Høgskulen på Vestlandet

Masteroppgave

MAKP601

Predefinert informasjon

Startdato:	17-04-2017 09:00	**Termin:**	2017 VÅR
Sluttdato:	15-05-2017 14:00	**Vurderingsform:**	Norsk 6-trinnsskala (A-F)
Eksamensform:	Masteroppgave	**Studiepoeng:**	45
SIS-kode:	MAKP601 1 MG		
Intern sensor:	(Anonymisert)		

Deltaker

Kandidatnr.:	102

Informasjon fra deltaker

Antall ord *:	9498	**Tro- og loverklæring *:**	Ja
Jeg godkjenner avtalen om tilgjengeliggjøring av masteroppgaven min *:	Ja		

Western Norway University of Applied Sciences

MASTER'S THESIS

Testing the best performing methodological search filters to retrieve health economic evaluations in Embase:
A filter validation study

Testing av søkefiltre for best gjenfinning av helseøkonomiske evalueringer i Embase:
En valideringsstudie

Ingrid Harboe

Master in Evidence-Based Practice for Health and Social Sciences
Department of Health and Social Care
Centre for Evidence Based Practice
Supervisor: Julie Glanville

Submission Date: May 15th 2017

I confirm that the work is self-prepared and that references/source references to all sources used in the work are provided, *cf. Regulation relating to academic studies and examinations at the Western Norway University of Applied Sciences (HVL), § **10**.*

Testing the best performing methodological search filters to retrieve health economic evaluations in Embase: A filter validation study (Title English)

Testing av søkefiltre for best gjenfinning av helseøkonomiske evalueringer i Embase: En valideringsstudie (Title Norwegian)

Master's student: Ingrid Harboe

Main supervisor: Julie Glanville

Co-supervisors: Lena V. Nordheim, Kjetil G. Brurberg

Submission Date: May 15th 2017

Number of words, introduction: 9,498

Preface

Working as a research librarian in a health economics and health care interventions unit, I have become increasingly aware of the importance of evidence-based practice. I have also experienced the many challenges of successfully implementing evidence-based practice. As part of my job, I perform systematic literature searches in medical bibliographic databases, to identify recent, high quality evidence about health care interventions. We use the output of the literature searches to inform systematic reviews and health technology assessments about the effect, safety and economic evaluations of health care interventions. My health economist colleagues and I have experienced that the literature searches often retrieve records irrelevant to our purpose. Thus, I wanted to develop a methodological search filter in order to reduce the proportion of irrelevant records to save time and reduce cost.

During my studies, I got interested in writing my master's thesis about an issue relevant to literature searches. The usual approach would have been writing a systematic review commissioned to my unit, which previous master's students at my unit has followed. I launched the idea of developing and testing an economic evaluation search filter to Julie Glanville, associate director of the York Health Economics Consortium Ltd., and Lena Nordheim, head of the master's degree program. They both supported the proposal, and I have been very fortunate to have Julie Glanville as my main supervisor, and Lena Nordheim as my co-supervisor. Kjetil Gundro Brurberg has supervised me in statistics. Thank you all kindly for guiding and encouraging me throughout this study.

My health economist colleagues Gunhild Hagen, Arna Desser and Vida Hamidi, supported me from the beginning, and have encouraged me throughout the process, for which I am very grateful. I am also very grateful for the many discussions, assistance and all support from my fellow master's students, especially Hanne Nordvik Ona and Karen Rosnes. Being part of such a network has been crucial for me during this study. I would also like to thank my research librarian colleagues and the head of my unit, Ingvil Von Mehren Sæterdal, for your cooperation, understanding and support during this period. Many thanks.

Oslo, May 2017

Ingrid Harboe

Sammendrag

Denne mastergradsoppgaven består av en innledningsdel som beskriver bakgrunnen for, metodene som ble brukt, drøfting av metodene og resultatene i valideringsstudien, og en artikkel. Innledningsdelen beskriver studien mer detaljert enn artikkelen.

Bakgrunn: Metodevurderinger som «health technology assessments» (HTAs) brukes i økende grad av norske helsemyndigheter som evidensgrunnlag ved prioritering av helsetjenester. Metodevurderinger inneholder vanligvis en kritisk vurdering og oppsummering av randomiserte kontrollerte studier av effekt og sikkerhet av medisinske tiltak, og en helseøkonomisk evaluering av tiltakene, som blant annet baseres på systematiske litteratursøk i databaser.

Hensikt og problemstilling: Å teste og analysere gjenfinningen av helseøkonomiske evalueringer, ved å sammenligne «the cost effectiveness analysis filter» (CEA) med seks publiserte søkefiltre i Ovid Embase, for å prøve å oppnå en sensitivitet på minst 0.90, en presisjon på 0,10, og spesifisitet på minst 0,95.

Metode: Søkefiltrene ble testet på en gullstandard av helseøkonomiske evalueringer fra the National Health Service Economic Evaluation Database (NHS EED) publisert i årene 2008-2013 (n=2248), og tilsvarende publikasjoner (n = 2198) i gjeldende versjon av Embase. Jeg sammenlignet gjenfinningen av søkefiltrene, og redigerte CEA-filteret for å teste om presisjonen kunne bedres uten samtidig å redusere sensitiviteten vesentlig.

Resultater: CEA filteret hadde en sensitivitet på 0,899, en presisjon på 0,029 og spesifisitet på 0,991. Filteret som kom nærmest målet til studien oppnådde en sensitivitet på 0,880, en presisjon på 0,075, og spesifisitet på 0,997. Filteret med lavest sensitivitet (0,702) hadde en presisjon på 0,141.

Konklusjon: Denne valideringsstudien viser at å utvikle søkefiltre for å identifisere helseøkonomiske evalueringer, med en god balanse mellom sensitivitet og presisjon, er mulig men utfordrende. Forskere bør enes om hvilket nivå av sensitivitet og presisjon som er ønskelig å oppnå i søkeresultatet, for å velge det best egnete søkefilteret for identifisering av helseøkonomiske evalueringer.

Nøkkelord (MeSH): Bibliografiske databaser; Informasjonslagring og –gjenfinning; Sensitivitet og spesifisitet; Kostnad-nytte-analyse;

Antall ord: 300

Abstract

This master's thesis consist of an introductory part describing the background for preforming the validation study, the methods used, and a discussion of the methods and results. The introductory part describes the filter validation study in more detail than possible in the article manuscript.

Background: Health technology assessments (HTAs) are increasingly used by Norwegian health authorities as the evidence base when prioritizing which health care services to offer. HTAs typically consist of a systematic review of the effects and safety of two or more health care interventions, and an economic evaluation of the interventions, based on systematic literature searches in bibliographic databases.

Objective: To identify the best performing of seven search filters to retrieve health economic evaluations used to inform health technology assessments (HTAs), by comparing the cost-effectiveness analysis (CEA) filter to six published filters in Ovid Embase, and achieve a sensitivity of at least 0.90 with a precision of 0.10, and specificity of at least 0.95.

Methods: In this filter validation study, the included filters' performances were compared against a gold standard of economic evaluations published in 2008-2013 (n=2,248) from the National Health Service Economic Evaluation Database (NHS EED), and the corresponding records (n=2,198) in the current version of Ovid Embase

Results: The CEA filter had a sensitivity of 0.899 and precision of 0.029. One filter had a sensitivity of 0.880 and a precision of 0.075, which was closest to the objective. The filter with lowest sensitivity (0.702) had a precision of 0.141.

Conclusion: Developing search filters for identifying health economic evaluations, with a good balance between sensitivity and precision, is possible but challenging. Researchers should agree on acceptable levels of performance before concluding on which search filter to use.

Keywords (MeSH**):** Databases, Bibliographic; Information storage and retrieval; Sensitivity and specificity; Cost-Benefit Analysis (used for Cost-Effectiveness Analysis)

Number of words: 299

Table of contents

Appendices to introductory part

Article manuscript **Intended journal, Journal of clinical epidemiology:** "Identifying the best performing search filters to retrieve health economic evaluations in Embase: A filter validation study"

List of tables, figures, appendices and supplementary material

1 Introduction

Decision makers increasingly use health economic evaluations to prioritize health care interventions at a national level (Ministry of Health and Care Services, 2012; Knowledge Centre for the Health Services 2015). These health economic evaluations (e.g. cost-effectiveness analyses or cost-utility analyses) are usually either published as is, or as part of a health technology assessment (HTA). HTAs typically consist of a systematic review of the effects and safety of two or more health care interventions (e.g. drugs, medical devices, or ways of organizing health services), and an economic evaluation of the interventions (Hagen et al., 2013).

Research librarians perform systematic literature searches in biomedical and other relevant databases to inform systematic reviews and HTAs with high quality research (Green S. et al., 2011; Knowledge Centre for the Health Services 2015). More than one database should be searched to to identify as many relevant studies as possible (Lefebvre et al., ch. 6),. The choice of databases depends on the research question. To identify health economic evaluations research librarians or researchers typically search the National Health Service Economic Evaluation Database (NHS EED), MEDLINE and Embase (Glanville et al., 2009a; Harbour et al., 2014). Updates of NHS EED were discontinued in 2014. Thus, MEDLINE and Embase, have become increasingly important sources to identify economic evaluations.

To identify economic evaluations in MEDLINE and Embase, one can limit the subject search (e.g search related to using mammography for early detection of breast cancer) by adding a methodological search filter. Methodological search filters are search strategies comprising specific methodological terms to identify publications using certain study designs, including economic evaluations (Haynes, R. Brian et al., 2005; Lefebvre C. et al., 2011, ch. 6.4.11). Despite their potential to identify health economic evaluations more efficiently, published methodological filters are too sensitive and retrieve many publications that are irrelevant to the context of the local (national) HTA (Wilczynski, Nancy L.et al., 2004; Glanville, J. et al., 2009a; Mathes, 2014). Moreover, health economists generally use the retrieved economic evaluations merely to inform the development of an economic model used to assess the cost-effectiveness of the intervention of interest. Because economic evaluations are rarely directly transferrable to other contexts, results from retrieved economic evaluations are often only included in the discussion chapter of the HTA (Hagen, 2013). Thus, in the Norwegian

context, making a sensitive literature search to identify "all" health economic evaluations is not always required. The aim of this study was to identify and validate the best performing methodological search filters to retrieve health economic evaluations in one major biomedical database, Ovid Embase.

1.1 Background

The Norwegian health authorities aim to provide high quality health care services to all citizens, given accepted criteria and values, and resources constraints. To achieve this aim, decision makers must prioritize which health care services to offer and which to reject (Norway Priority Committee (2013)) . Health technology assessments (HTAs) are increasingly used as the evidence base for these decisions (DiCenso et al., 2005). The systematic review in HTAs provides analyses, quality assessments, and summaries of available research about the clinical effectiveness and safety of interventions, usually drawn from randomized controlled trials. The HTAs might also include additional analyses of the interventions' ethical, judicial or social consequences. Moreover, they often include a review of economic evaluations and a model of the cost-effectiveness of the intervention (Green S. et al., 2011).

When undertaking an economic evaluation as part of an HTA, health economists often construct economic models. Health economic models include effect estimates from the systematic review, baseline epidemiological data, relevant aspects of the standard course of treatment for the disease group and comparative cost analyses of the health care interventions (Drummond et al., 2015, p. 278; Knowledge Centre for the Health Services 2015). The comparative cost analyses are typically cost-effectiveness analyses and cost-benefit analyses, although the latter is more often used in areas such as the environment and transport than in health care (Culyer, 2014; Drummond et al., 2015).

Cost-effectiveness analyses compare the opportunity costs and health effects of different interventions. These analyses usually report measures in natural units (e.g. life-years gained), and effects per unit of cost (life-years gained per amount of money spent) (Culyer, 2014; Drummond et al., 2015). Cost-utility analysis, a type of cost-effectiveness analysis, expresses health effects as quality-adjusted life-years (QALYs). The QALY measure combines length of life and self-reported levels of well-being (Quality of Life), and can be used to assess the

value for money of an intervention (Culyer, 2014; Drummond et al., 2015). QALYs combine quantity and quality of life by assigning a weight from 1 (perfect health) to 0 (a state judged equivalent to death), to each year of life dependent on a person 's health related quality of life during that year (Culyer, 2014; Drummond et al., 2015; Jiménez et al., 2015). Cost-utility analyses can simultaneously capture multiple treatment outcomes, e.g., life-years gained and treatment side effects, and allow for comparisons of treatments that capture the patient's total experience.

Cost-minimization analysis, another type of cost-effectiveness analysis, is used to decide which intervention is less costly when the health effects of interventions are equal.
Elements in health economic models, such as hospital expenditures and drug costs, can vary across countries. Thus, findings from health economic evaluations conducted in one country are rarely directly transferable to the context and development of economic models in other countries or regions (Hagen et al., 2013). The economic evaluations are therefore most often used to inform the development of the economic model. In Norway, they are typically described in the discussion chapter, not the results chapter, of the HTA. Accordingly, and in contrast to the systematic review evaluating the clinical effectiveness of the interventions, it is usually unnecessary to identify and include "all" health economic evaluations in the HTA.

The effectiveness studies and health economic evaluations included in an HTA are typically identified by conducting systematic searches in databases, hand-searching journals, checking references in known relevant publications or other methods (Booth, 2010; Lefebvre et al., 2011; Lefebvre et al., 2013). The National Health Service Economic Evaluation Database (NHS EED), hosted by the Centre for Reviews and Dissemination, provides access to economic evaluations of health and social care interventions. Until 2015, NHS EED was updated weekly by searches in MEDLINE, Embase, CINAHL, PsycINFO and PubMed (Centre for Reviews and Dissemination, 2015a). However, from 2015 updating of NHS EED was discontinued. Since then the large biomedical databases, MEDLINE and Embase, have become increasingly important sources for identifying economic evaluations (McKinlay, 2006). For HTAs comparing the effectiveness of different drugs for a specific condition, searching Embase is essential due to its coverage of pharmaceutical and drug research, toxicology and pharmacology (Ovid Technologies, 2017b). Embase includes unique records as well as records from all Ovid MEDLINE journals (Elsevier, 2016). Thus, one could search

Embase only to save time and resources, if access is provided, as Embase requires subscription.

A subject search in biomedical databases such as Embase is typically developed using different aspects of a research question, often based on the PICO-mnemonic Population (e.g. children, ages 12 – 18), Intervention (e.g. HPV-vaccines), Comparator (no vaccine), and Outcome (early detection, risk reduction, and cost-effectiveness) (Lefebvre et al., 2011, 6.4). To manage the volume of records retrieved from the subject search and identify studies with specific designs, one can limit the search by adding a methodological search filter.
In the next chapter, I describe typical characteristics of methodological search filters and the published filters included in this filter validation study for identifying economic evaluations.

1.2 Methodological search filters

Methodological search filters often combine relevant free text words, standardized subject headings (index terms) assigned by indexers (EMTREE terms in Ovid Embase), and publication types (such as "randomized controlled trial" in MEDLINE) to enhance retrieval of specific study types (Glanville et al., 2009a; Lefebvre et al., 2011). Subject headings can be useful as they can provide retrieval of publications describing a topic in other words than the free text terms used by the authors. In Ovid Embase, like in many other databases, free text words can be searched for in the records title (.ti), abstract (.ab), such as "cancer. ti,ab"), or any available search fields. Restricting the free text search to title and abstract increases the chances of retrieving a more precise search result than searching in all available search fields by using the field code .mp (title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, or floating subheading word). Synonyms within the same concept are usually combined with the Boolean "OR" operator (e.g. QALY OR QALYs OR quality-adjusted-life-year OR quality of life), and the concepts of a search can be joined together with the "AND" operator (Lefebvre et al., 2011). The "NOT" operator can be used to remove records with specified terms from a search. For example, the search "human NOT animal" removes records about animals (e.g. animal studies) as well as records about both animals and humans. Based on this example, NOT should be used with care.

The performance of a methodological search filter is measured by its sensitivity, specificity and precision. Sensitivity, specificity and precision are related concepts, and recommended measurements when testing and validating methodological filters (Glanville et al., 2008; Bak et al., 2009; Harbour et al., 2014; Health Information Research Unit, 2016). Sensitivity refers to the percentage (or proportion) of relevant publications retrieved by a search. Specificity is the percentage of irrelevant records not retrieved by the search, and precision is the percentage of records retrieved that are considered relevant (Glanville et al., 2009a; Lefebvre et al., 2011).

The benefits of search filters are that they can offer consistent performance by limiting a subject search to publications that have, for example, used the same research methodology (e.g. economic evaluation), a certain age group (e.g. adolescents), or geographical area. Search filters can also save time when creating the literature search and can save record processing time by reducing the number of records retrieved. However, many methodological filters for identifying economic evaluations contain several synonyms to cover inconsistencies in terminology, often leading to overly-sensitive searches with a high proportion of irrelevant records (Wilczynski, N.L. and Haynes, R.B., 2004; Glanville et al., 2009a; Mathes, 2014). Because it is possible to describe the same type of economic analysis with a more general rather than a more specific term, e.g. cost-effectiveness vs. cost-utility analysis, labelling of the economic methods used is rarely consistent. Accordingly, authors may not describe the methods used consistently, indexers may not index the publications using correct subject headings, or sufficient index terms might not be available (Glanville et al., 2009b; Lefebvre et al., 2011; Mathes, 2014).

Glanville and colleagues (2009a) evaluated different economic evaluation filters in MEDLINE and Embase using the Ovid interface. Their objective was to develop filters to maximize sensitivity and to achieve a level of precision considered satisfactory for meeting researchers' needs for HTAs. They analyzed term occurrence using statistical methods in a random collection of gold standard records identified from the NHS EED published in year 2000, 2003 and 2006. Their analysis resulted in eight methodological MEDLINE and Embase filters respectively. The filter with the highest precision (EMBASE F: 0.494) had a low sensitivity (0.570). No filter achieved more than 0.133 in precision keeping the sensitivity above 0.900 (the combined NHS EED OR EMBASE G: 0.931 sensitivity and 0.133 precision) (Glanville et al., 2009a, p. 20 table 12). A research team at McMaster University in

Canada tested filters for identifying economic evaluations in Embase (McKinlay, 2006). Their "best specificity" filter, which is included in Glanville and colleagues' test, is also available for use in Ovid Clinical Queries as "economics (maximizes specificity)" (Ovid Technologies, 2017a). This filter had a high precision (0.237) when tested by Glanville et al. and colleagues, but a sensitivity of 0.627, which I consider somewhat low for our HTAs

In summary, some economic evaluation filters are deliberately created to be sensitive, while others succeed in improving precision but not without affecting the sensitivity. I have created a new cost-effectiveness analysis (CEA) filter in an attempt to achieve 0.10 precision combined with at least 0.90 sensitivity and a specificity of at least 0.95 in retrieval of economic evaluations to inform our HTAs.

