis conducted with sufficient coverage of the identified sample?

A large number of dropouts, refusals, or not founds amongst selected subjects may diminish a study’s validity, as can low response rates for survey studies.

Did the authors describe the reasons for non-response and compare persons in the study to those not in the study, particularly with regards to their socio-demographic characteristics?

Could the not-responders have led to an underestimate of prevalence of the disease or condition under investigation?
If reasons for non-response appear to be unrelated to the outcome measured and the characteristics of non-responders are comparable to those in the study, the researchers may be able to justify a more modest response rate.

Did the means of assessment or measurement negatively affect the response rate? Measurement should be easily accessible, conveniently timed for participants, acceptable in length and suitable in content.

**Were objective, standard criteria used for measurement of the condition?**

Here we are looking for measurement or classification bias. Many health problems are not easily diagnosed or defined and some measures may not be capable of including or excluding appropriate levels or stages of the health problem. If the outcomes were assessed based on existing definitions or diagnostic criteria, then the answer to this question is likely to be yes. If the outcomes were assessed using observer reported, or self-reported scales, the risk of over- or under-reporting is increased, and objectivity is compromised. Importantly, determine if the measurement tools used were validated instruments as this has a significant impact on outcome assessment validity.

**Was the condition measured reliably?**

Considerable judgment is required to determine the presence of some health outcomes. Having established the objectivity of the outcome measurement instrument (see item 6 of this scale), it’s important to establish how the measurement was conducted. Were those involved in collecting data trained or educated in the use of the instrument/s? If there was more than one data collector, were they similar in terms of level of education, clinical or research experience, or level of responsibility in the piece of research being appraised?

Has the researcher justified the methods chosen?

Has the researcher made the methods explicit? (For interview method, how were interviews conducted?)

**Was there appropriate statistical analysis?**

As with any consideration of statistical analysis, consideration should be given to whether there was a more appropriate alternate statistical method that could have been used. The methods section should be detailed enough for reviewers to identify the analytical technique used and how specific variables were measured. Additionally, it is also important to assess the appropriateness of the analytical strategy in terms of the assumptions associated with the approach as differing methods of analysis are based on differing assumptions about the data and how it will respond. Prevalence rates found in studies only provide estimates of the true prevalence of a problem in the larger population. Since some subgroups are very small, 95% confidence intervals are usually given.

**Are all important confounding factors/ subgroups/ differences identified and accounted for?**

Prevalence studies often draw or report findings regarding the differences between groups. It is important that authors of these studies identify all important confounding factors, subgroups and differences and account for these.

**Were subpopulations identified using objective criteria?**

Objective criteria should also be used where possible to identify subgroups (refer to question 6).
Critical Appraisal of Studies Reporting Incidence Data

Unlike prevalence data, incidence data needs to be collected over time. Recall that the incidence of a disease indicates how many new cases of a disease occur in a defined period of time. Incidence data is therefore obtained through tracking of an initially disease-free cohort and determining the number of those who developed the disease over a period of time. To collect such data longitudinal studies are required, often in the form of cohort studies. Therefore, the use of the Joanna Briggs Institute tool for assessing cohort studies can be used when addressing incidence data.

Cohort Study Critical Appraisal Instrument

The nine criteria used to assess the methodological quality of studies reporting incidence data and an explanation are described below. These questions can be answered either with a yes, no, unclear or not applicable.

Is the sample representative of patients in the population as a whole?
This question relies upon knowledge of the broader characteristics of the population of interest. If the study is of women undergoing chemotherapy for breast cancer knowledge of at least the characteristics, demographics, and medical history is needed. The term population as a whole should not be taken to infer every individual from everywhere subject to a similar intervention or with similar disease or exposure characteristics. Instead, give consideration to specific population characteristics in the study, including age range, gender, morbidities, medications, and other potentially influential factors.

Are the patients at a similar point in the course of their condition/illness?
Check the paper carefully for descriptions of diagnosis and prognosis to determine if patients within and across groups have similar characteristics in relation to disease or exposure, for example tobacco use.

Has bias been minimized in relation to selection of cases and controls?
It is useful to determine if patients were included in the study based on either a specified diagnosis or definition. This is more likely to decrease the risk of bias. Characteristics are another useful approach to matching groups, and studies that did not use specified diagnostic methods or definitions should provide evidence on matching by key characteristics.

Are confounding factors identified and strategies to deal with them stated?
Confounding has occurred where the estimated intervention effect is biased by the presence of some difference between the comparison groups (apart from the intended intervention/s). Typical confounders include baseline characteristics, prognostic factors, or concomitant interventions. A confounder is a difference between the comparison groups and it influences the direction of the study results. A high quality study at the level of cohort or case-control design will identify the potential confounders and measure them (where possible). This is difficult for studies where behavioral, attitudinal or lifestyle factors may impact on the results.

Are outcomes assessed using objective criteria?
Refer back to item three of this appraisal scale and read the methods section of the paper again. If the outcomes were assessed based on existing definitions or diagnostic criteria, then the answer to this question is likely to be yes. If the outcomes were assessed using observer reported, or self-reported scales, the risk of over- or under-reporting is increased, and objectivity is compromised. Importantly, determine if the measurement tools used were validated instruments as this has a significant impact on outcome assessment validity.
Was follow-up carried out over a sufficient time period?

The appropriate length of time for follow-up will vary with the nature and characteristics of the population of interest and/or the intervention, disease or exposure. To estimate an appropriate duration of follow-up, read across multiple papers and take note of the range for duration of follow-up. The opinions of experts in clinical practice or clinical research may also assist in determining an appropriate duration of follow-up.

Were the outcomes of people who withdrew described and included in the analysis?

Any losses to follow up, particularly from prospective studies, can introduce bias to observational research and over- or underestimation of treatment effects, as it does with trials. This bias may result if subjects lost form a study group have a different health response from those who remain in the study. Here the reviewer should look for accurate reporting of loss to follow up and reasons for attrition. If loss to follow up is similar across comparison groups, despite losses, estimated effects may be unbiased.

Were outcomes measured in a reliable way?

Having established the objectivity of the outcome measurement instrument (see item 5 of this scale), it’s important to establish how the measurement was conducted. Were those involved in collecting data trained or educated in the use of the instrument/s? If there was more than one data collector, were they similar in terms of level of education, clinical or research experience, or level of responsibility in the piece of research being appraised?

Was appropriate statistical analysis used?

As with any consideration of statistical analysis, consideration should be given to whether there was a more appropriate alternate statistical method that could have been used. The methods section of cohort or case-control studies should be detailed enough for reviewers to identify the analytical technique used (in particular, regression or stratification) and how specific confounders were measured.

For studies utilizing regression analysis, it is useful to identify if the study identified which variables were included and how they related to the outcome. If stratification was the analytical approach used, were the strata of analysis defined by the specified variables? Additionally, it is also important to assess the appropriateness of the analytical strategy in terms of the assumptions associated with the approach as differing methods of analysis are based on differing assumptions about the data and how it will respond.

Chapter 4: Meta-analysis of Prevalence and Incidence Data

Introduction

The data synthesized within a systematic review are the results extracted from individual research studies relevant to the review question. As much as meta-analysis is preferred, it is not always possible in a systematic review if the included studies vary greatly from each other, either in terms of how they are conducted (different interventions), who they are performed on (different populations) or in their final result. When meta-analysis isn’t possible, common alternatives for the synthesis of quantitative data in a systematic review include narrative summary of results, vote counting, and presenting data via tables. Before discussing meta-analysis, alternative methods to synthesis are discussed below.
Narrative Summary
A narrative summary is commonly used where meta-analysis is not possible. A narrative summary describes the included studies and provides conclusions about the evidence. With a narrative summary, readers may not be able to discern how evidence was weighted and whether conclusions are biased. It is therefore important that when summarizing findings in narrative form, there is a clear structure to the summary, with an emphasis on reporting the characteristics of included studies along with data extracted relevant to the review outcomes (Lockwood & White, 2012). Narrative summary in systematic reviews should be rigorous and clear, and can utilize tables, graphs, and other diagrams to help convey how studies compare to each other and to assist in presentation of the data (Lockwood & White, 2012). A narrative summary should include the presentation of the quantitative results reported in individual studies; where available, the point estimates (one value that represent or best estimate of effects) and the interval estimates (usually presented as 95% confidence intervals) for the effects should be provided. Due to the flexibility of narrative summary in terms of the amount of data that can be conveyed textually, a structure that applies to each sequence or reporting of results from each study should be discussed beforehand and applied by the systematic review authors. This will ensure that there is consistency across the results section of a review. If a structure is not followed there may be substantial variability in reporting of results causing the data to appear incomplete or unreliable (Lockwood & White, 2012). Therefore adherence to this structure is critical; if studies do not provide the relevant information to comply with the structure this should be made clear in the summary (Lockwood & White, 2012). Bear in mind that there is no prescriptive guidance on presenting a narrative summary and it is recommended that the context of the review be taken into consideration.

Tabular
Tables where relevant should be included to aid in the presentation of the data. Mostly, these tables include individual studies with their raw data; for example, percentages, distribution of prevalence and incidence estimates and confidence intervals. The tables should also include other elements such as participant characteristics.

Graphical
The various graphs that may be useful in presenting include but are not limited to forest plot (for meta-analysis), funnel plot (for publication bias), L’abbe plot (explores heterogeneity and is applicable for meta-analysis of studies with binary outcomes), Galbraith plot (assesses the extent of heterogeneity between studies in a meta-analysis), and cumulative plot (incidence and prevalence estimates). However, it should be noted that the interpretation of the graphs is quite subjective and therefore should be interpreted with caution.

