Bibliographic databases for social care searching

social care
institute for excellence
scie

The Social Care Institute for Excellence (SCIE) was established by Government in 2001 to improve social care services for adults and children in the United Kingdom. We achieve this by identifying good practice and helping to embed it in everyday social care provision.

SCIE works to disseminate knowledge-based good practice guidance; involve service users, carers, practitioners, providers and policy makers in advancing and promoting good practice in social care; enhance the skills and professionalism of social care workers through our tailored, targeted and user-friendly resources.

Bibliographic databases for social care searching

Janet Clapton

First published in Great Britain in July 2010
by the Social Care Institute for Excellence

© SCIE 2010

Written by Janet Clapton

This PDF document was created in July 2010, and will not reflect any changes made since that date.

Please visit our website for the latest version at **www.scie.org.uk**

Contents

1 Summary

The main aim of this work is to help information professionals and researchers make an informed choice of databases when planning searches on social care topics. In light of lack of previous evidence, this study adds significantly to what is known about the relevance and coverage of social care databases, particularly on children's social care topics.

This study uses analysis of database search outputs to gain information about database coverage. Three searches were carried out on the topic of looked after children (="children in care"), specifically the subtopics of the children's education, their emotional health, and accommodation after leaving care. The searches were originally carried out in 2008 to support scoping reviews for the Centre for Excellence and Outcomes in Children's Services (C4EO).

713 references from 16 databases were analysed by database origin. This is a relatively large number of databases to use for systematic searching, however social care literature is widely dispersed across many topic fields and sources and not concentrated in a few locations such as MEDLINE. A literature review provided context for the study.

For each search on looked after children, relevant references were found in 14 out of 16 databases searched. The top databases for **relevance** varied with subtopic searches but included:

> **Social Care Online, PsycInfo, ChildData, Social Services Abstracts, Applied Social Science Information and Abstracts (ASSIA), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Social Work Abstracts and MEDLINE**.

Seven or eight databases per search contained unique references found in no other database. The top databases for **unique references** varied with subtopic searches but included:

> **Social Care Online, ChildData, PsycInfo, ASSIA, British Education Index and Social Services Abstracts**.

The references which occurred in more than one database were duplicated up to 8 times across 14 different databases for each search.

For social care searching, databases drawn from the broad fields of health, social science and social care are required, and UK-focussed databases should be included. A minimum of seven or eight databases needed to be searched to capture the relevant references for the topics examined in this study. The variation between coverage even for related topics suggests that the exact selection is highly topic dependent. However, the study adds to the sparse evidence base on social care database coverage and overlap. Using this method to analyse search output on other topics would extend the evidence base.

2 Introduction

Systematic searching is a time consuming and costly activity, both in terms of staff time and subscriptions to bibliographic databases (LISU, 2009). At the same time, the need to find key evidence means that searching should be exhaustive. Therefore information specialists and researchers should make informed decisions on which databases to use. In the social care field, the evidence on which to base these decisions is limited. The information that exists on database coverage and overlap is not plentiful in any field, but that which exists shows that the nature of database coverage and overlap is strongly topic dependent (Hood and Wilson, 2001).

Searching for literature on social care topics is challenging even for a search specialist. There are many potentially relevant sources across fields such as social science, health, criminal justice, and housing, and no single source contains all the relevant material. The terminology is varied and there is no single language tool such as the Medical Subject Heading thesaurus. The literature itself is very varied, with reports and unpublished literature making up a significant proportion. Finally, while the UK welfare context is distinct, many of the commonly used databases have predominantly US coverage. Taking these factors into consideration, it is particularly important that we have knowledge of database coverage and overlap.

The objective of the study was to gain evidence on coverage and overlap between bibliographic databases relevant for searching on children's social care topics. Coverage here means listing bibliographic details (=reference for a journal article, report, book or other written piece of work) with or without abstract). Overlap means that the bibliographic details for a piece of work appear in more than one database being compared.

A variety of approaches to describing database coverage have been used in the literature, including assessing journal title coverage and searching for key references in different databases. The general method used in the present study was to label a systematic search output according to source database at time of searching, and then examine the distribution pattern of sources of relevant references. Detailed searching was performed in 16 health, social science and social care databases on the topic of looked after children, to support scoping reviews in 2008 for C4EO (www.c4eo.org.uk).

The activity consisted of three overlapping searches on the subtopics of looked after children's education, emotional health, and accommodation after leaving care. Data from all three were used because this increased the pool of references for analysis (relative to using a single search output) and provided the opportunity to investigate differences in database coverage between three similar search topics.

The titles of the scoping reviews were:

> Improving educational outcomes for looked after children and young people (Brodie et al, 2009)

> Improving the emotional and behavioural health of looked-after children and young people (LACYP) (Fish et al, 2009)

> Increasing the number of care leavers in 'settled, safe accommodation' (Bostock et al, 2009)

The raw search output was screened for relevance to the search topic. Full details of the inclusion criteria for relevance are given in the scoping report appendices (see links on next page).

In the current study, the 713 records found to be relevant to at least one of the three subtopics were analysed according to distribution of source databases. EndNote® reference management software (Thomson Reuters, 2009) and Excel® spreadsheets (Microsoft Corporation, 2007) were used to manage the data, and simple statistical analysis was used to quantify database coverage and overlap. The breadth and number of search sources and total number of references analysed represent a relatively large dataset for this kind of comparison.

3 Method

3.1 Generating the search output data

The subject of looked after children was chosen for the current study because the literature review showed that there was no published research on bibliographic database coverage for children's social care topics.

The dataset of references for analysis came from the output of systematic searching carried out between September and December 2008 for three scoping reviews for C4EO (see www.c4eo.org.uk/themes/vulnerablechildren).

The scoping reviews were:

Improving educational outcomes for looked after children and young people (Brodie et al, 2009)
(www.c4eo.org.uk/themes/vulnerablechildren/scopingreview01.aspx)

Improving the emotional and behavioural health of looked-after children and young people (LACYP) (Fish et al, 2009)
(www.c4eo.org.uk/themes/vulnerablechildren/scopingreview02.aspx)

Increasing the number of care leavers in 'settled, safe accommodation' (Bostock et al, 2009)
(www.c4eo.org.uk/themes/vulnerablechildren/scopingreview03.aspx)

The detailed topic parameters, search strategy and relevance screening criteria are described in the scoping reviews' appendices 'Search parameters' and 'Scoping study process' (accessed via the links above).

The general search strategy combined index terms and freetext terms on the concepts of **population** (looked after children), residential childcare **settings**, **interventions** such as virtual school heads, **states** such as homelessness and **outcomes** such as achievement.

The relevance screening criteria included general concepts such as definition of looked after children, inclusion of material in English from developed welfare states, publication from 2000 onwards, and inclusion of empirical research from journal articles and reports only, as well as relevance to the specific topic.

The initial search output consisted of approximately 20,000 bibliographic references, most of which had abstracts.

After screening the search output titles and abstracts for relevance to at least one of the above scoping reviews, 713 references remained. These included duplication between different databases (but not within databases) and overlap between the three topics.

Discounting duplicates and overlap, there were 204 individual references.
Box 1 shows how individual and unique references are defined in this study.

Box 1 Defining individual and unique references

This worked example shows three search outputs:

Search output on Topic 1

Database A	Database B	Database C
Reference d	Reference d	
Reference e		Reference e
Reference f	Reference f	
Reference g		Reference g

Search output on Topic 2

Database A	Database B	Database C
Reference d		Reference d
Reference h		
Reference i	Reference i	
Reference j	Reference j	

Search output on Topic 3

Database A	Database B	Database C
Reference d	Reference d	Reference d
Reference e		
Reference i		Reference i
Reference k	Reference k	

After aggregating the data from the three searches and
removing duplicates within databases, there are **17 references** to analyse:

Searching
>   Database A gives us References d, e, f, g, h, i, j, k
>   Database B gives us References d, f, i, j, k
>   Database C gives us References d, e, g, i

There are **eight individual references:** Reference d, e, f, g, h, i, j, k
There is **one unique reference** (i.e. it is only found in one database):
>   Reference h

The databases searched (with the specific platform used for access) are listed below. The selection broadly followed the list of databases recommended for SCIE systematic reviews (Coren and Fisher, 2006) plus additional topic-specific sources (e.g. British Education Index).

