

# Lung cancer search filters: Connecting to the evidence

Damarell R<sup>1</sup>, Tieman J<sup>1</sup>, Olver I<sup>2</sup>, Currow D<sup>1</sup>

<sup>1</sup> Palliative and Supportive Services, Flinders University

<sup>2</sup> Cancer Council Australia

Contact email: [jennifer.tieman@flinders.edu.au](mailto:jennifer.tieman@flinders.edu.au)

There can be many challenges in finding needed literature and evidence in health activities. This study aimed to develop lung cancer search filters with experimentally determined rates of literature retrieval to facilitate access to the available evidence base.

## Introduction

Searching to support guideline development or systematic reviews is complex and can be difficult even for those with high level skills [1]. Search filters (also described as hedges or optimal search strategies) consist of combinations of search terms, both MeSH and textwords [2]. They are intended to support efficient retrieval of desired literature by searching for a characteristic such as a study type or a topic of medical interest [3-4]. The validated search filter can be used to help clinicians access the literature by running saved searches or to continuously harvest new literature through autoalerts.

## Aim

The specific objectives of this project were:

1. to develop and validate a search filter for lung cancer within Medline using a gold standard dataset of lung cancer references;
2. to explore the possibility of creating highly discriminatory filters for small cell lung cancer and non-small cell lung cancer; and
3. to translate and validate the search filters for open use in PubMed and combine them with a set of expert topic searches on lung cancer management.

## Method

The lung cancer filter were created following a variation of the classic methodology used to develop the methodological search filters established in PubMed as Clinical Queries [4]. The core of this approach is a 'gold standard set', or a set of references of known relevance to the topic of interest. The gold standard set for the project used all Medline-indexed lung cancer references included in the NHMRC endorsed CCA lung cancer guidelines (2004). It followed four phases:

§ the construction of a Gold Standard set;

§ term identification;

§ filter development; and

§ filter validation.

The steps for finalising the articles included in the Gold Standard set are outlined in Figure 1.

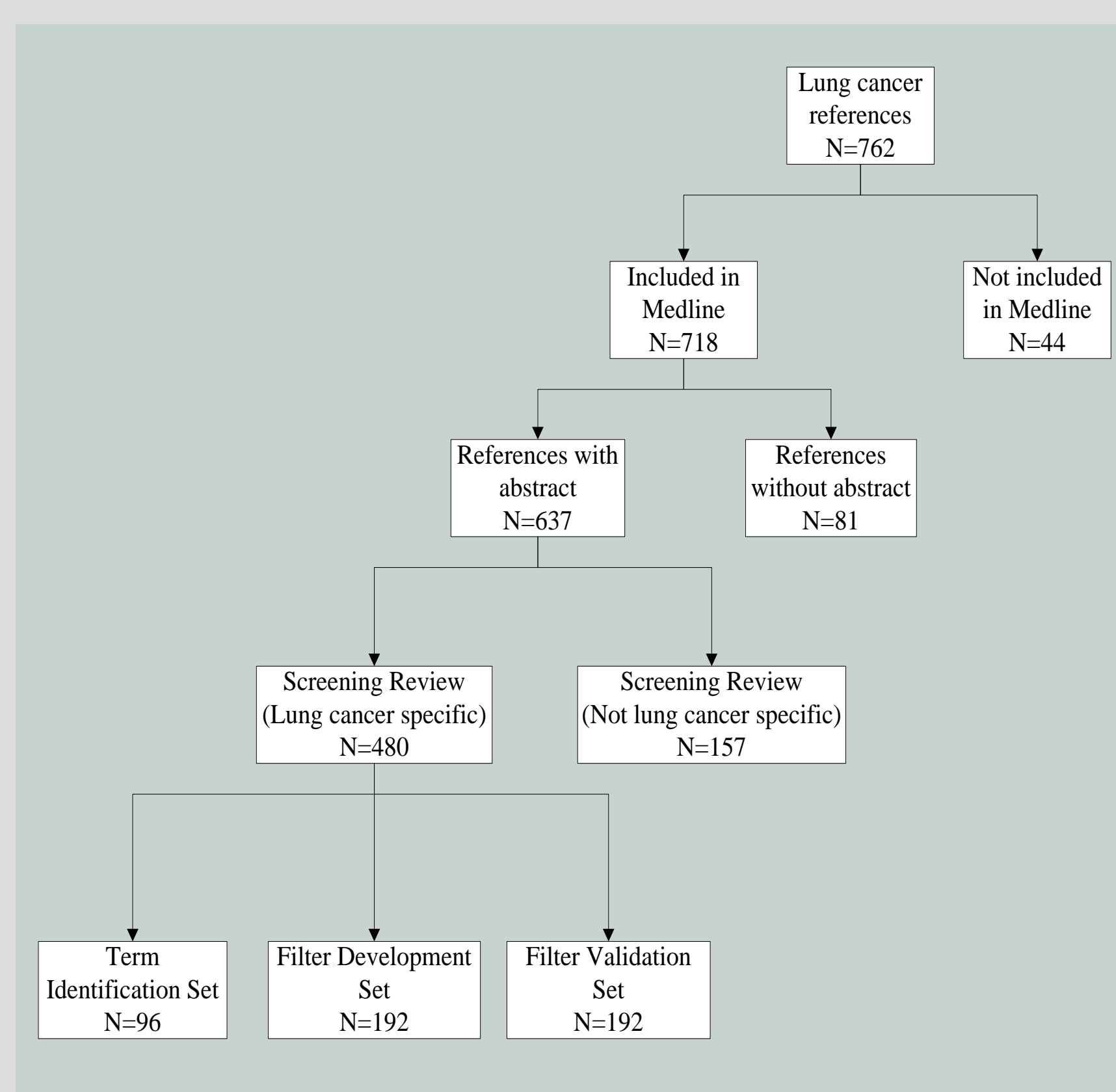


Figure 1: Schematic representation of the development of the Gold Standard Dataset

## Identifying candidate terms

The Gold Standard set was randomly allocated into three sets to avoid the bias inherent in validating a search filter within the same set of records used to create it. These sets were: a Term Identification Set (TIS) containing 20% of all references; a Filter Development Set (FDS) (40%); and a Filter Validation Set (FVS) (40%).