Meta-analysis
A meta-analysis is a statistical process that essentially calculates effect sizes for individual studies, converts them to a common metric, and then combines them to obtain an average effect size (Field, 2001). This statistical combination increases the power of the overall estimate from various small individual studies as a result of the overall increase in the sample size. In addition, meta-analysis also enables reviewers to explore the differences between individual studies. Meta-analysis should only be undertaken when the studies are sufficiently similar to combine; in the absence of this homogeneity, the conclusion from the meta-analysis may be invalid. The findings may also depend on the selection and quality of the studies included and the availability of relevant data.
Where meta-analysis is used, the statistical methods and the software used should be described. Prior to a meta-analysis to be undertaken, relevant data needs to be extracted. If the data is heterogeneous and is presented as a narrative summary, sources of heterogeneity should be discussed (e.g. clinical, methodological or statistical) as well as on what basis it was determined inappropriate to combine the data statistically (such as differences in populations, study designs or clinical or statistical heterogeneity).

There are established methods for conducting meta-analyses of RCTs and some observational study designs. However, no clear guidance exists on synthesizing frequency data from incidence and prevalence estimates. This section provides this guidance.

**Effect size**

The effect size statistically describes the relationship between two variables and is represented by a square on a forest plot. In traditional effectiveness reviews, this could be the impact of a new therapy on mortality rates or the effect of a new teaching method on exam scores. The effect size could be a single number such as for a prevalence study or a ratio such as a risk ratio. The effect size has been described as being the currency of the systematic review as the aim of the meta-analysis is to summarize the effect size of each included study to obtain a summary effect (Borenstein, Hedges, Higgins, & Rothstein, 2009). The summary effect is shown as a diamond on a forest plot. When effect sizes are statistically combined, the methods used make certain assumptions.

**Statistical combination of data**

In meta-analysis, the results of similar, individual studies are combined to determine the overall effect. In meta-analysis, the effect-size and weight of each study are calculated. The effect size indicates the direction and magnitude of the results of a particular study (i.e. do the results favor the treatment or control, and if so, how much), while the weight is indicative of how much information a study provides to the overall analysis when all studies are combined together.

It has been suggested that there are three important criteria for choosing a summary statistic for meta-analysis: consistency of effect across studies, mathematical properties, and ease of interpretation (Deeks & Altman, 2001).

Consistency of effect is important because the aim of meta-analysis is to bring together the results of several studies into a single result.

The main mathematical property required by summary statistics is the availability of a reliable variance estimate. Consensus about the other two mathematical properties (reliance on which of the two outcome states (e.g. mortality/survival) is coded as the event and odds ratio being the only statistic which is unbounded) has not yet been reached.

**Ease of interpretation.**

Essentially there are three popular approaches to conduct meta-analysis for all types of data: Hedge and Olkin (1985) technique, Rosenthal and Rubin (1986) technique and the Hunter and Schmidt (1982) technique. Hedge and Olkin developed both fixed- and random-effects models for pooling data, Rosenthal and Rubin developed a fixed-effects model only and Hunter and Schmidt developed a random-effects model.
Statistical assumptions in meta-analysis

Meta-analysis can be based on either one of two statistical assumptions – fixed or random effects. It is important to distinguish between fixed- and random-effects models when conducting meta-analysis, as it can lead to false assumptions about statistical significance of the pooled estimate.

The main difference between fixed and random effects models is in the calculation of standard errors associated with the combined effect size. Fixed effects models use only within-study variability in their error term because all other unknowns in the model are assumed not to affect the effect size. In contrast, in random effects models it is necessary to account for the errors associated with sampling from populations that themselves have been sampled from a super population. As such the error term contains two components: within-study variability and variability arising from differences between studies (Field, 2001).

The fixed effects model assumes that there is one true effect for the population underlying the studies in the analysis and that all the differences in the data are due to sampling error or chance within each study and that there is no heterogeneity between the studies. A fixed effect model is statistically stringent and should be used when there is little heterogeneity, as determined by Chi-square or I² tests. This model therefore assumes that the overall sample consists of samples that all belong to the same underlying population (Kock, 2009). The between-study variability will be zero in this model as it assumes that the population effect size is identical for all studies. In an analysis based on a fixed effects model, inference is applicable or generalizable (conditional) based on statistical justification only on the studies actually done (Petitti, 2000). The fixed effects model assumes that there is little interest or value in generalizing the results to other studies (Fleiss, 1993; Munn, Tufanaru, & Aromataris, 2014).

A random effects model allows more flexibility, assuming that there may be other factors influencing the data than error or chance, within and between studies. As a result, in an analysis based on a random effects model, inference relies on the assumption that the studies used in the analysis are a random sample of some hypothetical population of studies (Munn, Tufanaru, & Aromataris, 2014; Petitti, 2000). For example, the effect size may be influenced in studies where the participants are more educated, older, healthier, or if a more intense intervention is being used. The effect size is assumed to follow a normal distribution and consequently has a mean and variance. The random-effects model considers both between-study variability and within-study variability. The random-effects model enables generalizations beyond the population included in the studies.

There is no consensus about whether fixed- or random-effects models should be used in meta-analysis. In many cases when heterogeneity is absent, the two methods will give similar overall results. When heterogeneity is present, the random effects estimate provides a more conservative estimate of the overall effect size, and is less likely to detect significant differences. For this reason, random effects models are sometimes employed when heterogeneity is not severe; however, the random effects model does not actually analyze the heterogeneity away and should not be considered as a substitute for a thorough investigation into the reasons for heterogeneity. Additionally, random effects models give relatively more weight to the results of smaller studies – this may not be desirable because smaller studies are typically more prone to bias and are often lower quality than larger studies.

There are a number of meta-analytical techniques available – the selection of a particular technique is governed by three things: the study type, the nature of the data extracted and the assumptions underlying the meta-analysis.
Meta-analysis of prevalence and incidence data - Proportions

Prevalence and incidence data is often reported as a proportion. When pooling proportions for meta-analysis, a transformation of the data is required. There are two main ways to transform the data, the Freeman-Turkey transformation (arcsine square root transformation), and the Logit transformation, both of these are used to calculate the weighted summary proportion under the fixed and random effects model. The resultant meta-analysis will give pooled proportion with 95% CI both for the fixed effects model and the random effects model and in addition, will list the proportions (expressed as a percentage), with their 95% CI, found in the individual studies included in the meta-analysis. The results are then presented graphically in a forest plot. For all meta-analyses, prevalence estimates are transformed to logits to improve their statistical properties. These are then back-transformed to prevalence.

Converting proportions \( p \) to logits (Sutton, Abrams, Jonas, Sheldon, & Song, 2000):

\[
\text{Logit} = \log(\text{odds}) = \log \left( \frac{p}{1-p} \right).
\]

Using the number of cases with an event \( N_{\text{event}} \) and without an event \( N_{\text{-event}} \), the variance of logit is given by

\[
\text{Var}(\text{logit}) = \frac{1}{N_{\text{event}}} + \frac{1}{N_{\text{-event}}}.
\]

There are different models for performing the meta-analysis as mentioned above. Normally the reviewer is provided with a choice of using the Mantel-Haenszel (1959) model or the DerSimonian and Laird (1986) model. We recommend that the meta-analyses of the prevalence reported in the studies is grouped by a random-effects model and is presented with 95% confidence intervals (95% CI). Random effects model are used when there is sufficient information on standard errors. However, bear in mind that the random-effects model gives a conservative estimate with a wider confidence interval. The random effects model allows for between-study variation by assuming that the individual study prevalence estimates follow a normal distribution. The fixed model can be chosen but the reviewer should be aware of its underlying principles, particularly in relation to its assumption that there is one true effect, which may not hold for prevalence and incidence data.

Heterogeneity of the results is tested by the I-squared, Tau-squared, Cochran's Q test and Chi-squared \( (p > 0.05) \) tests. These tests of heterogeneity evaluate whether the differences in prevalence estimates across studies are higher than expected by chance. To identify the sources of heterogeneity across studies, subgroup analysis or meta-regression can be used to assess the contribution of each variable (i.e. year of study, geographic location, characteristic of countries, study population etc.) to the overall heterogeneity. Those variables significantly associated with the heterogeneity \( (p < 0.05) \) can be included in a multivariate hierarchical model. A \( p \) value of <0.05 is considered statistically significant in all the analyses.

Table 1 is an example of studies that were combined in a meta-analysis. These studies reported on overall termination rates for scans in the general MRI population.
Table 1: Studies Combined in a Meta-analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Events</th>
<th>Sample</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dantendorfer 1997</td>
<td>2</td>
<td>297</td>
<td>0.673400673</td>
</tr>
<tr>
<td>Dewey 2007 (all)</td>
<td>1004</td>
<td>55734</td>
<td>1.801413859</td>
</tr>
<tr>
<td>Eshed 2007</td>
<td>59</td>
<td>4821</td>
<td>1.223812487</td>
</tr>
<tr>
<td>Lang et al (2010)*</td>
<td>336</td>
<td>34521</td>
<td>0.973320587</td>
</tr>
<tr>
<td>Nawaz 2009*</td>
<td>58</td>
<td>2630</td>
<td>2.205323194</td>
</tr>
<tr>
<td>Sarji 1998</td>
<td>18</td>
<td>3324</td>
<td>0.541516245</td>
</tr>
<tr>
<td>Wiebe 2004</td>
<td>14</td>
<td>1790</td>
<td>0.782122905</td>
</tr>
</tbody>
</table>

Figure 3: Meta-analysis of Scan Termination Due to Claustrophobia in General Scan Types

Figure 3 represents meta-analysis of proportion data using random effects model from the seven studies. There was significant heterogeneity present in the studies, with a Cochran Q test reaching statistical significance and an $I^2$ value of 96.1%. The pooled proportion equaled 1.18% (95% CI 0.79 – 1.65). However, due to the significant heterogeneity, this value should be interpreted with caution.
There are limitations with conducting meta-analysis of frequency data, including (Saha, Chant, & McGrath, 2008):

Heterogeneity of data: If the data from the included studies are heterogenous, then the standard errors or confidence intervals for a pooled effect estimate will not adequately reflect the variability of underlying data.

