# The systematic review process: comparing databases for pet population searches

**CENTRE FOR EVIDENCE–BASED VETERINARY MEDICINE**
Putting research into practice

**The University of Nottingham**
UNITED KINGDOM · CHINA · MALAYSIA

**Martin Downes, Jenny Stavisky, Vicki Adams, Marnie Brennan and Rachel Dean**
Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science, The University of Nottingham, UK
Email: martin.downes@nottingham.ac.uk

## Background:

A systematic review aims to provide an exhaustive summary of literature (published and grey) relevant to a research question. The first step is a thorough search of the databases and citation indexes that are relevant to the topic area. Most databases currently use aids for indexing and retrieving publications. Subject headings (SH) are controlled terms or descriptors that are manually assigned to publications by the database indexers or librarians. They consist of a set of terms or descriptors used in a hierarchical structure which allows searching at various levels of specificity. The databases used in this review each have their own thesaurus of terms: Cab Abstracts™ uses the CAB Thesaurus, EMBASE™ uses a thesaurus known as Emtree® and both PubMed® and Medline® use the MeSH® (Medical Subject Headings) terms thesaurus formulated by the National Library of Medicine's (NLM) library.  While these databases are used extensively in systematic reviews, little has been done to demonstrate the difference between subject heading and keyword searching. Currently a systematic review is been undertaken to determine the best methods of determine the demographics and population size of the dog and cat population in the United Kingdom; we use this as a case study for this poster. The aim of the poster is to compare searching with keywords and subject headings using various publication databases.

## Methods:

Medline, CAB and EMBASE were accessed through the Ovid SP interface, whereas PubMed was accessed through the online interface provided by The National Center for Biotechnology Information. Search terms were entered into the respective databases for keyword searches and were as follows: population, population density, population dynamics, demography and census. The subject headings that were suggested by the database search engines were used as comparators for each of the keywords. The results were then tabulated in MS Excel™ and VennGenerator© was used to create Venn diagrams that were proportionally representative of the results, showing overlapping representation between some of the different search terms.
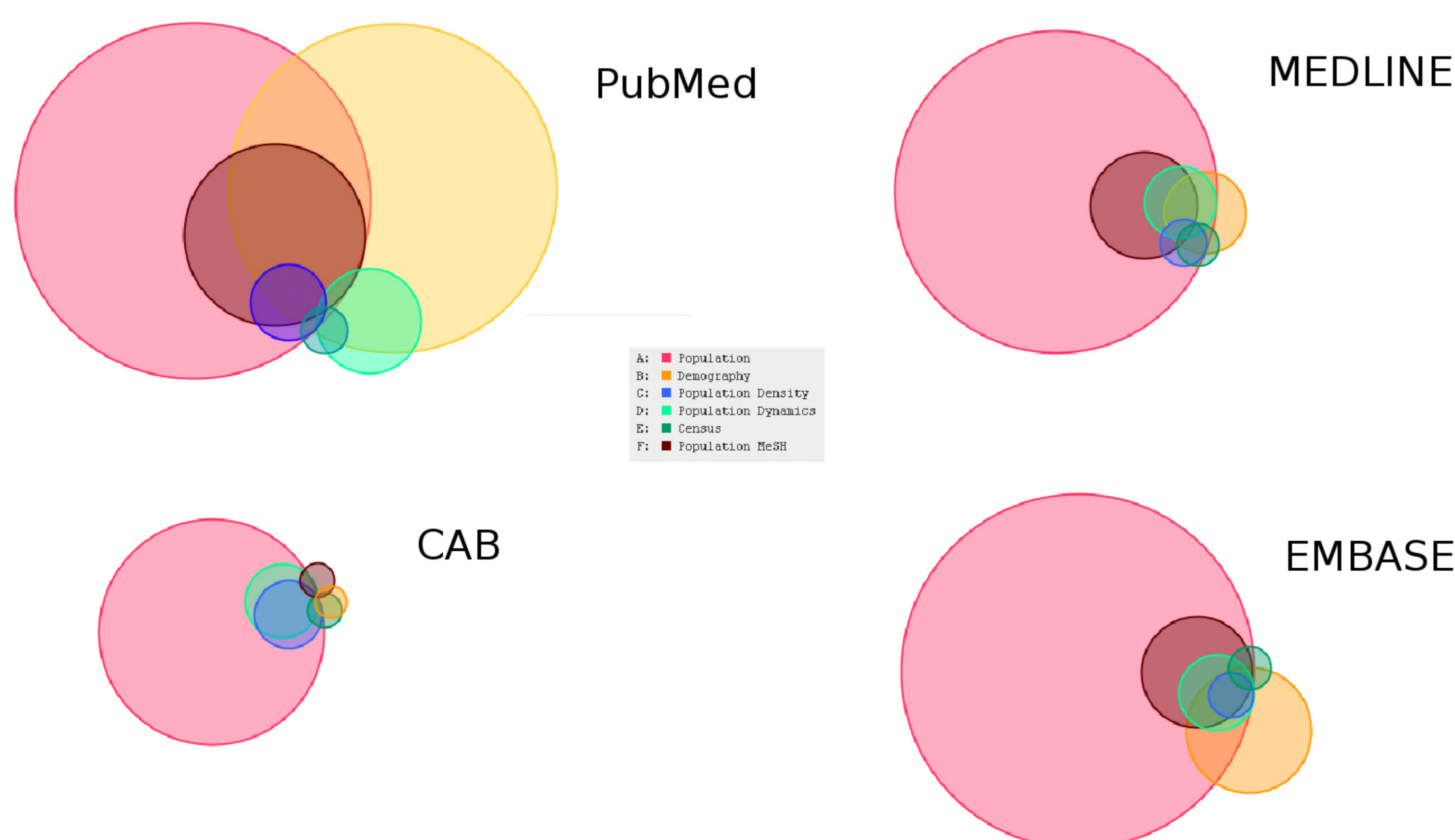
**Table 1. The number of publications found and the proportional differences between keywords and subject headings (SH), when using databases to search for publications for a systematic review.**