Applied Social Science Information and Abstracts (ASSIA), CSA Illumina

Australian Family and Society Abstracts, Informit

British Education Index, Dialog

Current Educational Research in the UK (CERUKplus), NFER

ChildData, NCB

Cumulative Index to Nursing and Allied Health Literature (CINAHL), EBSCO Host

Cochrane Library, Wiley Interscience. Includes the Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register, Health Technology Assessment Database, NHS Economic Evaluation Database.

EMBASE, Ovid SP

Health Management Information Consortium (HMIC), Ovid SP

International Bibliography of the Social Sciences (IBSS), EBSCO Host

MEDLINE, Ovid SP

PsycInfo, Ovid SP

Social Care Online, SCIE

Social Services Abstracts, CSA Illumina

Social Work Abstracts, Ovid SP

Zetoc, MIMAS

See Appendix 1 for description of the databases by sector or topics covered, geographical coverage, size and whether accessible for free or by subscription.

Additional search techniques were used for the C4EO work in 2008, such as drawing references from the bibliographies of highly relevant documents, and personal recommendations from experts in the field of looked after children, but the non-database contribution was discounted for the purpose of the current study.

During the search process, search output was imported to EndNote® reference management software. Output from each database was labelled with the name of that source database, using the Name of Database field. The total search output contained duplicates between databases, with each reference labelled according to its source.

The references were screened by relevance using the abstract (and later checked by reading the full text) according to the inclusion / exclusion parameters given in the

above reports. While the focus was on UK research, international studies were included if relevant and in the English language.

3.2 Analysing the data

Simple statistical analysis was carried out, using manual or semi-automated techniques, without using specialist statistics software.
The two software packages used to carry out the database comparison were:

EndNote® version X3 (Thomson Reuters, 2009): a bibliographic management programme, to manage and sort references by source and title, and list references by source database and by title.

Excel® (Microsoft Corporation, 2007): spreadsheets used to tabulate and calculate simple statistics for each database.

The following measures were calculated for each of the three topic searches:

- **Total number of relevant references** (including duplicates between different databases)

- Number of **individual references** (excluding all duplicates)

- **Distribution of database sources for relevant references**: (using subject bibliography function in EndNote® for the Name of Database)
  e.g.
  (Education search) relevant references in ChildData   x100
  total number of relevant references (from the Education
  search) across all databases

- **Distribution of database sources for unique references**: using manual counting.
  e.g.
  (Accommodation search) unique   x 100
  references in Social Care Online
  total of individual references (from the Accommodation
  search) across all databases

- **Database overlap** for three most frequent databases for each of the three searches: carried out by tabulating reference titles against occurrence in different databases as follows:

Database overlap

	Database A	Database B	Database C
Reference d	x	x	
Reference e	x		x
Reference f		x	

X in a database column indicates that the particular reference occurred in that database.

The next stage was to filter data in Excel according to the databases being compared and count overlapping records.
Percentage overlap was calculated as:

$$\frac{(\text{number of records in overlap})}{(A + B - \text{overlap})} \times 100$$

Where A = total number of records in Database A for this search output
       B = total number of records in Database B for this search output
See worked examples in the Results section.

- **Multi-database coverage** of relevant references
(using subject bibliography output from EndNote® for the Name of Database).
This measure describes the frequency of repetition of a reference in different databases e.g.

$$\frac{\text{(Emotional health search) Number of references occurring in any three different databases}}{\text{total relevant references for the Emotional health search}} \times 100$$

3.3 Processing duplicates

When carrying out database source analysis, it is important to process duplicates correctly. The objective is to work with a set of references in which you can identify the contribution of each database.

Removal of duplicates within databases was mostly automated during the search process by the database platforms; however, any further duplicates within a single databases' search output were deleted manually, taking care to retain all duplicates from different databases. Thorough data tidying is very important, this means that all data is in the correct fields, is identical between copies of a record (especially with respect to punctuation) and that author order is identical between records. Failure to carry out this step exhaustively means that later automation of analysis using the EndNote® bibliography function will give inaccurate results (see Box 2).

Box 2 Processing duplicates

Before any duplicate removal

Database A	Database B	Database C
Reference d	Reference d	Reference d
Reference d	Reference e	Reference f
Reference d	Reference f	Reference f
3 refs	3 refs	3 refs

Total number of references = 9, but this number contains duplicates within databases.

After duplicate removal within databases

Database A	Database B	Database C
Reference d	Reference d	Reference d
~~Reference d~~	Reference e	Reference f
~~Reference d~~	Reference f	~~Reference f~~
1 ref	3 refs	2 refs

Total number of references = 6: duplicates have been removed within databases but not between them.

     Database A contributes 1/6 coverage of relevant references
     Database B contributes 3/6 coverage
     Database C contributes 2/6 coverage

Indiscriminate duplicate removal from between databases
e.g. by failure to label all references with their source at time of searching, or by misuse of automatic duplicate removal within EndNote.

Database A	Database B	Database C
~~Reference d~~	Reference d	~~Reference d~~
~~Reference d~~	Reference e	~~Reference f~~
~~Reference d~~	~~Reference f~~	Reference f
0 refs	2 refs	1 refs

Total number of references = 3

     Database A's contribution has been accidentally deleted altogether
     Database B contributes 2/3 coverage
     Database C contributes 1/3 coverage

3.4 Literature review search strategy

The purpose of the literature review was to summarise previous approaches to database source analysis and give an overview of findings in previous studies. The range of sources searched included bibliographic databases, journal collections / full text hosts, bibliographies of relevant articles and internet search engine browsing.

The databases searched included ASSIA, Library and Information Science Abstracts, MEDLINE, Social Care Online / Social Policy and Practice Database, and Social Science Citation Index, using terms such as:

systematic search, database overlap, overlap between databases, source analysis, database relevance, relevant database, database comparison.

The names of the most commonly used social care databases were also searched in a representative range of sources. Similar terms were used to search journal and full text hosts, such as Emerald journals and EBSCOhost. Interesting references were followed up from the bibliographies of relevant full text articles.

4 Findings

4.1 Volume of search output data

The objective of this stage was to use systematic searching to generate three sets of search output on closely related topics for analysis by source database. These sets were used as samples to test database coverage and overlap. The references were all labelled according to their source database then screened for relevance to the search topics of looked after children's education, emotional health and accommodation.

The search output datasets were made up of:

Education search output: 178 references (including duplicates between databases) consisting of 63 individual references.

Emotional health search output: 338 references (including duplicates between databases) consisting of 112 individual references.

Accommodation search output: 197 references (including duplicates between databases) consisting of 73 individual references.

The combined dataset across the three searches was 713 references (including duplicates between databases and between searches), as shown in Fig. 1.