Inadequate reporting of frequency estimates: standard error (SE) for each estimate is required to weight the estimate when pooling the data. Standard errors can still be calculated if the data for the numerator, denominator and the duration of the study were available; however, these calculated SEs will not take into account various adjustments.

**How to interpret effect sizes?**

Once authors calculate effect sizes, they need to answer the question: What does the effect size mean?

An effect size is simply a number and its meaning and importance must be explained by the researcher. An effect size of any magnitude can mean different things depending on the research that produced it and the results of similar past studies. Therefore, it is the researcher's responsibility to discuss the importance of their findings and this information requires comparing current effects to those obtained in previous work in the same research area. Confidence Intervals (CIs) are an important way to evaluate the precision of a study's findings by providing a range of likely values around the obtained effect size.

**Heterogeneity**

When used in relation to meta-analysis, the term *heterogeneity* refers to the amount of variation in characteristics of included studies. For example, if three studies are to be included in a meta-analysis, does each of the included studies have similar demographics, and assess the same intervention? While some variation between studies will always occur due to chance alone, heterogeneity is said to occur if there are significant differences between studies, and under these circumstances meta-analysis is not valid and should not be undertaken.

There are three types of heterogeneity: clinical, methodological, and statistical heterogeneity (Higgins & Thompson, 2002). Differences in the characteristics of study populations and measurements represent clinical heterogeneity. Differences in study designs and methodological quality (risk of bias) represent methodological heterogeneity. Statistical heterogeneity is the variation of effects sizes between studies. Statistical heterogeneity may arise because of clinical heterogeneity, methodological heterogeneity, or simply by chance.

There is often heterogeneity amongst studies addressing prevalence and incidence. This is due to a number of reasons. Firstly, clinical heterogeneity may be present due to the measures used to determine the presence of a variable (Webb, Bain, & Pirozzo, 2005). For example, different scales exist to measure depression, and depending on the scale used, a person may be classified as suffering from depression whilst using one scale and not suffering based on a different scale. Additionally, prevalence and incidence studies often look at specific populations at a specific point of time, and the scope of the study may be limited by state or national borders. Another consideration with the population is whether those considered at risk or eligible for the disease have been included (Webb, Bain, & Pirozzo, 2005). For example, if look at the prevalence or incidence of breast cancer, have these studies reported the proportion out of the whole population, all females, only adult females, and so on. These different populations may contribute to clinical heterogeneity.
The Synthesis of Prevalence and Incidence Data

Methodological heterogeneity is also important to consider. Prevalence and incidence data can arise from various study designs with differing levels of methodological quality. This can also result in differences amongst studies.

But how does one tell whether or not differences are significant?

First, the studies should be assessed carefully to determine whether there is clinical or methodological heterogeneity present. If conducting a meta-analysis, then a visual inspection of the meta-analysis output (e.g. the forest plot) is the first stage of assessing heterogeneity. If the results are scattered across the forest plot and none of the confidence intervals overlap, this is a good indicator of heterogeneity.

A formal statistical test of the similarity of studies is provided by test of homogeneity. The test calculates a probability (P value) from a Chi-square statistic calculated using estimates of the individual study weight, effect size and the overall effect size. Note, however, that this test suffers from lack of power – and will often fail to detect a significant difference when a difference actually exists – especially when there are relatively few studies included in the meta-analysis. Because of this low power, some review authors use a significance level of \( P < 0.1 \), rather than the conventional 0.05 value, in order to protect against the possibility of falsely stating that there is no heterogeneity present. Often when combining the results from a series of observational studies, this is the default significance level due to the increased heterogeneity associated inherent with the study design.

The \( I^2 \) statistic is the percentage of observed total variation across due to heterogeneity and not due to chance. A value of 0% indicates no observed heterogeneity and larger values show increasing heterogeneity.

If there is statistically significant heterogeneity, a narrative synthesis or graphical representation is recommended.

Subgroup analysis (Analysis of subgroups or subsets)

Subgroup analysis is a means of investigating heterogeneous results and can be used to estimate the influence of various subsets including age group, gender, type of population and sampling strategy used to gather data (e.g. letter, phone, face-to-face). However, subgroups should be pre-specified a priori and should be few. Subgroup analysis may include by study design or by patient groups.

Meta-regression

Meta-regression investigates whether particular covariates explain any of the heterogeneity of treatment effects between studies. A meta-regression is either a linear or logistic regression and can be fixed-effect or random-effect model. The unit of analysis is a study and predictors in the regression are the study-level covariates.

Publication bias

The research that appears in the published literature may be systematically unrepresentative of the population of completed studies. File drawer problem or Publication bias is a term coined by Rosenthal to mean the number of statistically non-significant studies \( (p > 0.05) \) that remain unpublished (Rosenthal & Rubin, 1982). A Funnel plot is used to detect publication bias. This is a scatter plot of effect estimate \((x\text{-axis})\) against inverse of its variance \((y\text{-axis})\). If there is no bias then the funnel will appear symmetric and inverted and if there is bias, the funnel will be asymmetric or skewed in shape.
Chapter 5: Protocol Development

Introduction

A systematic review protocol is important because it pre-defines the objectives and methods of the systematic review. It is a systematic approach to the conduct and report of the review that allows transparency of the process, which in turn allows the reader to see how the findings and recommendations were arrived at. The protocol details the criteria the reviewers will use to include and exclude studies, to identify what data is important and how it will be extracted and synthesized. A protocol provides the plan or proposal for the systematic review and as such is important in restricting the presence of reporting bias. Any deviations between the protocol and systematic review report should be discussed in the systematic review report.

This section outlines the components of a systematic review protocol of prevalence and incidence evidence and provides guidance on the information that each component should contain. Specifically, it provides guidance on each of the following components: title page, title development, background, review objectives/questions, inclusion criteria, search strategy, critical appraisal, data extraction, data synthesis, narrative summary, conflict of interest, acknowledgements, references, and appendices. This guidance is based on the Joanna Briggs Institute approach to systematic reviews of prevalence and incidence.

Protocol Development

Prior to conducting a systematic review, it is recommended that reviewer’s receive training in systematic review and conduct. Organizations such as the Joanna Briggs Institute and the Cochrane Collaboration offer standardized training in the conduct of systematic reviews.

When deciding on a topic, some preliminary investigation of the literature is recommended to determine if studies are available on the topic of interest, while potential authors may also wish to consider the technical resources available to them. The conduct of a systematic review is greatly facilitated by access to extensive library and electronic databases and the use of citation management software.

Reviewers are encouraged to register their review title. This enables other groups and reviewers to identify topics that are currently in development and avoids accidental duplication of topics. At the Joanna Briggs Institute, once a title is registered it is valid for 6 months from the date of entry in the database. Should a protocol not be completed within that timeframe for a nominated topic, the topic becomes de-registered and available to any other JBI entity whose members may wish to conduct the review.

It is also recommended that reviewers search major electronic databases to determine that there have been no recently published systematic reviews on the same topic prior to registration of a review title. A search of the Joanna Briggs Institute Library of Systematic Review Protocols, Joanna Briggs Institute Library of Systematic Reviews, Cochrane Library, MEDLINE and DARE databases will assist to establish whether or not a recent review report exists on the topic of interest. The results of this search should be mentioned in the background of the systematic review protocol and review. If a systematic review on the topic of
interest has already been conducted, consider the following questions to establish if continuing with the review topic will be strategic:

- Is the date of last update longer than 3 years ago?
- Do the methods reflect the specific criteria of interest for your topic?
- Is there a specific gap in terms of population or intervention outcome that has not been addressed in the identified review?

**Title**

The title should be clear, explicit and reflect the core elements of the protocol. Titles should not be phrased as questions or conclusions and there should be congruency between the title, review objectives/questions and inclusion criteria. The title needs to include the phrase “A systematic review protocol.”

The title should give an indication on the type of data that will be reported (descriptive, analytical or a combination of both) by including the epidemiological indicator or a term that reflects the analysis that will be used to measure the variables of interest. Generally, measures of disease should appear in the title (prevalence, incidence).

The factors or events of interest (health condition or disease of interest) are defined by the time period, the place and the population at risk. Accordingly, the title should specify the defining characteristics of the population (i.e., gender, age) as well as the place and time of occurrence where relevant.

For example: Prevalence and incidence of depression amongst adolescents: A systematic review protocol.

**Review question/objective**

The review objective and any specific review question(s) must be clearly stated.

The overarching objective of reviews of prevalence and incidence data is to report on the frequency, distribution and determinants of specific factors, health states or conditions in a defined population. Reviews of this type are broadly classified as having two primary objectives.

Reviews that aim to describe the distribution of existing variables or seek to answer the question, “how common is a particular disease or condition in a specific group of individuals?” are often classified as descriptive and will utilize measures of prevalence and incidence to answer such lines of enquiry.

The objective of these reviews is to describe the health issue (what), those affected by it (who) as well as the location (where) and the time period (when) in which it occurred.

Accordingly, the review question should outline the factor, disease, symptom or health condition of interest, the epidemiological indicator used to measure its frequency (prevalence, incidence), the population or groups at risk, as well as the context/location (e.g., limited to specific geographic areas) and time period (e.g., peaks at a particular season) where relevant.

For example: The objective of this review is to assess the prevalence and incidence of perinatal depression among women in Australia.
Reviews focusing on how and why are predominantly analytic in nature. The objective of reviews of explanatory or analytic studies is to contribute to and improve our understanding of the causes of health-related events or outcomes by isolating the association between specific factors. This element is non-existent or lacking in studies that are purely descriptive. While studies that report prevalence and incidence only are broadly classified as descriptive and those that examine associations between exposures and outcomes are broadly classified as analytical a clear-cut distinction between analytical and descriptive study designs is not possible. Data generated from these studies can be measured and reported in different ways and the review question will indicate whether the review seeks to report data that is descriptive, analytical or a combination of both.