| | MEDLINE | | | | PubMed | | | | CAB | | | | EMBASE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Keyword | Subject headings | Proportion of SH not in Keyword | Proportion of keywords not in SH | Keyword | Subject headings | Proportion of SH not in Keyword | Proportion of keywords not in SH | Keyword | Subject headings | Proportion of SH not in Keyword | Proportion of keywords not in SH | Keyword | Subject headings | Proportion of SH not in Keyword | Proportion of keywords not in SH |
| Population | 750292 | 83458 | 0 | 0.89 | 921883 | 238855 | 0 | 0.74 | 367100 | 8196° | 0.33 | 0.99 | 908272 | 86129 | 0 | 0.91 |
| Population Density | 15989 | 13201 | 0 | 0.17 | 41614 | 13187 | 0 | 0.68 | 32230 | 23185 | 0 | 0.28 | 14692 | 15861 | 0.23 | 0.16 |
| Population Dynamics | 38071 | 56269 | 0.37 | 0.06 | 75637 | 56302 | 0 | 0.26 | 36798 | 40221 | 0.22 | 0.15 | 40695 | 41079 | 0.06 | 0.05 |
| Demography | 49034 | 773116 | 0.94 | 0.03 | 791645 | 772905 | 0 | 0.02 | 7084 | 4859 | 0 | 0.31 | 114224 | 112669 | 0 | 0.02 |
| Census | 12971 | 2146* | 0.34 | 0.88 | 15591 | 2141 | 0 | 0.86 | 8644 | 2346* | 0.31 | 0.79 | 13627 | 51799^ | 14.02 | 0.75 |
| All Together | 772369 | 1637247 | 0.82 | 0.62 | 1462966 | 910909 | 0 | 0.38 | 376161 | 73644 | 0.07 | 0.82 | 978801 | 253407 | 0.01 | 0.74 |
| Animal Population Groups | 1970 | 833720 | 0 | 0.99 | 836539 | 83299 | 0 | 0.01 | | | | | | | | |

*Censuses was the subject heading used
°Populations was the subject heading used
^Population research was the subject heading used

## Figure 1. Venn diagrams showing a proportional estimation of the overlap between the different search terms and subject headings



PubMed

MEDLINE

A: Population
B: Demography
C: Population Density
D: Population Dynamics
E: Census
F: Population MeSH

CAB

EMBASE

## Discussion:

The results show that Pubmed has the largest database of information. Both PubMed and the CAB abstracts indexing systems seem to be more accurate with both of them having the lower proportions of publications indexed outside of the keyword searches, with PubMed having all the SH publications checked in this study included in the keyword searches. Some of the databases have large discrepancies between which publications are indexed to a SH when compared with the keyword searches. What is of concern is that both PubMed and Medline are products of NLM and between them there seems to be substantial differences in the way that publications are indexed.  This study shows that great care is required when using SH searches as there can be huge variations between the SH and the keyword searches and publications are open to different librarian interpretations when been mapped to a SH. It also shows that if you want a more sensitive search strategy it is better to use both methods of searching whereas SH searches tend to be more specific.