2 Objective

The aim of this filter validation study was to identify the best performing methodological search filters to retrieve health economic evaluations from Ovid Embase. I compared my CEA filter to six published search filters for identifying economic evaluations in Embase.

2.1 Research questions

a) What is the performance of the methodological search filters CEA and selected published filters in retrieval of health economic evaluations Ovid Embase?
b) What is the performance of the CEA search filter in retrieval of health economic evaluations related to cancer treatment in Ovid Embase, when compared to published filters?

3 Methods

This study is partly based on the methods described in Glanville and colleagues' report "Development and testing of search filters to identify economic evaluations in MEDLINE and Embase"(2009a), although on a much smaller scale. Other research used in this study was mainly selected from the regularly updated sources the InterTASC Information Specialists' Sub-Group (ISSG) Search Filter Resource page "Filters to Identify Economic Evaluations", and the Summarized Research in Information Retrieval for HTA. Costs and economic evaluation (ISSG Search Filter Resource editorial team, 2017a; Kaunelis and Glanville, 2017). Both sources are regularly updated providing evidence-based information on aspects related to the development of search filters to identify economic evaluations.

I developed a gold (reference) standard set and a subject specific gold standard of known economic evaluations records (publications) that met the criteria for inclusion in the NHS EED (Centre for Reviews and Dissemination, 2015a). A gold standard is a set of known relevant records that meet specific eligibility criteria, and is determined by an extensive (sensitive) search in bibliographic databases, hand searching of journals, relative recall, or other methods (Sampson et al., 2006; Glanville et al., 2009a). Subsequently, I developed a pragmatic adaption of the NHS EED Embase filter, which was later split into the Cost-effectiveness analysis (CEA) filter and the Cost-effectiveness analysis Quality of Life (CEA QoL) filter, in the current version of Embase (years of coverage 1974 – Present, updated daily). I established the search filters' performance by testing them against the gold standard and subject specific gold standard set of economic evaluation records from the NHS EED published 2008-2013.

I analysed the search filters' performance, and the retrieved records' potential relevance to our HTAs, based on the records' title and abstracts. The screening and selection of relevant records was done independently by two health economists and myself, in order to discuss, amend (if necessary), and validate the filters' performance (Knowledge Centre for the Health Services 2015 p. 35). If a record's relevance was questioned, we included it without reading the full text of the publication. We considered cost-effectiveness analyses, cost-utility analyses and cost-benefit analyses as relevant to our context (Husereau et al., 2013). Cost-minimization analyses, which are only relevant if all important clinical effects are identical between treatment groups, and records reporting disability-adjusted life years (DALYs), were also included in the present study.

The CEA QoL filter was created to identify cost-utility analyses, and not economic evaluations in general. Because I developed the CEA QoL filter from the adapted NHS EED filter, I included it in the methods chapter but not as part of the filter testing. In order to test the CEA QoL filter I would have to develop an additional gold standard of cost-utility analysis.

3.1 Identification of gold standard

A range of methods exists to develop gold standards (Glanville et al., 2009a; Harbour et al., 2014; Frazier et al., 2015). I developed a gold standard set of economic evaluations records

(publications) from the NHS EED. The NHS EED has been an important free source for economic evaluations published internationally until updates were discontinued at the end of 2014, and is a good proxy (substitute) gold standard for available economic evaluations because of clear definitions and extensive searches in MEDLINE, Embase, CINAHL, PsycINFO and PubMed (Centre for Reviews and Dissemination, 2015a).

3.1.1 Sample size calculation

In order to establish the number of gold standard records required to achieve a robust and valid test result, I performed a sample size (power) calculation (Polit and Beck, 2012 , p. 283-285). If a sample (population) size is too small, the study will have inadequate power to reveal potential differences in the population (DiCenso et al., 2005 , p. 56; Hajian-Tilaki, 2014s, p. 195). The sample size calculation (equation below) showed that in order to achieve robust results with statistical validity within a 95 % confidence interval (CI), and a minimum of 80% sensitivity with a margin error of (±2,5%), the gold standard should consist of approximately 1300 records (population value). The CI indicates the upper and lower limits that cover the true (but unknown) population value (Polit and Beck, 2012, p. 283; Harbour et al., 2014, p. 6). Within a 99% CI (2,58 in equation) and a sensitivity of ≥ 80% (0,80 in the equation below) with a margin error of (±2,5%) (0,025 in equation), the calculation showed that the gold standard should consist of approximately 2,000 records as it is common to round up the result (see equation below).

Equation: Sample size calculation within 99% CI, 80% sensitivity and 2,5% margin error

$$n_{se} = \frac{2{,}58^2 x\ 0.80\ x\ 0.20}{0.025^2\ x\ 1} = \frac{6.6564\ x\ 0.16}{0.0006} = \frac{1.0650}{0.0006} = 1.775 \approx 2{,}000$$

My calculations corresponded with Glanville et al.'s (2009a, p. 4) which concluded that they needed a gold standard consisting of 2,070 NHS EED economic evaluation records to achieve robust results. Of these records, 1,873 corresponded to retrievable records in Ovid Embase from the publication years 2000, 2003, and 2006 (Glanville et al., 2009a, p. 18).

3.1.2 Development of NHS EEED and Embase gold standard

I identified the gold standard records by selecting the search options "NHS EED" and "CRD assessed economic evaluation (full abstract)", limited to publication years 2008-2013, in the NHS EED (CRD interface). The search retrieved 2,249 records including one duplicate (Table

1), which covered the sample size required to achieve a statistically robust result (greater than or equal to 2000) (see equation, section 3.1.1) (Centre for Reviews and Dissemination, 2015b).

Table 1: Search for gold standard records in NHS EED

Search	Hits
(*) and ((Economic evaluation:ZDT[1] and Abstract:ZPS[2])) IN NHSEED FROM 2008 TO 2013	2249

Table 1: [1]Document type (e.g. Economic Evaluation), [2]Publication status (e.g. Abstract or Bibliographic record) with a CRD economic evaluation critical assessment (abstract)

Accordingly, the final NHS EED gold standard set consisted of 2,248 unique records. I included records from complete publication years to enable the use of year limits when testing the filter's performance in Embase. Publication year 2013 was chosen as the last year of inclusion of gold standard records, because publication year 2014 in NHS EED may be incomplete due to time lags incorporating the records. The NHS EED time lag did not affect the filter testing in the current version of Embase, as the searches were limited to publication between the years 2008-2013. I downloaded the gold standard records to the reference software tool EndNote (Thomson Reuters, 2016). The records from NHS EED were assessed in accordance with the present study's eligibility criteria for economic evaluations when incorporated into the database (Centre for Reviews and Dissemination, 2015a). Thus, I concluded that additional review of the gold standard records' relevance was not necessary but my associates (HNO and KR) screened the records in order to control that all records were present.

I searched for the corresponding gold standard records in Embase by the records title or Digital Object Identifier (DOI). Ten of the 2,248 NHS EED records were not identified in Embase. The total number of gold standard records in Embase were 2,238 unique publications (2,239 records in total, including one duplicate record with identical accession number), of which 2,198 were from the publication years 2008-2013 in the version 1974 – Present, updated daily (Supplementary material I) (Figure 1, 1a). Subsequently, I downloaded the 2,198 Embase gold standard records to EndNote.

Figure 1: Development of gold standard from NHS EED and Ovid Embase

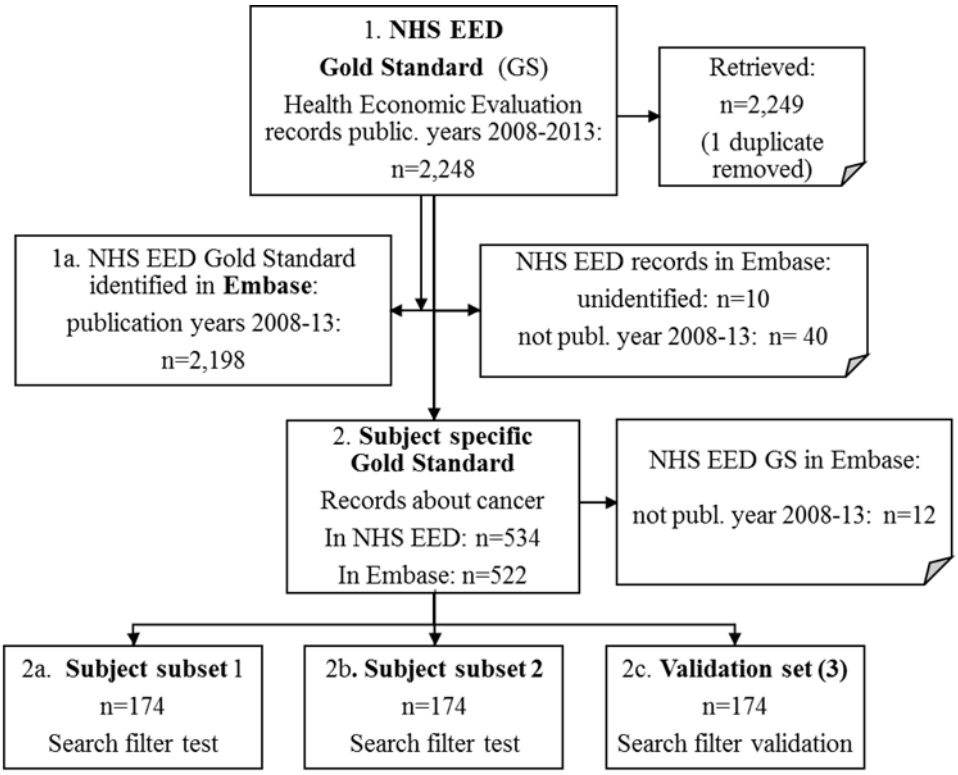


Figure 1: Subject specific gold standard (2): All search filters were tested on Subject subset 1 (2a). After amendments of the CEA and CEA QoL filter version 1, I re-tested the filters on Subject subset 1. Subsequently, all filters were tested on Subject subset 2 (2b), and Validation set (2c).

3.1.3 Subject specific gold standard

To provide an assessment of the search filters' performance (sensitivity, specificity and precision), I identified a subject specific subset gold standard set of 534 unique economic evaluation records related to cancer treatment (n=536 including 2 duplicates) from the NHS EED gold standard (Figure 1, section 2). I searched for the subject specific gold standard in NHS EED using a combination of subject headings and free text words limited to "CRD assessed economic evaluation (full abstract)" published in the years 2008 to 2013 (Table 2).

Table 2: Search for subject specific gold standard of cancer records

Line	Search	NHS EED
1	MeSH DESCRIPTOR Neoplasms EXPLODE ALL TREES	
2	(cancer* or carcinoma* or leukaemia or leukemia or lymphoma* or ur* or tumor* or neoplas* or sarcoma)	
3	#1 OR #2	
4	(*) and ((Economic evaluation:ZDT and Abstract:ZPS)) IN NHSEED M 2008 TO 2013	2249[1]
5	#3 AND #4	536[2]

[1]NHS EED gold standard including 2 duplicates. [2]Subject specific gold standard (in Embase n=534)

My associate (JG) peer reviewed the search strategy using the PRESS checklist (McGowan et al., 2016). The PRESS checklist was developed to improve the consistency in and quality of literature searches performed to inform HTAs and SRs. PRESS includes the most essential elements for creating and evaluating electronic search strategies, sufficient use of Boolean and proximity operators (e.g. AND, OR and adjacent), and relevant subject headings. Use of PRESS can identify errors in search strategies and improve the selection of search terms (Sampson et al., 2008a; Sampson et al., 2008b; McGowan et al., 2016). I identified 534 unique cancer records (536 retrieved in total, including 2 duplicates) that comprised the final subject specific gold standard. I selected the corresponding 534 unique records in Embase (535 including one duplicate), from the (already) downloaded Embase gold standard records in EndNote. Of these records, 522 records were published between the years 2008-2013 (Supplementary material II). By using the computer program QuickCalcs (GraphPad Software, 2016), my associate t (LN) randomly assigned the 522 unique cancer records into three groups, A, B and C, in three steps: (1) The computer program randomly assigned the record numbers to one of the groups. Afterwards, (2) the assignment of each record was swapped with the group assignment of a randomly chosen record number. Subsequently, (3) the 522 records were assigned into three subsets: Subject subset 1, 2 and a validation set (n=3x174) (Figure 1, section 2a-2c). I grouped and saved the subsets of cancer records in Embase in order be able to combine and test them with each of the included search filters in the database. HNO and I peer reviewed that the allocation of the records to the three subsets was in accordance with the randomization. Each subset, which contained the same number of cancer records, was used to test the methodological filters' performance.

By creating the subject subset gold standard, I was able to establish and compare the recommended standard measures for evaluating the search filter performance, sensitivity, specificity and precision, and (if preferred) number needed to read (NNR) (Glanville et al., 2008; Bak et al., 2009; Harbour et al., 2014). I could also alter the CEA filter, for example to try to improve precision, as a result of the testing (Haynes et al., 2005; Harbour et al., 2014). NNR indicates the number of retrieved records that must be read before a relevant article is revealed (Lee et al., 2012). NNR is calculated by 1 divided by precision. If a search retrieved 100 relevant records out of 1000 identified records, NNR = 1/(100/1000)=1/10=0.1, which indicates that there is 1 relevant article for every 10 records read.

3.2 Creating and selecting search filters

In an attempt to improve precision in retrieval without lowering sensitivity substantially, I adapted the NHS EED Embase filter acknowledging it as a sensitive filter. If I had simply added terms to the existing filter using OR the precision would probably have decreased as a result. I created the adapted NHS EED filter in Embase via the Ovid interface, and discussed the selected search elements relevance with my health economist colleagues. I will refer to this filter as "the adapted NHS EED filter" (Appendix I).

Subject headings also repeated as free text words from the NHS EED filter, included in the adapted NHS EED filter were:

- health economics
- economic evaluation (unexploded subject heading)
- cost-effectiveness analysis (narrower term of economic evaluation)
- cost utility analysis (narrower term of economic evaluation)
- cost benefit analysis (narrower term of economic evaluation)
- cost minimization analysis (narrower term of economic evaluation)
- pharmacoeconomics
- quality of life
- quality adjusted life year

Added free text words to the adapted NHS EED filter:

- Quality of life instruments (questionnaires):
    - Health-related quality of life instrument

- EuroQol 5D
- Short form-6 dimensions
- Health utility index
- Person trade-off
- Quality of well-being
- Standard gamble,
- Time-trade-off
- Assessment of quality of life (and abbreviations, see Appendix IV)

The search terms were selected based on our unit's experience from searching, and from analysing subject headings and free text words used in known relevant health economic evaluations (ISSG Search Filter Resource editorial team, 2017b). I chose not to explode the subject headings where this was an option, but included the selected terms one by one. By using the explode function, the search automatically includes the selected term and its associated narrower terms, although these are not visible as part of the search filter (Ovid Technologies, 2017c).

The free text words were searched for in the record's titles and abstracts, as is established practice in systematic literature searches (Lefebvre et al., 2011). After finalizing the adapted NHS EED filter, I discussed it with my health economist and any disagreements were resolved. I used the UK InterTASC Information Specialists' Sub-Group (ISSG) Search Filter Appraisal Checklist as a tool when creating the adapted NHS EED filter (Glanville et al., 2008 , table 1; Bak et al., 2009). Elements in the search filter appraisal checklist include identification of a gold standard of known relevant records, reporting of how search filter terms were identified, and the filters internal or external validity testing. Internal validity testing can be performed when the filter terms are derived from the gold standard, while external validity testing requires that filter testing is performed against records other than the records used to identify search terms. Other elements in the checklist are limitations of the search filter development, generalizability and obsolescence of search terms (ISSG Search Filter Resource editorial team, 2017b)

The first filter test on subject subset 1 (Figure 1, 2a) revealed that the adapted NHS EED filter was highly sensitive retrieving all records (n=174). Thus, I divided the filter into two parts,

the Cost-effectiveness analysis (CEA) filter and the Cost-effectiveness analysis Quality of Life (CEA QoL) filter respectively.

3.2.1 Cost-effectiveness analysis filter

The CEA filter was created to identify cost-effectiveness analyses as broadly defined, i.e. including cost-utility analyses. The CEA search filter version 1 (CEA 1) (Table 3) included subject headings and search terms such as health economics, economic evaluation, cost-effectiveness analysis and cost utility analysis, combined with the Boolean OR operator.

Table 3 Cost-effectiveness analysis (CEA) filter, version 1

LINE	SEARCH TERMS
1.	health economics/[1]
2.	economic Evaluation/[1]
3.	"cost Effectiveness Analysis"/[1]
4.	"cost Utility Analysis"/[1]
5.	"cost Benefit Analysis"/[1,2]
6.	"cost Minimization Analysis"/[1,3]
7.	pharmacoeconomics/[1,4]
8.	(health economic* or economic evaluation*).ti,ab[2]
9.	(cost* adj2 (analys* or effective* or utility or utilities or benefit* or minim*)).ti,ab[5]
10.	(cea or cua or cba).ti,ab[5]
11.	(pharmacoeconomic* or (pharmac* adj economic*)).ti,ab[5]
12.	OR/1-11[6]

[1]EMTREE terms (subject headings). [2]Express both costs and outcomes of the interventions in monetary terms. Is infrequently used to evaluate health interventions but was included because it can be used for cost-utility analysis and quality of life (QoL). [3]Describes interventions with approximately the same effect but different costs. It was included to identify publications that, in fact, are cost effectiveness analyses. [4] Labels economic evaluations of drug therapy and includes cost analysis, treatment outcome and quality of life studies".[5]Free text words searched for in title and abstract.[6] The search lines are combined with OR..

Cost-benefit analyses, i.e. analyses in which both costs and benefits are measured in monetary terms, are rarely used to evaluate health care interventions in our context. However, the subject heading was included in the CEA search filter version 1 because it is sometimes used as a key word to capture all types of cost-effectiveness analysis. Cost-minimization analysis, a type of CEA, which describes interventions with approximately the same effect but different costs, is not commonly used in our HTAs. It was included because authors sometimes use inconsistent terminology describing publications that in fact are cost-effectiveness analyses. The subject heading "Pharmacoeconomics" was included because it is used for economic evaluations of drug therapy and includes cost analysis, treatment outcome and quality of life studies. (Ovid Technologies, 2017d). I did not include the narrower subject heading "drug costs" as it usually describes global costs. Unfortunately, I have not succeeded in detecting a scope note that defines the term in Embase.

