BMJ Open

PubMed search filters for the study of putative outdoor air pollution determinants of disease

Stefania Curti,[1] Davide Gori,[2] Valentina Di Gregori,[2] Andrea Farioli,[1] Alberto Baldasseroni,[3] Maria Pia Fantini,[2] David C Christiani,[4] Francesco S Violante,[1] Stefano Mattioli[1]

To cite: Curti S, Gori D, Di Gregori V, *et al*. PubMed search filters for the study of putative outdoor air pollution determinants of disease. *BMJ Open* 2016;**6**: e013092. doi:10.1136/bmjopen-2016-013092

► Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2016-013092).

Received 17 June 2016
Revised 3 October 2016
Accepted 19 October 2016

For numbered affiliations see end of article.

Correspondence to
Professor Francesco S Violante; francesco.violante@unibo.it

ABSTRACT

Objectives: Several PubMed search filters have been developed in contexts other than environmental. We aimed at identifying efficient PubMed search filters for the study of environmental determinants of diseases related to outdoor air pollution.

Methods: We compiled a list of Medical Subject Headings (MeSH) and non-MeSH terms seeming pertinent to outdoor air pollutants exposure as determinants of diseases in the general population. We estimated proportions of potentially pertinent articles to formulate two filters (one 'more specific', one 'more sensitive'). Their overall performance was evaluated as compared with our gold standard derived from systematic reviews on diseases potentially related to outdoor air pollution. We tested these filters in the study of three diseases potentially associated with outdoor air pollution and calculated the number of needed to read (NNR) abstracts to identify one potentially pertinent article in the context of these diseases. Last searches were run in January 2016.

Results: The 'more specific' filter was based on the combination of terms that yielded a threshold of potentially pertinent articles ≥40%. The 'more sensitive' filter was based on the combination of all search terms under study. When compared with the gold standard, the 'more specific' filter reported the highest specificity (67.4%; with a sensitivity of 82.5%), while the 'more sensitive' one reported the highest sensitivity (98.5%; with a specificity of 47.9%). The NNR to find one potentially pertinent article was 1.9 for the 'more specific' filter and 3.3 for the 'more sensitive' one.

Conclusions: The proposed search filters could help healthcare professionals investigate environmental determinants of medical conditions that could be potentially related to outdoor air pollution.

Strengths and limitations of this study

- We evaluated the overall performance of the proposed topic-based filters in terms of sensitivity and specificity, as compared with our gold standard derived from systematic reviews on diseases potentially related to outdoor air pollution.
- The two search filters (one 'more sensitive', one 'more specific') can be evoked in PubMed by entering the shortened Uniform Resource Locators (URLs).
- We formulated these two search filters based on the abstracts, while we did not evaluate the main body of the sampled articles nor the quality of the individual studies; we cannot exclude the loss of some information reported in articles without a summary.
- The proposed search filters cannot exclude studies not yet indexed in PubMed as animal studies (eg, recently published articles).
- The present study was restricted to PubMed as any medical database has its own syntax and key terms and needs to be studied separately.

INTRODUCTION

Environmental exposure has become a prominent issue in the study of aetiology of acute and chronic diseases in the last few years. Many different adverse effects have been linked to exposure to air pollution, including an increased risk of respiratory and cardiovascular diseases among all ages.[1–4] There is considerable evidence that exposure to different air pollutants may be associated with an increase in morbidity and short-term and long-term mortality.[5 6]

Topics related to environmental exposure have become a major issue for public health throughout the world and increasingly popular in the press including medical literature as evidenced by indexing in medical databases.

Nowadays, the standard practice includes the use of medical databases in order to retrieve all the possible relevant articles and select them through the use of filters and specific search strategies.[7–9] The methodological aspects for a ready-to-use tool aimed to efficiently retrieve all the possible

pertinent articles have already been explored for the PubMed database, in contexts other than environmental.[10–13]

The aim of this study is to create two PubMed search filters (one 'more sensitive', one 'more specific') in order to efficiently retrieve all the relevant articles for the study of putative environmental determinants of diseases related to outdoor air pollution.

METHODS

The four stages of search filter development are summarised in figure 1.

Selection of terms to be tested

The USA National Library of Medicine (NLM) maintains a controlled vocabulary thesaurus—namely, the Medical Subject Headings (MeSH)—for indexing biomedical journals for the Medline/PubMed database. PubMed users may construct a query using MeSH terms to retrieve abstracts related to a particular field or matter. Using MeSH terms allows highly specific searches. However, recent journal articles are provisionally indexed in PubMed as supplied by the publisher, without MeSH terms. Furthermore, the indexing process may be slow and inaccurate, particularly for advanced/innovative topics. Hence, PubMed queries that include free-text words are usually more sensitive than those containing MeSH terms only.

For the purpose of the present study, we collected a list of MeSH terms (and their subheadings) pertinent to the field of environmental diseases related to outdoor air pollution from the Medline MeSH database. We then verified that the definition of the selected MeSH terms matched the topic under study. At first, we identified a group of 'core terms' including those MeSH terms that appeared to be strictly related to the field of exposure to outdoor air pollution and environmental diseases. We further identified other—apparently less specific—MeSH terms. To evoke a more specific search, these MeSH terms were combined with the term 'AND air pollut*', where necessary.

