Sensitivity and specificity of electronic databases: the example of searching for evidence on child protection issues related to pregnant women

Helena Mc Elhinney1 MSc, BSc, Brian Taylor2 Phd, PGCert, PGDip, BSc. Marlene Sinclair3 Phd, RM, RN, RNT. Mary Rose Holman4 MSc, BA, PGChap.

1. PhD student, Institute of Nursing and Health Research, Maternal Fetal and Infant Research Centre, Ulster University, Belfast BT37 0QB Northern Ireland. Email: mc_elhinney@ulster.ac.uk
2. Professor of social work, Institute for Research in Social Sciences, Ulster University, Belfast BT37 0QB Northern Ireland. Email: bj.taylor@ulster.ac.uk
3. Professor of midwifery research, Institute of Nursing and Health Research, Maternal Fetal and Infant Research Centre, Ulster University, Belfast BT37 0QB Northern Ireland. Email: m.sinclair1@ulster.ac.uk
4. Subject assistant librarian, Life and Health Sciences, Ulster University, Belfast BT37 0QB Northern Ireland. Email: mr.holman@ulster.ac.uk

The authors would like to acknowledge the PhD scholarship received by Helena Mc Elhinney from the Department of Employment and Learning in Northern Ireland.

Abstract

Background. There are increasing demands on health and social care (HSC) professionals to make decisions based on best evidence to inform their practice. To do this, they must be skilled in searching the literature. A robust approach to literature reviewing that results in optimal outcomes is highly desirable in a climate where time and resources are limited.

Aim. This paper explores the processes of undertaking a structured literature search and measuring the effectiveness of five commonly used health and social care databases.

Method. A review question was posed using the qualitative version of PICO (Population, Interest, Context and Outcome): ‘How do HSC professionals (P) make decisions (I) in relation to pregnant women (C) where there is a safeguarding concern (O) regarding an unborn child?’ Databases selected for review were: ASSIA, CINAHL Plus, Ovid MEDLINE, PsycINFO and Social Care Online. Searches were undertaken from October 2014 to April 2015. A rapid update was undertaken in March 2016 prior to publication. Papers were screened for their suitability for inclusion using a screening tool developed by the research team. Papers were required to report empirical research; to have been published in peer-reviewed journals, as an indicator of a measure of quality; and to be available in the English language. Full-text papers were chosen if the data were gathered from or about decision-making regarding safeguarding in pregnancy by midwives, nurses, social workers and professional managers. The quality of the chosen databases was determined by sensitivity (capacity to retrieve a satisfactory number of papers), precision (to prevent the retrieval of too many irrelevant papers) and Numbers Needed to Read (NNR) – number of papers needed to read to find one paper to include.

Results. A total of 866 papers were identified, titles and abstracts were reviewed by the researcher and full-text papers were further reviewed by the research team, both using a screening tool. These results were discussed and nine papers were identified for review. Sensitivity was greatest on CINAHL Plus and Ovid MEDLINE. Precision scores were generally low; CINAHL Plus scored the highest at 4%. CINAHL Plus was found to be most effective with an NNR score of 26%, followed by PsycINFO with an NNR score of 36% and Ovid MEDLINE was the lowest precision with an NNR score of 45%.

Implications. The challenges of robust searching for literature indicate that if evidence-based practice is to become a reality, regular training for midwives, nurses, social workers and other healthcare professionals in database searching is essential

Key words: Search methodology, systematic literature searching, sensitivity and precision, decision-making, child safeguarding, pregnancy, evidence-based midwifery

Introduction

This paper discusses the methodology for searching the literature to retrieve papers using an example of professional decision-making around child safeguarding in pregnancy. It specifically focuses on sensitivity and precision measures of database quality, appropriateness and effectiveness of the search strategy and its ability to be replicated by future researchers in this area of study (Taylor et al, 2007). With the increasing demand on midwives and social workers to make decisions based more explicitly on best evidence, effective and efficient strategies need to be available to facilitate access to online resources for research (Beall, 2007).