**Inclusion criteria**

This section of the protocol details the basis on which studies will be considered for inclusion into the systematic review and should be as clear and unambiguous as possible.

When determining the inclusion criteria, the CoCoPop mnemonic (*Condition, Context and Population*) can be used for reviews assessing prevalence and incidence data.

**Population**

The population should be appropriate for the review objectives. The reasons for the inclusion or exclusion of participants should be explained in the background.

It is important that the population or study subjects are clearly defined and described in detail. This includes outlining the specific or defining characteristics of the population such as age, sex, race, gender, educational status, individual behavior, socio-demographic factors etc.

For example, we will include studies involving adult pregnant women aged 18 – 45 years at any trimester up to delivery and for 6 months post-birth.

Exclusion criteria should also be outlined where relevant. For example, studies examining pregnancies with neural tube defects, intra-uterine growth retardation and early pregnancy loss; and those involving adolescent pregnancies and anemic mothers.

**Condition**

This refers to the variable of interest and may refer to a health condition, disease, symptom, event or factor. It is important that the variable of interest is clearly stated and defined. For example, malaria could be *P. falciparum* infection, *P. vivax* infection or disease due to malarial infection. This may include providing information on how the condition will be measured, diagnosed, or confirmed.

**Context**

Environmental factors can have a substantial impact on the prevalence or incidence of a condition. Accordingly, it is important that authors define the context or specific setting relevant to their review question. For example, this may include defining the geographic area or country, specific community or setting (inpatient vs outpatient) and the time period given that some conditions may peak at a particular season (e.g. the incidence of influenza in different seasons and years).
Types of studies

Reviews of prevalence and incidence are predominantly derived from observational studies. A cross-sectional study is the appropriate study design to determine the prevalence of a particular health problem. Cross-sectional surveys are typically used to estimate the point prevalence of common conditions of long duration and are generally not appropriate for rare or temporary diseases. As incidence is the number of new cases of a particular illness within a population over time study participants need to be followed up. Therefore, cohort studies that have a prospective or longitudinal design and follow up each subject over a suitable period of time are the best way to establish the incidence of a disease or the natural history of a condition. However many study designs may provide prevalence and incidence information.

Search strategy

This section details how the reviewers plan to search for relevant papers. A review should consider papers published in both commercial and in non-commercially operated databases (grey literature). The timeframe chosen for the search should be justified and any language restrictions stated (e.g. only studies published in English will be considered for inclusion). The databases to be searched must be listed along with the initial keywords to be used for the search. Appropriate databases to search should be included, including specification from the outset of the platform used to search a particular database.

Prevalence and incidence data are reported within the published, peer-reviewed literature and accordingly the standard JBI 3-step search strategy can be applied to locating this type of evidence.

There are also many and various sources of epidemiological data, within the grey literature, particularly for estimates of prevalence and incidence.

Some examples include:

- administrative sources (clinical records, insurance data),
- vital statistics data, government surveillance data and reports, Centers for Disease Control and Prevention data, population censuses and surveys (i.e., national or state health survey data), health care utilization records, and disease registries (population-based disease registries established to record cases of certain serious diseases),
- disease associations (e.g., American Diabetes Association), and
- medical books, grey literature, and reports from experts.

What sources are chosen will obviously depend on the specific research question and its scope. For example, estimating the worldwide prevalence of a common condition (chronic disease) will need to include many more sources than a review examining the prevalence of a condition within a specific regional setting.

Refer to chapter 6 for further information on searching.
Assessment of methodological quality

The protocol should detail the criteria considered when determining methodological quality of papers to include in the review. Critical appraisal tools must be appended to the protocol. For questions assessing incidence, the critical appraisal tool should be selected based on the type of study design retrieved from the search process. However, as prevalence data may be sourced from a number of study designs (including RCTs), a critical appraisal checklist specifically for prevalence studies has been developed as mentioned in Chapter 3. Critical appraisal must be conducted by two reviewers independently of each other. The reviewers should then meet to discuss the results of their critical appraisal for their final appraisal. If the two reviewers disagree on the final critical appraisal and this cannot be resolved through discussion, a third reviewer may be required.

Data collection

Standardized data extraction tools allow the extraction of the same types of data across the included studies and are required for systematic reviews. The protocol should detail what data the reviewers plan to extract from the included studies and the data extraction tool should be appended to the protocol.

The data extraction sheet should be adapted to suit the collection and stratification of the variables of interest from the included studies. It is important to extract data that reflects points of difference/heterogeneous characteristics between studies that affect the interpretation of the findings and synthesis of data.

Whether data synthesis can be performed will depend on the heterogeneity of the variables of interest across included studies. To facilitate such a comparison it is critical data extraction details the variables that will be extracted and compared.

The description of disease patterns often includes analysis of demographic, geographical, social, seasonal and other risk factors.

It is also likely to include the setting/location, dates of survey, definitions of conditions and populations, inclusion and exclusion criteria, mean age, sex, sample size, estimates of prevalence, incidence, etc.

Gender categorization, while important for sexually transmitted diseases and other diseases with a large gender gap may not be important for numerous other diseases. Geographical distribution is important to describe diseases linked to environmental conditions but may not be useful for other diseases.

Data synthesis

The protocol should detail how the reviewers plan to synthesize data extracted from included studies. The types of data it is anticipated will be synthesized should be consistent with the methods used for data collection and the included study designs.

Conflicts of interest

A statement which either declares the absence of any conflicts of interest or which describes a specified or potential conflict of interest should be made by the reviewers in this section.
Acknowledgements
Any acknowledgements should be made in this section (e.g. sources of external funding or the contribution of colleagues or institutions). It should also be noted if the systematic review is to count towards a degree award.

References
All references should be listed in full using the required referencing style.

Appendices
The critical appraisal and data extraction tools should be appended in this section.

Chapter 6: Searching for Prevalence and Incidence Data

Introduction
Developing a comprehensive search strategy that seeks to capture all relevant and eligible studies in relation to a particular question is critical to the quality and validity of a systematic review.

Given that systematic reviews of prevalence and incidence data seek to establish an accurate measurement of the proportion or occurrence of a disease in a given population it is important that a search for this data type is comprehensive. Any omission of relevant studies could skew the final estimates and the conclusions that are drawn from the data presented.

As highlighted earlier, accurate information regarding measures of disease assist in planning management of disease services (by ensuring resources are available to cope with the burden of disease), set priorities regarding public health initiatives, and evaluate changes and trends in diseases over time.

The following section offers some suggestions and guidance on how to devise a search strategy as well as the types of considerations and resources that may be helpful when constructing such a search.

The development of a comprehensive search strategy for a systematic review involves a range of processes. These include: developing a search string using a combination of free-text words and index terms; adapting the devised search and implementing it across a range of bibliographic citation databases that index relevant literature; searching across various sources of grey or unpublished literature; and hand searching for relevant publications.

Where to search?
There is insufficient evidence to suggest that searching a particular number or even particular databases will identify all of the evidence on a particular topic, therefore JBI recommend that a comprehensive systematic review search should be as broad and as inclusive as possible and should consider papers published in both commercial (black literature) and in non-commercially operated databases (grey literature).

Estimates of prevalence and incidence relating to a particular disease, condition or factor can be extracted from both research studies published in the peer-reviewed literature or from
Section 3
Conducting a Review

various organizations that collect and report such estimates in the unpublished or grey literature.

Bibliographic databases

The bibliographic databases that are chosen for searching will vary depending on the nature of the topic being reviewed or the resources that are available to the review author. Most databases are subject specific and only index bibliographic records related to a specific discipline or field of enquiry. For example, PsycINFO, the database produced by the American Psychological Association (APA) is the largest resource devoted to providing access to international peer-reviewed literature in psychology and therefore should be included as source when searching for studies that relate to the prevalence or incidence of mental health issues.

While prevalence and incidence indicators can be used to measure the proportion or occurrence of a broad range of factors they are traditionally used broadly within the health sciences to measure the health status of a given population.

Some of the most widely searched medical and health science databases for reviews of prevalence and incidence data are: MEDLINE, EMBASE, PsycINFO and CINAHL. The most commonly utilized search engine is PubMed; however, MEDLINE and PubMed are often used interchangeably. There are in fact some important differences. PubMed is updated more quickly than MEDLINE and PubMed indexes more journal titles as well as including studies that have been indexed in the Old MEDLINE database.

MEDLINE

(Medical Literature Analysis and Retrieval System Online) is the US National Library of Medicine's® (NLM) main bibliographic database with references to journal articles in biomedicine and the life sciences. This is the main component of PubMed, which provides access to MEDLINE and some other resources, including articles published in MEDLINE journals, which are beyond the scope of MEDLINE, such as general chemistry articles. Approximately 5200 journals published in the United States and more than 80 other countries have been selected and are currently indexed for MEDLINE. A distinctive feature of MEDLINE is that the records are indexed with NLM's controlled vocabulary, the Medical Subject Headings (MeSH®).

In addition to MEDLINE citations, PubMed also contains:

- process citations which provide a record for an article before it is indexed with MeSH and added to MEDLINE or converted to out-of-scope status,
- records that precede the date that a journal was selected for MEDLINE indexing (when supplied electronically by the publisher),
- OLDMEDLINE citations that have not yet been updated with current vocabulary and converted to MEDLINE status,
- articles that are out-of-scope (e.g. covering plate tectonics or astrophysics) from certain MEDLINE journals, primarily general science and general chemistry journals, for which the life sciences articles are indexed with MeSH for MEDLINE,
- some life science journals that submit full text to PubMed Central® and may not yet have been recommended for inclusion in MEDLINE although they have undergone a review by NLM, and some physics journals that were part of a prototype PubMed in the early to mid-1990s, and
• citations to author manuscripts of articles published by National Institutes of Health-funded researchers.