3.2.2 Cost-effectiveness analysis Quality of Life filter

The CEA QoL filter (Table 3) was created to identify cost-utility analyses reporting QALYs, and consisted of search terms describing QALYs, the concept Quality of Life (QoL), and values or instruments (questionnaires) used to measure quality of life (e.g. health-related quality of life instrument, EuroQol 5D (D=dimensions), Short form-6 dimensions, Quality of well-being, 15D). Identifying QoL values can be challenging because they are not always described in a publication's title, abstract, keywords or subject headings. QALYs might be confused with disability-adjusted life year (DALYs), a measure of overall disease burden that is usually not relevant to our HTAs. I did not include the subject heading "Quality of life index" because it was not considered relevant for the identification of cost-utility analyses.

Table 4: Cost-effectiveness analysis Quality of Life (CEA QoL) filter

LINE	SEARCH TERMS*
1.	Quality of life/[1]
2.	Quality adjusted life year/[1]
3.	(QALY or QALYs or quality-adjusted-life-year* or quality of life or utility or utilities).ti,ab[2]
4.	(15D or HRQOL3 or health-related quality of life instrument* or EuroQol 5D* or EQ-5D*).ti,ab[2,4]
5.	(SF-6D or Short form-6 dimensions or HUI or Health utility index or PTO or Person trade-off or QWB or Quality of well-being or SG or Standard gamble or TTO or Time-trade-off or AQOL or assessment of quality of life).ti,ab[2,4]
6.	OR/1-5[5]

[1]EMTREE terms (the EMTREE term "Quality of life index" was not included because I consider it too general to our use.. [2]Free text words searched for in title and abstract, [3]Health-related quality of life, includes a person's Health utility index (physical, mental and social well-being). [4]In general these instruments are not yet covered by a specific subject heading (e.g. 15D (15-dimensional), EuroQol-5D descriptive system (EQ-5D), SF-6D (Short Form-6 Dimension),[6]The search lines are combined with OR.

The CEA QoL filter differed to some extent from the rest of the included search filters in terms of its focus on retrieval of cost-utility analyses. Thus it was not tested on the gold standard, as testing of this filter required a separate gold standard of cost-utility analysis. Irrelevant publication types (e.g. letter, editorial, or note) were not excluded by the CEA filter. Using exclusion terms in a search by adding, for example the query "NOT letter or editorial or note", may cause exclusion of a relevant record that includes both the relevant and the exclusion term if both terms appears as part of the abstract search term (Glanville et al., 2009a).

3.2.3 Peer review of the CEA and CEA QoL filter

My associate (LVN), not involved in developing the adapted NHS EED filter, the CEA or the CEA QoL search filter, appraised the adapted NHS EED filter using the Peer Review of Electronic Search Strategies (PRESS) checklist (Appendix II) before it was divided into two search filters (Sampson et al., 2008a; McGowan et al., 2016). As a result of the peer review process, some terms and combinations of terms using the Ovid proximity operator adjacent (adj), were amended in the CEA filter version 1 (Table 3, line 9). The search query in the adapted NHS EED filter (Appendix I), line 13 ((pharmacoeconomic? or pharmac*) adj economic?) was changed to the more suitable way of searching for pharmacoeconomic* as a

compound word combined by the Ovid proximity operator adj (pharmac* adj economic*)) in the CEA filter (table 3, line 11).

The free text terms utility or utilities were added to the CEA QoL filter (Table 3, line 3). The proximity operator can be used in a search query to identify records with two search terms near each other, allowing for some or no words between the terms (Lefebvre et al., 2011, ch. 6.4.8).

3.2.4 Published filters

The six comparator filters in this study were EMBASE A, EMBASE F, EMBASE G, EMBASE H, McKinlay Best specificity, and the combination of NHS EED OR EMBASE G filter (Appendix 2) (Glanville et al., 2009a, p. 20 table 12). Glanville and colleagues developed the EMBASE A, EMBASE F, EMBASE G, and EMBASE H filters based on a word occurrence analysis identifying the terms best distinguishing the gold standard set of economic evaluation records from a comparator set of other economic records. They used the reference software tool EndNote's subject bibliography feature and a statistical data-led approach to analyze the results and develop the search filters (Glanville et al., 2009a, p. 4-7; Thomson Reuters, 2016). McKinlay et al. developed the McKinlay Best specificity filter (McKinlay, 2006, p. 4 table 2; Glanville et al., 2009a, p. 20 table 12), while the Centre for Reviews and Dissemination (CRD) developed the NHS EED filter (Appendix 2) to capture economic evaluations relevant to the NHS (Glanville et al., 2009a, p. 20 table 12; Centre for Reviews and Dissemination, 2014b).

The comparator filters reached the highest precision (except EMBASE G) when tested by Glanville et al., from 0.494 (EMBASE F) to 0.133 (NHS EED OR EMBASE G). I included the EMBASE G although it had a slightly lower precision (0.130), to compare its performance to the CEA filter and the combined NHS EED OR EMBASE G filter. The comparator filters' reported sensitivity varied from 0.570 (EMBASE F) to 0.931 (NHS EED OR EMBASE G, and EMBASE G). The McKinlay Best specificity filter performed the fourth-highest (sensitivity 0.627, precision 0.237) (Glanville et al., 2009a, p. 20 table 12).

3.3 Testing and amending search filter performance

3.3.1 CEA filter

The first test of the CEA filter on Subject subset 1 (n=174) identified 169 records. The CEA filter was therefore amended to improve precision by removing search terms that did to affect the search results (McGowan et al., 2016). For example, I revised the proximity operator in search line 9 from adj2 to adj (removing the opportunity of allowing one word between the search terms) (Table 5a). In search line 9, the proximity operator "adj2" allowed for one word between "cost* analys*" or "cost* effective*" etc., regardless of word order (Ovid Technologies, 2017c). I did this in order to narrow the search, as it no longer allowed for a word between the free text words. I also removed the text word abbreviation cba (cost-benefit analysis) from search line 10, because we primarily were interested in cost-effectiveness and cost-utility analyses (cea and cua).

Truncation (asterisk *) is an Ovid search command used to expand the search query by allowing for different word endings (suffixes) (Appendix 2a) (Lefebvre et al., 2011, ch. 6.4.8). The expression "cost analys*" in single or plural (or other suffix) did not focus on a specific type of cost analysis, and might result in a more sensitive search result. I removed the truncated text word "analys*", and the truncation from the remaining words, leaving the search (cost adj (effectiveness or utility or utilities or benefit or minimization)).ti,ab (Table 5a, line 9). Neither of these amendments affected the search results. Because cost-effectiveness analysis and cost-utility analysis are the typical foci of our reports, I removed "cost-benefit analysis" and cost-minimization analysis" as subject headings and free text words from search line 9 (Table 5b). This final amendment retrieved 161 of 174 records from subject subset 1, and retrieved 487 of 522 records from the subject specific gold standard (n=522).

Table 5: Cost-effectiveness analysis (CEA) search filter, final version

LINE	A) CEA FILTER VERSION 1	B) CEA FILTER FINAL VERSION
1.	health economics/[1]	health economics/[1]
2.	economic Evaluation/ [1]	economic Evaluation/[1]
3.	"cost Effectiveness Analysis"/[1]	"cost Effectiveness Analysis"/[1]
4.	"cost Utility Analysis"/[1]	"cost Utility Analysis"/[1]
5.	"cost Benefit Analysis"/[1]	pharmacoeconomics/[1]
6.	"cost Minimization Analysis"/[1]	(health economic* or economic evaluation*).ti,ab[2]
7.	pharmacoeconomics/[1]	(cost adj (effectiveness or utility or utilities)).ti,ab[2]
8.	(health economic* or economic evaluation*).ti,ab[2]	(cea or cua).ti,ab[2]
9.	(cost* adj2 (analys* or effective* or utility or utilities or benefit* or minim*)).ti,ab[2]	(pharmacoeconomic* or (pharmac* adj economic*)).ti,ab[2] OR/1-9
10.	(cea or cua or cba). ti,ab[2]	
11.	(pharmacoeconomic* or (pharmac* adj economic*)).ti,ab[2].	
12.	OR/1-11	

Table 5: [1]EMTREE terms (subject headings). [2]Free text words searched for in title and abstract.

Sensitivity

Sensitivity was defined as the proportion of relevant (gold standard) records retrieved by a search filter (equation below: a), divided by the total number of gold standard records in the subject subset (equation below: a+c) in the current edition of Embase (Glanville et al., 2009a; Lee et al., 2012; Health Information Research Unit, 2016). Sensitivity was calculated as:

$$Sensitivity = \frac{a\ (number\ of\ relevant\ records\ \ retrieved\ by\ a\ search\ filter)}{a+c\ (total\ number\ of\ relevant\ records\ in\ gold\ standard)}$$

Table 6: Procedure testing sensitivity in Embase[2]

Search 1 (S1)	Subject specific gold standard subsets publication year 2008-2013
Search 2 (S2)	Each search filter (one at the time)

Search 3 (S3)	S1 AND[1] S2 = the number of relevant records identified by a filter
Search 4 (S4)	S3 divided by S1 (the total number of subject subset records)

[1]Boolean AND operator, [2]version 1974 – Present, updated daily

To calculate the sensitivity of the search filters, I combined the subject specific subset limited to publication year 2008-2013 (S1, Table 6) with each included search filter (S2), using the Boolean AND operator (S3) in the current edition of Embase. Subsequently, I divided the relevant records (S3) by the total number of subject subset records (S1). I imported the identified records from each search to EndNote and the internet-based screening tool Rayyan (Ouzzani et al., 2016), and assessed the records' relevance in order to decide whether to amend the filters completely.

Precision

Precision was defined as the proportion of relevant records retrieved by a search filter (equation below: a), divided by the total number of relevant (a) and irrelevant (b) records retrieved by a filter (equation below: a+b) (Glanville et al., 2009a; Lee et al., 2012; Health Information Research Unit, 2016). Precision was calculated as:

$$Precision = \frac{a\ (number\ of\ relevant\ \ records\ retrieved\ by\ a\ filter)}{a+b\ (total\ number\ of\ records\ retrieved\ by\ a\ filter)}$$

Table 7: Procedure testing precision in Embase[2]

Search 1 (S1)	Sin Embase limited to publication year 2008-13
Search 2 (S2)	Each search filter (one at the time)
Search 3 (S3)	S1 AND[1] S2 = Records considered relevant identified by a filter
Search 4 (S4)	Gold standard records (n=2238)
Search 5 (S5)	S4 minus S3 = Irrelevant records identified
Search 6 (S6)	S3 / (S3 AND S5) = Precision

[1]Boolean AND operator, [2]version 1974 – Present, updated daily

I combined the cancer search query (Table 7, S1) (Appendix 3) with each search filter (S2) using the Boolean AND operator limited to publication year 2008-13 (S3). Subsequently, I deducted S3 from the gold standard (S4) using the Boolean NOT operator (S5). Finally, I divided the records considered relevant (S3) by the total number of identified records (S3/S3

AND S5). The potentially irrelevant records were not screened due to the large number of records, and thus the precision was calculated based on a rough estimate of relevance which means that the results might not represent the true precision (Waffenschmidt et al., 2016).

Specificity

Specificity is the proportion of irrelevant publications (d) that were correctly not identified by a search filter (Table 7) (Glanville et al., 2009a; Lee et al., 2012; Health Information Research Unit, 2016). Specificity was calculated as:

$$Specificity = \frac{d\ (number\ of irrelevant\ records\ not\ retrieved\ by\ a\ filter)}{b+d\ (total\ number\ of\ irrelevant\ reords)}$$

Table 8: Procedure testing specificity in Embase[2]

Search 1 (S1)	All Embase records 2008-2013
Search 2	Gold standard records (relevant records)
Search 3	S1 NOT[3] S2 = irrelevant records (true negatives)
Search 4	Search filter
Search 5	S1 AND[1] S4 = records retrieved by filter
Search 6	S2 AND S4 = relevant records retrieved by filter (true positives)
Search 7	S5 NOT S6 = false positives
Search 8	S3/ S3 AND S7

[1]Boolean AND operator, [2]version 1974 – Present, updated daily, [3]Boolean NOT operator

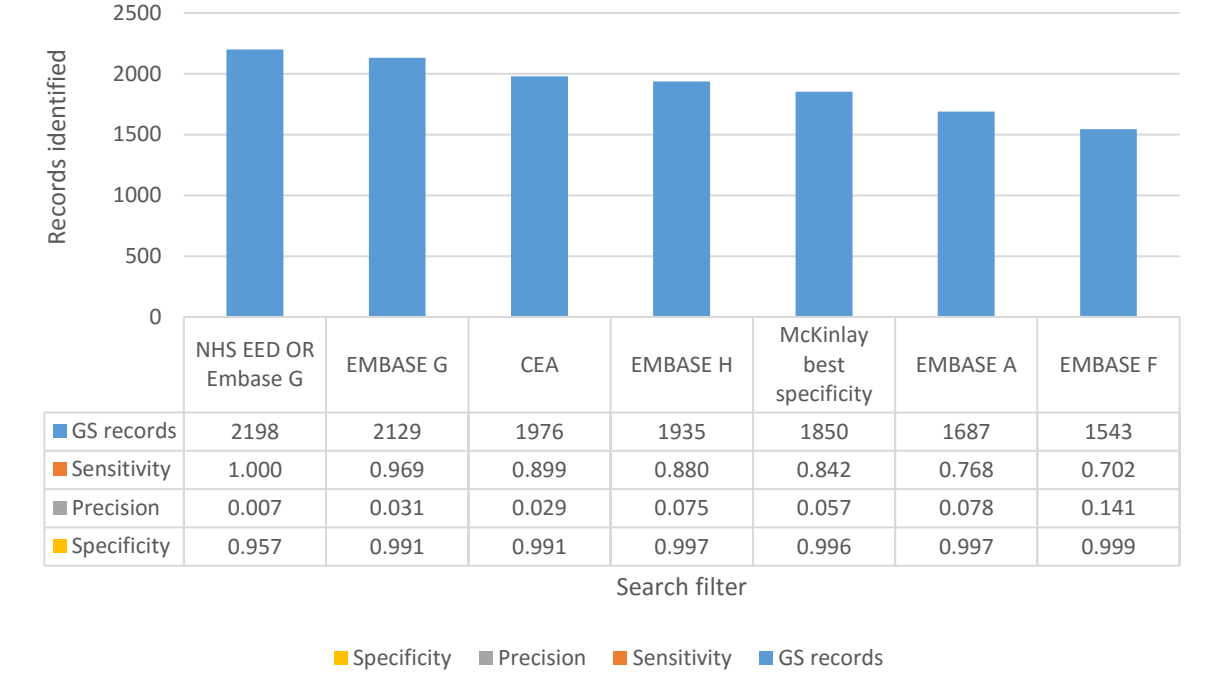
To calculate specificity I identified the total number of records in Embase publication between the years 2008-2013 (S1). Then I deducted the gold standard records (S2) from S1 to identify the irrelevant records (true negatives) (S3). Subsequently I identified the false positives (S7) and divided the true negatives (S3) by the sum of the true negatives (S3) and the false positives (S7) (Table 8).

4 Results

4.1 Testing search filter performance on Gold standard

The combined NHS EED OR EMBASE G filter (Appendix IIIA) had highest sensitivity (1.00) and lowest precision (0.007) when tested on the gold standard (n=2,198) (Fig. 2) (Supplementary material SI and SIII). The EMBASE G filter had a sensitivity of 0.969 and a precision of 0.031, whereas the CEA filter had a sensitivity of 0.899 and a precision of 0.029. The Embase H filter had a sensitivity of 0.880 and precision of 0.075, followed by the McKinlay Best specificity filter with sensitivity of 0.842 and precision of 0.057. The EMBASE A filter performed with a sensitivity of 0.768, and a precision of 0.078, while the EMBASE F filter had the lowest sensitivity (0.702) and the highest precision (0.141).
All filters had a specificity of more than 0.950. The EMBASE F had a specificity of 0.999, EMBASE A and EMBASE H, both 0.99, The McKinlay Best specificity a specificity of 0.996, while the CEA filter and the EMBASE G filter both achieved a specificity of 0.991. The NHS EED OR EMBASE G filter had the lowest specificity of 0.957.

Figure 2: Filter performance on Gold Standard[1], descending numbers of records



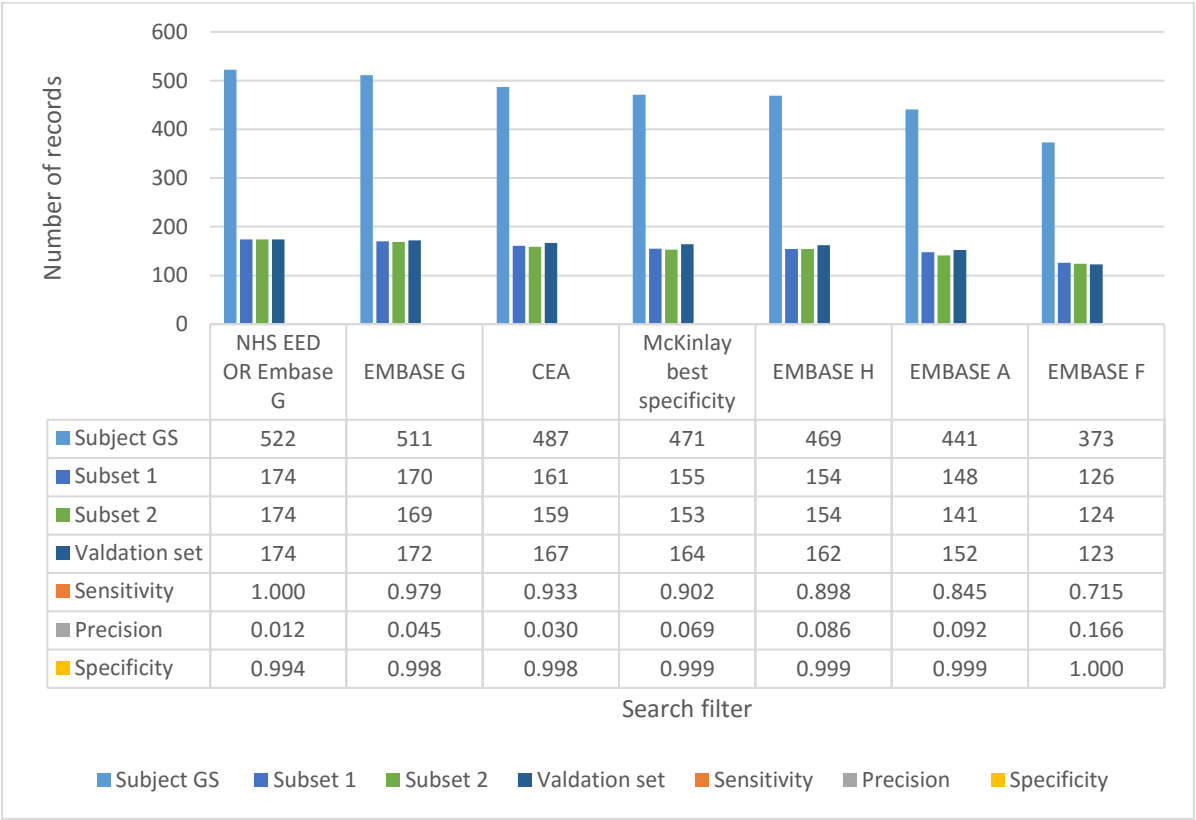
	NHS EED OR Embase G	EMBASE G	CEA	EMBASE H	McKinlay best specificity	EMBASE A	EMBASE F
GS records	2198	2129	1976	1935	1850	1687	1543
Sensitivity	1.000	0.969	0.899	0.880	0.842	0.768	0.702
Precision	0.007	0.031	0.029	0.075	0.057	0.078	0.141
Specificity	0.957	0.991	0.991	0.997	0.996	0.997	0.999

Figure 2: [1]Embase Gold standard (GS) 2008-2013 (n=2198). EMBASE A-H, McKinlay best specificity, and NHS EED OR Embase G filter (Glanville et al., 2009a). CEA=Cost-effectiveness analysis filter.