Background

Service provision is grounded in a solid foundation of evidence-based practice. The term ‘evidence-based practice’ indicates the quality, robustness or validity of evidence and is applied to current issues within the health service (Hoagwood et al, 2001). The importance of combining clinical and research knowledge is imperative in professional decision-making in order to deliver effective services or indeed improve current service provision (McCullough et al, 2014) as many research questions emerge from clinical practice issues (Politi and Tatano Beck, 2014). There is growing demand on those in the health and social care professions to further develop their skills in collating, synthesising and critiquing information for the advancement of their profession and for use within clinical practice (Rees, 2011). There is increasing pressure from government, taxpayers and managers of health and social care services for evidence of money well spent and efficient delivery of services (Taylor and Campbell, 2011). Both the code (NMC, 2015)
for midwives and nurses and the code of ethics for social work (BASW, 2012) for social workers make it clear that professionals are responsible for maintaining and improving their knowledge, skills and practice and must act with the best evidence possible at that time. Therefore, there is an assumption that all professionals should learn core skills, such as literature searching of databases, and regularly update their knowledge in this area.

In the process of undertaking scientifically robust research, careful consideration must be given to analysing the evidence that exists within the literature. A systematic approach to literature searching has the potential to provide robust data and this evidence will subsequently inform clinical practice (Taylor et al, 2007). This approach must be undertaken in an unambiguous, transparent and replicable manner, beginning with a comprehensive literature search strategy (Arthur et al, 2012). Careful scrutiny of the literature will aid in the identification of the body of knowledge that currently exists on a topic and also highlight the gaps where further investigation is needed (Ford and Pearce, 2010).

However, the enormity of this task should not be underestimated. The advancement of technology has seen a move away from the use of traditional print journals towards journals based on electronic bibliographic databases that are accessible online (Best et al, 2014). Ironically, accessibility of journals may be improved through their availability online but the process of identifying and wading through material can be time consuming, and a daunting task (Rowley and Johnson, 2013). This requires a skilled researcher, librarian or search coordinator with a carefully designed search strategy and the ability to identify the relevant papers from those that are irrelevant (Creaser et al, 2006).

**Method**

**Data sources**

In consultation with a subject librarian at the university, five electronic databases (ASSIA, CINAHL Plus, Ovid MEDLINE, PsycINFO and Social Care Online) were accessed for the purpose of this study. All databases provide abstracts of journal papers to professions including midwifery/nursing and social work, academia and managers and were deemed appropriate for use with this study topic. CINAHL Plus, Ovid MEDLINE and PsycINFO use individualised controlled vocabulary thesaurus for indexing papers. CINAHL Plus (subject headings which are adapted from Medical Subject Headings (MeSH), Ovid MEDLINE (MeSH) and PsycINFO (American Psychological Association (APA) thesaurus of Psychological Index Terms).

ASSIA and Social Care Online use phrase searching and are aimed at those in social science. Nevertheless, using a wide variety of databases ensures a comprehensive search across both professions (Finfgeld-Connett and Johnson, 2013). To compete with health and social care (HSC) professionals’ time demands, it is crucial that a search undertaken in these databases produces relevant papers in a replicable approach. Good sensitivity (retrieving a high number of relevant papers from a database) and good precision (low number of irrelevant papers retrieved) scores depend upon the researcher’s ability to design an effective search strategy, the effectiveness of the indexing against the thesaurus, the number of journals abstracted and how well the database supports the searchers in their searching (Taylor et al, 2007). High sensitivity and precision scores are indicative of an effective search strategy which emphasises the importance of securing this from the outset (Lee et al, 2012).

The majority of journal papers, PhD theses and research reports rest on a foundation of a solid review undertaken of previous literature to ascertain gaps in knowledge that also informs evidence-based practice within the HSC setting (Taylor et al, 2007). The Cochrane and Campbell Collaboration reviews focus on questions of effectiveness and mainly on research designs that are experimental in nature. However, different types of research questions, similar to this study on decision-making, require different study designs. The methodology and systems for identifying studies, appraising quality and synthesis are not as well developed as they are for questions of effectiveness. This paper hopes to complement existing knowledge by adding an additional layer of scrutiny to the process, which includes sensitivity and precision measurements of databases.