One of the ways users can limit their retrieval to MEDLINE citations in PubMed is by selecting MEDLINE from the subsets menu on the limits screen. MEDLINE® is the NLM’s premier bibliographic database that contains approximately 18 million references to journal articles in life sciences with a concentration on biomedicine. PubMed is provided free of charge by the National Library of Medicine. PubMed includes MEDLINE, as well as Pre-MEDLINE and select online publications provided directly from publishers.

**Grey or Gray Literature, Deep Web searching**

Since the mid-1980s and particularly since the explosion of the Internet and the opportunity to publish all kinds of information electronically, there has been an information revolution. This revolution is making it increasingly impossible for people to read everything on any particular subject. In this case medicine, health care, nursing, or any other EBPs are no exception. There is such a vast amount of data being written, published and cited that Internet search engines and medical specialist databases such as MEDLINE, EMBASE, CINAHL, the Cochrane Library and PsycINFO cannot hope to catalogue or index everything. There are bound to be valuable sources of medical evidence, which can nonetheless prove useful when doing systematic reviews but have not been ‘captured’ by commercial electronic publishers.

Rather than compete with the published literature, grey literature has the potential to complement and communicate findings to a wider audience. Grey (or gray – alternative spelling) literature is also known as deep or hidden web material and refers to papers that have not been commercially published.

The US Interagency on Gray Literature Working Group (1995) defined grey literature (or greylit as it is sometimes referred to in the information management business) as foreign or domestic open source material that usually is available through specialized channels and may not enter normal channels or system of publication, distribution, bibliographical control or acquisition by booksellers or subscription agents.

Grey literature has also been defined as information produced at all levels of government, academic, business, and industry in electronic and print formats not controlled by commercial publishing (i.e. where publishing is not the primary activity of the producing body) (4th International Conference on Grey Literature, cited in Tillett & Newbold, 2006).

Grey (or gray – alternative spelling) literature includes documents such as:

- conference papers and proceedings,
- theses and dissertations,
- newsletters,
- blogs,
- raw data such as census and economic results or ongoing research results, and
- non-independent research or other documents produced and published by government agencies, academic institutions and other groups that are not distributed or indexed by commercial publishers.

When building a search strategy for grey literature, it is important to select terms specifically for each source. In using mainstream databases, or Google-type searches (including Google Scholar), it is best to draw from a list of keywords and variations developed prior to starting the
search. To be consistent and systematic throughout the process, using the same keywords and strategy is recommended. It is important to create a strategy, compile a list of keywords, wildcard combinations and identify organizations that produce grey literature that is relevant to the research question at hand. If controlled vocabularies are used, record the index terms, qualifiers, keywords, truncation, and wildcards.

Searching the medical grey literature can be time-consuming because there is no ‘one-stop shopping’ database or search engine that indexes materials the way, for example as CINAHL does for nursing and allied health, or MEDLINE does for the biomedical sciences. It should be remembered that access to bibliographic databases may depend on the subscriptions taken by your library service and the search interface may also vary depending on the database vendor, for example Ovid, EBSCO, ProQuest, etc. or whether you access MEDLINE via the free PubMed interface.

Deciding which sources should be chosen will depend on the research question, the availability, suitability and quality of the data available, as well as the scope of the enquiry. For example, a question that seeks to determine the global prevalence of a condition such as dementia, which is associated with more than 100 different diseases, will require a strategy that searches across a far greater number of resources than a question examining the prevalence of a specific disease within a single regional setting.

**Grey literature sources**

The following search engines are very useful for finding health-based scientific literature:

- www.scirus.com
- www.metacrawler.com
- www.hon.ch/Medhunt/Medhunt.html
- www.medworld_stanford.edu/medbot/
- http://sumsearch.uthscsa.edu/cgi-bin/SUMSearch.exe/
- www.intute.ac.uk/healthandlifesciences/omnilost.html
- www.mdchoice.com/index.asp
- www.science.gov/
- http://medworld.stanford.edu/medbot/
- http://omnimedicalsearch.com/
- http://www.ingentaconnect.com/
- http://www.medical-zone.com/

Scirus (www.scirus.com), for example, is a science-specific search engine with access to over 410 million science-related web pages (as of February 2011), and it indexes sites that other search engines do not. Its medical sites include ArXiv.org, Biomed Central, Cogprints, DiVa, LexisNexis, and PsyDok. PsyDok is a disciplinary Open Access repository for psychological documents.

PsyDok is operated by Saarland University and State Library (SULB), which also hosts the special subject collection psychology and the virtual library psychology. PsyDok is a free, full-text e-print archive of published, peer-reviewed journal post-prints plus pre-publications, reports, manuals, grey literature, books, journals, proceedings, dissertations and similar document types.

Search the World Wide Web for higher level – usually government-affiliated – funding bodies, for instance Australia’s NHMRC (National Health and Medical Research Council) or MSAC
The Synthesis of Prevalence and Incidence Data

(Medical Services Advisory Committee) for pointers to reports such as clinical trials or reviews from funded research programmes.

Be aware that there are health information gateways or portals on the Internet containing links to well organized websites containing primary research documents, clinical guidelines, other sources, and further links. For example:

- National Electronic Library for Health (UK), http://www.nelh.nhs.uk/

**Finding grey literature on a government website**

Generally, most health or medicine-related government-sponsored or maintained websites will go to the trouble of showing:

- how or if their documents are organized alphabetically, topically, or thematically;
- how individual documents are structured, i.e. contents pages, text, executive summary, etc.;
- database-type search strategies to find them;
- links to other web sites or other documents that are related to the documents that they produce;
- when their collection of grey literature has been updated; and
- documents in PDF or Microsoft Word downloadable form.

**Sources of prevalence and incidence data**

There are a wide variety of sources of epidemiological information within the grey literature. Two main types exist: data that is population based and collected through personal interviews or examinations; and data collected from vital and medical records (Parkin & Bray, 2005).

Parkin and Bray (2005) outline various sources of epidemiological data including:

- population based censuses and surveys (e.g., national or state health survey data) are a good source of information on the prevalence of common conditions or symptoms, which may otherwise be overlooked in clinical settings;
- community based surveys (such as those conducted within the primary care setting) can capture all rates of disease and the general morbidity of a sample of individuals residing within a particular community;
- hospital activity statistics and records generally summarize admissions and discharges from hospital. This data is typically event based and may be affected by the hospital system and its policies;
- disease registries are often population-based disease registries established to record new cases of certain serious diseases in defined populations. They capture incidence data and are often used as notification systems for infectious diseases. The quality of the data from disease registries may be variable but likely to be better for serious or highly contagious diseases. For example, cancer registries are the most commonly and widely used registries for collecting incidence data. This is because most patients
with this condition will present for diagnosis and treatment. A worldwide network of cancer registries has been established, which provides data on defined populations. Incidence data from cancer registries worldwide that meets predetermined quality standards are also published every five years in the Cancer Incidence in Five Continents series;

- national administrative data sources (clinical records, insurance data);
- vital statistics data (especially death certificate data), government surveillance data and reports, Centers for Disease Control and Prevention data, healthcare utilization records, etc.;
- disease associations (e.g., American Diabetes Association);
- medical books, grey literature, and reports from experts;
- census data, or population-based registers;
- notification systems (especially for infectious diseases);
- primary care contacts; and
- diagnostic services such as pathology laboratories.

Prevalence estimates may vary across sources and it is important to understand how variations in data collection may impact on reported estimates. For example, in Australia five national data sources were identified by the Australian Institute of Health and Welfare (AIHW, 2009) that could be used to assess diagnosed diabetes prevalence: the Australian Diabetes and Obesity Lifestyle Study (AusDiab), the National Diabetes Services Scheme (NDSS) database, the National Health Survey (NHS), the Medicare Benefits Schedule (MBS) database and the Pharmaceutical Benefits Scheme (PBS) database. Differential diagnosed diabetes prevalence estimates were reported by each data source.

Accordingly, three criteria were established to describe and compare the estimates (e.g., the coverage and currency of the data, and frequency of data updates) derived from these sources and to determine which was preferential in relation to the specific investigation at the time.

Review authors must be aware of limitations in each of the data sources. For example, some of data sources were a decade old, so the more recent data source may be preferred if more recent prevalence estimates were desired. Conversely, some of alternative data sources provided prevalence estimates for specific disorders and subtypes, which the more recent data sources did not provide. This again highlights importance of selecting databases and sources that can best provide the data that addresses the research question and the topic at hand.

**Constructing a search strategy**

Defining the scope of a research question and understanding the existing literature base, including the gaps and uncertainties surrounding the definitions and conceptualizations of key terms is an important step in the systematic review process and one that will inform the scope of the systematic search strategy and the selection of electronic databases and grey literature sources to be searched.

As mentioned in a previous section, a research question that seeks to capture the prevalence estimates of a broad and differentially defined condition within the general population across multiple geographical locations will require a search that is considerably greater in scope than a research question examining the prevalence of a narrowly defined condition within a specific population at a single location.
It is advisable that an initial, quick and dirty scoping search of the literature is undertaken on the topic of interest at the protocol development stage to ensure that sufficient evidence and data are available and accessible to answer the research question(s) posed.

A well-formulated research question and clearly defined inclusion criteria provide the foundation for the development of an efficient and effective search strategy (Aromataris & Riitano, 2014). As prevalence and incidence data are reported within the published, peer-reviewed literature the standard JBI three-step approach to searching for literature can be applied to locate this kind of empirical data.

**JBI three-step search strategy**

Once there is sufficient clarity around the review topic and the research question, phase one of the three-step search strategy can be implemented. Phase one is composed of two steps:

The first step broadly involves identifying the key terms and concepts that will ultimately comprise the final search string. This process involves identifying key terms from the review question and the inclusion criteria in the review protocol and mapping them into a logic grid, which is made up of columns; each of which represents a discrete concept.

In a standard review of effectiveness this would typically involve identifying each of the key concepts that align with the population, intervention, comparator and outcome(s) of interest (PICO). For reviews of prevalence and incidence, the columns of the logic grid can be adapted to reflect and capture each of the key concepts that relate to the CoCoPop mnemonic.