**Figure 1. Search output volumes per search**
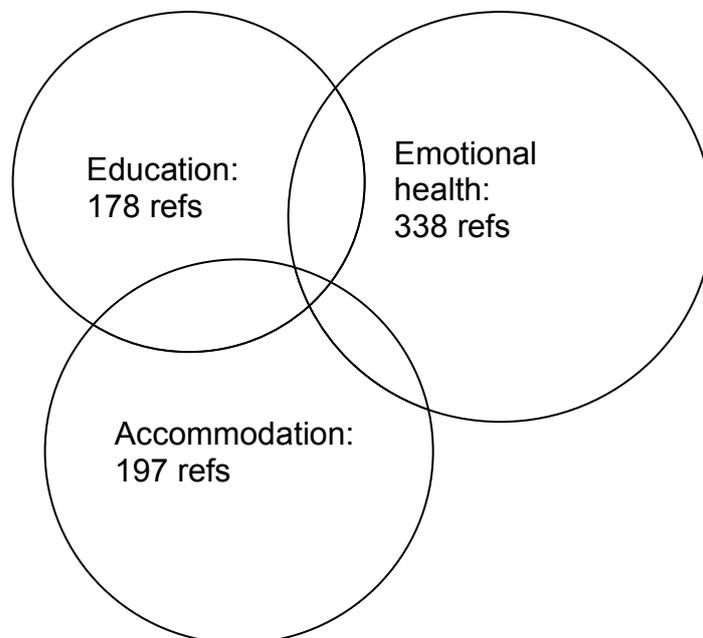
To describe the volume of overlap between the searches:
- 248 individual references came from the three sets (63+112+73).
- 204 individual references remained after removal of duplicates **between** the searches.
- 44 references were found in more than one subset (248 individual references in aggregated search output minus 204 individual references in combined and deduped search output)

4.2 Database relevance

Based on the search output samples, databases with the highest proportion of relevant records varied between the subtopics, but overall the most relevant databases (contributing 5 per cent or over of relevant references or over, in order of decreasing relevance) were:

Social Care Online
PsycInfo
ChildData,
Social Services Abstracts
ASSIA
CINAHL
Social Work Abstracts
MEDLINE
Zetoc

Decreasing relevance
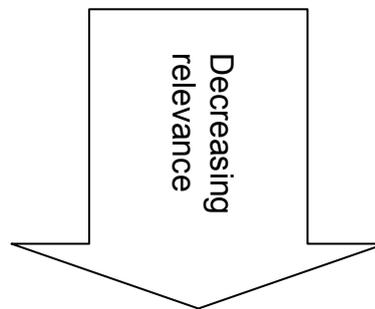
The maximum relevant references percentage contribution was 25 per cent for Social Care Online in the Accommodation search. Variation in percentage points per database between searches varied by only up to several per cent.

The less relevant databases overall (in decreasing order) were

EMBASE
HMIC
IBSS
British Education Index
Australian Family and Society Abstracts
Cochrane Library
CERUK

Decreasing relevance
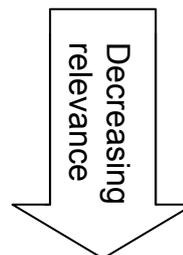
See Table 1 and Figure 2 for data representation.

Table 1. Databases ranked by relevance for each subtopic search output

Relevance is shown below as percentages:

(Search subtopic specific) relevant references in a specific database   x 100
total number of relevant references (for that search) across all databases

Results are shown for databases which contributed at least 5 per cent of relevant references to each subtopic search.
Totals do not add up to 100 per cent as database contributions less than 5 per cent have been excluded.

	Education (n=178)	Emotional health (n=338)	Accommodation (n=197)
Social Care Online	18%	16%	25%
PsycInfo	15%	15%	12%
ChildData	13%	12%	16%
Social Services Abstracts	13%	13%	11%
ASSIA	10%	10%	6%
CINAHL	7%	10%	8%
Social Work Abstracts	5%	5%	9%
MEDLINE		5%	6%
Zetoc	6%		

**n= relevant references per subtopic search output**

Figure 2. Databases ranked by relevance for each subtopic search output
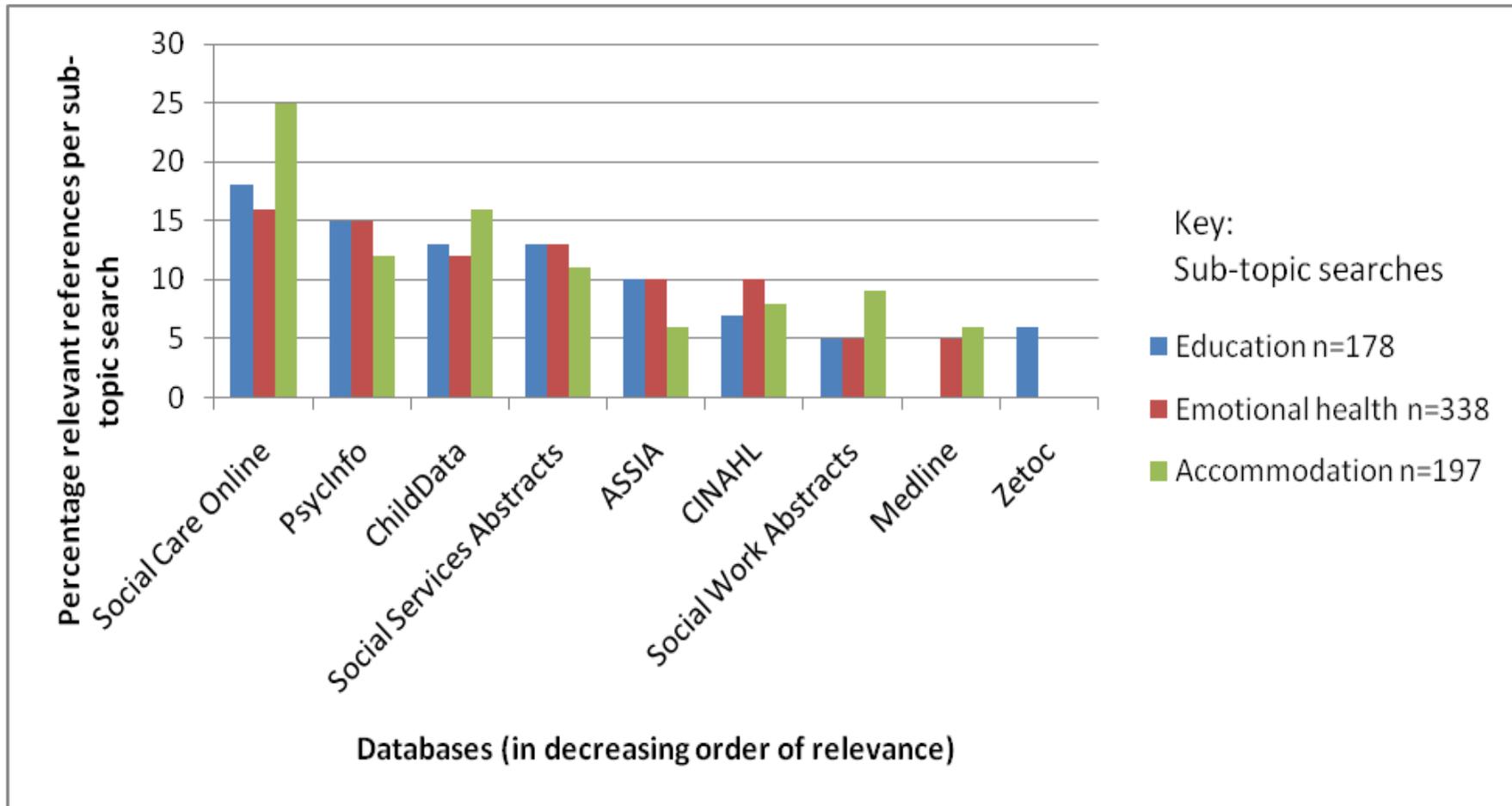
## 4.3 Database unique content (not found in another database)

Databases with the highest proportion of unique records in the search output varied between the subtopics, but overall the greatest percentage of unique coverage (in decreasing order, 5 per cent or over for any one search) was found in:

Social Care Online,
ChildData,
PsycInfo,
ASSIA,
British Education Index,
Social Services Abstracts,
MEDLINE,

Decreasing uniqueness
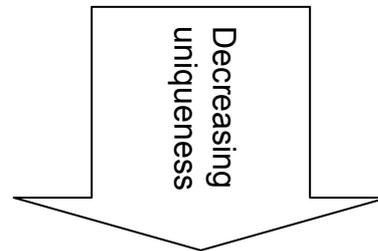
The databases with least proportion of unique records in search output (in decreasing order) were:

Zetoc,
CINAHL,
Australian Family and Society Abstracts
CERUK
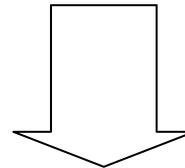
The maximum unique contributions were around 24 per cent per search apart from the contribution of Social Care Online to the Accommodation search at 52 per cent.