4.2 Testing search filter performance on Subject specific Gold standard

The results of the subject specific gold standard (n=522) largely corresponded to the gold standard (Fig. 3) (Supplementary material SII and SIII). The NHS EED OR EMBASE G filter had the highest sensitivity (1.00) and the lowest precision (0.012), followed by the EMBASE G with a sensitivity of 0.979 and precision of 0.045. The CEA filter's sensitivity was 0.933 and the precision 0.023. The McKinlay best specificity filter's sensitivity was 0.902 and the precision was 0.069. EMBASE H had a sensitivity of 0.898 and a precision of 0.086, while the EMBASE A filter's sensitivity was 0.845 and precision 0.092. The EMBASE F had the lowest sensitivity (0.715) and the highest precision EMBASE F (0.166), with a specificity of 1.00. The CEA and EMBASSE G filter had a specificity of 0.998, whereas the NHS EED OR EMBASE G filter's specificity was 0.994.

Figure 3: Filter performance on Subject specific Gold standard[1], descending numbers of records



	NHS EED OR Embase G	EMBASE G	CEA	McKinlay best specificity	EMBASE H	EMBASE A	EMBASE F
Subject GS	522	511	487	471	469	441	373
Subset 1	174	170	161	155	154	148	126
Subset 2	174	169	159	153	154	141	124
Valdation set	174	172	167	164	162	152	123
Sensitivity	1.000	0.979	0.933	0.902	0.898	0.845	0.715
Precision	0.012	0.045	0.030	0.069	0.086	0.092	0.166
Specificity	0.994	0.998	0.998	0.999	0.999	0.999	1.000

Figure 3: [1]Embase Subject specific Gold standard 2008-2013 (n=522). EMBASE A-H, McKinlay best specificity, and NHS EED OR Embase G filter (Glanville et al., 2009a). CEA=Cost-effectiveness analysis filter.

Each of the included search filters retrieved similar numbers of records from the two subject specific subsets and the validation set (Figure 3), although there were some differences. The EMBASE H filter retrieved, for example, retrieved a difference of nine records from the subject subset 1 and 2 and the validation set., while the McKinlay best specificity and EMBASE A filter retrieved a difference of 11 records (McKinlay best specificity 155, 153, 164 records, EMBASE A 148, 141, and 152 records). The other filters performed with a variation of zero to three records.

5 Discussion

The aim of this filter validation study was to identify the best performing methodological filters to retrieve economic evaluations to inform HTAs. I compared the performance of my CEA filter (and amended versions) to six selected search filters to identify economic evaluations, that performed with high precision when tested by Glanville et al. (2009a, p. 20 table 12). The number of records required to achieve a robust test result was approximately 2000 records (see section 3.1.1). Thus, I developed a gold standard set of 2,198 economic evaluation records assessed for inclusion in the NHS EED and published between the years 2008-2013. In this section, I present a summary of findings and discuss the strengths and weaknesses of the filter validation study.

5.1 Summary of findings

5.1.1 Filter performance

The adapted NHS EED filter identified all records from Subject subset 1. In an attempt to increase precision, and to be able to search separately for cost-utility analyses, I divided the adapted NHS EED filter into two, and created the CEA filter and the CEA QoL filter (Table 2 and 3). The CEA filter version 1 (Table 2) retrieved 165 of 174 records from publication year 2008-2013, reducing the sensitivity from 100% to 94% (0.948). Two of the 165 records were indexed with the subject heading “cost-benefit analysis” and “cost-minimization analysis”, and was not of typical interest to HTAs. Thus, I made some amendments (section 3.3.1) which retrieved 161 records from subject subset 1 (Figure 3). This amendment decreased the sensitivity of the CEA filter from 0.948 to 0.899 (Figure 2), while increasing precision from 0.019 to 0.029. Of the 13 records not identified, three reported on quality of life (QoL) or quality-adjusted life-years (QALYs), the others were cost-benefit analyses.

When tested on the gold standard (n=2198), the CEA filter's final version (Table 4) performed with a sensitivity of 0.899, a precision of 0.029, and a specificity of 0.991. The increase in precision constitute a difference of 222 gold standard records not retrieved. Compared to the specified objectives of this study, the filter preformed had with a higher sensitivity, a lower precision, but and a specificity within the objective the study the specified range. The EMBASE G filter performance was close to the CEA filter, with higher sensitivity (0.969) and precision (0.031), but the same specificity of 0.991.

Both the EMBASE H, McKinlay best specificity, EMBASE A and EMBASE F, had a lower sensitivity (EMBASE F: 0.702, McKinlay best specificity: 0.842, and EMBASE H: 0.880), and higher precision (0.057 to 0.141) than the CEA and the EMBASE G filter. The McKinlay best specificity and EMBASE H filter performed with a sensitivity above 0.800, and a precision of 0.057 and 0.075 respectively, which was higher than the objective of this filter validation study. Thus, I could use these filters if a more precise search result is required. As the filters included in this study represent a range of different levels of performance, there might not be a need for further amendment of the CEA filter.

The purpose of the CEA QoL filter was to identify cost-utility analyses and not economic evaluations in general. Unfortunately, this filter could not be tested on the NHS EED gold standard since that was developed to identify different kinds of economic evaluations. Thus, in order to test the CEA QoL filter I would have to develop a separate gold standard of cost-utility analysis. I could do this, for example, by using the relative recall method identifying relevant SRs and HTAs included cost-utility analyses reporting health state utility values (HSUV) (Golder et al., 2006; Arber, M.; et al., 2016; Waffenschmidt et al., 2016). HSUVs are used to inform QALY measures but are not only used to inform cost-utility analyses (Drummond et al., 2015)A cost-utility gold standard could also allow for comparing the CEA QoL filter's performance to the HSUV filter developed by Arber and colleagues to identify studies reporting health state utility values (HSUVs) (Arber, M. et al., 2015; Arber, M.; et al., 2016).

Additional testing and validation of relevant published filters' performance on some of our new HTA commissions, could help determine if further amendment of the CEA is needed. In order to agree on the optimal balance between sensitivity and precision in retrieval of health economic evaluations for our use, we would need to assess the full text of the records

retrieved and discuss whether further amendments would be useful. The information relevant to developing economic models is not always available in the record's abstract. This might also give additional indication regarding whether the level of precision intended to achieve was higher than what is sufficient to inform our HTAs.

5.2 Strengths of this filter validation study

5.2.1 Gold standard development

This filter validation study describes the development of a gold standard set of known relevant economic evaluation records from NHS EED published year 2008-2013 (n=2,198). The number of gold standard records was established based on a sample size calculation in order to be able to achieve robust results. The study describes a procedure for creating, amending, testing and validating search filters for identifying economic evaluations in Ovid Embase.

5.2.2 Development and testing of search filter

I created the search filter using combination of elements adapted from the published NHS EED Embase search strategy (Centre for Reviews and Dissemination, 2014a), terms suggested by health economists, and terms from the Embase thesaurus.

I compared the performance of the CEA filter and six published search filters (Appendix 2b) that performed with high precision when developed and tested by Glanville et al. (2009a). The filters were tested on the gold standard, and on a subject specific gold standard of records about cancer published in the years 2008-2013 (n=522), derived from the gold standard (Figure 1). The gold standard subject subsets that were used for testing and amending the adapted NHS EED and CEA filter were randomly allocated to two subject subsets and a validation set of gold standard records. These are design features included in the ISSG search filter checklist (Glanville et al., 2008; ISSG Search Filter Resource editorial team, 2017b).

5.2.3 Peer review

I tried to be as rigorous as possible by using peer review at key stages of the filter validation study. The search strategy for identifying the gold standard of economic evaluation records in NHS EED was checked by JG, and the downloaded gold standard set of economic evaluations records were managed in Rayyan by HNO and KR to control that all records were present.

The development of the corresponding gold standard set in Embase, and the randomization and allocation of the gold standard records into the three subject specific subsets, was peer reviewed by HNO. LNV peer reviewed the adapted NHS EED filter and the CEA filter. The peer review of the CEA QoL filter was performed before concluding that a separate gold standard of cost-utility analysis was required in order to test and amend the search filter.

5.2.4 Transparency

Tailoring a search filter to specific needs of retrieval demands transparency and clarification of the advantages and disadvantages of using a particular search filter. Achieving a good (optimal) balance between sensitivity and precision can be challenging, and researchers' opinions about what is an optimal search result might depend on factors as the scope of the search, and time and resources available for producing an HTA (Haynes et al., 2005; Booth, 2010; Lee et al., 2012). Thus, it is important to be transparent regarding the likely consequences of choosing a particular filter. I have tried to follow these principles in describing the processes in detail.

5.3 Limitations of this filter validation study

5.3.1 Creating and amending the CEA search filter

The peer review of the CEA filter was performed using the PRESS checklist and not the ISSG search filter checklist used in the development of the filter (Sampson et al., 2008b; ISSG Search Filter Resource editorial team, 2017b). Using the ISSG checklist might have revealed additional elements that could have improved the development of the search filter. For example, I could have performed a statistical analysis of free text terms and subject headings from the gold standard set of records when developing the CEA filter (internal validity, section 3.2) (Glanville et al., 2009b, p. 4-5). It might be that development of the search filter based on such methods would contribute to the development of a more precise search filter which was my aim when creating the filter.

In an attempt to increase the precision on the CEA filter, I could have repeated the process of testing and re-testing the filter against the subject specific subsets. For example, a test on subject subset 1 revealed that removing the subject headings "heath economics", "economic evaluation only reduced the number of retrieved records by 1. Using free text words only

might increase precision in retrieval. Free text words could also be convenient when performing a federated search in MEDLINE and Embase, as adaptation of subject headings from MeSH to EMTREE terms will not be necessary. This requires a sufficient choice of text words searches in adequate search fields such as title, abstract and, for example, "subject heading word" (.hw) as was done in EMBASE F and EMBASE H developed by Glanville et al. (2009a, p. 13, table 7) (Appendix III). Using the field code subject heading word (e.g. effectiveness.hw) would retrieve EMTREE subject headings that includes effectiveness. This is one amendment that could be tested in the CEA filter to see if precision could be improved. The filter testing was performed with great caution and intention to avoid bias of any sort or critical errors. Nevertheless errors might have occurred in either of the steps accomplishing this filter validation study.

6 Conclusions

6.1 Implications for practice

The process of developing a gold standard of known relevant records, and subject specific gold standard in order to test, amend and compare search filters, is described clearly and could be used by others. The CEA filter might be relevant for research librarians, researchers, or others, to use when searching for health economic evaluations in Ovid Embase to inform HTAs. The CEA filter's performance did not differ much from the performance of the EMBASE G filter, and was shown to be useful to identify economic evaluations to inform our HTAs as the majority of the records not retrieved did not sees typically relevant to our HTAs.

Both the EMBASE H, McKinlay best specificity, EMBASE A and EMBASE F, had a lower sensitivity and higher precision than the CEA and the EMBASE G filter (Table 2). The EMBASE H filter performed with a sensitivity of 0.880 and a precision of 0.075, and was closest to the objective of this filter validation study. Thus, this filter could be useful if a more precise search result than achieved by the CEA, is required. This result corresponded to the results from Glanville et al.'s test (Glanville et al., 2009a p. iv). As the filters included in this study represent a range of different levels of performance, there might not be a need for further additional amendment of the CEA filter.

6.2 Implications for research

Like other studies concludes (Glanville et al., 2009a), this filter validation study shows that developing a search filter for identifying health economic evaluations with a precision of at least 10% and a sensitivity of at least 90% is challenging. Now that Ovid MEDLINE is included in Embase, it might reveal new opportunities to develop better performing search filters although it could also result in a more sensitive retrieval, as EMTREE terms provides a more detailed indexing than the MeSH thesaurus used in MEDLINE. By analysing differences in indexing of health economic evaluations, and by performing text analyses of a set of known relevant records from the two databases, a more precise way of searching might be possible. Performing a federated search in the two databases provides access to a wider range of records through the same source.

Except from Arber et al.'s HSUV filter research (Arber, M.; et al., 2016), there are to my knowledge, few published precise search filters for identifying cost-utility analyses and health state utility values (HSUV) available. The inclusion of MEDLINE in Embase could also reveal better approaches for identifying this kind of research. In order to improve precision in retrieval without substantially reducing the sensitivity, further testing and validation of methodological search filters for cost-effectiveness analyses, cost-utility analyses and HSUVs, using a checklist such as the ISSG Search Filter Appraisal Checklist, is needed.

7 References

Arber, M., Garcia, S., Veale, T., Edwards, M., Shaw, A. and Glanville, J. (2015) *Sensitivity of a Search Filter Designed to Identify Studies Reporting Health State Utility Values. HTAi Annual Conference 2015 [Poster]* [Internett]. York: York Health Economics Consortium. Available from: <http://www.yhec.co.uk/yhec-content/uploads/2015/06/Poster-374-Sensitivity-Of-A-Search-Filter.pdf> [Read 24.04.2017].

Arber, M., Garcia, S., Veale, T., Edwards, M., Shaw, A. and Glanville, J. (2016) *PERFORMANCE OF SEARCH FILTERS TO IDENTIFY HEALTH STATE UTILITY STUDIES (PMR 183)* [Internett]. ISPOR 19th Annual European Congress The Ispor Scientific Presentations Database. Available from: <https://www.ispor.org/ScientificPresentationsDatabase/Presentation/67681> [Read 24.04.2017].

Bak, G., Mierzwinski-Urban, M., Fitzsimmons, H., Morrison, A. and Maden-Jenkins, M. (2009) A pragmatic critical appraisal instrument for search filters: introducing the CADTH CAI. *Health Info Libr J*, 26 (3), p. 211-9.

Booth, A. (2010) How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. *Int J Technol Assess Health Care*, 26 (4), p. 431-5.

Centre for Reviews and Dissemination (2014a) *Search strategies* [Internett]. York: Centre for Reviews and Dissemination Available from: <http://www.crd.york.ac.uk/crdweb/searchstrategies.asp> [Read 06.09.2015].

Centre for Reviews and Dissemination (2014b) *Search strategies. NHS EED EMBASE using OvidSP* [Internett]. York: Centre for Reviews and Dissemination Available from: <http://www.crd.york.ac.uk/crdweb/searchstrategies.asp> [Read 06.09.2015].

Centre for Reviews and Dissemination (2015a) *About the databases. About NHS EED* [Internett]. York: Centre for Reviews and Dissemination Available from: <http://www.crd.york.ac.uk/CRDWeb/AboutPage.asp> [Read 06.09.2015].

Centre for Reviews and Dissemination (2015b) *NHS Economic Evaluation Database (NHS EED)* [Internett]. York: Centre for Reviews and Dissemination Available from: <http://www.crd.york.ac.uk/crdweb/HomePage.asp> [Read 06.09.2015].

Culyer, A.J. (2014) *The Dictionary of Health Economics*. 3rd. utg. Cheltenham: Edward Elgar Publishing.

DiCenso, A., Guyatt, G. and Ciliska, D. (2005) *Evidence-based nursing: a guide to clinical practice*. St. Louis, MO: Elsevier Mosby.

Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L. and Torrance, G.W. (2015) *Methods for the economic evaluation of health care programmes*. 4th. utg. Oxford: Oxford University Press.

Elsevier (2016) *Embase Fact Sheet* [Internett]. Elsevier B.V. Available from: <https://www.elsevier.com/__data/assets/pdf_file/0016/59011/R_D_Solutions_Embase_Fact_Sheet-Web.pdf > [Read 01.09.2016].

Frazier, J.J., Stein, C.D., Tseytlin, E. and Bekhuis, T. (2015) Building a gold standard to construct search filters: a case study with biomarkers for oral cancer. *J Med Libr Assoc*, 103 (1), p. 22-30.

Glanville, J., Bayliss, S., Booth, A. and Dundar, Y., et al. (2008) So many filters, so little time: the development of a search filter appraisal checklist. *J Med Libr Assoc*, 96 (4), p. 356-61.

Glanville, J., Fleetwood, K., Yellowlees, A., Kaunelis, D. and Mensinkai, S. (2009a) *Development and testing of search filters to identify economic evaluations in MEDLINE and EMBASE*. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH).

Glanville, J., Kaunelis, D. and Mensinkai, S. (2009b) How well do search filters perform in identifying economic evaluations in MEDLINE and EMBASE. *International Journal of Technology Assessment in Health Care*, 25 (4), p. 522-9.

Golder, S., McIntosh, H.M., Duffy, S. and Glanville, J. (2006) Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Info Libr J*, 23 (1), p. 3-12.

GraphPad Software (2016) *QuickCalcs*. CA, USA: GraphPad Software, Inc.

Green S., Higgins, J.P.T., Alderson P., Clarke, M., Mulrow, C.D. and Oxman, A.D. (2011) Chapter 1: Introduction. In: Higgins, J., Green S. red. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration. Available from: <http://handbook.cochrane.org/> [Read 01.10.2015].

Guyatt, G. (2008) *Users' guides to the medical literature : a manual for evidence-based clinical practice*. 2nd ed. utg. New York: McGraw-Hill Medical.

Hagen, G., Wisløff, T. and Klemp, M. (2013) [Health economic evaluation at the Knowledge Centre for the Health Services]. *Norwegian journal of epidemiology*, Vol 23, Iss 2.

Hajian-Tilaki, K. (2014) Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform*, 48, p. 193-204.