The second step involves conducting an initial and limited search using a selected database such as PubMed to find any alternative terms or synonyms to the key concepts that have already been identified. This involves analyzing the text words contained in the titles and abstracts of any relevant papers to identify any alternative key terms being used in the literature.

The index terms used by the bibliographic database to describe the content of these articles should also be examined and added to the logic grid along with any newly identified key concepts or synonyms. A search strategy should always be constructed using a combination of both free text words (those found in the title and abstract of articles) and index terms (standard terms used by a database to categorize articles) to ensure that the approach is comprehensively capturing all potentially relevant articles.

The process to identify appropriate index terms will need to be repeated for each included database as most major databases utilize a unique set of standardized terms (i.e. thesaurus terms) to categorize their indexed studies. In other words, to build a comprehensive search, each search string will need to be individually tailored to requirements of the database being searched.

Phase two of the search process involves implementing the database-specific searches for each included database. Once all the search terms have been collected and finalized they can be grouped together in most bibliographic databases using some form of Boolean logic. First, each set of concepts (index terms and free-text words) are combined through the use of the Boolean operator OR. Each set of concepts or themes are then subsequently combined using the Boolean operator AND.
Boolean operators such as “OR” and “AND” are used to connect key terms and parentheses are used to maintain the logical sequence of a search query. Without parentheses, a search is executed from left to right. Words that you enclose in parentheses are searched first. Why is this important? Parentheses allow you to control and define the way the search will be executed. The left phrase in parentheses is searched first; then based upon those results the second phrase in parentheses is searched. Search field qualifiers will need to be added to key terms to ensure the database searches in the correct field code (i.e., in the title/abstract etc.). Wildcard characters and shortcuts may also need to be applied to variations in spelling or terminology.

Example of PubMed search field descriptors and wildcard characters:

- ab = words in abstract
- exp = before an index term indicates that the term was exploded
- hw = word in subject heading
- mp = free text search for a term
- pt = publication type
- sh = subject heading
- ti = words in title
- tw = textwords in title/abstract
- ? = in middle of term indicates use of a wildcard
- / = MeSH subject heading (and includes all subheadings being selected)
- $ = truncation symbol
- adj = two terms where they appear adjacent to one another (so adj4, for example, is within four words)

Typically, the selection and combination of key terms for a search string requires careful planning and consideration, and involves a process with multiple iterations until a combination of search terms is found that successfully captures all relevant known studies (sensitivity) while simultaneously excluding enough illegible studies (specificity) to make the search results manageable.

Once the search has been conducted in each database, the search results can be limited or refined by selecting a filter from the range that is available. Search results are often filtered by publication date and language; however, any filter can be used as long as it is relevant to the review question and its utilization has been adequately justified or explained in the body of the review. Reviewers should be aware however that filters are not fail proof and their application can cause some potentially relevant studies to be omitted from the search results.

**Limiting by date:**

Limiting the search by date may be used where the focus of the review is on a more recent condition (such as video game addiction) or factor (such as electronic health records). However, potentially relevant studies as well as seminal, early studies in the field may be missed if the limit set is too recent thus date limits should be used in an informed way, based on knowledge of key papers relevant to the review question.

**Limiting by language:**

Limiting by language is a common practice in settings with lack of ready access to translators. The caveat associated with excluding papers based upon language is that important cultural contexts or findings may be missed. The exclusion of selected languages also means the
review audit trail is incomplete. If limiting by language is required, it is preferable to search inclusively, and keep a record of numbers of studies per language group. This allows the reader to identify how many studies have been identified, but are not included, therefore promoting transparency in the process.

Searching is an iterative and potentially infinite process. The total number of relevant studies is never known from the outset so it is important that the search results are constantly examined to determine whether a relevant and manageable set of results are being retrieved. One way to establish whether the search is sensitive (retrieving relevant articles) is to check whether the already identified studies are being captured by the search. If this is not found to be the case, the search can be modified and refined accordingly.

Differences in thesaurus or index terms, search term qualifiers or descriptors (e.g., [tiab] in PubMed meaning "title/abstract") and wildcard characters across various databases as well as the many different rules that govern how a search should be undertaken within each platform means that reviewers need to be cognizant of the variations and limitations of each database if they are to achieve accurate results. The help of an experienced research librarian or information scientist who can examine the search strategy for errors or omissions is highly recommended.

Phase three involves commonly used manual means of locating potentially relevant studies that may have been otherwise missed through the database search. One such method involves searching for studies that may be relevant to the review by manually screening the reference lists of all studies that have been already selected for inclusion in the review. Hand searching is another method that can be used to identify relevant studies. This involves a page-by-page inspection of the entire contents of a preselected journal issue or conference proceedings to identify potentially eligible reports or studies.

Managing references

Commonly, electronic databases used to search for papers allow users to construct complex search strategies, save their results and export their findings using a text file or other alternative format. The file containing the search results can then be imported into a bibliographic software program such as Endnote, which can be extremely helpful for managing study records and removing duplicate references.

Example search strategy

The selection of key terms and synonyms is equally important for reviews of prevalence and incidence as it is for other types of reviews. However, finding the perfect combination of key terms to maximize the sensitivity and specificity of the search is less critical for reviews of prevalence and incidence.

Unlike other review types, search strategies for reviews of prevalence and incidence can utilize the epidemiological indicator of interest as a key concept within the search string. This type of data (prevalence, incidence) is used in a consistent manner and often explicitly referred to by study authors in the abstracts and titles of their publications.

By including the epidemiological indicator of interest as a key concept in the search string, review authors are able to automatically improve the sensitivity and specificity of the search by returning a yield that drops studies that do not report prevalence and incidence data while simultaneously retaining those that do. A caveat to this approach is if the prevalence or
incidence data is not the primary purpose of the research study and is only reported in the full-text of the article, which may be the case in RCTs that incidentally report the prevalence and incidence of a factor following a particular intervention alongside other outcome data. Therefore, careful consideration to including the epidemiological indicator is required, and may depend on the review question and scope of research.

Notwithstanding this consideration, a simple search string that utilizes a few key concepts together with the epidemiological indicator of interest will typically be sufficient for reviews of prevalence and incidence.

For example, a relatively simple search string was developed using key inclusion criteria terms for a systematic review that sought to establish the global incidence and prevalence of type-2 diabetes in children and adolescents (Table 2).

Table 2: Search Terms and Search Strategy

<table>
<thead>
<tr>
<th>Search terms</th>
<th>Search terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>('Infant' OR 'Infants' OR 'Toddler' OR 'Toddlers' OR 'Child' OR 'Children' OR 'Adolescent' OR 'Adolescents' OR 'Teens' OR 'Teen' OR 'Teenagers' OR 'Teenager' OR 'Youth' OR 'Youth' OR 'Adolescence')</td>
</tr>
<tr>
<td>Condition</td>
<td>('Diabetes Mellitus, Non Insulin Dependent' OR 'Diabetes Mellitus, Non-Insulin-Dependent' OR 'Type 2 Diabetes Mellitus' OR 'Diabetes Mellitus, Slow-Onset' OR 'Diabetes Mellitus, Slow Onset' OR 'Diabetes Mellitus, Stable' OR 'Stable Diabetes Mellitus' OR 'Diabetes Mellitus, Type II' OR 'NIDDM' OR 'Diabetes Mellitus, Adult-Onset' OR 'Diabetes Mellitus, Adult Onset' OR 'Diabetes Mellitus, Noninsulin Dependent' OR 'T2DM' OR 'T2D')</td>
</tr>
<tr>
<td>Outcome (i.e., epidemiological indicator)</td>
<td>('Incidence' OR 'Incidences' OR 'Prevalence' OR 'Prevalences')</td>
</tr>
</tbody>
</table>

Documenting a search strategy

One of the major strengths of a systematic review is the systematic approach to identifying relevant studies. An important factor in this process is the transparent reporting of the search strategy and the findings of the search.

The documentation of search strategies contributes to the scientific validity and credibility of the systematic review by allowing readers to evaluate the steps taken by the reviewer, the decisions made and the comprehensiveness of the search strategy, including the type and number of literature sources searched. Another key characteristic of a systematic review is its reproducibility - other researchers should be able to apply the same methods in a given review and arrive at similar conclusions.
Accordingly, the review should include the details of the search strategy, such as the databases and sources searched and any limits or filters applied to the search strings (i.e., language or publication period etc.) in the appendix of the review.

The methods section of the review should also provide a more detailed account of all limitations to the scope of the search in terms of dates and languages etc., and provide a list all of the grey literature sources and bibliographic citation databases searched, including specification of the platform used to search a particular database. Details of the hand searches that were conducted (e.g. the searching of journals that are not indexed in electronic databases) and the sources subjected to manual review should also be included. A discussion of the limitations and implications of any restrictions should also be detailed in the discussion section of the review so the reader can glean an understanding of how the results of the search should be interpreted.

JBI endorses the PRISMA statement, which provides guidance on items for reporting in systematic reviews and meta-analyses. PRISMA stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses. The full PRISMA statement can be located at the PRISMA website: http://www.prisma-statement.org/. These guidelines dictate that the full search strategy for a least one major database should be appended and published alongside the systematic review.

Online repositories of systematic reviews, such as the JBI Database of Systematic Reviews and Implementation Reports and the Cochrane Database of Systematic Reviews, allow for publication of all the search strategies and filters applied to the various databases and resources included in the review.

The results at each stage of the search process such as the number of titles identified through the database and grey literature search should also be reported in the systematic review report and presented in a flow diagram in line with the PRISMA guidelines.

Chapter 7: Systematic Review Development

This section provides information on how to synthesize evidence relating to prevalence and incidence data. It provides guidance on the components that should comprise a systematic review of prevalence and incidence data and the information that each component should contain. Specifically, guidance is provided on the following components: layout of the report, inclusion criteria (i.e. PICO), search strategy, critical appraisal, data extraction, data synthesis, results, and conclusions. The section also presents a series of questions designed to prompt the reviewer to check that certain key information or requirements have been adequately addressed in the review. This guidance is based on the Joanna Briggs Institute approach to systematic reviews of prevalence and incidence.