See Table 2 and Figure 3.

Table 2. Databases ranked by unique content for each subtopic search output

Unique content is shown below as percentages:
<u>(Search subtopic specific) unique references in a specific database</u>   x 100
total of individual references (for that search) across all databases

	Education (n=26)	Emotional health (n=32)	Accommodation (n =31)
Social Care Online	23%	19%	52%
ChildData	19%	22%	23%
PsycInfo	27%	25%	6%
ASSIA	8%	6%	3%
British Education Index	12%	0%	3%
Social Services Abstracts	4%	9%	0%
MEDLINE	0%	0%	10%
Zetoc	4%	3%	3%
CINAHL	0%	9%	0%
Australian F & S Abstracts	0%	6%	0%
CERUK	4%	0%	0%

n= unique references per subtopic

NB Emotional health percentages add up to 99 per cent due to rounding up or down

Figure 3. Databases ranked by unique content for each subtopic search output
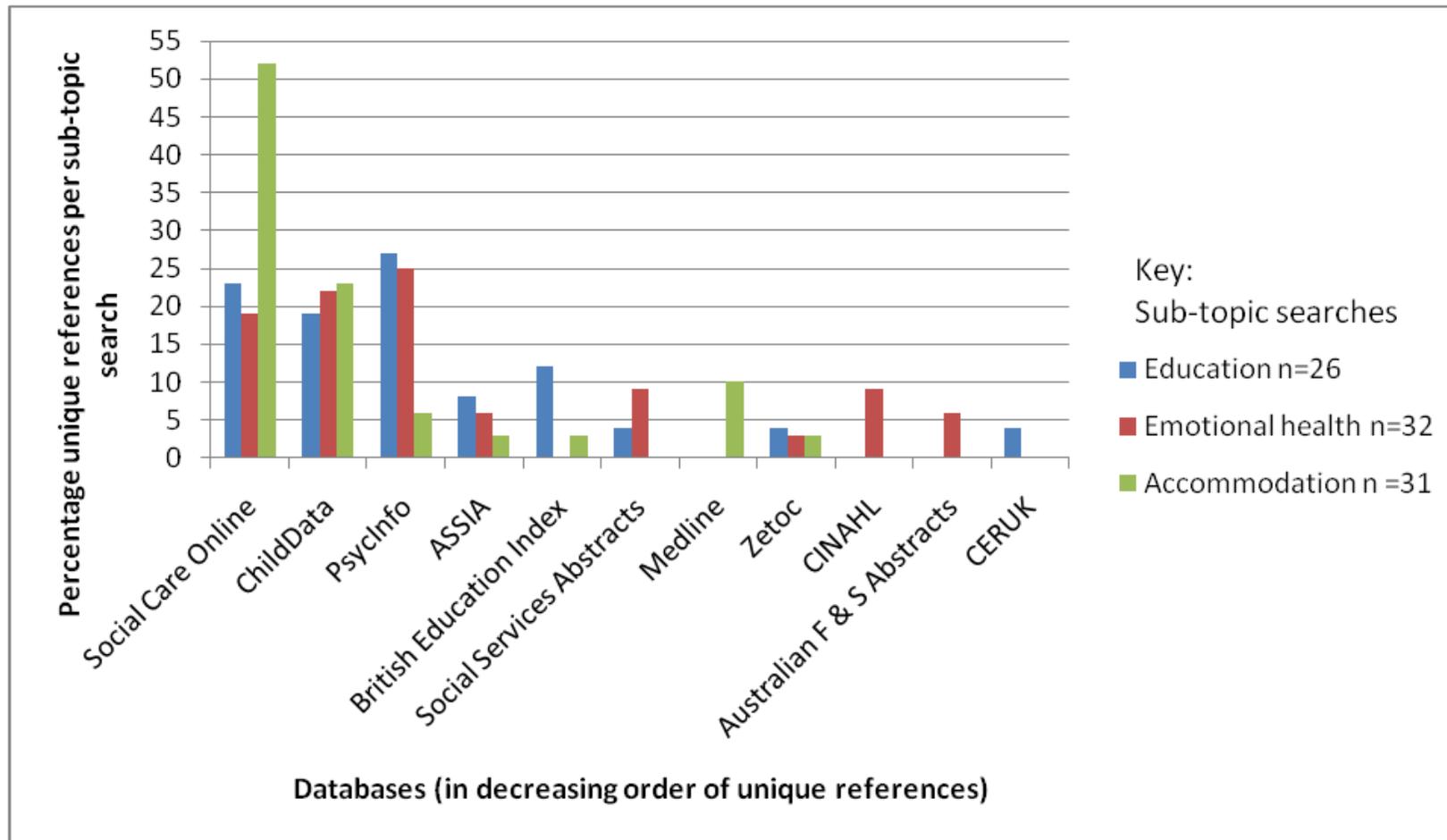
4.4 Multi-database coverage

How frequently references occurred across multiple databases varied by only a few percentage points between the subtopics.

Occurrence in a single database only was most frequent (around 38 per cent) with occurrence in two to six databases having similar frequencies at between 16 and 8 per cent.

For the Education and Emotional health searches, the maximum number of databases a reference appeared in was seven, while for the Accommodation search it was eight.

4.5 Overlap in coverage

To recap, overlap means that the bibliographic details for a piece of work appear in search outputs of more than one database. The whole database is not being assessed, only a small sample as represented by the specific search output.

In contrast with Section 4.4, which describes frequencies, here the percentage overlap was calculated for search output from the three most relevant databases for each search. The most relevant databases were chosen because they provided the largest pool of references for analysis. It is difficult to extend overlap analysis to more than three databases without more complex calculations and automation with specialist software.

The implication of finding identical references in search output from different databases is that the degree of database overlap is high. However, we are analysing what could be found in particular databases with a particular search strategy. References could be present but not found due to limitations of the search strategy, the available search utility, the indexing or incorrect reference details.
The three most **relevant** databases varied slightly between subtopics but included:
Social Care Online, PsycInfo, ChildData and Social Services Abstracts.

For the respective search outputs, the three most relevant databases were:

Education: ChildData, PsycInfo and Social Care Online.
Emotional: Social Care Online, PsycInfo and Social Services Abstracts.
Accommodation: Social Care Online, ChildData and PsycInfo.

The raw figures for output overlap are summarised in tables 3–5.

Table 3. Education search: Numbers of relevant records in database overlap

Education subtopic (N=63 individual references)	ChildData	PsycInfo	SCO	Overlap between all 3 databases
ChildData	24	9	14	
PsycInfo	9	26	10	
SCO	14	10	32	
Overlap between all 3 databases				6

Table 4. Emotional health search: Numbers of relevant records in database overlap

Emotional health subtopic (N=112 individual references)	Social Care Online	PsycInfo	Social Services Abstracts	Overlap between all 3 databases
Social Care Online	54	26	29	
PsycInfo	26	52	26	
Social Services Abstracts	29	26	45	
Overlap between all 3 databases				21

Table 5. Accommodation search: Numbers of relevant records in database overlap

Accomm-odation subtopic (N=83 individual references)	Social Care Online	ChildData	PsycInfo	Overlap between all 3 databases
Social Care Online	50	17	15	
ChildData	17	31	10	
PsycInfo	15	10	23	
Overlap between all 3 databases				7

Raw number overlap does not take into account the total number in both search outputs from two comparator databases.
Therefore a percentage formula is used.
Percentage overlap of relevant references between two databases is expressed as:

$$\frac{\text{number of records in overlap}}{(A + B - \text{overlap})} \times 100$$

where:
A= number of relevant records in Database A
B= number of relevant records in Database B
and overlap = relevant records in Database A and in database B

This formula can be expanded to three databases (A, B, C) as

$$\frac{\text{number of records in overlap}}{(A + B + C - \text{overlap})} \times 100$$

where overlap = relevant records in Databases A, B and C

To illustrate the calculation of percentage overlaps, two examples are worked for one subtopic (Education).