Harbour, J., Fraser, C., Lefebvre, C., Glanville, J., Beale, S., Boachie, C., et al. (2014) Reporting methodological search filter performance comparisons: a literature review. *Health Information & Libraries Journal*, 31 (3), p. 176-194.

Haynes, R.B., McKibbon, K.A., Wilczynski, N.L., Walter, S.D. and Werre, S.R. (2005) Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *BMJ*, 330 (7501), p. 1179.

Health Information Research Unit (2016) *Hedges* [Internett]. McMaster University. Available from: <http://hiru.mcmaster.ca/hiru/HIRU_Hedges_home.aspx > [Read 14.01.2017].

Husereau, D., Drummond, M., Petrou, S. and al., e. (2013) Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health*, 16 (2), p. 010.

ISSG Search Filter Resource editorial team (2017a) *Filters to Identify Economic Evaluations* [Internett]. Available from: <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/filters-to-find-i> [Read 23.04.2017].

ISSG Search Filter Resource editorial team (2017b) *ISSG search filter checklist proforma* [Internett]. Available from: <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/issg-search-filter-checklist-proforma> [Read 23.04.2017].

Jiménez, E., Torkilseng, E. and Klemp, M. (2015) *Cost-effectiveness of HPV-vaccination of boys aged 12 in a Norwegian setting*. Oslo: Knowledge Centre for the Health Services. Available from: <http://www.kunnskapssenteret.no/en/publications/Cost-effectiveness+of+HPV-vaccination+of+boys+aged+12+in+a+Norwegian+setting> [Read 10.05.2017].

Kaunelis, D. and Glanville, J. (2017) *Summarized Research in Information Retrieval for HTA. Costs and economic evaluation* [Internett]. Available from: <http://vortal.htai.org/?q=node/336> [Read 23.04.2017].

Knowledge Centre for the Health Services (2015) *[How to summarize research. Handbook for the Knowledge Centre for the Health Services]* 4. ed. Oslo: Knowledge Centre for the Health Services. Available from: <https://www.fhi.no/globalassets/kss/filer/filer/verktoy/2015_handbok_slik_oppsummerer_vi_forskning.pdf > [Read 01.04.2017].

Lee, E., Dobbins, M., Decorby, K., McRae, L., Tirilis, D. and Husson, H. (2012) An optimal search filter for retrieving systematic reviews and meta-analyses. *BMC medical research methodology*, 12, p. 51.

Lefebvre, C., Glanville, J., Wieland, L.S., Coles, B. and Weightman, A.L. (2013) Methodological developments in searching for studies for systematic reviews: past, present and future? *Syst Rev*, 2, p. 78.

Lefebvre, C., Manheimer, E. and Glanville, J. (2011) Chapter 6: Searching for studies. In: Higgins JPT., G.S. red. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration. Available from: <http://handbook.cochrane.org/> [Read 14.01.2017].

Mathes, T.W., M.; Antoine, S. L.; Pieper, D.; Eikermann, M. (2014) Methods for systematic reviews of health economic evaluations: a systematic review, comparison, and synthesis of method literature. *Med Decis Making*, 34 (7), p. 826-40.

McGowan, J., Sampson, M., Salzwedel, D.M., Cogo, E., Foerster, V. and Lefebvre, C. (2016) PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol*, 75, p. 40-6.

McKinlay, R.J.W., N. L.; Haynes, R. B. (2006) Optimal search strategies for detecting cost and economic studies in EMBASE. *BMC Health Services Research*, 6, p. 67.

Ministry of Health and Care Services (2012) *[Good quality-safe services: quality and patient safety in the health and care services]*. Oslo: Ministries service center, information management.

Norway Priority Committee (2013) *[Open and fair. Priority setting in the health services]* Oslo: Ministry of Security and Service Organization, Information Management. Available from: <https://www.regjeringen.no/contentassets/16a0834c9c3e43fab452ae1b6d8cd3f6/no/pdfs/nou201420140012000dddpdfs.pdf> [Read 01.04.2017].

Ouzzani, M., Hammady, H., Fedorowicz, Z. and Elmagarmid, A. (2016) Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5, p. 210.

Ovid Technologies (2017a) *Embase 1974 to 2017 April* [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 04.04.2017].

Ovid Technologies (2017b) *Embase: Excerpta Medica Database Guide* [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 14.01.2017].

Ovid Technologies (2017c) *Ovid Help. Advanced Search Techniques* [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 14.01.2017].

Ovid Technologies (2017d) *Ovid Help. Search Tools* [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 14.01.2017].

Ovid Technologies (2017e) *Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R)* [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 06.04.2017].

Polit, D.F. and Beck, C.T. (2012) *Nursing research: generating and assessing evidence for nursing practice*. 9. utg. Philadelphia, PA: Wolters Kluwer Health.

Sampson, M., McGowan, J., Cogo, E., Grimshaw, J., Moher, D. and Lefebvre, C. (2008a) *PRESS: Peer Review of Electronic Search Strategies*. Ottawa, Canada. Available from: <https://www.cadth.ca/sites/default/files/pdf/477_PRESS-Peer-Review-Electronic-Search-Strategies_tr_e.pdf>

Sampson, M., McGowan, J., Cogo, E., Grimshaw, J., Moher, D. and Lefebvre, C. (2008b) *PRESS: Peer Review of Electronic Search Strategies Appendices*. Ottawa, Canada. Available from: <https://www.cadth.ca/sites/default/files/pdf/477_PRESS-Peer-Review-Electronic-Search-Strategies_tr_Appendices.pdf>

Sampson, M., Zhang, L., Morrison, A., Barrowman, N.J., Clifford, T.J., Platt, R.W., et al. (2006) An alternative to the hand searching gold standard: validating methodological search filters using relative recall. *BMC Med Res Methodol*, 6, p. 33.

Thomson Reuters (2016) *EndNote X7.5 [Computer software]*. Philadelphia (PA): Thomson Reuters.

Waffenschmidt, S., Hermanns, T., Gerber-Grote, A. and Mostardt, S. (2016) No suitable precise or optimized epidemiologic search filters were available for bibliographic databases. *J Clin Epidemiol*, 82, p. 112-118.

Wilczynski, N. and Haynes, B. (2004) Optimal search strategies for retrieving scientifically strong studies of treatment and diagnosis from MEDLINE: an analytical survey [abstract]. *12th Cochrane Colloquium: Bridging the Gaps; 2004 Oct 2-6; Ottawa, Ontario, Canada*, p. 70-71.

Wilczynski, N.L. and Haynes, R.B. (2004) Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Medicine*, 2, p. 23.

Appendix I Adapted NHS EED filter

Search filter: the adapted NHS EED[1] economic evaluations, for Ovid Embase

1.	Health economics/[1]
2.	Economic Evaluation/
3.	"Cost Effectiveness Analysis"/
4.	" Cost Benefit Analysis"/[2]
5.	"Cost Utility Analysis"/
6.	"Cost Minimization Analysis"/[3]
7.	Pharmacoeconomics/,[4]
8.	(health economic? or economic evaluation?). ti,ab[5]
9.	(cost* adj2 (analys* or benefit* or effective* or minim* or utilit*)).ti,ab
10.	cba.ti,ab
11.	cea.ti,ab
12.	cua.ti,ab
13.	((pharmacoeconomic? or pharmac*) adj economic?).ti,ab
14.	(15D or HRQoL or health-related quality of life instrument*[6] or EuroQol 5D* or EQ-5D*).ti,ab
15.	(SF-6D or Short form-6 dimensions or HUI or Health utility index or PTO or Person trade-off or QWB or Quality of well-being or SG or Standard gamble or TTO or Time-trade-off or AQOL or Assessment of quality of life). ti,ab[7]
16.	Quality of life/
17.	Quality adjusted life year/
18.	(QALY or QALYs or quality-adjusted-life-year? or quality of life).ti,ab
19.	or/1-18

Appendix 8 a: [1]EMTREE terms (subject headings). 2Express both costs and outcomes of the interventions in monetary terms, and is infrequently used to evaluate health interventions. It was included in the search filter because it can be confused with cost utility, and quality of life (QoL). [3]Describes interventions with approximately the same effect but different costs. It was included to identify publications that in fact is a cost effectiveness analysis. [4]Pharmacoeconomics labels economic evaluations of drug therapy and includes cost analysis, treatment outcome and quality of life studies".[5]Free text words searched for in title and abstract. [6]Health-related quality of life, includes a person's Health utility index (physical, mental and social well-being). [7]In general these instruments are not yet covered by a specific subject heading (e.g. 15D (15-dimensional), EuroQol-5D descriptive system (EQ-5D), SF-6D (Short Form-6 Dimension).

Appendix II: Peer review of the adapted NHS EED filter

The peer review was written in Norwegian as comments in the search strategy

Date: 31.8.2016

#	Search terms
1	Health economics/
2	Economic Evaluation/ **Peer review**: I tillegg til linje 3-6 er det to underordnede termer som ikke er tatt med i dette søket (1) «cost control» og (2) «cost of illness» **Kommentar**: Ikke aktuelle - for generelt for våre prosjekter som sammenligner kostnadseffektivitet to tiltak, ikke for andre, mer generelle (samfunns)kostnader **Peer review:** EMBASE mangler definisjoner for termer generelt, noe som gjør det litt vanskelig å vurdere omfanget av de utelatte termene. Jeg sjekket MeSH-definisjonen for de samme termene, og synes det virker rimelig å utelate disse to termene fra filteret.
3	"Cost Benefit Analysis"/ **Kommentar**: forveksles ofte med cost utility, f.eks i indeksering, benefit (nytte) også brukt i vurdering av samfunnskostnader feks samferdsel)
4	"Cost Effectiveness Analysis"/ - **Kommentar**: brukes oftest dersom metode A og B har ulik effekt
5	"Cost Minimization Analysis"/ **Kommentar**: brukes dersom metode A og B har tilnærmet lik effekt, men ulik pris
6	"Cost Utility Analysis"/ **Kommentar**: dekker QALY, leveår og evt livskvalitet
7	Pharmacoeconomics/ **Peer review:** Jeg lurer på om de underordnede termene kan være relevante her? (1)«drug costs» **Kommentar**: utenlandske priser og for generelt for våre prosjekter (2) «utilization review» **Kommenater**: for generelt, våre prosjekter sammenligner to tiltak
8	(health economic* or economic evaluation?).tw.
9	(cost* adj2 (analys* or benefit* or effective* or minim* or utilit*)).tw.
10	cba.tw.
11	cea.tw.
12	cua.tw.
13	(pharmacoeconomic? or (pharmac* adj economic*)).tw
14	(15D or HRQoL or health-related quality of life instrument* or EuroQol 5D* or EQ-5D*).tw. **Peer review:** Er det andre QoL-instrumenter som kan være relevante **Kommentar**: EQ-5D brukes i ca 80% av helseøk. eval i dag. De andre som er nevnt her ifølge våre helseøkonomers oppfatning de som brukes i dag. Peer reviewer: EMBASE har emneord for noen skjemaer som (til dels) måler QoL, og vurderer (1) «Quality of life index» **Kommentar**: for generell, lister opp instrumenter

	(2) «Short Form 36» **Kommentar**: våre helseøkonomer bruker konvertert versjon som kalles SF-6D for norske forhold som spesielt relevante. Dere har med den forkortede varianten av SF 36 som gjelder helseøkonomi, SF-6D, som tekstord. Kan hende holder det fint.
15	(SF-6D or Short form-6 dimensions or HUI or Health utility index or PTO or Person trade-off or QWB or Quality of well-being or SG or Standard gamble or TTO or Time-trade-off or AQOL or Assessment of quality of life).tw
16	Quality of life/
17	Quality adjusted life year/
18	(QALY or QALYs or quality-adjusted-life-year? or quality of life or utility).tw **Peer review**: Søke «*quality adj2 life*» for å få med synonymet 'life quality' også? **Kommentar**: Ikke vanlig å bruke ifølge vår HØK
19	or/1-18

Appendix III: Published search filters and syntax guide

A. Search filters tested by Glanville et al.[1]

Filter name	Search terms
1. EMBASE A	1. Cost adj effectiveness.ab.
2. EMBASE F	1. Cost adj effectiveness.ti. 2. (Costs.ab. and controlled study/ and cost.hw.) AND (effectiveness.hw. or randomized controlled trial/) 3. 1 OR 2
3. EMBASE G	1. Cost adj effectiveness.ab. 2. Cost adj effectiveness.ti. 3. Life adj years.ab. 4. Life adj year.ab. 5. Qaly.ab. 6. (Cost or costs).ab. and Controlled Study/ 7. (Cost and costs).ab. 8. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7
4. EMBASE H	1. Cost adj effectiveness.ti. 2. (cost adj effectiveness).ab. AND (costs or cost).ab. 3. (cost AND costs).ab. AND cost effectiveness analysis/ 4. (Costs.ab. AND controlled study/ and cost.hw[2]) AND (effectiveness.hw. OR randomized controlled trial/) 5. 1 OR 2 OR 3 OR 4
5. McKinlay Best specificity	1. Cost effectiveness.tw. OR sensitivity analys*.tw.

6. NHS EED EMBASE[2]	1. Health Economics/ 2. exp Economic Evaluation/ 3. exp Health Care Cost/ 4. pharmacoeconomics/ 5. 1 or 2 or 3 or 4 6. (econom$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic*).ti,ab. 7. (expenditure* not energy).ti,ab 8. (value adj2 money).ti,ab. 9. Budget*.ti,ab. 10. 6 or 7 or 8 or 9 11. 5 or 10 12. (metabolic adj cost).ti,ab. 13. ((energy or oxygen) adj cost).ti,ab. 14. ((energy or oxygen) adj expenditure).ti,ab. 15. 12 or 13 or 14 16. 11 not 15

[1](Glanville et al., 2009a), for search filed codes (e.g. .ti, .ab, .hw) see Appendix III B: Syntax guide, [2]The search filter is reported without exclusion strategy and doesn't include costs: letter.pt; editorial.pt; note.pt (Centre for Reviews and Dissemination, 2014)

III B: Syntax guide

Code	Description
/ (slash)	Subject heading (e.g. Economic Evaluation /)
.ab.	Abstract
.fs.	Floating subheading
.hw.	Heading Word - retrieves every EMTREE subject heading that includes a particular word
.mp.	Title, abstract and indexing
.pt.	Publication Type
.sh.	Subject heading
.ti.	Title.
tw,.	Textword
.yr.	Publication year
* (asterisk)	Truncation symbol when following a term (e.g. economic*)
?	Indicates there can be zero or one additional characters (e.g. cost? Identifies cost and also costs)
Adj	Adjacent terms
Adj1	Terms within one word of each other
Exp	Explode (subject heading)
Or/1-3	Combine sets 1 to 3 using OR (1 OR 2 OR 3)

Article manuscript

Identifying the best performing search filters to retrieve health economic evaluations in Embase: A filter validation study

Master's student: Ingrid Harboe

Master's program: Master in Evidence-Based Practice for Health and Social Sciences

Faculty: Faculty of Health and Social Sciences

Institution: Bergen University College

Intended journal: Journal of clinical epidemiology

Number of words: 4,999

Abstract

Objective: To identify the best performing of seven search filters to retrieve health economic evaluations used to inform health technology assessments (HTAs), by comparing the cost-effectiveness analysis (CEA) filter to six published filters in Ovid Embase, and achieve a sensitivity of at least 0.90 with a precision of 0.10, and specificity of at least 0.95.

Study Design and Setting

HTAs are increasingly used by Norwegian health authorities as the evidence base when prioritizing which health care services to offer. In this filter validation study, the included filters' performances were compared against a gold standard of economic evaluations published in 2008-2013 (n=2,248) from the National Health Service Economic Evaluation Database (NHS EED), and the corresponding records (n=2,198) in the current version of Ovid Embase.

Results: The CEA filter had a sensitivity of 0.899 and precision of 0.029. One filter had a sensitivity of 0.880 and a precision of 0.075, and was closets to the objective. The filter with lowest sensitivity (0.702) had a precision of 0.141.

Conclusion: Developing search filters for identifying health economic evaluations, with a good balance between sensitivity and precision, is challenging. Researchers should agree on acceptable levels of performance before concluding on which search filter to use.

Keywords: Information storage and retrieval (MeSH); Databases, Bibliographic (MeSH); Sensitivity and specificity (MeSH); Cost-benefit analysis (MeSH);

Title: Identifying the best performing methodological search filters to retrieve health economic evaluations in Embase: a validation study.

Word count: 200 words

1 Introduction

Decision makers increasingly use health economic evaluations to prioritize health care interventions at a national level (Norway Priority Committee (2013); Knowledge Centre for the Health Services 2015). The Norwegian health authorities aim to provide high quality health care services to all citizens and must prioritize which health care services to offer (Norway Priority Committee (2013)). Health technology assessments (HTAs) are increasingly used as the evidence base for these decisions (DiCenso et al., 2005; Guyatt, 2008; Ministry of Health and Care Services, 2012). HTAs typically consist of a systematic review (SR) of the effects and safety of two or more health care interventions (e.g. drugs, medical devices, or organization of health services), and an economic evaluation of the interventions (Hagen et al., 2013). Health economic evaluations can be published as part of a health technology assessment HTA or as separate articles.

Systematic literature searches are performed in biomedical or other relevant databases to inform SRs and HTAs with high quality research (Green S. et al., 2011; Lefebvre et al., 2011). Databases like the National Health Service Economic Evaluation Database (NHS EED), MEDLINE and Embase are typical sources to identify health economic evaluations of health care interventions (Centre for Reviews and Dissemination, 2015a; Ovid Technologies, 2017a; Ovid Technologies, 2017e). The NHS EED has been an important free source for this kind of research evidence but since updates of NHS EED were discontinued in 2014, MEDLINE and Embase have become increasingly important sources. (Glanville et al., 2009a; Harbour et al., 2014; Centre for Reviews and Dissemination, 2015a).

For HTAs comparing the effectiveness of different drugs for a specific condition, searching Embase is essential due to its coverage of pharmaceutical and drug research, toxicology and pharmacology (Ovid Technologies, 2017b). Embase, which is available by subscription, includes unique records as well as records from all Ovid MEDLINE journals (Lefebvre et al., 2011; Elsevier, 2016).