We also explored the associated entry terms of the MeSH terms (ie, synonyms, alternate forms, and other closely related terms generally used interchangeably with the preferred term for the purposes of indexing and retrieval) to draw up a list of non-MeSH terms (ie, free-text keywords) to be tested. Subsequently, this list was expanded by the analysis of the terms or keywords used in relevant articles (including systematic reviews) or in reports on air pollution of the US Environmental Protection Agency (EPA) and WHO.

To consider all the possible variations, we checked for synonyms and acronyms of the selected non-MeSH terms and took into account differences between British and American spelling. In addition, we used the truncation symbols to create searches that took into consideration multiple spellings and various endings.

Finally, for each MeSH and non-MeSH term, we calculated the proportion of abstracts retrieved by the selected search term which was not retrieved by the 'core terms'.

Estimating proportions of pertinent articles

For each studied term (MeSH term or non-MeSH term), we collected a sample of English-language abstracts added to PubMed by 31 December 2010. The language restriction was introduced as the availability of an English-language abstract can be of practical importance when assessing the relevance of an article. In order to evaluate only those articles dealing with human diseases, we added to each search filter the words 'NOT (animals [MH] NOT humans [MH])'. The purpose of this was to exclude from the search those articles indexed with the MeSH term 'animals', but without the MeSH term 'humans'. When the search query produced more than 1000 citations, we collected a systematically recruited sample of 100 abstracts. For terms that retrieved <1000 citations, we evaluated a sample of <100 available abstracts. The number of abstracts to be sampled was calculated assuming an α error of 0.05 and a precision level of 90% (sampling error of 10%).[14]

To obtain systematically recruited samples, we used PubMed 'show' function in order to retrieve a number of pages closest to 100 or multiples. Then, according to the number of abstracts to be read, for each page the 'top-of-the-page' article was kept, to make sure that abstracts across the considered time period were included.

Two pairs of authors (SC, SM; DG, VDG) independently assessed the pertinence of each collected abstract based on the subject of the article. An abstract was judged as pertinent in the case it reported some kind of relation between the exposure to outdoor air pollution and the disease taken into account in the abstract, irrespective of study design and quality. Another author (AF) resolved any disagreements. Inter-rater agreement, explored in a preliminary assessment of 100 abstracts, was 'good' (κ=0.68 for both pairs).[15]

Formulation of search filters

To formulate two different search filters—one 'more specific', one 'more sensitive'—we classified the studied search terms on the basis of an a priori cut-off of 40% of retrieval rate of pertinent articles corresponding to a number needed to read (NNR) value of 2.5. The NNR defines the total number of articles that must be read to find each relevant one and corresponds to the inverse of precision.[16] Hence, a cut-off of 40% means that of every 2.5 retrieved abstracts, 1 is pertinent.

We then included only the terms with a proportion of pertinent abstracts ≥40% in the 'more specific' search filter. All the tested terms were instead included in the 'more sensitive' one, as every term contributed to the retrieval of the pertinent literature, even if with different proportion of pertinent abstracts.