**Search question**

Prior to choosing appropriate databases, it is important to have a clear review question using the PICO framework. This framework is used to structure clinical questions for systematic review and to increase the likelihood of retrieving papers which are relevant to the question (Schartd et al, 2007). There are two versions of PICO, depending upon whether the search is for qualitative or quantitative reviews. Searching for quantitative studies uses the Patient, Intervention, Comparison and Outcome (PICO) version. However, the qualitative version, Population, Interest, Context and Outcome (PICO) was used to frame the search question for this paper (Lewensn and Truglio-Londrigan, 2015). Consequently, the question was framed as ‘How do HSC professionals (P) make decisions (I) in relation to pregnant women (C) where there is a safeguarding concern (O) regarding an unborn child?’

**Inclusion criteria for search**

This paper focuses on the rigour of the searching process but does not include the synthesis of literature to create a systematic narrative review on the search question posed. Predetermined criteria were established for the purpose of this. Papers retrieved from this search were screened for their suitability for inclusion using a screening tool developed by the research team. Papers were required to report empirical research; to have been published in peer-reviewed journals as an indicator of a measure of quality (Best et al, 2014); and to be available in the English language. Full-text papers were chosen if the data were gathered from or about decision-making regarding safeguarding in pregnancy by midwives, nurses, social workers and professional managers. Papers focusing on the risk factors identified by professionals’ child safeguarding in pregnancy were excluded. In addition, papers...
that focused on assessment tools used in practice to assess the risk of harm to an unborn child were also excluded.

Selection of databases

In the process of systematically identifying and retrieving literature, it is recommended that more than one database should be used to ensure a comprehensive and rigorous search process (Whiting et al, 2008). Five academic and professional databases were chosen in consultation with an experienced subject librarian for use within this review. Applied Social Sciences and Abstracts (ASSIA), Cumulative Index of Nursing and Allied Health Literature (CINAHL) Plus, Ovid MEDLINE, PsycINFO and Social Care Online. These databases were selected based on their availability at the university and their suitability to capture the relevant subject fields as indexing of papers varies between databases. Papers retrieved from CINAHL Plus, PsycINFO and Ovid MEDLINE are indexed using either Medical Subject Headings (MeSH), an adaptation of MEDLINE (MeSH) or APA Thesaurus of Psychological Index Terms. However, searches within ASSIA and Social Care Online, which do not have an indexing system, relied heavily upon retrieving papers through the use of relevant phrase searching. All these databases are large interdisciplinary international databases, available in both the UK and the US. CINAHL Plus, Ovid MEDLINE and PsycINFO include papers from the professions of psychology, nursing and medicine and ASSIA and Social Care Online (provided by the Social Care Institute for Excellence) include papers from social care and social work. A piloting exercise to test and refine the search formula was undertaken which revealed sufficient numbers of papers were available.

Search formulae and filters

The use of Boolean operators within the search forms relationships between concepts or words for the purpose of establishing search parameters (Best et al, 2014). The most frequently used Boolean operators are AND (used between terms to capture papers containing both terms), OR (used to retrieve papers using either term) and NOT (to narrow or refine a search) (Houser, 2012). The search formula used for CINAHL Plus, Ovid MEDLINE and PsycINFO is shown in Figure 1. The use of truncation facilitated the retrieval of papers using singular and plural words with different endings (Reznowski, 2011).

For the purpose of this review, truncation was used on several terms including midwife* which returned words such as midwives, midwife and midwifery. Proximity operators, which vary through the databases but are generally represented using N for Near or adjn for adjacent were used to search for terms within a specific number of words from each other, for example, significant N1 harm and was used in CINAHL Plus, Ovid MEDLINE and PsycINFO to retrieve a paper which contains the term ‘significant’ within one word of ‘harm’ (Dresch et al, 2015).

The search formula was adapted for use on the databases, but notably on ASSIA and Social Care Online as searches were undertaken using phrase searching such as ‘child protection’ OR ‘child abuse’ OR ‘child neglect’ OR ‘child welfare AND nur* OR midwi* OR ‘health visitor*’ OR ‘social work*’ AND pregnant* OR perinatal* OR maternity* OR fetus OR foetus OR ‘unborn child*’ AND measure* OR assess* OR decision* OR decide OR judgment* OR criter* OR regist* OR diagnos* OR threshold OR ‘harm reduction’ OR ‘significant harm’ OR prevent. Search filters were applied to narrow the number of papers retrieved to include those papers available in the English language only and exclude books to narrow the search strategy (Schneider et al, 2013).