To identify economic evaluations in MEDLINE and Embase, one can limit the subject search by adding a methodological search filter (search strategy), comprising specific methodological terms to identify publications using certain study designs (Haynes et al., 2005; Lefebvre et al., 2011). Despite their potential to identify health economic evaluations

more efficiently, published methodological filters are often sensitive and retrieve many publications that are irrelevant to the context of the HTA (Wilczynski, N. and Haynes, B., 2004; Glanville et al., 2009a; Mathes, 2014). Health economists generally use the retrieved economic evaluations merely to inform the development of an economic model used to assess the cost-effectiveness of the intervention of interest (Culyer, 2014; Drummond et al., 2015). Elements in economic models (e.g. hospital expenditures and drug costs) can vary across countries. Thus, economic evaluations are rarely directly transferrable to other contexts, and results from retrieved economic evaluations are often only included in the discussion chapter of the HTA (Hagen et al., 2013). Subsequently, in the Norwegian context, making a sensitive literature search to identify "all" health economic evaluations is not always required.

1.1 Methodological search filters

Methodological search filters often combine relevant free text words searched for in the records title, subject headings (EMTREE terms in Embase) and publication types to enhance retrieval of different study types (Glanville et al., 2009a; Lefebvre et al., 2011).

Search filters can offer consistent performance by limiting a subject search to publications that have, for example, used the same research methodology (e.g. economic evaluation), or a specified age group. Search filters can also save time creating the literature search, and can save record processing time by reducing the number of records retrieved. Many economic evaluation search filters contain several synonyms to cover inconsistencies in terminology, often leading to overly-sensitive searches with a high proportion of irrelevant records (Glanville et al., 2009a; Lefebvre et al., 2011; Mathes, 2014). Accordingly, authors may not describe the methods used consistently and database indexers may not index the publications using correct subject headings (Glanville et al., 2009b; Lefebvre et al., 2011; Mathes, 2014).

The performance of a methodological search filter is measured by its sensitivity, specificity and precision, which are recommended measurements when testing and validating methodological filters (Bak et al., 2009; Harbour et al., 2014; Health Information Research Unit, 2016). Sensitivity refers to the percentage (or proportion) of relevant publications retrieved by a search. Specificity is the percentage of irrelevant records not retrieved by the search, and precision is the percentage of records retrieved that are considered relevant (Glanville et al., 2009a; Lefebvre et al., 2011).

Glanville and colleagues (2009a) developed and evaluated eight economic evaluation filters in MEDLINE and Embase using the Ovid interface, which were compared to other published search filters. Their goal was to maximize sensitivity and achieve a satisfactory precision to meet researchers' needs when composing HTAs. The EMBASE F filter had the highest precision (0.494) and a sensitivity of 0.570, no filter achieved precision greater than 0.133 while maintaining a sensitivity above 0.900 (Glanville et al., 2009a). To meet our health economists' need of increasing the precision in retrieval of economic evaluations to inform our HTAs, I created a new cost-effectiveness analysis filter (CEA).

2 Objective

The aim of this filter validation study was to identify the best performing methodological search filters to retrieve health economic evaluations from Ovid Embase. I compared the CEA filter to six published search filters for identifying economic evaluations attempting to achieve a sensitivity of at least 0.90, a precision of at least 0.10 and specificity of at least 0.95.

3 Methods

3.1 Identification of gold standard

This filter validation study was partly based on the methods used by Glanville et al. (2009a). In order to establish the search filters' performance, they developed a gold standard consisting of 2,070 NHS EED economic evaluation records, and the corresponding 1,873 retrievable records in Ovid Embase from the publication years 2000, 2003, and 2006 (Glanville et al., 2009a). A gold standard is a set of relevant records that meet specific eligibility criteria, and is determined by an extensive (sensitive) search in bibliographic databases, hand searching of journals, relative recall, or other methods (Sampson et al., 2006; Glanville et al., 2009a; Harbour et al., 2014; Frazier et al., 2015). The gold standard set of NHS EED records is a good proxy gold standard for available economic evaluations, because of NHS EED's clear definitions and extensive searches in MEDLINE, Embase, CINAHL, PsycINFO and PubMed (Sampson et al., 2006; Centre for Reviews and Dissemination, 2015a; Frazier et al., 2015). Other research used in this study was mainly selected from the regularly updated sources the InterTASC Information Specialists' Sub-Group (ISSG) Search Filter Resource page "Filters to Identify Economic Evaluations", and the Summarized Research in Information Retrieval for HTA. Costs and economic evaluation (ISSG Search Filter Resource editorial team, 2017a; Kaunelis and Glanville, 2017).

3.2 Sample size calculation

In order to establish the number of gold standard records required to achieve a robust test result with statistical validity, I performed a sample size calculation (DiCenso et al., 2005; Polit and Beck, 2012; Hajian-Tilaki, 2014). The calculation (Appendix A) showed that a robust result within a 99% confidence interval (CI), and a sensitivity of at least 80% with a margin error of (±2,5%),required a gold standard consisting of approximately 2000 records (Polit and Beck, 2012; Harbour et al., 2014). The CI indicates the upper and lower limits that cover the true (but unknown) population. This result is equivalent to Glanville et al.'s gold standard (2009a).

3.3 Development of NHS EEED and Embase gold standards

I developed a gold (reference) standard set and a subject specific gold standard set of economic evaluation records (publications) that met the criteria for inclusion in the NHS EED (Centre for Reviews and Dissemination, 2015a). I identified the gold standard records of NHS EED assessed economic evaluation (full abstract) limited to publication between the years 2008-2013, in the NHS EED interface, which retrieved 2,248 unique records (2,249 including one duplicate) (Appendix B) (Fig. 1). Complete publication years were chosen to enable the use of year limits when testing the filter's performance in Embase, and 2013 was chosen as the last year because publication year 2014 in NHS EED may be incomplete due to time lags incorporating the records.

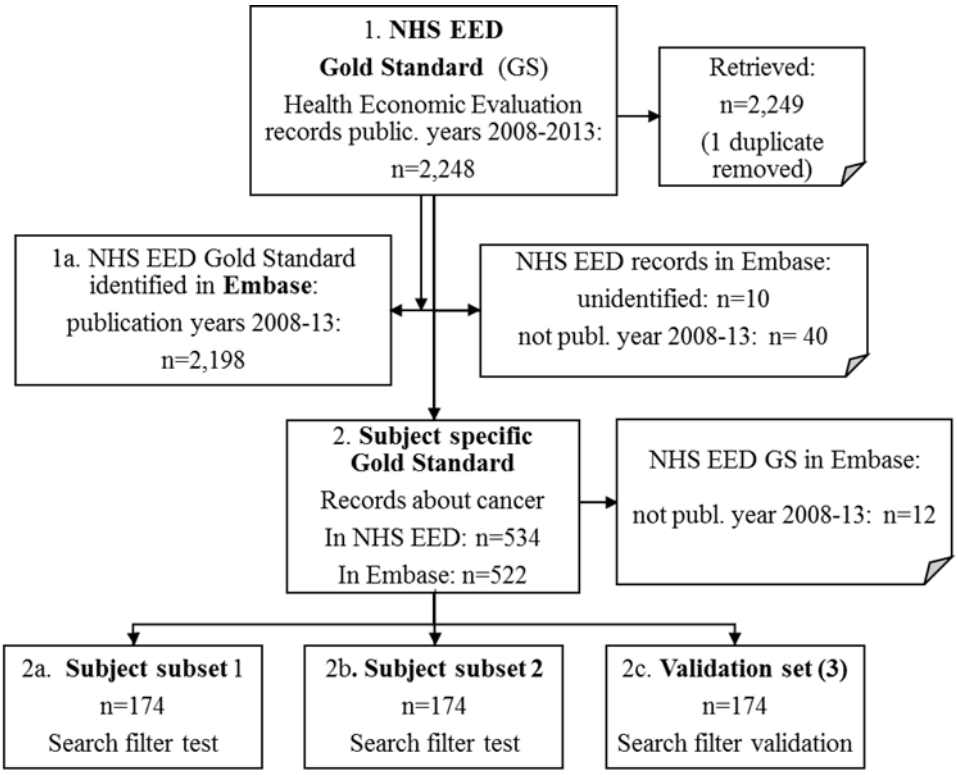


Fig. 1: Development of gold standard using NHS EED and Ovid Embase
All search filters were tested on Subject subset 1 (2a). After amendments of the CEA version 1 filter, the filter was re-tested on Subject subset 1. Subsequently, all filters were tested on Subject subset 2 (2b), and the Validation set (2c).

I downloaded the retrieved records to EndNote (Thomson Reuters, 2016). I searched for the corresponding gold standard records in Embase using the records title or Digital Object Identifier (DOI), and downloaded 2,198 unique records published between the years 2008-2013 in the version 1974 – Present, updated daily (Fig. 1, 1a). The downloaded Embase gold standard set was managed in the screening tool Rayyan (Ouzzani et al., 2016).

3.3.1 Subject specific gold standard

The recommended standard measures for evaluating search filters' performance are sensitivity, specificity and precision, and (if preferred) number needed to read (NNR) (Glanville et al., 2008; Bak et al., 2009; Harbour et al., 2014). To establish the search filter performance, I identified a subject specific subset gold standard set of economic evaluation records related to cancer treatment from the NHS EED gold standard (n=534) (Appendix C) (Fig. 1, 2.). NNR is defined as the inverse of precision and indicates how many retrieved records that must be read before a relevant article is revealed (Lee et al., 2012). If a search

retrieved 100 relevant records out of 1000 identified records, NNR = 1/(100/1000) = 1/0.1 = 10, (10 records read needed to find 1 relevant article (Lee et al., 2012). I selected the corresponding 534 unique records in Embase, of which 522 records were published between the years 2008-2013. Using QuickCalcs (GraphPad Software, 2016), my associate (LN) randomly assigned the 522 unique cancer records into subject subset 1 and 2, and a validation set (n=3x174) (Fig. 1, section 2a-2c). Subsequently, my associate (HNO) checked that the allocation of the records was in accordance with the randomization. Each subject subset was used to test the search filters' performance, and alter the CEA filter attempting to improve precision (Haynes et al., 2005; Harbour et al., 2014).

3.4 Creating and selecting search filters

In an attempt to improve the precision in accordance with the objective (section 2), I adapted the NHS EED Embase filter (Centre for Reviews and Dissemination, 2014b) in Embase via the Ovid interface, acknowledging it as a sensitive filter. I will refer to this filter as the "adapted NHS EED filter" (Appendix D.1). Search terms were selected in consultation with health economists based on my experience performing searches, and from by analysing subject headings and free text words used in known relevant health economic evaluations (ISSG Search Filter Resource editorial team, 2017b).

Subject headings also repeated as free text words from the NHS EED filter, included in the adapted NHS EED filter:

- health economics
- economic evaluation (unexploded subject heading)
- cost-effectiveness analysis (narrower term of economic evaluation)
- cost utility analysis (narrower term of economic evaluation)
- cost benefit analysis (narrower term of economic evaluation)
- cost minimization analysis (narrower term of economic evaluation)
- pharmacoeconomics

The free text words were searched for in the records' titles and abstracts as is customary in order to retrieve a more precise search result (Lefebvre et al., 2011). I used the ISSG Search Filter Appraisal Checklist when creating the adapted NHS EED filter (Glanville et al., 2008; ISSG Search Filter Resource editorial team, 2017b). Elements in the Search Filter Appraisal

Checklist include identification of a gold standard of known relevant records, reporting of identification of search filter terms, and the search filters internal (filter terms derived from the gold standard) or external validity testing (filter testing against records other than the records used to identify search terms). Other elements were limitations of the search filter development, generalizability, and obsolescence of search terms (ISSG Search Filter Resource editorial team, 2017b).

The first filter test on Subject subset 1 (section 3.3, Fig. 1, 2a) revealed that the adapted NHS EED filter was highly sensitive retrieving all records (n=174). Thus, I divided the filter into two filters, the Cost-effectiveness analysis (CEA) filter (Appendix D.2) and the Cost-effectiveness analysis Quality of Life (CEA QoL) (Appendix D.3) filter respectively. The CEA filter was created to identify cost-effectiveness analyses as broadly defined, i.e., including cost-utility analyses, whereas the CEA Quality of Life (CEA QoL) filter was created to identify cost-utility analyses reporting quality-adjusted life-years (QALYs), and not economic evaluations in general. The CEA QoL filter differed to some extent from the rest of the included search filters in terms of its focus on retrieval of cost-utility analyses. Testing of the CEA QoL filter required a separate gold standard of cost-utility analysis.

3.4.1 Cost-effectiveness analysis filter

The CEA filter version 1 (CEA 1) (Appendix D.2), was created to identify cost-effectiveness analyses as broadly defined, i.e., including cost-utility analyses, and included relevant subject headings and free text terms combined with the Boolean OR operator. Cost-benefit analyses are rarely used to evaluate health care interventions in our context. The subject heading was included in the CEA 1 filter because it is sometimes used as a key word (and medical subject heading (MeSH) used in MEDLINE)) to capture all types of cost-effectiveness analysis. Cost-minimization analysis, a type of cost-effectiveness analyses, is not commonly used in our HTAs but was included because authors or indexers sometimes mislabel publications that in fact are cost effectiveness analyses. The subject heading “pharmacoeconomics” was included because it is used for economic evaluations of drug therapy and includes cost analysis, treatment outcome and quality of life studies. (Ovid Technologies, 2017d). Exclude terms such as irrelevant publication types (e.g. letter, editorial, or note) were not included in either of the filters (Glanville et al., 2009a).

3.4.2 Peer review of the CEA and CEA QoL filter

My associate (LVN) who was not involved in developing the search filters, appraised them using the Peer Review of Electronic Search Strategies (PRESS) checklist (McGowan et al., 2016). The PRESS checklist was developed to improve the consistency in, and quality of, literature searches performed to inform HTAs and SRs. PRESS includes the most essential elements for creating and evaluating electronic search strategies, sufficient use of Boolean and proximity operators, and relevant subject headings (Sampson et al., 2008a; McGowan et al., 2016). As a result of the peer review process, some terms and combinations of terms using the Ovid proximity operator adjacent (adj), were amended in the CEA 1 filter (Lefebvre et al., 2011).

3.4.3 Published filters

The six comparator filters were EMBASE A, EMBASE F, EMBASE G, EMBASE H, McKinlay Best specificity, and the combination of NHS EED OR EMBASE G filter (Appendix E.1) (Glanville et al., 2009a). The EMBASE A, F, G, and H filters, were developed based on a word occurrence analysis identifying the terms best distinguishing the gold standard set of economic evaluation records from a comparator set of other economic records. The included filters were among those achieving the highest precision when tested by Glanville et al, ranging from 0.494 (EMBASE F) with a sensitivity of 0.570, to 0.130 (EMBASE G) with a sensitivity of 0.931. The McKinlay Best specificity filter performed the fourth-highest (sensitivity 0.627, precision 0.237) (Glanville et al., 2009a).

3.5 Testing and amending search filter performance

To assess the search filters' performance and alter the CEA 1 filter if needed, I tested the filters' retrieval on subject subset 1 in the current edition of Embase limited to publication between the years 2008-2013. I calculated the sensitivity, specificity and precision using a 2 x 2 table (Table 1) (Health Information Research Unit, 2016). I imported the identified records from each search to EndNote and Rayyan, and we assessed the records' relevance in order to decide whether to amend the CEA 1 filter.

		Relevant records (true positives)	Irrelevant records (true negatives)	Total
Search filter	Retrieved	a	b	a + b
	Not retrieved	c	d	c + d
	Total	a + c	b + d	a+b+c+d

Table 1. Testing search filter performance

Sensitivity=a/(a+c), Precision=a/(a+b), Specificity=d/(b+d), all records retr.=(a+b+c+d)

Sensitivity was defined as the proportion of relevant records retrieved by a search filter (a) divided by the total number of relevant records (a+c) (Table 1, Fig. 2 and Fig. 3) (Lee et al., 2012; Health Information Research Unit, 2016). Precision was defined as the proportion of relevant records (true positives) retrieved by a search filter (a) (Table 1, Fig. 2 and Fig. 3), divided by the total number of relevant and irrelevant records retrieved (a+b). The potentially irrelevant records retrieved were not screened due to the large number of records in Embase published between the years 2008-2013 (n=7,172,075). Thus, precision was calculated based on a rough estimate of relevance which means that the results might not represent the true precision (Waffenschmidt et al., 2016). Specificity was defined as the proportion of irrelevant publications (d) that were correctly not found by a search filter (Lee et al., 2012; Health Information Research Unit, 2016).

3.5.1 CEA filter amendment

The CEA 1 filter performed with a sensitivity of (sensitivity of 0.971) on subject subset 1 (169/174 records). As we considered this as being too sensitive, I amended the filter in order to improve precision (McGowan et al., 2016). For instance, I replaced the proximity operator adj2 with adj, removing the opportunity of allowing one word between the search terms no matter in which word order (Table 2a, line 9) (Ovid Technologies, 2017c).

LINE	A. FILTER VERSION 1	B. FILTER FINAL VERSION
1.	health economics/[1]	health economics/[1]
2.	economic Evaluation/ [1]	economic Evaluation/[1]
3.	"cost Effectiveness Analysis"/[1]	"cost Effectiveness Analysis"/[1]
4.	"cost Utility Analysis"/[1]	"cost Utility Analysis"/[1]
5.	"cost Benefit Analysis"/[1]	pharmacoeconomics/[1]
6.	"cost Minimization Analysis"/[1]	(health economic* or economic evaluation*).ti,ab[2]
7.	pharmacoeconomics/[1]	(cost adj (effectiveness or utility or utilities)).ti,ab
8.	(health economic* or economic evaluation*).ti,ab[2]	(cea or cua).ti,ab[2]
9.	(cost* adj2 (analys* or effective* or utility or utilities or benefit* or minim*)).ti,ab[2]	(pharmacoeconomic* or (pharmac* adj economic*)).ti,ab[2] OR/1-9
10.	(cea or cua or cba). ti,ab[2]	
11.	(pharmacoeconomic* or (pharmac* adj economic*)).ti,ab[2].	
12.	OR/1-11	

Table 2: Cost-effectiveness analysis (CEA) search filter, version 1 and final version

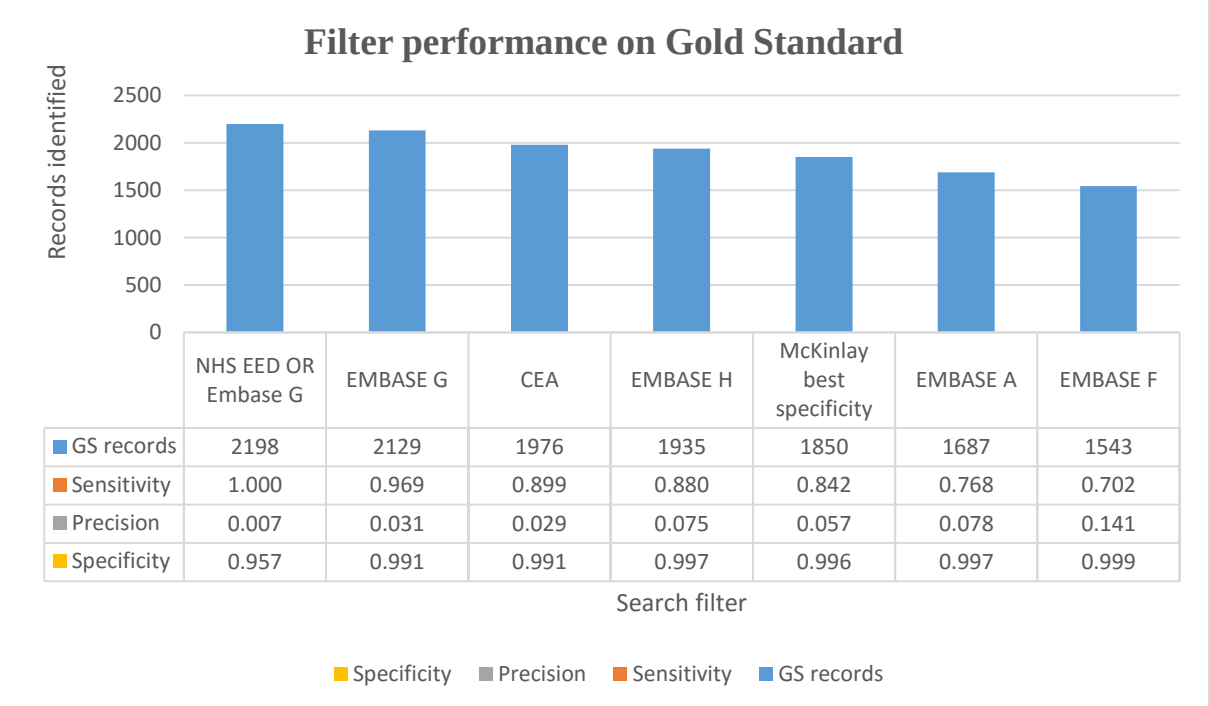
[1]EMTREE terms (subject headings). [2]Free text words searched for in title and abstract.