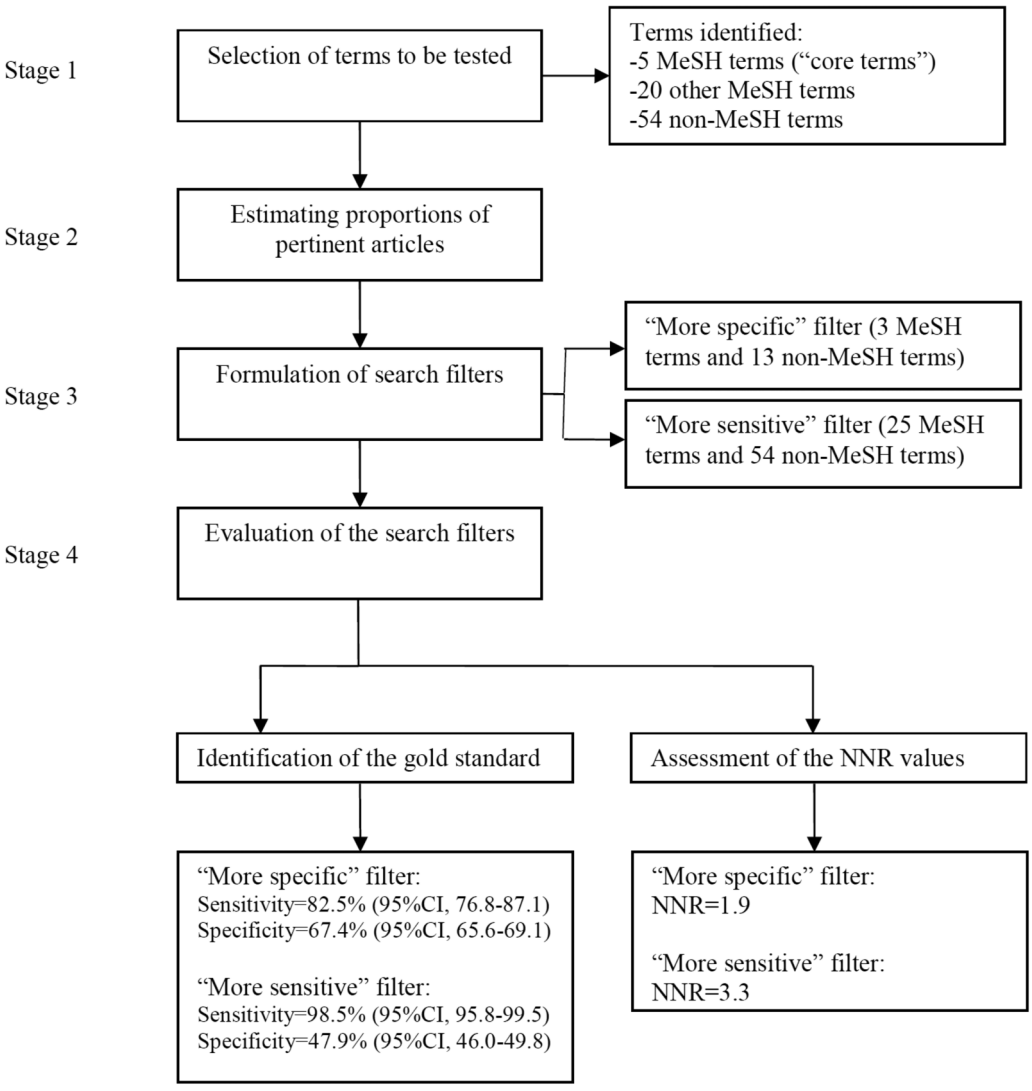

Figure 1 Stages of search filter development.

To limit the retrieval of those articles not dealing with human diseases, we added to each search filter the words 'NOT ((animals [MH] OR plants [MH]) NOT humans [MH])'.

We regarded the 'more specific' filter as a method to minimise the number of 'false positive' abstracts, even if losing some possible pertinent abstracts. On the other hand, the 'more sensitive' filter—at the cost of a higher number of 'false positive' abstracts—was conceived to retrieve a larger amount of pertinent literature.

Evaluation of the search filters

Identification of the gold standard

To build the gold standard, we searched Medline using a mini-filter composed of the 'core terms' and adding the terms (systematic rev* OR metanal* OR meta anal* OR metaanal*) limiting the search to abstract availability. This search was run on 10 November 2015.

Two authors (SC, SM) independently screened titles and abstracts of the 100 most recent references to select those systematic reviews that addressed a specific medical condition potentially related to outdoor air pollution. A third author (DG) resolved any disagreements.

Based on journal rankings (impact factor (IF)) as supplied by the Journal Citation Reports database in 2014,[17] we retained only those reviews published in journals categorised as Q1 (top 25% of the IF distribution) or Q2 (between top 50% and top 25%) in the subject categories relevant for the journal. These inclusion criteria were used to include high-impact journals in the gold standard.

Then, we evaluated the selected reviews using the R-AMSTAR checklist for quality assessment for Systematic Reviews.[18] Only reviews with a score of 20 or more (out of a maximum score of 44) were included.

For those systematic reviews with a medium/high score, we intended to reproduce the original search strategy executed for PubMed by the authors in order to identify the amount of pertinent (articles included in the reviews) and not pertinent (articles retrieved by the

original search strategies, but not included in the reviews) literature. To reproduce the PubMed searches of the selected reviews, we used the same strategies as reported by the authors applying the same limits such as date of searches or language restrictions if any.

The gold standard was built combining all the references indexed in PubMed of the studies included in those reviews whose search strategies were sufficiently described to properly allow their replication. Only the articles on health studies related to outdoor air pollution were used for the creation of the gold standard. Articles covering topics other than outdoor air pollution were excluded from the gold standard.

Hence, an ad hoc filter was built combining all the PMIDs of the included references. A PMID (PubMed identifier or PubMed unique identifier) is a unique number assigned to each PubMed record that does not change over time or during processing and is never reused.[19] This ad hoc filter containing the PMIDs of the included references together with the replication of the PubMed search strategies of the identified reviews enabled us to appropriately identify the total amount of pertinent or not pertinent references of our gold standard.

The references retrieved by the two proposed search filters (ie, 'more specific' and 'more sensitive') were compared with the gold standard. For each filter, we were able to correctly calculate sensitivity and specificity along with 95% CIs produced with the Wilson score method.[20]

The searches for both filters were run on 16 December 2015.

Assessment of the number needed to read values

We assessed the performance of the two proposed search filters in the study of three diseases that have been associated with the exposure to outdoor air pollution: arrhythmia, sudden death and congenital heart defects. Of note, the selected diseases were searched as text word and MeSH term (eg, congenital heart defects [MH] OR congenital heart defect*).