Using data in Table 3, where:

A=Number of relevant records in the Education search output from Social Care Online
B= Number of relevant records in the Education search output from PsycInfo
C= Number of relevant records in the Education search output from ChildData

Overlap between Social Care Online and PsycInfo:

$$\frac{10}{32+26-10} \times 100 = 21\%$$

Overlap between Social Care Online, PsycInfo and ChildData

$$\frac{6}{32+26+24-6} \times 100 = 8\%$$

The calculated percentage overlaps for all three search outputs are summarised in Table 6:

Table 6: Percentage overlaps between the three most relevant databases

Search topic	2 databases with biggest overlap		2 databases with second largest overlap		2 databases with least overlap		3 database overlap	
Education	ChildData-SCO	33%	ChildData-PsycInfo	22%	PsycInfo - SCO:	21%	ChildData-PsycInfo-SCO	8%
Emotional Health	SCO-Social Services Abstracts	41%	PsycInfo-Social Services Abstracts	33%	PsycInfo-Social Care Online	37%	PsycInfo-SCO-Social Services Abstracts	16%
Accommodation	ChildData-SCO	27%	PsycInfo-SCO:	26%	ChildData-PsycInfo:	23%	ChildData-PsycInfo-SCO:	7%

The percentage overlaps are shown in the Venn diagrams of Figures 4–6.

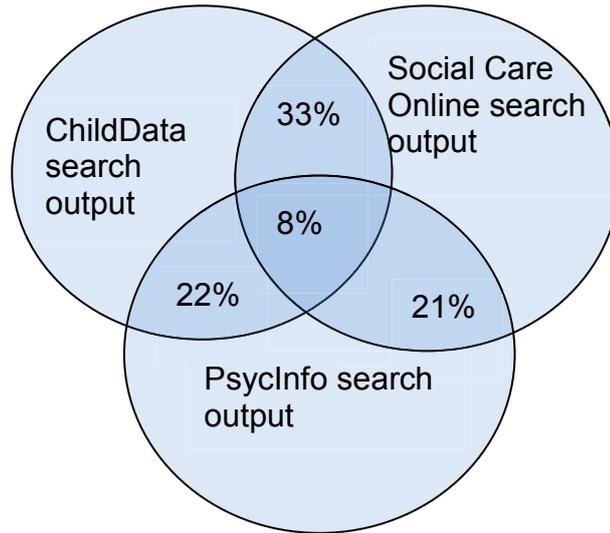**Figure 4. Database output percentage overlaps on Education of looked after children**

Figure 5. Database output percentage overlaps on Emotional health of looked after children
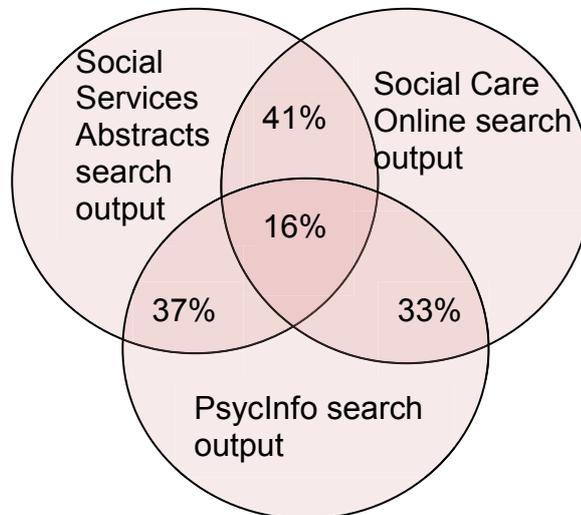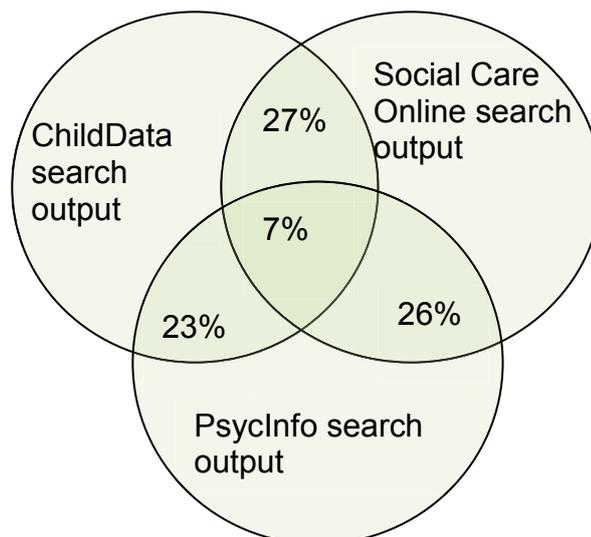

Figure 6. Database output percentage overlaps on Accommodation of care leavers

Overlap between the outputs of the three most relevant databases varied slightly for each subtopic. This was particularly apparent when the same pairs were compared in different searches:

      ChildData-Social Care Online (27% cf. 33%)
      ChildData-PsycInfo (22% cf. 23%)
      PsycInfo-Social Care Online (26% cf. 33%)

Overlap between outputs of two databases varied between 21 and 41%
Overlap between outputs of three databases varied between 7 and 16%.
The volume of search overlap was similar for the Education and Accommodation search outputs and larger for the Emotional health search output.

Exploration of the nature of materials inside or outside overlaps is outside the scope of this study, but further investigation is needed, e.g. were the journals inside or outside overlap highly cited? What was the overlap of grey literature? A broad research aim could be to work towards a wider description of databases.

4.6 Findings from previous studies

The search for similar work revealed very few studies, a striking point being lack of comparative work on this range of databases and sparse studies on social care topics. If you cast the net wider to include studies on broader social welfare, psychiatry and health topics, there is more material for consideration, but relevance to the current study becomes diluted. Probably a richer seam is the methods used in such studies. There is no single way to analyse database contents, although several approaches seem fairly well established.

The most relevant nine studies on database content coverage are summarised in Appendix 2. There are two aspects of interest in these:
- where the topic is of most interest
- a social care database or set of databases is the focus.

None of the studies concerned children's social care: the closest related to the care of older people (Golder et al, 2008; Taylor et al, 2003; Taylor et al, 2006), mental health (Brettle and Long, 2001), and bereavement (Long et al, 2002). None of the studies used the same set of databases as the current study. The databases in common tended to be MEDLINE, CINAHL and PsycInfo, with three studies using CareData (predecessor of Social Care Online) (Taylor et al, 2003; Taylor et al, 2006; Golder et al, 2008). One study drew on ASSIA, IBSS and HMIC (Golder et al, 2008). Three studies focused specifically on Social Work Abstracts (Holden et al, 2008; Holden et al, 2009; Shek, 2008).

Findings which support the current study relate to distribution of relevant and unique records across different databases and the low relevance of Social Work Abstracts (via the proxy measure of missing journal coverage).

Findings which contrast with the present study include the relatively high relevance ranking of health databases (Brettle and Long, 2001; Golder et al, 2008; Taylor et al, 2003; Taylor et al, 2006), the relatively low relevance of two social care specific-databases (AgeInfo and CareData, (Golder et al, 2008; Taylor et al, 2003; Taylor et al, 2006)) and of general social science databases (Golder et al, 2008).

Methodology in the literature: Database output analysis

Methods of search output comparison mainly consist of analysing a set of systematically searched and screened records to describe which databases they came from. This is either done at time of searching (conduct systematic searching, label each reference according to its source, screen for relevance then analyse the final set for source overlap) or retrospectively (i.e. obtain a set of relevant records then search multiple databases to determine coverage).