Title

The title should be clear, explicit and reflect the core elements of the review. Titles should not be phrased as questions or conclusions and there should be congruency between the title, review objectives/questions and inclusion criteria. The title should include the phrase “A systematic review.” Refer to the guidance above for title development.
Executive summary
This section forms a structured abstract of the main features of the systematic review. It must be no longer than 500 words and should contain no abbreviations or references. The executive summary must accurately reflect and summarize the systematic review. The executive summary should include the following headings:

Background
This section should briefly describe the issue under review including the population, condition and context that are documented in the literature. The background should be an overview of the main issues. It should provide sufficient detail to justify why the review was conducted and the choice of the various elements such as the condition and context.

Objectives
The review objectives should be stated in full, as detailed in the protocol section.

Inclusion criteria
Population: The report should provide details about the types of participants included in the review. Useful details include: age range, gender, profession, etc. Information supporting the decisions about the types of participants should be explained in the background.
Condition: This section should present all the conditions examined, as detailed in the protocol.
Context: This section should present all the contexts examined, as detailed in the protocol.
Types of studies: As per the protocol section, the types of studies that were considered for the review should be included. There should be a statement about the target study type and whether or not this type was not found. The types of study identified by the search and those included should be detailed in the report.

Search strategy
A brief description of the search strategy should be included. This section should detail search activity (e.g. relevant databases searched, initial search terms or keywords, and any limitations) for the review, as predetermined in the protocol.

Methodological quality
Reviewer’s should make mention of how the studies included in the review were appraised.

Data extraction
This section should include a brief description of the types of data collected and the instrument (as specified in the protocol) used to extract data.

Data synthesis
This section should include a brief description of how the data was synthesized –as a meta-analysis or as a narrative summary or in a graphical or tabular form.

Results
This section should include a brief description of the findings of the review.

Conclusions
This section should include a brief description of the conclusions of the review.
Implications for practice
This section should include a brief description of how the findings and conclusions of the review may be applied in practice, as well as any implications that the findings may have on current practice.

Implications for research
This section should include a brief description of how the findings of the review may lead to further research in the area—such as gaps identified in the body of knowledge.

Background
The background section should be comprehensive and cover all the main elements of the topic under review. Many reviewers will find that the background provided with the protocol needs modification or extension following the conduct of the review proper. The background should detail any definitions important to the review. The information in the background section must be sufficient to put the inclusion criteria into context. The background section should conclude with a statement that a preliminary search for previous systematic reviews on the topic was conducted (state the databases searched e.g. JBI Library, Cochrane Library, CINAHL, PubMed, PROSPERO). If there is a previous systematic review on the topic, it should be specified how the proposed review differs. JBI places significant emphasis on a comprehensive, clear and meaningful background section to every systematic review particularly given the international circulation of systematic reviews, variation in local understandings of clinical practice, health service management and client or patient experiences. It is recommended that all JBI systematic reviews should contain a sentence clearly indicating:

The objectives, inclusion criteria, and methods of analysis for this review were specified in advance and documented in a protocol.

This sentence should appear as the final line of the background/introduction section of the review report and complies with the recommendations for reporting of systematic reviews detailed in the PRISMA guidelines.

Review objectives/review questions
As discussed previously in the protocol section, the objective(s) of the review should be clearly stated. This should be followed by specific question(s) that have been addressed in the review.

Inclusion criteria
As detailed in the protocol, this section of the review should detail the basis on which studies were considered for inclusion in the systematic review and should be as clear and unambiguous as possible. For a systematic review of prevalence and incidence studies, aspects include: population, condition, and context.

Search strategy
This section should detail how the reviewers searched for relevant papers. The databases that were searched must be listed along with the search dates. A detailed search strategy for at least one of the major databases searched should be appended to the review. The documentation of search strategies is a key element of the scientific validity of a systematic review. It enables readers to look at and evaluate the steps taken, decisions made to consider the comprehensiveness and exhaustiveness of the search strategy for each included database.
A JBI review should consider papers published in both commercial (e.g. PubMed, Cochrane, JBI Library etc.) and non-commercially operated databases (grey literature).

Each electronic database is likely to use a different system for indexing key words within their search engines. Hence the search strategy will be tailored to each particular database. These variations are important and need to be captured and included in the systematic review report. The timeframe chosen for the search should be justified and any language restrictions stated (e.g. only studies published in English were considered for inclusion).

**Method of the review**

**Assessment of methodological quality**

This section should detail the approach to critical appraisal, not the assessment results, and should be consistent with the protocol. Any deviations from the protocol must be reported and explained. The report should detail the criteria that were considered when determining the methodological quality of papers considered for inclusion in the review. Critical appraisal tools must be appended to the review.

The primary and secondary reviewer should discuss each item of appraisal for each study included in their review. The discussions should focus on what is considered acceptable to the needs of the review in terms of the specific study characteristics. The reviewers should be clear on what constitutes acceptable levels of information to allocate positive appraisal compared with a negative, or response of unclear. This discussion should take place before independently conducting the appraisal. The critical appraisal tool should be appended to the review.

**Data collection**

This section of the review should include details of the types of data extracted from the included studies. Standardized data extraction tools allow the extraction of the same types of data across the included studies and are recommended for JBI systematic reviews. The included studies may include several conditions; however, the review should focus on extracting information related to the research questions and condition of interest. Information that may impact upon the generalizability of the review findings such as study methods, setting and population characteristics should also be extracted and reported. Population characteristics include factors such as age, past medical history, co-morbidities, complications or other potential confounders. Systematic reviewers should aim to reduce errors in data extraction by using two independent reviewers and a standardized data extraction instrument. The data extraction tool used must be appended to the review.

The data collection should include the following items and a brief description is provided for each item:

**Study details**

- **Reviewer** – Mostly includes details or ID of the primary reviewer
- **Study ID/Record Number** - is a numeric code to identify the study from which the effect size estimate was obtained
- **Date** – the date when this data extraction form was filled
- **Study title** – the full title of the study
The Synthesis of Prevalence and Incidence Data

- Author - This is an alphabetic or character code which is usually the first few characters of the primary study author's name. This serves as an easy way to identify the study in the bibliography
- Year – the year of publication
- Journal – the journal in which the article was published

Study Method
- Aims of the study – as stated in the report
- Setting – may refer to hospital or community or aged care facility. May also refer to rural/urban etc.
- Study design – briefly describing the type of study design. For e.g. if it is a cohort study or cross-sectional study
- Follow-up or study duration – any details on the duration of the study or follow-up of the participants
- Subject characteristics – Includes age, sex, country/location, sample size, diagnosis and other relevant characteristics.
- Dependent variable
- Outcomes – the primary outcome measured and where relevant includes associated secondary outcomes.
- Outcome measurements – describe the scales or tools used to measure the outcomes. For e.g. a standardized pain scale to measure pain.
- Ethical approval – yes/no
- Method of data analysis

Results
- Prevalence n/N (%)
- Proportion and 95% Confidence Intervals
- Incidence n/N (%)
- Proportion and 95% Confidence Intervals and duration of recruitment or the study
- Authors’ comments
- Reviewer comments

Data synthesis
This section should detail the approach to data synthesis, not the results of the synthesis. The review should detail how the reviewers synthesized the data extracted from the included studies and how it was applied consistently across all included studies. The types of data detailed in this section should be consistent with the methods used for data collection and the included study designs.

Different methods exist for presenting a synthesis and these are outlined below:
- narrative
- tabular
- graphical
- meta-analysis
If the data is heterogeneous and is presented as a narrative summary, sources of heterogeneity should be discussed (e.g. clinical, methodological or statistical) as well as on what basis it was determined inappropriate to combine the data statistically (such as differences in populations, study designs or clinical or statistical heterogeneity). Where meta-analysis was used, the statistical methods and the software used should be described.

**Results**

This section should allow the reader to clearly follow how the included studies were identified and selected for inclusion in the review. In addition, the number of papers excluded should also be stated. There should be a narrative description of the process accompanied by a flowchart of review process (from the PRISMA statement) detailing the flow from the search, through study selection, duplicates, full text retrieval, and any additions from 3rd search, appraisal, extraction and synthesis (Figure 4).

Details of full-text articles retrieved for critical appraisal should be given. There should be separate appendices for details of included and excluded studies and for excluded studies; reasons should be stated on why they were excluded.

**Figure 4: Flowchart Detailing Identification and Selection of Studies for Inclusion in the Review**
Description of studies
This section of the results should also include an overall description of the included studies (with reference to the table in the appendices), with the main aim to provide some context to the results section and sufficient detail for the reader to determine if the included studies are similar enough to combine in meta-analysis. Specific items/points of interest from individual studies may also be highlighted here. Additional details may include the assessment of methodological quality, characteristics of the participants and types of interventions and outcomes.

Where a systematic review has several foci, the results should be presented in a logical, structured way, relevant to the specific questions. The roles of tables and appendices should not be overlooked. Adding extensive detail on studies in the results section may crowd the findings, making them less accessible to readers, hence the use of tables and in text reference to specific appendices is encouraged.

Methodological quality
This section should focus on methodological quality as determined by the relevant critical appraisal checklist. There should be a narrative summary of the overall methodological quality of the included studies, which can be supported (optional) by a table showing the results of the critical appraisal (see Table 3 for example). Where only few studies are identified, or there are specific items of interest from included studies, these should be addressed in the narrative also, particularly where studies were deficient, or particularly good. Use of N/A should also be justified in the text.

Table 3: Critical Appraisal Results for Included Studies Using the JBI-Prevalence Critical Appraisal Checklist

<table>
<thead>
<tr>
<th>Study</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s) ref</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>U</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>U</td>
</tr>
</tbody>
</table>

Y - Yes, N - No, U - Unclear

Findings of the review
Although there is no defined structure for this section, the findings of the review should flow logically from the review objection/question i.e. they must ultimately answer the question! Findings should be presented and a narrative, tabular, graphical or meta-analysis should constitute part of this section.