I removed the truncated text word "analys*" and the truncation from the remaining words, leaving the search (cost adj (effectiveness or utility or utilities or benefit or minimization)).ti,ab. Truncation (asterisk *) is an Ovid search command used to expand the search query by allowing for different word endings (suffixes) (Lefebvre et al., 2011). Neither of these amendments affected the search result on subject subset 1. As cost-effectiveness analysis and cost-utility analysis were the typical focus of our reports, I removed "cost-benefit analysis" and cost-minimization analysis" as subject headings and free text words in title or abstract from search line This amendment retrieved 161 records (sensitivity 0.925) from subject subset 1 (n=174), and 487 records (sensitivity 0.933) from the subject specific gold standard (n=522). Some of these records were potentially relevant to our HTAs, and thus we agreed that this was a satisfactory result, knowing that additional amendment might cause missing some relevant records. I tested the CEA filter on subset 2 and the validation set.

4 Results

4.1 Search filter performance on Gold standard

The combined NHS EED OR EMBASE G filter (Appendix E1) had highest sensitivity (1.00) and lowest precision (0.007) when tested on the gold standard (n=2198) (Fig. 2) (Supplementary material SI and SIII). The EMBASE G filter had a sensitivity of 0.969 and a precision of 0.031, while the CEA filter had a sensitivity of 0.899 and a precision of 0.029. The Embase H filter had a sensitivity of 0.880 and precision of 0.075, followed by the McKinlay Best specificity filter with sensitivity of 0.842 and precision of 0.057. The EMBASE A filter performed with a sensitivity of 0.768, and a precision of 0.078, while the EMBASE F filter had the lowest sensitivity (0.702) and the highest precision (0.141). All filters had a specificity of more than 0.950. The EMBASE F had a specificity of 0.999, EMBASE A and EMBASE H, both 0.99, The McKinlay Best specificity a specificity of 0.996, while the CEA filter and the EMBASE G filter both achieved a specificity of 0.991. The NHS EED OR EMBASE G filter had the lowest specificity of 0.957.



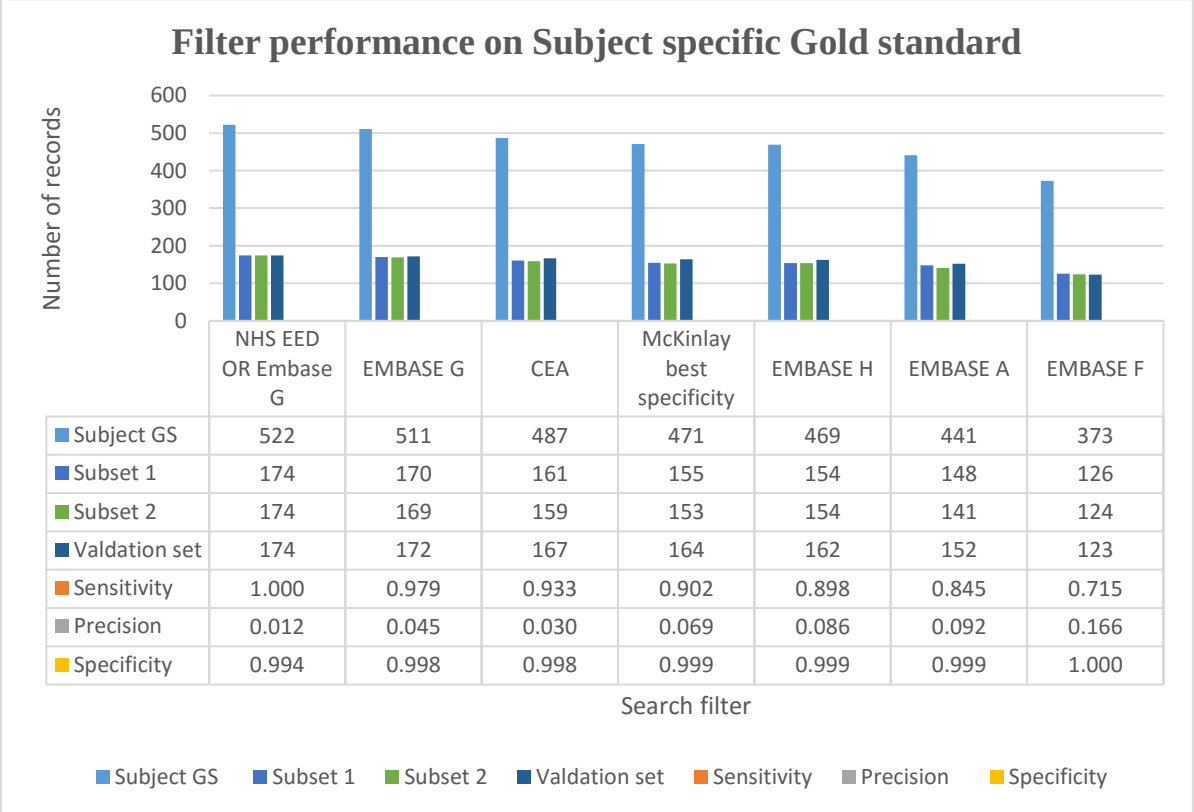
	NHS EED OR Embase G	EMBASE G	CEA	EMBASE H	McKinlay best specificity	EMBASE A	EMBASE F
GS records	2198	2129	1976	1935	1850	1687	1543
Sensitivity	1.000	0.969	0.899	0.880	0.842	0.768	0.702
Precision	0.007	0.031	0.029	0.075	0.057	0.078	0.141
Specificity	0.957	0.991	0.991	0.997	0.996	0.997	0.999

Fig. 2: Filter performance on Gold Standard[1], descending numbers of records

[1]n=2198. EMBASE A-H, McKinlay best specificity, and NHS EED OR Embase G filter (Glanville et al., 2009a). CEA=Cost-effectiveness analysis filter.

4.2 Search filter performance on Subject specific Gold standard

The results of the subject specific gold standard (n=522) largely corresponded to the gold standard (Fig. 3) (Supplementary material SII and SIII). The NHS EED OR EMBASE G filter had the highest sensitivity (1.00) and the lowest precision (0.012), followed by the EMBASE G with a sensitivity of 0.979 and precision of 0.045. The CEA filter's sensitivity was 0.933 and the precision 0.023. The McKinlay best specificity filter's sensitivity was 0.902 and the precision was 0.069. EMBASE H had a sensitivity of 0.898 and a precision of 0.086, while the EMBASE A filter's sensitivity was 0.845 and precision 0.092. The EMBASE F had the lowest sensitivity (0.715) and the highest precision EMBASE F (0.166), with a specificity of 1.00. The CEA and EMBASSE G filter had a specificity of 0.998, whereas the NHS EED OR EMBASE G filter's specificity was 0.994.



	NHS EED OR Embase G	EMBASE G	CEA	McKinlay best specificity	EMBASE H	EMBASE A	EMBASE F
Subject GS	522	511	487	471	469	441	373
Subset 1	174	170	161	155	154	148	126
Subset 2	174	169	159	153	154	141	124
Valdation set	174	172	167	164	162	152	123
Sensitivity	1.000	0.979	0.933	0.902	0.898	0.845	0.715
Precision	0.012	0.045	0.030	0.069	0.086	0.092	0.166
Specificity	0.994	0.998	0.998	0.999	0.999	0.999	1.000

Fig. 3: Filter performance on Subject specific Gold standard[1], descending numbers of records

[1]Embase Subject specific Gold standard 2008-2013 (n=522). EMBASE A-H, McKinlay best specificity, and NHS EED OR Embase G filter (Glanville et al., 2009a). CEA=Cost-effectiveness analysis filter.

Each of the included search filters retrieved similar numbers of records from the two subject specific subsets and the validation set (Fig. 3), although there were some differences. The EMBASE H filter, for example, retrieved a difference of nine records from the subject subset 1 and 2 and the validation set., while the McKinlay best specificity and EMBASE A filter retrieved a difference of 11 records (McKinlay best specificity 155, 153, 164 records, EMBASE A 148, 141, and 152 records). The other filters performed with a variation of zero to three records.

5 Discussion

The aim of this filter validation study was to identify the best performing methodological filters to retrieve economic evaluations to inform HTAs, by comparing the performance of the CEA filter (and amended versions) to six selected search filters to identify economic evaluations (Glanville et al., 2009a). The number of records required to achieve a robust test result was approximately 2000 records (section 3.1.1). Accordingly, I developed a gold standard set of n=2,198 economic evaluation records assessed for inclusion in the NHS EED published between the years 2008-2013.

5.1 Summary of findings

5.1.1 Filter performance

The adapted NHS EED filter was sensitive (1.0), and in an attempt to increase precision, and be able to search separately for cost-utility analyses, I divided it into two filters, creating the CEA filter and the CEA QoL filter. The first amendments of the CEA 1 filter (Table 2) reduced the sensitivity from 1.000 to 0.948. I made some additional amendments of the free text terms and proximity operator, and removed “cost-benefit analysis” and “cost-minimization analysis” as subject headings and free text words, since they generally were not relevant to HTAs. This decreased the sensitivity of the CEA filter to 0.899 (Fig. 2), while increasing the precision to 0.029. Of the 13 records not identified from subject subset 1, three reported on quality of life (QoL) or quality-adjusted life-years (QALYs), the others were cost-benefit analyses. This indicates that increasing the precision might cause missing relevant records (Glanville et al., 2009a; Booth, 2010).

When tested on the gold standard (n=2198), the CEA filter's final version (Table 2) performed with a sensitivity of (0.899), a precision of 0.029, and a specificity of 0.991. The increase in precision constitutes a difference of 222 gold standard records not retrieved. Compared to the specified objectives of this study, the filter preformed had with a higher sensitivity, a lower precision, but and a specificity within the objective the study the specified range.. The EMBASE G filter performance was close to the CEA filter.

Both the EMBASE H, McKinlay best specificity, EMBASE A and EMBASE F, had a lower sensitivity and higher precision than the CEA and the EMBASE G filter (Table 2). The McKinlay best specificity and EMBASE H filter had a sensitivity above 0.800, and a precision of 0.057 and 0.075 respectively, which was higher than the objective of this filter validation study. Thus, they could be used if a more precise search result is required. As the filters included in this study represent a range of different levels of performance, there might not be a need for further additional amendment of the CEA filter.

The purpose of the CEA QoL filter was to identify cost-utility analyses and not economic evaluations in general. Unfortunately, this filter could not be tested on the NHS EED gold standard which was developed to identify different kinds of economic evaluations. Thus, in order to test the CEA QoL filter I would have to develop a separate gold standard of cost-utility analysis. I could do this, for example by using the relative recall method, and identifying relevant SRs and HTAs that included cost-utility analyses reporting health state utility values (HSUV) (Golder et al., 2006; Arber, M.; et al., 2016; Waffenschmidt et al., 2016). HSUVs are used to inform QALY measures but are not only used to inform cost-utility analyses (Drummond et al., 2015). A cost-utility gold standard could also allow for comparing the CEA QoL filter's performance to the HSUV filter developed by Arber and colleagues to identify studies reporting health state utility values (HSUVs) (Arber, M. et al., 2015; Arber, M.; et al., 2016).

Additional testing and validation of relevant published filters' performance on some of our new HTA commissions could help determine if further amendment of the CEA filter is needed. In order to agree on the optimal balance between sensitivity and precision in retrieval of health economic evaluations for our use, we would need to assess the full text of the records retrieved. The information relevant to developing economic models is not always available in the record's abstract. This might also give additional indication regarding whether

the level of precision intended to achieve was higher than what is sufficient to inform our HTAs.

5.2 Strengths of this filter validation study

5.2.1 Gold standard development

This filter validation study describes the development of a gold standard set of known relevant economic evaluation records from NHS EED published year 2008-2013 (n=2198). The number of gold standard records was established based on a sample size calculation in order to be able to achieve robust results. The study describes a procedure for creating, amending, testing and validating search filters for identifying economic evaluations in Ovid Embase.

5.2.2 Development and testing of search filter

I created the search filter using a combination of elements adapted from the published NHS EED Embase search strategy (Centre for Reviews and Dissemination, 2014a), terms suggested by health economists, and terms from the Embase thesaurus.

I compared the performance of the CEA filter and six published search filters that performed with high precision when developed and tested by Glanville et al. (2009a). The filters were tested on the gold standard, and on a subject specific gold standard of records about cancer published in the years 2008-2013 (n=522), derived from the gold standard (Fig. 1). The gold standard subject subsets that were used for testing and amending the adapted NHS EED and CEA filter were randomly allocated to two subject subsets and a validation set of gold standard records. These are design features included in the ISSG search filter checklist (Glanville et al., 2008; ISSG Search Filter Resource editorial team, 2017b).

5.2.3 Peer review

I tried to be as rigorous as possible by using peer review at key stages of the filter validation study. The search strategy for identifying the gold standard of economic evaluation records in NHS EED was checked by JG, and the downloaded gold standard set of economic evaluations records were managed in Rayyan. The development of the corresponding gold standard set in Embase, and the randomization and allocation of the gold standard records into the three subject specific subsets, was peer reviewed by LN. LNV peer reviewed the adapted NHS EED filter and the CEA filter.

5.2.4 Transparency

Tailoring a search filter to specific needs of retrieval demands transparency and clarification of the advantages and disadvantages of using a particular filter. Achieving a good (optimal) balance between sensitivity and precision can be challenging, and researchers opinions about what is an optimal search result might depend on such factors as the scope of the search, and time and resources available for producing an HTA (Haynes et al., 2005; Booth, 2010; Lee et al., 2012). Thus, it is important to be transparent regarding the likely consequences of choosing a particular filter. I have tried to follow these principles in describing the processes in detail.

5.3 Limitations of this filter validation study

5.3.1 Creating and amending the CEA search filter

The peer review of the CEA filter was performed using the PRESS checklist and not the ISSG search filter checklist used in the development of the filter (Sampson et al., 2008b; ISSG Search Filter Resource editorial team, 2017b). Using the ISSG checklist might have revealed elements that could have improved the development of the search filter. For example, I could have performed a statistical analysis of terms or extracted terms from the gold standard set of records when developing the filters (Glanville et al., 2009a; Glanville et al., 2009b). It might be that development of the search filters based on such methods would contribute to the development of a more precise search filter which was my aim when creating the filters.

In an attempt to increase the precision on the CEA filter, I could have repeated the process of testing and re-testing the filter against the subject specific subsets. For example, a test on subject subset 1 revealed that removing the subject headings “heath economics” and “economic evaluation” only reduced the number of retrieved records by 1.

6 Conclusions

The process of developing a gold standard set of known relevant records, and subject specific gold standard in order to test, amend and compare search filters, is described in sufficient detail to be reproduced by others. Although the CEA filter’s performance does not differ much from the performance of the EMBASE G filter, it could be useful for information specialists, researchers, or others, in identifying economic evaluations in Ovid Embase to

inform HTAs. Both the EMBASE H, McKinlay best specificity, EMBASE A and EMBASE F filters, had a lower sensitivity and higher precision than the CEA and the EMBASE G filters (Table 2). Thus, they could be used if a more precise search result is required. As the filters included in this study represent a range of different levels of performance, there might not be a need for further amendment of the CEA filter.

This filter validation study, like other studies (Glanville et al., 2009a), demonstrates that developing a search filter for identifying health economic evaluations with a precision of at least 10% and a sensitivity of at least 90% is challenging. Except for Arber et al.'s HSUV filter research (Arber, M.; et al., 2016), there are to my knowledge, few published precise search filters available for identifying cost-utility analyses and health state utility values (HSUVs). Further research aimed at developing and testing these types of methodological search filters, using a search filter appraisal checklist such as the ISSG Search Filter Appraisal Checklist, is needed.

6.1 Acknowledgements

The author would like to thank the supervisors Julie Glanville, Lena V. Nordheim and Kjetil G. Brurberg for their helpful comments and suggestions. I also thank Arna Desser, Lien Nguyen, Ingvil Von Mehren Sæterdal and other colleagues at the Division for Health Services in the Norwegian Institute of Public Health, as well and Hanne N. Ona and Karen Rosnes has for their helpful comments and suggestions.

7 References

[1] Knowledge Centre for the Health Services (2015) [How to summarize research. Handbook for the Knowledge Centre for the Health Services] 4. ed. Oslo: Knowledge Centre for the Health Services. Available from: <https://www.fhi.no/globalassets/kss/filer/filer/verktoy/2015_handbok_slik_oppsummerer_vi_forskning.pdf > [Read 01.04.2017].

[2] Norway Priority Committee (2013) [Open and fair. Priority setting in the health services] ed. Oslo: Ministry of Security and Service Organization, Information Management. Available from: <https://www.regjeringen.no/contentassets/16a0834c9c3e43fab452ae1b6d8cd3f6/no/pdfs/nou201420140012000dddpdfs.pdf> [Read 01.04.2017].