First, we collected all potentially pertinent abstracts retrieved for these diseases by the proposed search filters restricting to abstracts up to 31 December 2010. Then, two authors (SC, SM) independently evaluated the pertinence of the retrieved abstracts considering whether a relation between exposure to outdoor air pollution and the disease has been reported (κ=0.74). A third author (DG) resolved any disagreements. Finally, we calculated the NNR values for each filter. NNRs were defined as the ratio of the number of retrieved abstracts to the number of pertinent ones.[16] The searches for both filters were run on 18 January 2016.

Comparative analysis

We evaluated whether the method we used to formulate the proposed search filters could be efficiently carried out by other researchers when creating new filters on another topic.

For this purpose, we compared the characteristics (sensitivity, specificity and NNR) of the proposed search filters with those of a 'conventional' search filter developed in a recent systematic review on outdoor air pollution. We searched Medline using a mini-filter composed of the 'core terms' and adding the terms (systematic rev* OR metanal* OR meta anal* OR metaanal*) limiting the search to abstract availability. Two authors (SC, SM) independently screened titles and abstracts of the most recent references to select that systematic review that addressed a specific medical condition potentially related to outdoor air pollution and published in a top-ranked journal (classified as Q1 or Q2).

The references retrieved by the 'conventional' search filter of the recent systematic review (using only the part of the filter on outdoor air pollution) were compared with the gold standard identified with the same methodology described in the previous section. Sensitivity and specificity along with 95% CI produced with the Wilson score method[20] were then calculated.

Furthermore, we assessed the performance of this 'conventional' filter on outdoor air pollution in the study of the same three diseases mentioned above (ie, arrhythmia, sudden death and congenital heart defects). The same methodology for the evaluation of the pertinence of the retrieved abstracts was applied and NNR values were then calculated. All the searches were run on 22 September 2016.

Stata V.14.1 SE (Stata Corporation, Texas, Texas, USA) was used for analysis with a significance level of 0.05.

RESULTS

Selection of terms to be tested

The MeSH terms 'air pollutants', 'air pollution', 'disorders of environmental origin', 'environmental exposure' and 'particulate matter' appeared to be the more related to the field of environmental diseases related to outdoor air pollution with respect to the definitions reported in the Medline MeSH database; these five MeSH terms were hence defined as 'core terms' (see online supplementary table S1).

According to the definitions provided in the MeSH database, we also evaluated the contribution of other MeSH terms, which were combined with the term 'AND air pollut*', where necessary. Therefore, we added 20 other MeSH terms to the pool of search terms to be studied (see online supplementary table S1).

In addition, we explored the entry terms associated with the selected MeSH terms and compiled a list of non-MeSH terms to be evaluated. This list was extended by non-MeSH terms retrieved in pertinent articles, keywords found in reports on air pollution of the EPA and WHO, and terms suggested by coauthors. Furthermore, we checked for all the possible variations of the proposed terms including synonyms and acronyms. Every

term was tested as free-text either using truncation or inverted commas in order to select the most comprehensive search term and to take into consideration multiple spellings as well. Finally, we included 54 non-MeSH terms in our study (see online supplementary table S1).

The searches conducted using the proposed terms (MeSH and non-MeSH terms) were then compared with those performed with the 'core terms' in order to calculate the proportion of abstracts retrieved by the selected search terms which was not retrieved by the 'core terms' (see online supplementary tables S2 and S3).

Estimating proportions of pertinent articles

The 'core terms' (ie, 'air pollutants', 'air pollution', 'disorders of environmental origin', 'environmental exposure' and 'particulate matter' searched as MeSH terms) identified 127 296 abstracts added to PubMed by 31 December 2010 and limited by the mini-filter NOT (animals [MH] NOT humans [MH]). For each of the other 20 MeSH terms considered, overlaps with the aforementioned 'core terms' ranged from 18% to 100%.

The 54 non-MeSH search terms evoked 154 617 abstracts (almost 1.7% of all articles listed in PubMed) using the cited limits. For each non-MeSH term, the overlapping with the 'core terms' ranged from 8% to 95%. Data on the proportion of pertinent abstracts for the 25 MeSH and 54 non-MeSH terms are reported in online supplementary tables S2 and S3.

Formulation of search filters

The 'more specific' search filter included 3 MeSH terms and 13 non-MeSH terms which retrieved an estimated proportion of pertinent articles ≥40% (corresponding to an NNR value ≤2.5). All the other terms (22 MeSH terms and 41 non-MeSH terms) were included in the 'more sensitive' filter, which also included all the search terms of the 'more specific' filter.

The two proposed PubMed search filters are presented in box 1.

Evaluation of the search filters

Identification of the gold standard

Based on titles and abstracts, we screened the 100 most recent references of potentially eligible systematic reviews. Of these, we identified 16 systematic reviews which addressed a specific medical condition potentially related to outdoor air pollution. We further excluded five reviews which did not fulfil the inclusion criteria (ie, four of them were not published in journals categorised as Q1 or Q2 and one was not a proper systematic review). Then, 11 systematic reviews were assessed for quality using the R-AMSTAR checklist. Of these, only one reported a score <20 (out of a maximum score of 44).