Sensitivity and precision

Determining the quality of a database can be measured by its capacity to retrieve a satisfactory number of published papers available on the study topic (McFadden et al, 2012). The capacity of the database to do this is generally measured in terms of sensitivity. Sensitivity is calculated by identifying the number of relevant papers retrieved by a database, which is then divided by the total number of relevant papers identified by all searches (Watson and Richardson, 1999). However, adding more terms to a particular search makes it more sensitive, resulting in the retrieval of papers that are less relevant to the study question (Taylor et al, 2007). Consequently, a second measure of quality is necessary to prevent retrieving too many irrelevant papers, as this would involve a lengthy task of elimination.

Therefore, a calculation of precision was undertaken to determine the number of relevant articles identified by a search (Taylor et al, 2007) using the formulae of relevant number of database hits divided by the total number of database hits (Taylor et al, 2003). Number Needed to Read (NNR) is a further measurement of precision of databases (Best et al, 2014). This measurement indicates the number of database hits divided by the total number of relevant papers retrieved by a database, which is then divided by the total number of relevant papers identified by all searches (Watson and Richardson, 1999). However, adding more terms to a particular search makes it more sensitive, resulting in the retrieval of papers that are less relevant to the study question (Taylor et al, 2007). Consequently, a second measure of quality is necessary to prevent retrieving too many irrelevant papers, as this would involve a lengthy task of elimination.

Therefore, a calculation of precision was undertaken to determine the number of relevant articles identified by a search (Taylor et al, 2007) using the formulae of relevant number of database hits divided by the total number of database hits (Taylor et al, 2003). Number Needed to Read (NNR) is a further measurement of precision of databases (Best et al, 2014). This measurement indicates the number of database hits divided by the total number of relevant papers retrieved by a database, which is then divided by the total number of relevant papers identified by all searches (Watson and Richardson, 1999). However, adding more terms to a particular search makes it more sensitive, resulting in the retrieval of papers that are less relevant to the study question (Taylor et al, 2007). Consequently, a second measure of quality is necessary to prevent retrieving too many irrelevant papers, as this would involve a lengthy task of elimination.

Figure 1. Search formula CINAHL Plus, Ovid MEDLINE and PsycINFO (all circles are linked with the operator AND)
of papers that must be read in order to find one paper for inclusion (Terwee et al, 2009). Calculating NNR involved dividing the total number of hits retrieved by the total number of included studies on a database (Golder et al, 2008). Therefore, a low NNR score measured on a database is an indication of high quality.

Identifying unique hits
Within the retrieved searches, unique hits were identified – that is a relevant paper sourced from one database only. This process was undertaken using a method of identifying the database(s) in which each of the final nine papers for review were located. The papers were then listed in chronological order indicating the databases from which they were retrieved.

Results
A systematic search of the databases retrieved 866 papers. Titles and abstracts were reviewed by the researcher using a screening tool. A smaller number of papers were reviewed by the research team and nine papers were identified for review.

Unique hits retrieved
Of the five databases, only two returned unique hits (Table 1). The search on Ovid MEDLINE returned the largest number of unique hits retrieved (three papers) and CINAHL Plus returned one unique paper. ASSIA, PsycINFO and Social Care Online did not return any unique papers in this search.

Sensitivity and precision
The ability of each database to retrieve relevant items was measured producing sensitivity scores. CINAHL Plus (36%) recorded the highest sensitivity, followed by Ovid MEDLINE (28%) and PsycINFO (14%). ASSIA and Social Care Online both scored 0% indicating that they were ineffective for use within this study and retrieved no relevant items. In general, precision scores between all databases were extremely low. CINAHL Plus had the highest precision score (4%) of all the databases indicating that it was the most effective database at avoiding retrieving irrelevant papers. However, PsycINFO had a marginally lower precision score (3%) and Ovid MEDLINE had a precision score of 2%. ASSIA, Social Care Online and the hand search had poor precision scoring between 0% and 1%. The NNR scores are a measure of how many papers need to be read to retrieve one relevant hit on each database. CINAHL Plus was found to be most effective with a NNR score of 26%, followed by PsycINFO with a NNR score of 36% and Ovid MEDLINE was the weakest database with an NNR score of 45%. No relevant hits were found in either ASSIA or Social Care Online.