[3] DiCenso A, Guyatt G, Ciliska D. Evidence-based nursing: a guide to clinical practice. St. Louis, MO: Elsevier Mosby; 2005.

[4] Guyatt G. Users' guides to the medical literature : a manual for evidence-based clinical practice. 2nd ed. ed. New York: McGraw-Hill Medical; 2008.

[5] Ministry of Health and Care Services. [Good quality-safe services: quality and patient safety in the health and care services]. Oslo: Ministries service center, information management; 2012.

[6] Hagen G, Wisløff T, Klemp M (2013) [Health economic evaluation at the Knowledge Centre for the Health Services]. *Norwegian journal of epidemiology* [Internett], Vol 23, Iss 2.

[7] Green S., Higgins JPT, Alderson P., Clarke M, Mulrow CD, Oxman AD (2011) Chapter 1: Introduction. In: Higgins J, Green S. red. Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011). utg.: The Cochrane Collaboration, p. Available from: <http://handbook.cochrane.org/> [Read 01.10.2015].

[8] Lefebvre C, Manheimer E, Glanville J (2011) Chapter 6: Searching for studies. In: Higgins JPT. GS red. Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011). utg.: The Cochrane Collaboration, p. Available from: <http://handbook.cochrane.org/> [Read 14.01.2017].

[9] Ovid Technologies (2017) Embase 1974 to 2017 April [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 04.04.2017].

[10] Ovid Technologies (2017) Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 06.04.2017].

[11] Centre for Reviews and Dissemination (2015) About the databases. About NHS EED [Internett]. York: Centre for Reviews and Dissemination Available from: <http://www.crd.york.ac.uk/CRDWeb/AboutPage.asp> [Read 06.09.2015].

[12] Glanville J, Fleetwood K, Yellowlees A, Kaunelis D, Mensinkai S. Development and testing of search filters to identify economic evaluations in MEDLINE and EMBASE. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH); 2009.

[13] Harbour J, Fraser C, Lefebvre C, Glanville J, Beale S, Boachie C, et al. (2014) Reporting methodological search filter performance comparisons: a literature review. *Health Info Libr J* [Internett], 31 (3), p. 176-94. DOI: 10.1111/hir.12070

[14] Ovid Technologies (2017) Embase: Excerpta Medica Database Guide [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 14.01.2017].

[15] Elsevier (2016) Embase Fact Sheet [Internett]. Elsevier B.V. Available from: <https://www.elsevier.com/__data/assets/pdf_file/0016/59011/R_D_Solutions_Embase_Fact_Sheet-Web.pdf > [Read 01.09.2016].

[16] Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR (2005) Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *BMJ* [Internett], 330 (7501), p. 1179. DOI: 10.1136/bmj.38446.498542.8F [Read 2005-05-19 22:03:54].

[17] Wilczynski N, Haynes B (2004) Optimal search strategies for retrieving scientifically strong studies of treatment and diagnosis from MEDLINE: an analytical survey [abstract]. *12th Cochrane Colloquium: Bridging the Gaps; 2004 Oct 2-6; Ottawa, Ontario, Canada* [Internett], p. 70-1.

[18] Mathes TW, M.; Antoine, S. L.; Pieper, D.; Eikermann, M. (2014) Methods for systematic reviews of health economic evaluations: a systematic review, comparison, and synthesis of method literature. *Med Decis Making* [Internett], 34 (7), p. 826-40. DOI: 10.1177/0272989x14526470 [Read Oct].

[19] Culyer AJ. The Dictionary of Health Economics. 3rd ed. Cheltenham: Edward Elgar Publishing; 2014.

[20] Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. 4th ed. Oxford: Oxford University Press; 2015.

[21] Glanville J, Kaunelis D, Mensinkai S (2009) How well do search filters perform in identifying economic evaluations in MEDLINE and EMBASE. *Int J Technol Assess Health Care* [Internett], 25 (4), p. 522-9.

[22] Health Information Research Unit (2016) Hedges [Internett]. McMaster University. Available from: <http://hiru.mcmaster.ca/hiru/HIRU_Hedges_home.aspx > [Read 14.01.2017].

[23] Bak G, Mierzwinski-Urban M, Fitzsimmons H, Morrison A, Maden-Jenkins M (2009) A pragmatic critical appraisal instrument for search filters: introducing the CADTH CAI. *Health Info Libr J* [Internett], 26 (3), p. 211-9. DOI: 10.1111/j.1471-1842.2008.00830.x [Read Sep].

[24] Sampson M, Zhang L, Morrison A, Barrowman NJ, Clifford TJ, Platt RW, et al. (2006) An alternative to the hand searching gold standard: validating methodological search filters using relative recall. *BMC Med Res Methodol* [Internett], 6, p. 33. DOI: 10.1186/1471-2288-6-33

[25] Frazier JJ, Stein CD, Tseytlin E, Bekhuis T (2015) Building a gold standard to construct search filters: a case study with biomarkers for oral cancer. *J Med Libr Assoc* [Internett], 103 (1), p. 22-30. DOI: 10.3163/1536-5050.103.1.005 [Read Jan].

[26] ISSG Search Filter Resource editorial team (2017) Filters to Identify Economic Evaluations [Internett]. Available from: <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/filters-to-find-i> [Read 23.04.2017].

[27] Kaunelis D, Glanville J (2017) Summarized Research in Information Retrieval for HTA. Costs and economic evaluation [Internett]. Available from: <http://vortal.htai.org/?q=node/336> [Read 23.04.2017].

[28] Polit DF, Beck CT. Nursing research: generating and assessing evidence for nursing practice. 9 ed. Philadelphia, PA: Wolters Kluwer Health; 2012.

[29] Hajian-Tilaki K (2014) Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* [Internett], 48, p. 193-204. DOI: 10.1016/j.jbi.2014.02.013 [Read Apr].

[30] Thomson Reuters. EndNote X7.5 [Computer software]. Philadelphia (PA): Thomson Reuters; 2016.

[31] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A (2016) Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews* [Internett], 5, p. 210. DOI: DOI: 10.1186/s13643-016-0384-4

[32] Glanville J, Bayliss S, Booth A, Dundar Y, et al. (2008) So many filters, so little time: the development of a search filter appraisal checklist. *J Med Libr Assoc* [Internett], 96 (4), p. 356-61. DOI: 10.3163/1536-5050.96.4.011 [Read Oct].

[33] Lee E, Dobbins M, Decorby K, McRae L, Tirilis D, Husson H (2012) An optimal search filter for retrieving systematic reviews and meta-analyses. *BMC Med Res Methodol* [Internett], 12, p. 51.

[34] GraphPad Software. QuickCalcs. CA, USA: GraphPad Software, Inc.; 2016.

[35] Centre for Reviews and Dissemination (2014) Search strategies. NHS EED EMBASE using OvidSP [Internett]. York: Centre for Reviews and Dissemination Available from: <http://www.crd.york.ac.uk/crdweb/searchstrategies.asp> [Read 06.09.2015].

[36] ISSG Search Filter Resource editorial team (2017) ISSG search filter checklist proforma [Internett]. Available from: <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/issg-search-filter-checklist-proforma> [Read 23.04.2017].

[37] Ovid Technologies (2017) Ovid Help. Search Tools [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 14.01.2017].

[38] McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C (2016) PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol* [Internett], 75, p. 40-6. DOI: 10.1016/j.jclinepi.2016.01.021 [Read 10.05.2017].

[39] Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. PRESS: Peer Review of Electronic Search Strategies. Ottawa, Canada: Canadian Agency for Drugs and Technologies in Health; 2008.

[40] Waffenschmidt S, Hermanns T, Gerber-Grote A, Mostardt S (2016) No suitable precise or optimized epidemiologic search filters were available for bibliographic databases. *J Clin Epidemiol* [Internett], 82, p. 112-8.

[41] Ovid Technologies (2017) Ovid Help. Advanced Search Techniques [Internett]. Wolters Kluwer. Available from: <http://ovidsp.uk.ovid> [Read 14.01.2017].

[42] Booth A (2010) How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. *Int J Technol Assess Health Care* [Internett], 26 (4), p. 431-5. DOI: 10.1017/s0266462310000966 [Read Oct].

[43] Arber M, Garcia S, Veale T, Edwards M, Shaw A, Glanville J (2016) PERFORMANCE OF SEARCH FILTERS TO IDENTIFY HEALTH STATE UTILITY STUDIES (PMR 183) [Internett]. ISPOR 19th Annual European Congress The Ispor Scientific Presentations Database. Available from: <https://www.ispor.org/ScientificPresentationsDatabase/Presentation/67681> [Read 24.04.2017].

[44] Golder S, McIntosh HM, Duffy S, Glanville J (2006) Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Info Libr J* [Internett], 23 (1), p. 3-12. DOI: 10.1111/j.1471-1842.2006.00634.x [Read Mar].

[45] Arber M, Garcia S, Veale T, Edwards M, Shaw A, Glanville J (2015) Sensitivity of a Search Filter Designed to Identify Studies Reporting Health State Utility Values. HTAi Annual Conference 2015 [Poster] [Internett]. York: York Health Economics Consortium. Available from: <http://www.yhec.co.uk/yhec-content/uploads/2015/06/Poster-374-Sensitivity-Of-A-Search-Filter.pdf> [Read 24.04.2017].

[46] Centre for Reviews and Dissemination (2014) Search strategies [Internett]. York: Centre for Reviews and Dissemination Available from: <http://www.crd.york.ac.uk/crdweb/searchstrategies.asp> [Read 06.09.2015].

[47] Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. PRESS: Peer Review of Electronic Search Strategies Appendices. Ottawa, Canada: Canadian Agency for Drugs and Technologies in Health; 2008.

Appendix A: Sample size calculation

Eq. (A.1)

Equation: Sample size calculation within 99% Confidence Interval, minimum 80% sensitivity, and marin error of 2,5%

$$n_{se} = \frac{2{,}58^2 x\ 0.80\ x\ 0.20}{0.025^2\ x\ 1} = \frac{6.6564\ x\ 0.16}{0.0006} = \frac{1.0650}{0.0006} = 1.775 \approx 2{,}000 \text{ records}$$

Appendix B: Search for gold standard records in NHS EED

Search	**Hits**
(*) and ((Economic evaluation:ZDT[1] and Abstract:ZPS[2])) IN NHSEED FROM 2008 TO 2013	2249

Table 1: [1]Document type (e.g. Economic Evaluation), [2]Publication status (e.g. Abstract or Bibliographic record) with a CRD economic evaluation critical assessment (abstract)

Appendix C: Cancer search

Database: NHS EED

Date: 2017.03.17

Line	Search	NHS EED
1	MeSH DESCRIPTOR Neoplasms EXPLODE ALL TREES	
2	(cancer* or carcinoma* or leukaemia or leukemia or lymphoma* or ur* or tumor* or neoplas* or sarcoma)	
3	#1 OR #2	
4	(*) and ((Economic evaluation:ZDT and Abstract:ZPS)) IN NHSEED M 2008 TO 2013	2249[1]
5	#3 AND #4	**536[2]**

[1]NHS EED gold standard including 1 duplicate. [2]Subject specific gold standard (in Embase n=534)

Appendix D: Adapted NHS EED filter, CEA filter and CEA OqL filter

Search interface: Ovid

Version: Embase 1974 to May 12

D.1: Adapted NHS EED filter

20.	Health economics/[1]
21.	Economic Evaluation/[1]

22.	"Cost Effectiveness Analysis"/[1]
23.	" Cost Benefit Analysis"/[1,2]
24.	"Cost Utility Analysis"/[1]
25.	"Cost Minimization Analysis"/[1,3]
26.	Pharmacoeconomics/[1,4]
27.	(health economic? or economic evaluation?).ti,ab[5]
28.	(cost* adj2 (analys* or benefit* or effective* or minim* or utilit*)).ti,ab[5]
29.	cba.ti,ab[5]
30.	cea.ti,ab[5]
31.	cua.ti,ab[5]
32.	((pharmacoeconomic? or pharmac*) adj economic?).ti,ab
33.	(15D or HRQoL or health-related quality of life instrument*[6] or EuroQol 5D* or EQ-5D*).ti,ab[5]
34.	(SF-6D or Short form-6 dimensions or HUI or Health utility index or PTO or Person trade-off or QWB or Quality of well-being or SG or Standard gamble or TTO or Time-trade-off or AQOL or Assessment of quality of life).ti,ab[5,7]
35.	Quality of life/[1]
36.	Quality adjusted life year/[1]
37.	(QALY or QALYs or quality-adjusted-life-year? or quality of life).ti,ab[5]
38.	or/1-18

Appendix D.1: [1]EMTREE terms (subject headings). [2]Express both costs and outcomes of the interventions in monetary terms, and is infrequently used to evaluate health interventions. It was included because it can be confused with cost utility, and quality of life (QoL). [3]Describes interventions with approximately the same effect but different costs. It was included to identify publications that in fact is a cost effectiveness analysis. [4]Pharmacoeconomics labels economic evaluations of drug therapy and includes cost analysis, treatment outcome and quality of life studies.[5]Free text words searched for in title and abstract. [6]Health-related quality of life, includes a person's Health utility index (physical, mental and social well-being). [7]These instruments are not yet covered by a specific subject heading (e.g. 15D (15-dimensional), EuroQol-5D descriptive system (EQ-5D), SF-6D (Short Form-6 Dimension).

Appendix D.2: Cost-effectiveness analysis filter, version 1 (CEA 1)

Line	Search terms
1.	Health economics/[1]
2.	Economic Evaluation/[1]
3.	"Cost Effectiveness Analysis"/[1]
4.	"Cost Utility Analysis"/[1]
5.	"Cost Benefit Analysis"/[1,2]
6.	"Cost Minimization Analysis"/[1,3]
7.	Pharmacoeconomics/[1,4]
8.	(health economic* or economic evaluation*).ti,ab[2]
9.	(cost* adj2 (analys* or effective* or utility or utilities or benefit* or minim*)).ti,ab
10.	(cea or cua or cba). ti,ab
11.	(pharmacoeconomic* or (pharmac* adj economic*)).ti,ab.
12.	OR/1-11

Appendix D.2: [1]EMTREE terms (subject headings).[2E]xpress both costs and outcomes of the interventions in monetary terms, and is infrequently used to evaluate health interventions. It was included in the search filter because it can be confused with cost utility, and quality of life (QoL). [3]Describes interventions with approximately the same effect but different costs. It was included to identify publications that, in fact, are cost effectiveness analyses. [4]Pharmacoeconomics labels economic evaluations of drug therapy and includes cost analysis, treatment outcome and quality of life studies".[5]Free text words searched for in title and abstract.

Apppendix D.3: Cost-effectiveness analysis Quality of Life (CEA QoL)

Line	Search terms
1.	Quality of life/[1]
2.	Quality adjusted life year/[1]
3.	(QALY or QALYs or quality-adjusted-life-year* or quality of life or utility or utilities).ti,ab[2]
4.	(15D or HRQOL[3] or health-related quality of life instrument* or EuroQol 5D* or EQ-5D*).ti,ab[2,4]
5.	(SF-6D or Short form-6 dimensions or HUI or Health utility index or PTO or Person trade-off or QWB or Quality of well-being or SG or Standard gamble or TTO or Time-trade-off or AQOL or assessment of quality of life).ti,ab[2,4]
6.	OR/1-5

Appendix D.3: [1]EMTREE terms (the EMTREE term "Quality of life index" was not included because I consider it too general to our use. [2]Free text words searched for in title and abstract, [3]Includes a person's Health utility index (physical, mental and social well-being). [4]In general these instruments are not yet covered by a specific subject heading (e.g. 15D (15-dimensional), EuroQol-5D descriptive system (EQ-5D), SF-6D (Short Form-6 Dimension).

Appendix E: Published search filters and syntax guide

E.1 Search filters tested by Glanville et al.[1]

Filter name	Search terms
1. EMBASE A	1. Cost adj effectiveness.ab.
2. EMBASE F	4. Cost adj effectiveness.ti. 5. (Costs.ab. and controlled study/ and cost.hw.) AND (effectiveness.hw. or randomiz 6. 1 OR 2
3. EMBASE G	9. Cost adj effectiveness.ab. 10. Cost adj effectiveness.ti. 11. Life adj years.ab. 12. Life adj year.ab. 13. Qaly.ab. 14. (Cost or costs).ab. and Controlled Study/ 15. (Cost and costs).ab. 16. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7
4. EMBASE H	6. Cost adj effectiveness.ti. 7. (cost adj effectiveness).ab. AND (costs or cost).ab. 8. (cost AND costs).ab. AND cost effectiveness analysis/ 9. (Costs.ab. AND controlled study/ and cost.hw[2]) AND (effectiveness.hw. OR randomized controlled trial/) 10. 1 OR 2 OR 3 OR 4

5. McKinlay Best specificity	2. Cost effectiveness.tw. OR sensitivity analys*.tw.
6. NHS EED EMBASE[2]	17. Health Economics/ 18. exp Economic Evaluation/ 19. exp Health Care Cost/ 20. pharmacoeconomics/ 21. 1 or 2 or 3 or 4 22. (econom$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic*).ti,ab. 23. (expenditure* not energy).ti,ab 24. (value adj2 money).ti,ab. 25. Budget*.ti,ab. 26. 6 or 7 or 8 or 9 27. 5 or 10 28. (metabolic adj cost).ti,ab. 29. ((energy or oxygen) adj cost).ti,ab. 30. ((energy or oxygen) adj expenditure).ti,ab. 31. 12 or 13 or 14 32. 11 not 15

[1](Glanville et al., 2009a), for search filed codes (e.g. .ti, .ab, .hw) see Appendix E.2: Syntax guide, [2]The search filter is reported without exclusion strategy and doesn't include costs: letter.pt; editorial.pt; note.pt (Centre for Reviews and Dissemination, 2014)

Appendix E.2: Syntax guide

Code	Description
/ (slash)	Subject heading (e.g. Economic Evaluation /)
.ab.	Abstract
.fs.	Floating subheading
.hw.	Heading Word - retrieves every EMTREE subject heading that includes a particular word
.mp.	Title, abstract and indexing
.pt.	Publication Type
.sh.	Subject heading
.ti.	Title.
tw,.	Textword
.yr.	Publication year
* (asterisk)	Truncation symbol when following a term (e.g. economic*)
?	Indicates there can be zero or one additional characters (e.g. cost? Identifies cost and also costs)
Adj	Adjacent terms
Adj1	Terms within one word of each other
Exp	Explode (subject heading)
Or/1-3	Combine sets 1 to 3 using OR (1 OR 2 OR 3)