To correctly identify the amount of pertinent/not pertinent references, we intended to reproduce the search strategy executed for PubMed (using the same limits) by the authors of the 10 systematic reviews with a medium/high score (ie, 20 or more). Out of these 10 systematic reviews, we were able to reproduce the original PubMed search strategy for six of them[21–26] (see online supplementary table S4).

These six systematic reviews identified altogether 244 pertinent references in PubMed (of these, 11 were duplicates). Therefore, the gold standard was built combining all the PMIDs of the included references together with the records extracted from the proper replication of the PubMed search strategies of the six systematic reviews included in our study. Hence, the gold standard was composed of 206 pertinent references and 2736 not pertinent references.

This gold standard was used as a comparison set for the two proposed search filters. The 'more specific' filter correctly identified 170 pertinent references out of 206, whereas the 'more sensitive' filter only missed three pertinent references. The 'more specific' filter was able to correctly identify 1844 not pertinent references out of 2736, whereas the 'more sensitive' filter correctly identified 1310 not pertinent ones.

The 'more specific' filter yielded a sensitivity of 82.5% (95% CI 76.8% to 87.1%) along with a specificity of 67.4% (95% CI 65.6% to 69.1%). The 'more sensitive' filter reported a sensitivity of 98.5% (95% CI 95.8% to 99.5%) together with a specificity of 47.9% (95% CI 46.0% to 49.8%).

Assessment of the number needed to read values

To test the two proposed search filters, we explored the available literature on three diseases potentially associated with exposure to outdoor air pollution, namely, arrhythmia, sudden death and congenital heart defects. The results of these searches for both filters are shown in table 1. As expected, the 'more sensitive' filter showed a higher NNR in comparison to the 'more specific' one.

In the overall search, the 'more specific' search filter retrieved 260 articles out of 180 859 articles indexed for arrhythmia, sudden death and congenital heart disease in the Medline database. Out of 260 articles, 140 were judged as pertinent, accounting for 54% of pertinence and hence an NNR of 1.9.

In the overall search, the 'more sensitive' search filter retrieved 895 articles out of 180 859 articles indexed for arrhythmia, sudden death and congenital heart disease in the Medline database. Out of 895 articles, 271 were judged as pertinent, accounting for 30% of pertinence and hence an NNR of 3.3.

Comparative analysis

We screened the titles and abstracts of the most recent references of potentially eligible systematic reviews on outdoor air pollution. We were able to identify the most recent one that satisfied the inclusion criteria. The selected systematic review evaluated the association between cognitive functioning and exposure to air pollution and it was published on a journal ranked Q1 in the

Box 1 Proposed PubMed search filters for identifying potentially pertinent articles for the field of diseases potentially related to outdoor air pollution