With detail on the studies reported, the results section then focuses on providing a detailed description of the results of the review. For clarity and consistency of presentation, JBI recommends that the reviewer, in discussion with their review panel, give consideration to whether the specific review question used to structure the results section, or whether the findings can be reported under the conditions specified in the protocol. When a systematic review seeks to address multiple questions, the results may be structured in such a way that particular conditions are reported according to the specific questions.
Given there is no clear international standard or agreement on the structure or key components of this section of a review report, and the level of variation evidence in published systematic reviews, the advice here is general in nature. In general, findings are discussed textually and then supported with meta-graphs, tables, figures as appropriate. Graphs may be particularly useful for presenting prevalence and incidence data where meta-analysis is not possible.

The focus should be on presenting information in a clear and concise manner. Any large or complex diagrams/tables/figures should be included as appendices so as not to break the flow of text. Meta-view graphs represent a specific item of analysis that can be incorporated into the results section of the review. However, the results are more than meta-view graphs, and whether this section is structured based on the intervention of interest, or some other structure, the content of this section needs to present the results with clarity.

A narrative summary is commonly used where meta-analysis is not possible. A narrative summary describes the included studies and provides conclusions about the evidence. With a narrative summary, readers may not be able to discern how evidence was weighted and whether conclusions are biased. It is therefore important that when summarizing findings in narrative form, there is a clear structure to the summary, with an emphasis on reporting the characteristics of included studies along with data extracted relevant to the review outcomes (Lockwood & White, 2012). A narrative summary should include the presentation of the quantitative results reported in individual studies; where available, the point estimates (one value that represent or best estimate of effects) and the interval estimates (usually presented as 95% confidence intervals) for the effects should be provided. Due to the flexibility of narrative summary in terms of the amount of data that can be conveyed textually, a structure that applies to each sequence or reporting of results from each study should be discussed beforehand and applied by the systematic review authors. This will ensure that there is consistency across the results section of a review. If a structure is not followed there may be substantial variability in reporting of results causing the data to appear incomplete or unreliable (Lockwood & White, 2012). Therefore adherence to this structure is critical; if studies do not provide the relevant information to comply with the structure this should be made clear in the summary (Lockwood & White, 2012).

**Meta-analysis**

It is important to combine the studies in an appropriate manner; otherwise the conclusions that are drawn will not be valid. Study results should be pooled in statistical meta-analysis where possible. All results must be double entered in order to avoid data entry errors. Where statistical pooling is not possible the findings can be presented in narrative summary or graphical form, as previously discussed.

This section of the report should describe the data type, the required effects model used (random/fixed), the statistical method of meta-analysis required and the size of confidence limits to be included in the calculations. The method used will depend on the data type. In terms of confidence intervals, the default setting of most meta-analysis software is to calculate 95% confidence intervals; however this can be adjusted to either 90% or 99% as required. Once all the appropriate settings have been selected, the forest plot summarizing the results of individual studies and their combined meta-analysis can be generated.
Discussion
This section should discuss the results of the synthesis as well as any limitations of the primary studies included in the review and of the review itself (i.e. language, access, timeframe, study design, etc.). The results should be discussed in the context of current literature, practice and policy.

The aim of this section is to minimize and discuss the main findings – including the strength of the evidence, for each main outcome. It should address the issues arising from the conduct of the review including limitations and issues arising from the findings of the review (such as search limitations). The discussion does not bring in new literature or information that has not been reported in the results section. The discussion does seek to establish a line of argument based on the findings regarding the effectiveness of an intervention, or its impact on the outcomes identified in the protocol. The application and relevance of the findings to relevant stakeholders (e.g. healthcare providers, patients, and policy makers) should also be discussed in this section.

Points to consider this section include:

- Where any problems identified undertaking the search (perhaps there is little primary research on this topic or perhaps it is poorly indexed by the databases that were searched or perhaps the search was insufficient)?
- What limitations were found in the included primary research (e.g. were there inconsistencies or errors in reporting)?
- How do the review findings fit with what is currently known on the topic (from issues highlighted in the Background section)?
- Are the findings generalizable to other populations of participants/healthcare settings etc?

Conclusions
This section should begin with an overall conclusion based on the results. The conclusions drawn should match with the review objective/question.

The conclusion section of a systematic review should provide a general interpretation of the findings in the context of other evidence and provide a detailed discussion of issues arising from the findings of the review and demonstrate the significance of the review findings to practice and research. Areas that may be addressed include:

- a summary of the major findings of the review,
- issues related to the quality of the research within the area of interest,
- other issues of relevance, and
- potential limitations of the systematic review.

Implications for practice
It should be stated how the findings of the review impact on clinical practice in the area. If there is sufficient evidence to make specific recommendations for practice, then the appropriate JBI Grades of Recommendation should be assigned to each recommendation based on the study design that led to the recommendation.

Grades of Recommendation are used to assist healthcare professionals when implementing evidence into practice.
The JBI grades of recommendation are informed by the GRADE working party, which has a binary system for recommendations with only the two options: strong or weak. The benefit of such a system is its ease of interpretation by both clinicians and patients. When forming a recommendation, GRADE recommends the following four key factors be considered: the balance between desirable and undesirable effects, the quality of the evidence, values and preferences, and costs (Andrews et al., 2013; Guyatt et al., 2008). Recommendations can be made for or against particular management approaches (Andrews et al., 2013; Guyatt et al., 2008). Due to negative connotations associated with the term weak, GRADE have provided the alternative terms of conditional, discretionary or qualified recommendations which can substitute for the term weak (Andrews et al., 2013).

Recommendations should be actionable. When wording recommendations, the following need to be specified as much as possible, as the more specific a recommendation is, the easier it is to implement and the more likely it is that it will be acted upon (Andrews et al., 2013; Guyatt et al., 2008; Michie & Johnston, 2004; Michie & Lester, 2005; Woolf, Schunemann, Eccles, Grimshaw, & Shekelle, 2012):

- the population (i.e. age, sex, condition),
- intervention (i.e. dose, timing, intensity, professional),
- any comparator (where applicable), and
- the setting (where applicable).

Wording for strong recommendations should be in the active voice. This can be achieved by using phrases such as we recommend/ Health professionals should/ or Do, or must (Andrews et al., 2013; Guyatt et al., 2008; Hillier et al., 2011) For weak recommendations, phrases such as we suggest/health professionals might (could/may) /we conditionally recommend can be used. (Andrews et al., 2013; Guyatt et al., 2008; Hillier et al., 2011) An example of a strong recommendation is: Health professionals should provide written information detailing methods of self-management of blood glucose levels for patients with type 2 diabetes living in the community. An example of a weak recommendation is: Health professionals may provide information regarding self-management of blood glucose levels for patients with type 2 diabetes living in the community. As mentioned above, recommendations can be made for or against particular management approaches. When making strong recommendations against a certain strategy, terms such as we recommend against, health professionals should not, or don’t can be used.

The use of the term consider has been advised against due to its difficulty of interpretation when determining if a certain activity was considered (Lomotan, Michel, Lin, & Shiffman, 2010). Other terms to avoid include the use of phrases such as where necessary or when clinically indicated (Woolf et al., 2012). Whatever words are chosen to convey the recommendation, the connection between the strength of the recommendation and the wording needs to be explicit, which can be achieved through consistent use of the same wording structure (Akl et al., 2012).

See table 4 for the JBI Grades of Recommendation.
Table 4: The JBI Grades of Recommendation

<table>
<thead>
<tr>
<th>JBI Grades of Recommendation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade A</td>
<td>A strong recommendation for a certain health management strategy where it is clear that desirable benefits outweigh undesirable benefits of the strategy; where there is evidence of sufficient quality supporting its use; there is a benefit or no impact on resource use, and values, preferences and the patient experience have been taken into account.</td>
</tr>
<tr>
<td>Grade B</td>
<td>A weak recommendation for a certain health management strategy where desirable benefits appear to outweigh undesirable benefits of the strategy, although this is not as clear; where there is evidence supporting its use, although this may not be of high quality; there is a benefit, no impact or minimal impact on resource use, and values, preferences and the patient experience may or may not have been taken into account.</td>
</tr>
</tbody>
</table>

**Implications for research**

This section should include clear, specific recommendations for future research based on gaps in knowledge identified from the results of the review. Implications for research should avoid generalized statements calling for further research, but should be linked to specific issues.

**Conflicts of interest**

A statement which either declares the absence of any conflicts of interest or which describes a specified or potential conflict of interest should be made by the reviewers in this section.

**Acknowledgements**

Any acknowledgements should be made in this section e.g. sources of external funding or the contribution of colleagues or institutions. It should also be noted if the systematic review is to count towards a degree award.

**References**

The references should be appropriate in content and volume and include background references and studies from the initial search.

**Appendices**

Appendices should be numbered in the order in which they have been referred to in the body of the text. There are several required appendices for a JBI review:

- **Appendix I: Search strategy**
  
  A detailed search strategy for at least one of the major databases searched must be appended.

- **Appendix II: Critical appraisal instrument**
  
  The critical appraisal instrument used must be appended i.e. Prevalence Critical Appraisal Checklist

- **Appendix III: Data extraction form**

- **Appendix IV: Table of included studies**
  
  A table of included studies is crucial to allow a snapshot of the studies included in the review.
Appendix V: List of excluded studies

At a minimum, a list of studies excluded at the critical appraisal stage must be appended and reasons for exclusion should be provided for each study (these reasons should relate to the methodological quality of the study, not study selection). Studies excluded following examination of the full-text may also be listed along with their reason for exclusion at that stage (i.e. a mismatch with the inclusion criteria). This may be as a separate appendix or itemized in some fashion within the one appendix.
References