Methodology of retrieved papers
The relevant papers retrieved comprised of qualitative studies (using semi-structured interviews, 11%), surveys (22.4%), mixed methods, including both qualitative and quantitative methods (questionnaires, interviews and focus groups, 33.3%), and quantitative (surveys, 33.3%).

Discussion

Summary of approach
The systematic retrieval of papers from databases is an important aspect of evidence-based practice (Taylor et al, 2007). Consequently, it is essential that identification of relevant databases and accurate search formulae and filters are applied from inception. The development of a sophisticated search formula was key to the retrieval of relevant papers (Best et al, 2014). A search question was established, a screening tool developed to determine inclusion/exclusion of papers, the search strategy piloted to determine the appropriateness and effectiveness of search terms and modified when necessary. The database searches were methodical and the initial search was repeated to ensure it was up to date (Ramlaul, 2010). A preliminary search of the databases provided a good indication of the relevant information available and the correct search terms to be used.

Effectiveness of the databases
The study topic of child safeguarding in pregnancy from the perspective of decisions made by social work and midwifery staff indicated that databases had to be chosen which encompassed research from both professional groups. In general, the databases for social sciences were more difficult to navigate than those aimed at professionals in healthcare fields. This may be due to the variation in terminology internationally and also due to the facilities on the databases (McFadden et al, 2012). It could be assumed that the databases predominately aimed at those in social sciences – ASSIA and Social Care Online, in this instance – would generate most of the relevant papers.

However, this study showed the contrary, as no relevant hits were identified on either database. It could be argued that social science databases, ASSIA and Social Care Online are less effective than their counterparts – Ovid MEDLINE, PsycINFO and CINAHL Plus – possibly due to inconsistency in language, as identifying papers depends upon how well they are indexed within databases. The latter three databases are aimed at the nursing (including midwifery) and allied health professionals and index their papers using MeSH headings, which are believed to produce greater specificity than phrase searching used in ASSIA and Social Care Online (McIntosh, 2011).

Table 1. Number needed to read (NNR) and unique hits

<table>
<thead>
<tr>
<th>Databases</th>
<th>Total Hits Retrieved</th>
<th>Relevant Hits Retrieved</th>
<th>NNR</th>
<th>Unique Hits</th>
<th>Sensitivity %</th>
<th>Precision %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSIA</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>0*</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CINAHL Plus</td>
<td>267</td>
<td>10</td>
<td>26</td>
<td>1*</td>
<td>36%</td>
<td>4%</td>
</tr>
<tr>
<td>Ovid MEDLINE</td>
<td>361</td>
<td>8</td>
<td>45</td>
<td>3*</td>
<td>28%</td>
<td>2%</td>
</tr>
<tr>
<td>PsycINFO</td>
<td>144</td>
<td>4</td>
<td>36</td>
<td>0*</td>
<td>14%</td>
<td>3%</td>
</tr>
<tr>
<td>Social Care Online</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0*</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Hand Search</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>866</td>
<td>28*</td>
<td>28%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Number asterisked (*) include total number of hits after duplicates removed.

© 2016 The Royal College of Midwives. Evidence Based Midwifery 14(1): 29-34
Measures of sensitivity, precision and NNR