'More specific' filter:
(air pollution [MH] OR particulate matter [MH] OR (air pollutants/adverse effects [MH] NOT air pollutants [PA]) OR ('air pollution' AND exposure) OR ('air pollution' AND 'health effects') OR ('diesel exhaust' NOT vehicle emissions [MH]) OR environmental toxicant* OR exhaust part* OR ((fossil fuel* NOT fossil fuels [MH]) AND air pollut*) OR ('fossil fuels' AND exposure) OR gaseous pollut* OR ((motor vehicle* NOT motor vehicles [MH]) AND air pollut*) OR ((motorway* OR roadway* OR highway* OR freeway*) AND air pollut*) OR ('particulate matter' AND exposure) OR ('PM2,5' OR 'PM 2,5' OR 'PM2.5' OR 'PM 2.5') OR ultrafine particle*) NOT ((animals [MH] OR plants [MH]) NOT humans [MH]) AND name(s)-of-the-disease
'More sensitive' filter:
(air pollutants [MH] OR air pollution [MH] OR disorders of environmental origin [MH] OR environmental exposure [MH] OR particulate matter [MH] OR air pollutants [PA] OR (air pollutants/adverse effects [MH] NOT air pollutants [PA]) OR (aromatic hydrocarbons [MH] AND air pollut*) OR (benzene derivatives [MH] AND air pollut*) OR (carbon monoxide [MH] AND air pollut*) OR (dioxins [MH] AND air pollut*) OR dust [MH] OR environmental health [MH] OR environmental medicine [MH] OR environmental pollutants [MH] OR environmental pollution [MH] OR fires [MH] OR fossil fuels [MH] OR gasoline [MH] OR (hydrogen sulfide [MH] AND air pollut*) OR motor vehicles [MH] OR (nitrogen oxides [MH] AND air pollut*) OR ozone [MH] OR sulfur dioxide [MH] OR vehicle emissions [MH] OR air contamin* OR (air pollutant* NOT air pollutants [MH]) OR ('air pollution' AND exposure) OR ('air pollution' AND 'health effects') OR ('air pollution' NOT air pollution [MH]) OR 'air quality' OR atmospheric contamin* OR atmospheric pollut* OR (automobile* AND air pollut*) OR ('diesel exhaust' NOT vehicle emissions [MH]) OR ((dust OR dusts) NOT dust [MH]) OR (emission* AND air pollut*) OR environmental contamin* OR environmental disease* OR (environmental exposure* NOT environmental exposure [MH]) OR (('environmental health' NOT environmental health [MH]) AND air pollut*) OR (('environmental medicine' NOT environmental medicine [MH]) AND air pollut*) OR ((environmental pollutant* NOT environmental pollutants [MH]) AND air pollut*) OR ('environmental pollution' NOT environmental pollution [MH]) OR environmental toxicant* OR 'environmental toxicology' OR exhaust part* OR ((fossil fuel* NOT fossil fuels [MH]) AND air pollut*) OR ('fossil fuels' AND exposure) OR gaseous pollut* OR ((gasoline NOT gasoline [MH]) AND air pollut*) OR (incinerator* AND air pollut*) OR (industr* AND air pollut*) OR (industrial pollut* OR industry pollut*) OR (lead AND air pollut*) OR ((metal OR metals) AND air pollut*) OR ((motor vehicle* NOT motor vehicles [MH]) AND air pollut*) OR ((motorway* OR roadway* OR highway* OR freeway*) AND air pollut*) OR ((Nox NOT nitrogen oxides [MH]) AND air pollut*) OR ((ozone NOT ozone [MH]) AND air pollut*) OR (ozone AND exposure) OR (PAH NOT PAH [AU] NOT 'pulmonary arterial hypertension') OR (particle* AND air pollut*) OR particulate* OR ('particulate matter' AND exposure) OR persistent organic pollut* OR ('PM2,5' OR 'PM 2,5' OR 'PM2.5' OR 'PM 2.5') OR ('PM 10' OR 'PM10') OR pollut* OR 'polycyclic aromatic hydrocarbon' OR POPS OR (road* AND air pollut*) OR (smog NOT ('particulate matter' OR particulate matter [MH])) OR (traffic AND air pollut*) OR ultrafine particle* OR (urban AND air pollut*) OR (vehicle emission* NOT vehicle emissions [MH]) OR VOCS OR 'Volatile Organic Compounds') NOT ((animals [MH] OR plants [MH]) NOT humans [MH]) AND name(s)-of-the-disease
AU, Author; MH, Medical Subject Heading; PA, Pharmacological Action; the asterisk (*) represents the PubMed truncation symbol.
Notes:
1. It is possible to 'copy and paste' each of the two filters into PubMed from a .doc file. Alternatively, the filters can be evoked in PubMed by entering the following shortened Uniform Resource Locators (URLs) in the browser address box: http://tinyurl.com/pollution-specific for the 'more specific' filter; http://tinyurl.com/pollution-sensitive for the 'more sensitive' filter.
2. The name-of-the-disease should be entered. For diseases that have more than one name, the various 'names-of-the-disease' should be entered in brackets, connected by the OR operator: for example, … AND (epicondylitis OR tennis elbow).

public, environmental and occupational health category.[27]

The already identified gold standard was used as a comparison set for the 'conventional' search filter on outdoor air pollution developed in the selected systematic review. This filter correctly identified 175 pertinent references out of 206 and 1839 out of 2741 not pertinent ones. The sensitivity of this 'conventional' filter on outdoor air pollution was 85.0% (95% CI 79.4% to 89.2%) together with a specificity of 67.1% (95% CI 65.3% to 68.8%).

With respect to the assessment of the NNR values, this 'conventional' filter on outdoor air pollution developed in a recent systematic review retrieved 79 articles for arrhythmia, 42 for sudden death and 22 for congenital heart defects. Out of the 143 articles retrieved in the overall search, 57 were judged as pertinent, with an NNR of 2.5.

DISCUSSION

We applied a systematic approach to identify efficient PubMed search strategies to retrieve information on environmental determinants of disease related to outdoor air pollution. We created two readily applicable search filters for use by health professionals: one 'more specific' and the other 'more sensitive' (see box 1). An easy and fast 'copy and paste' tool could be particularly relevant as it could help fill the gap between the standard practice and research practice.

We evaluated the overall performance of these topic-based filters in terms of sensitivity and specificity, as compared with our gold standard derived from systematic reviews on diseases potentially related to outdoor air pollution. As expected, the 'more specific' filter reported the highest specificity (67.4%), while the 'more sensitive' one reported the highest sensitivity (98.5%).

Moreover, to give a practical perspective of these two topic-based filters, we evaluated them through the investigation of three diseases selected a priori together with the use of the NNR. On the one hand, 54% of the abstracts retrieved by the 'more specific' filter provided information on environmental determinants of disease related to outdoor air pollution (NNR 1.9), suggesting that this search filter can be applied to aetiological questions encountered in routine practice. In particular, the 'more specific' filter would allow researchers and professionals for a better overview over new evidence on the health effects of outdoor air pollution by monitoring newly published systematic reviews.

On the other hand, the 'more sensitive' filter yielded almost twice the number of potentially pertinent articles —although, as expected, the overall NNR was as high as 3.3 (table 1).

Therefore, we consider the 'more sensitive' filter as a second-line approach useful to investigate those diseases that have been little studied as for their putative environmental origin.