Term identification for the filter was based on separate frequency analyses of the Medical Subject Headings (MeSH) terms and title/abstract textwords of the TIS items; review of the relevant concept 'trees' in the Medline MeSH thesaurus; examination of other search strategies (e.g. National Cancer Institute); and clinical advice.

The number of times a MeSH term was found in a TIS record was noted and tallied. (See Table 1.)

The titles and abstracts of all records in the TIS were saved as a text file. This text file was then imported into Concordance, a text analysis program, which created a list of single textwords by frequency. Each term was searched across the title and abstract fields of the TIS records to determine the number of unique records each term retrieved. (Table 1)

Table 1: Terms most frequently identifying Gold Standard Dataset items

Terms	Record occurrence N=96	% Record Occurrence
<b>MeSH</b>		
Lung neoplasms	88	91.7
Carcinoma, Non-Small-Cell Lung	36	37.5
Carcinoma, Small Cell	26	27.1
Carcinoma, Squamous Cell	12	12.5
Adenocarcinoma	11	11.4
Carcinoma, Bronchogenic	6	6.2
<b>Textwords/phrases</b>		
Lung cancer	81	84.4
Small cell lung cancer	46	47.9
Non small cell lung cancer	33	34.4
NSCLC	21	21.9
Small cell carcinoma	9	9.4
Adenocarcinoma	8	8.3
SCLC	7	7.3
Bronchogenic carcinoma	6	6.2
Lung cancers	5	5.2
Lung carcinoma	4	4.2

## Filter Development

An objective method for building the filter was adhered to. The term with the highest recall was then combined individually with each other candidate term and run in the Gold Standard dataset. Terms not retrieving additional records were eliminated. The most successful combination was then combined with all remaining candidate terms. This process was repeated until no further candidate terms remained.

The final best performing search in the FDS was a two-term search: Lung neoplasms/ OR lung cancer\$.mp. The two-term lung cancer filter was then run in the FVS to determine its retrieval rate where it retrieved 188/192 references (recall = 97.9%). Although it demonstrated a high level of recall, two additional lung cancer terms [bronchial neoplasms/ exploded' and Pancoast syndrome/] were added from the Medline MeSH thesaurus along with their narrower concepts because of their unique association with lung cancer. The final Ovid Medline version of the lung cancer filter is therefore:

**Lung neoplasms/ OR exp Bronchial neoplasms/ OR Pancoast syndrome/ OR lung cancer\$.mp.**

## NSCLC and SCLC Search Filters

The same methodology was used to create a non-small cell lung cancer (NSCLC) filter and a small cell lung cancer filter. The final 10-term NSCLC filter is:

**Carcinoma, Non-Small-Cell Lung/ OR Non small cell lung cancer\$.mp. OR NSCLC.mp. OR ((Lung neoplasms/ or Bronchial neoplasms/ or Carcinoma, Bronchogenic/) AND (Adenocarcinoma/ or Adenocarcinoma, Bronchiolo-alveolar/ or Carcinoma, Large cell/ or Carcinoma, Squamous cell/))**

The final search filter for SCLC is:

**Small cell lung carcinoma/ OR SCLC.mp. OR (Carcinoma, Small cell/ AND (Lung neoplasms/ OR Bronchial neoplasms/ OR Carcinoma, Bronchogenic/))**

## PubMed Variants

A translation and validation exercise led to the creation of optimal versions of the Ovid Medline search filters for use in PubMed. The three PubMed variants of the lung cancer, NSCLC and SCLC search filters are now publicly available through the CareSearch website at [www.caresearch.com.au](http://www.caresearch.com.au). Hyperlinks of the searches enable 'one-click' searching.



## Conclusion

Experimentally developed search filters with high sensitivity are now available on the CareSearch website. Cancer practitioners can use these searches to facilitate the retrieval of literature relevant to lung cancer, small cell lung cancer and non-small cell lung cancer.

## References

1. Sampson, M., McGowan, J. Errors in search strategies were identified by type and frequency 2006 Journal of Clinical Epidemiology 59:1057-1063
2. Deurenberg, R., Vlayen†, J., et al. Standardization of search methods for guideline development: an international survey of evidence-based guideline development groups 2007 Health Information and Libraries Journal 25:23-30
3. Damarell RA, Tieman J et al. Development of a heart failure filter for Medline: an objective approach using evidence-based clinical practice guidelines as an alternative to hand searching. BMC Med Res Methodol. 2011 Jan 28;11:12.
4. Haynes RB, McKibbon KA, et al. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey BMJ 2005 330(7501):1179.

## Acknowledgements

This project was made possible by funding from the Cancer Council Australia.

CareSearch is funded by the Australian Government Department of Health and Ageing.