The measure of sensitivity of a database is significantly important to ensure that the papers retrieved from the search are relevant (Haynes et al, 2005). Aiming for a high sensitivity score may reduce the chance of missing papers that are relevant (Pack, 2014). In this study, the highest sensitivity score was calculated for CINAHL Plus and Ovid MEDLINE scored marginally lower. PsycINFO scored disappointingly lower in sensitivity. CINAHL Plus, which indexes considerably fewer papers than PsycINFO, had the highest sensitivity score and included papers from the midwifery profession, but not social work, unlike its counterpart Ovid MEDLINE, which source papers from both. The final two databases – ASSIA and Social Care Online – had a sensitivity score of 0%, indicating that they were ineffective for use within this study. However, they are predominately social work and social science based databases and the previous three databases weighed heavily towards the medical and midwifery professions. Although ASSIA does include social service topics, it does not include midwifery, therefore, papers may be retrieved connected with child safeguarding, but not necessarily in pregnancy. Additionally, Social Care Online is predominately aimed at practitioners and policy-makers with the core material sourced from the UK, therefore restricting its scope. Similarly to ASSIA, Social Care Online covers child safeguarding but, again, does not include pregnancy or maternity research. However, it would be advisable to include both those databases to ensure that a comprehensive search is undertaken (Kemp and Brustman, 1997). Similar to CINAHL Plus, Ovid MEDLINE and PsycINFO, ASSIA is an international database that indexes fewer journals than the other databases but may be aimed at those in academia rather than social work professionals (McFadden et al, 2012).

An effective literature search is reflected through the precision scores of a particular database and the number of relevant hits retrieved (Schardt et al, 2007). Precision within this study was quite low, identifying many papers that were not relevant to the study question (Gough et al, 2012). The highest precision score was recorded on CINAHL Plus and the lowest on Ovid MEDLINE. These higher precision scores could be attributed to the scope of the journal coverage on this specific topic area (Best et al, 2014).

The retrieval of papers that are not found on other databases (unique hits) can give an indication of the best choice of databases to use in the search. Two of the five databases searched within this study retrieved a small number of unique hits. Ovid MEDLINE recovered the highest number of unique hits (3%) followed by CINAHL Plus (1%). ASSIA, PsycINFO and Social Care Online failed to retrieve any unique hits. The use of these latter databases within future searches around the topic of professional decision-making around child safeguarding in pregnancy may be unproductive. It could have been presumed that ASSIA and Social Care Online, both social science databases, would retrieve some unique hits considering the study topic but, in this instance, proved fruitless. As both database searches use phrase searching, developing a standard vocabulary for searching literature, consistent throughout all databases and across HSC professions may be beneficial (Curran et al, 2007).

In an almost alternative measurement to precision, NNR refers to the number of papers that must be read to find one relevant paper from a database search (Hersh, 2009). A low NNR is an indication of a good and efficient search string narrowing the search of the literature, ensuring that the task it less time-consuming for a busy professional (Pillastrini et al, 2015). However, this depends upon the correct use of language and indexing terms within those databases (Stewart et al, 2014). The relevant papers retrieved for final review included a wide range of research methods confirming the validity of the search strategy and appraisal of papers (McFadden et al, 2012).

Limitations of the review

Every good literature search and retrieval has its limitations that must be acknowledged to assist future research in this study area. Two of the databases in this study, ASSIA and Social Care Online, were less than optimal for this topic area. This was partly due to the limited number of journals they index in comparison with the more medically-based journal databases CINAHL Plus, PsycINFO and Ovid MEDLINE. On reflection, it may have been beneficial to add a further social science database to the search to test its effectiveness, but the results might still have been limited in comparison with those found on more sophisticated databases with advanced indexing facilities and superior user interface.

Implications for practice

Literature searching is a common core skill that HSC professionals are expected to be able to conduct with confidence and skill. This paper provides a new insight for those in management, policy-making, midwifery, nursing and social work with further methods for undertaking a robust literature search, which can inform practice and policy to improve services. Employers and professional bodies need to ensure that employees have access to regulated training in literature searching skills for staff undertaking research and clinical projects designed to improve maternal and child care.

Conclusion

There are increasing demands on the interdisciplinary team to remain updated with research developments in their fields of practice. The task of locating relevant research now relies substantially on the quality of bibliographic databases. This study highlighted low precision in searching five databases on this topic, despite a detailed search formula and the expertise of a specialist librarian. The ‘information age’ requires investment in systems that are efficient as well as effective if professionals are to make the most use of available knowledge and increase their chances of retrieving relevant literature in an efficient manner. Support from leaders in the HSC organisations and professions to develop the quality of databases is a priority. The development of expertise in identifying relevant research and education on the most robust methods of database searching is a priority if the high ideals of evidence-based practice are to become a reality.
of searching for evidence on child protection issues related to pregnant women. *Evidence Based Midwifery* 14(1): 29-34

---

**References**