In comparison with the proposed 'more sensitive' filter, the 'conventional' one on outdoor air pollution developed in a recent systematic review lacks in sensitivity (98.5% for the 'more sensitive' filter vs 85.0% for the 'conventional' one) and reports characteristics more similar to the 'more specific' one.

The 'conventional' filter on outdoor air pollution developed in the selected systematic review retrieved only 143 articles for the investigation of three diseases selected a priori. Of these, 57 were judged as pertinent (NNR 2.5). Specifically, the number of retrieved abstracts and of pertinent ones were found to be substantially lower than those calculated for the 'more sensitive' filter (895 retrieved, 271 pertinent abstracts) and even lower than those of the 'more specific' one (260 retrieved, 140 pertinent abstracts) (table 1).

It is striking to note that the 'conventional' filter of the selected systematic review reported roughly the same sensitivity of the proposed 'more specific' one, probably because of the 'conventional' way to develop search filters used by the authors of the six systematic reviews included in the gold standard. On the other hand, the 'conventional' filter developed in the selected systematic review explores only a limited subset of the total amount that should be explored, as shown when assessing the NNR of the three diseases selected a priori.

Although our procedure seems to be rather complex and time-consuming, the proposed search filters provide a very efficient tool to retrieve more pertinent articles and appear to be highly sensitive—two essential characteristics when conducting a systematic review.

Owing to feasibility issues, we only assessed the pertinence of articles with available (English-language) abstracts. Hence, we cannot exclude that we lost some information reported in articles without a summary, like research letters or brief reports. However, a previous study pointed out that this factor should not represent a

Table 1 Application of search filters to three pathologies: number of citations retrieved, proportion of potentially pertinent articles and overall NNR values

	Arrhythmia (n=100 538)			Sudden death [MH] OR 'sudden death' (n=22 187)			Congenital heart defects [MH] OR congenital heart defect* (n=58 134)			Overall (n=180 859)		
	Retrieved†	Pertinent		Retrieved†	Pertinent		Retrieved†	Pertinent		Retrieved†	Pertinent	
	n	n (%)	NNR	n	n (%)	NNR	n	n (%)	NNR	n	n (%)	NNR
PubMed query	91	72 (79)	1.3	153	57 (37)	2.7	16	11 (69)	1.5	260	140 (54)	1.9
'More specific' filter	411	136 (33)	3.0	305	67 (22)	4.6	179	68 (38)	2.6	895	271 (30)	3.3
'More sensitive' filter	320	64 (20)	5.0	152	10 (7)	15.2	163	57 (35)	2.9	635	131 (21)	4.8
'More sensitive' filter NOT 'more specific' filter (incremental contribution of the 'more sensitive' filter)												

*signifies the PubMed truncation symbol.
Searches run on 18 January 2016.
†Filters for date (31 December 2010) and abstract availability.
NNR, number needed to read value.

major source of bias.[10] We formulated our search filter based on the abstracts, while we did not evaluate the main body of the sampled articles. Hence, we may have underestimated the proportion of potentially pertinent articles, especially in the absence of widespread implementation of more informative abstracts. It should be stressed that we did not assess the quality of the individual studies.

Using the words 'NOT ((animals [MH] OR plants [MH]) NOT humans [MH])', we intended to limit our search filters to human studies, including those animal studies that have human relevance. Nevertheless, this filter cannot exclude from the search those animal studies that have not been indexed yet (eg, recently published articles). With the purpose to help further to restrict to human studies only, we added to our proposed search filters a limit regarding 'plants' as well.

Our selection of non-MeSH terms was to some extent arbitrary, but at the same time it was so extensively performed that the identification of such search terms with computational techniques (like text-mining) was not considered worthwhile. In addition, considering the specificity and sensitivity values of the proposed filters together with NNR values, this a priori limitation did not appear to greatly affect the end product.

Even though it has been reported that most of the high-quality articles, like those included in Cochrane Reviews, are indexed in PubMed,[28] it is strongly recommended to consult more than one relevant database when performing systematic reviews of the literature.[29] The present study was restricted to PubMed as any medical database has its own rules (eg, syntax, map of key terms) and needs to be studied separately (including the assessment of sensitivity and specificity). When carrying out a systematic review on the health effects of outdoor air pollution, it could be possible to translate the proposed 'more sensitive' filter into other database syntax; however, it has to be underlined that it is likely to suppose that the characteristics of the translated filter—in terms of sensitivity and specificity—would not be the same as those calculated for PubMed.

Improvements of reporting practices—like the implementation of the STROBE guidelines[30]—should facilitate the retrieval of pertinent literature in the future. Also, the bibliometric properties of our topic-based filter might be influenced by changes in the indexing process that occur over the years.[9]

CONCLUSIONS

We present two PubMed search filters—one 'more specific', one 'more sensitive'—which may be easily applied to collect evidence on diseases possibly associated with exposures to outdoor air pollution. Both filters can be copied and pasted into the PubMed search box together with the name of the disease under study. Alternatively, the filters can be evoked in PubMed by entering the shortened Uniform Resource Locators (URLs) provided at the bottom of box 1. The 'more specific' filter can be used as a first-line approach, while the 'more sensitive' filter could help practitioners when deeper searches are necessary. Health professionals could take advantage of these topic-based filters in contexts ranging from evidence-based patient evaluation to original research.