The majority of approaches in the literature seem prospective (27 were found in the current review, e.g. Aker et al, 1996; Burnham and Shearer, 1993; Crumley et al, 2005; Minozzi et al, 2000; Papaioannou et al, 2009; Stokes et al, 2009; Taylor et al, 2006; Topfer et al, 1999) whereas only two were identifiable as retrospective (Moseley et al, 2009; Shek, 2008). The prospective approach is probably more like 'real life' searching, and is used in the current study.

Methodology in the literature: Database coverage overlap

The repetition of a reference in different databases was measured by Brettle and Long (Brettle and Long, 2001). Another option is to calculate overlap – this can be extended into the realm of complex mathematics or dealt with quite simply (Egghe and Goovaerts, 2007). The important thing is to take account of the volume of search output from both or all databases being compared, as well as the intersection between them. Database studies often mention overlap but lack details of the calculations. Theoretical articles gave more information (Gluck, 1990; Egghe and Goovaerts, 2007; Hood and Wilson, 2003).

Methodology in the literature: Database contents description

Moving from analysis of outputs to description of database contents, assessment of journal title coverage is commonly used in comparison studies. Holden et al, for example, highlighted journal issue coverage in Social Work Abstracts (Holden et al, 2008; Holden et al, 2009). Most databases contain abstracted details of journal articles which match selection criteria of journal titles or organisations, in specific topic areas and within other limits such as geographic coverage and publication dates.

Social care research, however, contains a wide variety of types of work, such as reports, guides, events, online resources and other material which is not formally published (or grey literature). Reliance on journal title-listing would miss this type of coverage. Social Care Online contains a broad description of its grey literature sources (see www.scie-socialcareonline.org.uk/about/about.asp) but this does not list specific organisations. It is also difficult to determine the proportion of grey literature in other databases due to lack of clearly identifying record field, e.g. in AgeInfo and ChildData. Therefore, coverage description for whole social care databases would be difficult.

Methodology in the literature: Other output measures

Several studies use measures such as sensitivity, precision, specificity and recall for database comparison (Aker et al, 1996; Crumley et al, 2005; Wong et al, 2006). However, in practice, many practitioners are not clear on their definitions. Korphage thought precision and recall were not useful because recall cannot be calculated as not all possible relevant documents can be known (Stokes et al, 2009). Recall and sensitivity are often assumed to be the same thing, as are specificity and precision. Possibly such measures are more appropriate for analysing performance of search strategies than for database comparison. For the current study, total output statistics were not available which made it impossible to calculate sensitivity and precision.

5 Conclusion

As information professionals working for a social care research organisation, we aim to improve the evidence base in our own work and justify our approach to partner agencies in collaborative projects. This is particularly true when discussing the evidence for choosing specific databases for social care searching

Conscious of the need for research into database coverage, it made sense to use existing work for raw data, in order to save staff time and use existing subscription resources and software. Future analysis could also take place in a resource-efficient way, extending the evidence base for relatively little cost or staff input.

When drawing conclusions from this study's findings, we can celebrate gathering the first evidence on database coverage on a children's social care topic. This is in the context of how little prior work exists to inform evidence based choice of social care databases. However, we are only working with small extracts of the databases concerned, in one aspect of the field of social care. Therefore this study should be seen as an indication of which databases to use and a call to further evidence gathering on social care databases, not an end point.

5.1 How the results compared with previous studies

Previous studies varied widely on the number of databases used for systematic searching; the 16 used here represent more than average. Of these, seven or eight databases contributed unique content to each topic search and so, by implication, that is the minimum required. In contrast, McDonald et al (1999) found that only four databases were needed for psychiatry topics. Approximately 38 per cent of the current study's references were only found in one database: this matches Brettle and Long's findings on mental health rehabilitation (Brettle and Long, 2001).

Database performance on relevance, unique contribution and overlap was topic dependent for three related searches on different aspects of the lives of looked after children. Prior studies found that database coverage and comparison is very topic dependent (Hood and Wilson, 2001; Taylor et al, 2003). Reasons for coverage variation by topic are not explored here, but could be due to deliberately specific topic coverage within databases or poor standards of record selection and indexing. The range of overlap was similar to that found by Taylor (2003).

There is little previous research on specific databases against which to judge the current findings.

Health databases

The relevance of health databases (e.g. CINAHL and MEDLINE) found in this study largely agrees with studies such as Long et al (2002) on the topic of preventing complicated bereavement in older people and Taylor et al (2003) on care services for older people . Relatively few studies have concentrated on the PsycInfo, but Shek (2008) found that its coverage of social care journals compared favourably with Social Work Abstracts.

Social care-specific databases

No previous analysis of Social Care Online's coverage or overlap with other databases has been carried out, however before 2005 a few studies used its predecessor, CareData. Taylor et al (2006) found that CareData's relevance contribution on the topic of care services for older people was relatively small, but that it had high precision .

Social Services Abstracts was among the most relevant databases in the current study. Flatley et al (2007) found that Social Services Abstracts and Social Work Abstracts seemed to have complementary coverage, however the contribution by Social Work Abstracts was so small in the current study that difference is implied rather than proved. Previous findings on incomplete journal coverage of Social Work Abstracts (Holden et al, 2008; Holden et al, 2009; Shek, 2008) seem borne out by its relatively poor performance on relevance and contribution to unique references, especially relative to Social Services Abstracts and PsycInfo. However, it is important to bear in mind that exhaustive journal coverage analysis has not been carried out on every bibliographic database.

Population-specific databases

No prior research has been found on coverage of ChildData, a key database for this population and topic.

5.2 Limitations of the study

The methodology used in this study is simple and could be extended with more sophisticated statistical analysis and automation of overlap analysis, using software such as SPSS® (SPSS Inc., 2009). Despite the attempt in this study to use a wide range and relatively large number of databases, database selection is open to criticism. For practical reasons relating to source identification, ChildData and Social Care Online were used instead of the larger Social Policy and Practice, which incorporates these databases plus Planex and Acompline. Zetoc was used in preference to other British Library interfaces such as BL Direct and BL Inside web as Zetoc offers more years of

coverage than BL Direct and was already available to us on subscription, unlike BL Inside web. Although this study's sample sizes are relatively large in comparison to most previous studies, a few hundred references represent a very small part of databases which can run to several million records in size. Database coverage is very topic specific, so the potential to generalise the results is limited.

6 Discussion

6.1 Social care – challenging to search

There are several reasons why social care is a difficult field to search and make informed decisions on search sources and strategies. Like other practical disciplines, social care cuts across several fields such as health, social policy, social science, psychology and economics. Therefore, when conducting systematic searching on social care topics, databases are chosen from these diverse fields, as illustrated by the scope of the 20 databases recommended in SCIE guidance for carrying out systematic reviews in social care (Coren and Fisher, 2006).

The general characteristics of social care literature relative to published health research are diversity of literature types and methodological approaches, location of relevant literature in many different disciplines, high proportion of grey literature, lack of controlled language tools and lack of standardised structure within articles. This reflects the practical nature of social care, and the contribution to and use made of knowledge by practitioners.

Locating the evidence is hampered by lack of standardisation of controlled language: there is no equivalent of health's MeSH thesaurus, which is an important aid to searching. Some databases have limited indexing, and title searches may depend on searching a title which lacks words suggestive of the topic. This means that when faced with absent / non-standard indexing and free text which does not contain commonly used phrases or words, the chances of retrieving relevant material when searching are reduced, relative to the health literature.

6.2 Why database coverage overlap is important

From a cost and staff resources perspective, database overlap sounds wasteful. However, Sampson et al (2006) noted that overlap could allow searchers to find an article in one database when it could not be tracked down in another, due to differences in indexing. Minozzi et al (2000) presented evidence that, for a cross discipline topic such as rehabilitation, a high proportion of the shared references in EMBASE and MEDLINE were indexed differently. Hood and Wilson (2005) commented that overlap could help identify highly cited (and therefore key) items.

We can't assume that perfect searching is possible. Choice of search topic affects ability to retrieve evidence. Wong et al (2006) reported that overlap between EMBASE and MEDLINE varied from 10% to 85% depending on the topic searched. Quality of indexing or search utilities or absence of abstracts (Othman and Halim, 2004) and the skill of the searcher are also important. Misleading details creep into databases, sometimes due to deliberate indexing policy such as ordering author surnames alphabetically (as in ASSIA) – an estimated one to three per cent of database-derived records in

the current dataset required correction. Databases are not fail-safe boxes of evidence.

6.3 UK databases' importance

Drawing on UK-specific content is essential, as the welfare context is specific to the UK. However, the majority of material in large databases comes from the US. To overcome this bias, it is important to search UK databases such as ChildData and Social Care Online. Such sources are smaller and more specific, and may lack comprehensiveness or the sophistication of commercial interfaces. Grayson and Gomersall (2003) noted that geographical and coverage bias within research sources tends to be neglected by those searching for social science evidence. Another area of difficulty is subscription access – small providers are less likely to be included in research institutions' subscription packages. The Social Policy and Practice database is an attempt to solve this problem: this Ovid product combines the UK databases from Acompline, AgeInfo, ChildData, Planex, and Social Care Online under single subscription access.

Looking to the future, portals seem to be becoming popular with evidence providers, NHS Evidence (www.evidence.nhs.uk) being an example. Several social care portal ideas are currently under discussion by information providers - would a portal approach solve some of the problems of database choice for the searcher?

6.4 Messages for practice

The implications of findings for LIS professionals and researchers are
- search a wide range of databases (from large, health- and social science-orientated ones to smaller, social-care specific ones) for social care relevant material (Mehdyzadeh, 2004).
- Large health and social science databases (such as PsycInfo) tend to contain a high proportion of relevant material,
- Select different databases with overlapping content. This may increase the chances of finding relevant references if indexing tools lack social care specificity.
- Also use smaller, more specific, particularly UK-based databases (such as Social Care Online and ChildData), which are more likely to have topic-specific indexing (Grayson and Gomersall, 2003).

Overall, the evidence is sparse with respect to coverage of most social care topics; further database source analysis would improve this situation.

Databases are not fail-safe boxes of evidence. On close examination, every aspect of databases' contents and usage is affected by subjective decisions and potential for error. Awareness of these limitations and selection of further sources to insure against such weaknesses are required when searching for

evidence on a topic. In practice, bibliographic database choice is probably influenced by familiarity, subscription budgets, perceptions of ease of use, and capacity and skill of the searching team. However, the searcher should maintain a critical approach to assessing database coverage and overlap.

7 References

Aker, P. D., McDermaid, C., Opitz, B. G. and White, M. W. (1996) Searching chiropractic literature: a comparison of three computerized databases. *Journal of Manipulative and Physiological Therapeutics,* 19 (8) 518–524.

Bostock, L., Brodie, I., Clapton, J., Fish, S., Fisher, M., Morris, M., Kearney, P. and Rutter, D. (2009) Increasing the number of care leavers in 'settled, safe accommodation'. *Vulnerable children scoping review 3.* London: Centre for Excellence and Outcomes in Children and Young People's Services.

Brettle, A. J. and Long, A. F. (2001) Comparison of bibliographic databases for information on the rehabilitation of people with severe mental illness. *Bulletin of the Medical Library Association,* 89 (4) 353–362.

Brodie, I., Bostock, L., Clapton, J., Fish, S., Fisher, M., Morris, M., Kearney, P. and Rutter, D. (2009) Improving educational outcomes for looked after children and young-people. *Vulnerable children scoping review 1.* London: Centre for Excellence and Outcomes in Children and Young People's Services.

Burnham, J. and Shearer, B. (1993) Comparison of CINAHL, EMBASE, and MEDLINE databases for the nurse researcher. *Medical Reference Services Quarterly,* 12 (3) 45–57.

Coren, E. and Fisher, M. (2006) The conduct of systematic research reviews for SCIE knowledge reviews *SCIE Research resource 1.* London: Social Care Institute for Excellence.

Crumley, E. T., Wiebe, N., Cramer, K., Klassen, T. P. and Hartling, L. (2005) Which resources should be used to identify RCT/CCTs for systematic reviews: a systematic review. *BMC Medical Research Methodology,* 5–24.

Egghe, L. and Goovaerts, M. (2007) A note on measuring overlap. *Journal of Information Science,* 33 (2) 189–195.

Fish, S., Fisher, M., Rutter, D., Clapton, J., Bostock, L. and Brodie, I. (2009) Improving the emotional and behavioural health of looked-after children and young people (LACYP). *Vulnerable children scoping review 2.* London: Centre for Excellence and Outcomes in Children and Young People's Services.

Flatley, R. K., Lilla, R. and Widner, J. (2007) Choosing a database for social work: A comparison of Social Work Abstracts and Social Services Abstracts. *The Journal of Academic Librarianship,* 33 (1) 47–55.

Gluck, M. (1990) A review of journal coverage overlap with an extension to the definition of overlap. *Journal of the American Society for Information Science,* 41 (1) 43–60.

Golder, S., Mason, A. and Spilsbury, K. (2008) Systematic searches for the effectiveness of respite care. *Journal of the Medical Library Association,* 96 (2) 147–152.

Grayson, L. and Gomersall, A. (2003) A difficult business: Finding the evidence for social science reviews. *Working paper 19.* London: Evidence Network.

Holden, G., Barker, K., Covert-Vail, L., Rosenberg, G. and Cohen, S. A. (2008) Does social work abstracts work? *Research on Social Work Practice,* 18 (5) 487–499.

Holden, G., Barker, K., Covert-Vail, L., Rosenberg, G. and Cohen, S. A. (2009) Social Work Abstracts fails again. *Research on Social Work Practice,* 19 (6) 715–721.

Hood, W. W. and Wilson, C. S. (2001) The scatter of documents over databases in different subject domains: How many databases are needed? *Journal of the American Society for Information Science and Technology,* 52 (14) 1242–1254.

Hood, W. W. and Wilson, C. S. (2003) Overlap in bibliographic databases. *Journal of the American Society for Information Science and Technology,* 54 (12) 1091–1103.

Hood, W. W. and Wilson, C. S. (2005) The relationship of records in multiple databases to their usage or citedness. *Journal of the American Society for Information Science and Technology,* 56 (9) 1004–1007.

LISU (2009) Periodical prices (Great Britain). *Libraries, Archives, Museums and Publishing Online Statistics Tables.* LISU.

Long, A. F., Godfrey, M., Randall, T., Bretle, A. and Grant, M. (2002) Feasibility of undertaking systematic reviews in social care. *Developing evidence based social care policy and practice.* Health Care Practice R&D Unit, University of Salford / Nuffield Institute for Health, University of Leeds.

McDonald, S., Taylor, L. and Adams, C. (1999) Searching the right database. A comparison of four databases for psychiatry journals. *Health Libraries Review,* 16 (3) 151–156.

Mehdyzadeh, H. (2004) Searching for the Evidence: An Introduction to Social Science Information Retrieval. London: Department for Culture Media and Sport.

Microsoft Corporation (2007) Excel Version 2007. Redmond, WA, Microsoft Corporation.

Minozzi, S., Pistotti, V. and Forni, M. (2000) Searching for rehabilitation articles on MEDLINE and EMBASEEMBASE. An example with cross-over design. *Archives of Physical Medicine & Rehabilitation,* 81 (6) 720–722.

Moseley, A. M., Sherrington, C., Elkins, M. R., Herbert, R. D. and Maher, C. G. (2009) Review: Indexing of randomised controlled trials of physiotherapy interventions: a comparison of AMED, CENTRAL, CINAHL, EMBASE, Hooked on Evidence, PEDro, PsycINFO and PubMed. *Physiotherapy,* 95 (3) 151–156.

Othman, R. and Halim, N. S. (2004) Retrieval features for online databases: common, unique, and expected. *Online Information Review,* 28 (3) 200–210.

Papaioannou, D., Sutton, A., Carroll, C., Booth, A. & Wong, R. (2009) Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information and Libraries Journal,* 27 (2) 114–122.

Sampson, M., McGowan, J., Cogo, E. and Horsley, T. (2006) Managing database overlap in systematic reviews using Batch Citation Matcher: case studies using Scopus. *Journal of the Medical Library Association,* 94 (4) 461–463.

Shek, D. T. L. (2008) Comprehensiveness of Social Work Abstracts as a Database for Researchers and Practitioners. *Research on Social Work Practice,* 18 (5) 500–506.

SPSS Inc. (2009) SPSS. Chicago, Illinois, SPSS Inc.

Stokes, P., Foster, A. and Urquhart, C. (2009) Beyond relevance and recall: testing new user-centred measures of database performance. *Health Information and Libraries Journal,* 26 (3) 220-231.

Taylor, B., Dempster, M. and Donnelly, M. (2003) Hidden gems: Systematically searching electronic databases for research publications for social work and social care. *British Journal of Social Work,* 33 423–439.

Taylor, B. J., Wylie, E., Demster, M. and Donnelley, M. (2006) Systematic searching on the AgeInfo database. *SCIE Report 13.* London: Social Care Institute for Excellence.

Thomson Reuters (2009) EndNote Version X3. New York: Thomson Reuters.

Topfer, L. A., Parada, A., Menon, D., Noorani, H., Perras, C. and Serra-Prat, M. (1999) Comparison of literature searches on quality and costs for health technology assessment using the MEDLINE and EMBASE databases. *International Journal of Technology Assessment in Health Care,* 15 (2) 297–303.

Wong, S. S. L., Wilczynski, N. L. and Haynes, R. B. (2006) Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. *Journal of the Medical Library Association,* 94 (4) 451–455.

8 Acknowledgements

Appendix 1 Description of databases searched in this study

Name of Database, and host / platform	Sector / topics covered	Geographical coverage	Size: number of records	Access
Applied Social Science Information and Abstracts (ASSIA), CSA Illumina	Covers a wide range of social science topics.	US and UK	375,000	Subscription (available at British Library)
Australian Family and Society Abstracts, Informit	About families and the social issues that affect them.	Australia	70,000	Subscription
British Education Index, Dialog	All aspects of education.	UK and Europe	150,000	Subscription
CERUKplus, NFER	Research project and output database on education and children's services.	UK	4,000	Freely available
ChildData, NCB	Issues affecting children.	UK	70,000	Subscription (available at British Library)
Cumulative Index to Nursing and Allied Health Literature (CINAHL), EBSCO Host	Nursing and allied health professions.	US-based	2m	Subscription (available at British Library)

Bibliographic databases for social care searching

Name of Database, and host / platform	Sector / topics covered	Geographical coverage	Size: number of records	Access
Continued from previous page				
Cochrane Library, Wiley Interscience. Includes the Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register, Health Technology Assessment Database, NHS Economic Evaluation Database.	Evidence-based health.	UK	4,000 records in Cochrane Database of Systematic Reviews	Freely available
EMBASE, Ovid SP	Biomedical database with emphasis on drug discovery.	Coverage staged as US / Canada: 31%, Europe: 54%, Australasia: 10%, Rest of World: 5%.	12m	Subscription (available at British Library)

Name of Database, and host / platform	Sector / topics covered	Geographical coverage	Size: number of records	Access
Continued from previous page				
Health Management Information Consortium (HMIC), Ovid SP	Health and social care services, management and economics. Content from Department of Health and The King's Fund.	UK	250,000	Subscription (available at British Library)
International Bibliography of the Social Sciences (IBSS), EBSCO Host	Major social science database.	International	2.5m	Subscription (available at British Library)
MEDLINE, Ovid SP	Major health database.	International	11m	Subscription, but PubMed is freely available
PsycInfo, Ovid SP	Major database on psychology and related disciplines.	International	2.7m	Subscription (available at British Library)
Social Care Online, SCIE	Social care.	UK	110,000	Freely available
Social Services Abstracts	Social services.	US-based	155,000	Subscription (available at British Library)
Social Work Abstracts	Social work.	US-based	74,000	Subscription
Zetoc, MIMAS	British Library electronic table of contents.	UK	20m	Subscription (available at British Library)

Appendix 2. Findings from other relevant studies

Study	Topic	Databases	Findings	In comparison with present study this example had:
Brettle and Long (2001)	Community-based rehabilitation of people with severe mental illness	CINAHL, Cochrane, MEDLINE, PsycLit, Sociofile Social Science Citation Index	96 records analysed ~1/3 records unique to each database 42% records found in only one database PsycLit most relevant with 44% records MEDLINE was second most relevant with 29% records Records duplicated in up to 4 databases Overlap was largest between PsycLit and SSCI at 20%	Similar methodology Greater concentration of records in fewer databases Proportionally lower percentage overlap
Long, et al. (2002)	Preventing complicated bereavement in older people	CINAHL, Cochrane, MEDLINE, PsycLit Sociofile	60 records analysed PsycLit was most relevant with 66% of records, and highest proportion unique records with 54%	Greater concentration of records in fewer databases
Taylor et al. (2003)	Decision making about community care for older people	Caredata (the precedessor of Social Care Online), CINAHL, MEDLINE Social Science Citation Index	276 individual records were analysed. SSCI was most relevant with 55% records MEDLINE: 41% Cinahl: 19% CareData: 1% CareData performed relatively poorly on relevance However, it enabled precise searches All four databases retrieved unique records	CareData, as a UK social care specific database performed relatively poorly on relevance

Study	Topic	Databases	Findings	In comparison with present study this example had:
Continued from previous page				
Taylor et al. (2006)	Decision making about community care for older people	AgeInfo CareData CINAHL MEDLINE PsycInfo Social Science Citation Index Social Services Abstracts	363 records were analysed Relevance ranked by sensitivity: MEDLINE was most relevant, then SSCI CINAHL AgeInfo Precision was high for AgeInfo and CareData as above	CareData and AgeInfo, as UK social care focus databases performed relatively poorly on relevance
Golder, et al. (2008)	Respite care for carers of frail older people	36 databases, including ASSIA CareData IBSS MEDLINE PsycInfo	44 individual records were analysed The most relevant databases were: MEDLINE Ageline SSCI / Cochrane PsycInfo Six databases were considered essential, drawn from health, social science and specialist subjects: MEDLINE and EMBASE SSCI AgeLine, HMIC and PsycInfo	Similar large number of databases searched – 36 Relevance findings were dissimilar to current study Fewer databases were needed, but the message of breadth of disciplines agrees with current findings

Study	Topic	Databases	Findings	In comparison with present study, this example:
Continued from previous page				
Flatley et al. (2007)	Study not topic-based except social work focus of databases– analysed journal title coverage	Social Work Abstracts Social Services Abstracts	Comparison of journal titles coverage showed complementary content	Reinforces contrasting content between Social Work Abstracts and Social Services Abstracts
Holden et al. (2008) Holden et al. 2009)	Study not topic-based except social work focus of Social Work Abstracts– analysed journal title coverage	Social Work Abstracts PsycInfo	23 journal titles' coverage was compared: Social Work Abstracts lacked proportionately more journal issues than PsycInfo. The study was later extended for SWA with 33 journal titles – of which 82% had missing issues	Reinforces relatively low relevance of Social Work Abstracts
Shek (2008)	Study not topic based except focus of Social Work Abstracts– analysed journal title coverage	Social Work Abstracts, PsycInfo, MEDLINE Sociological Abstracts	Two journal titles' coverage was compared Social Work Abstracts was missing more issue coverage than the other databases	Ditto

The main aim of this work is to help information professionals and researchers make an informed choice of databases when planning searches on social care topics. In light of lack of previous evidence, this study adds significantly to what is known about the relevance and coverage of social care databases, particularly on children's social care topics.