Finally, researchers and professionals could use these filters to be periodically up-to-date on this specific field. For this purpose, users can take advantage of 'My NCBI' tool provided by the NLM that allows to permanently save searches and set automatic email updates of the results.[31]

Author affiliations
[1]Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy
[2]Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy
[3]Tuscany Regional Centre for Occupational Injuries and Diseases (CeRIMP), Florence, Italy
[4]Department of Environmental Health, Harvard School of Public Health, Harvard University, Boston, Massachusetts, USA

Contributors SC and SM contributed to the conception and study design; acquisition, analysis and interpretation of the data; drafting of the manuscript; critical revision of the manuscript for important intellectual content as whole. DG contributed to the conception and study design; acquisition, analysis and interpretation of the data; drafting of the manuscript. VDG contributed to the acquisition; critical revision of the manuscript for important intellectual content. AF, MPF, DCC and FSV contributed to the analysis and interpretation of the data; critical revision of the manuscript for important intellectual content. AB contributed to the conception and study design; critical revision of the manuscript for important intellectual content. All the authors read and approved the final version of the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Shah AS, Langrish JP, Nair H, *et al*. Global association of air pollution and heart failure: a systematic review and meta-analysis. *Lancet* 2013;382:1039–48.
2. Scheers H, Jacobs L, Casas L, *et al*. Long-term exposure to particulate matter air pollution is a risk factor for stroke: meta-analytical evidence. *Stroke* 2015;46:3058–66.
3. Lu F, Xu D, Cheng Y, *et al*. Systematic review and meta-analysis of the adverse health effects of ambient PM2.5 and PM10 pollution in the Chinese population. *Environ Res* 2015;136:196–204.
4. Pope CA III, Burnett RT, Thun MJ, *et al*. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 2002;287:1132–41.
5. Anderson HR, Spix C, Medina S, *et al*. Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the APHEA project. *Eur Respir J* 1997;10:1064–71.
6. Biggeri A, Bellini P, Terracini B. [Meta-analysis of the Italian studies on short-term effects of air pollution—MISA 1996–2002]. *Epidemiol Prev* 2004;28:4–100.

7. Damarell RA, Tieman J, Sladek RM, *et al*. Development of a heart failure filter for Medline: an objective approach using evidence-based clinical practice guidelines as an alternative to hand searching. *BMC Med Res Methodol* 2011;11:12.
8. Sladek RM, Tieman J, Currow DC. Improving search filter development: a study of palliative care literature. *BMC Med Inform Decis Mak* 2007;7:18.
9. Haynes RB, Wilczynski N, McKibbon KA, *et al*. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *J Am Med Inform Assoc* 1994;1:447–58.
10. Mattioli S, Zanardi F, Baldasseroni A, *et al*. Search strings for the study of putative occupational determinants of disease. *Occup Environ Med* 2010;67:436–43.
11. Mattioli S, Gori D, Di Gregori V, *et al*. PubMed search strings for the study of agricultural workers' diseases. *Am J Ind Med* 2013;56:1473–81.
12. Guaraldi F, Grottoli S, Arvat E, *et al*. PubMed search strategies for the identification of etiologic associations between hypothalamic-pituitary disorders and other medical conditions. *Pituitary* 2013;16:471–82.
13. Pillastrini P, Vanti C, Curti S, *et al*. Using PubMed search strings for efficient retrieval of manual therapy research literature. *J Manip Physiol Ther* 2015;38:159–66.
14. Cochran WG, ed. *Sampling techniques*. 2nd edn. New York: John Wiley and Sons, Inc, 1963.
15. Altman DG. Some common problems in medical research. In: *Practical statistics for medical research*. London: Chapman and Hall, 1991:403–9.
16. Bachmann LM, Coray R, Estermann P, *et al*. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. *J Am Med Inform Assoc* 2002;9:653–8.
17. *2014 Journal Citation Reports® Science Edition*. Thomson Reuters, 2015.
18. Kung J, Chiappelli F, Cajulis OO, *et al*. From systematic reviews to clinical recommendations for evidence-based health care: validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for grading of clinical relevance. *Open Dent J* 2010;4:84–91.
19. National Center for Biotechnology Information. PubMed Help: Search Field Descriptions and Tags. http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Search_Field_Descrip (accessed 9 May 2016).
20. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 1927;22:209–12.
21. Shah AS, Lee KK, McAllister DA, *et al*. Short term exposure to air pollution and stroke: systematic review and meta-analysis. *BMJ* 2015;350:h1295.
22. Jaacks LM, Staimez LR. Association of persistent organic pollutants and non-persistent pesticides with diabetes and diabetes-related health outcomes in Asia: a systematic review. *Environ Int* 2015;76:57–70.
23. Liu JC, Pereira G, Uhl SA, *et al*. A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. *Environ Res* 2015;136:120–32.
24. Song Q, Christiani DC, XiaorongWang, *et al*. The global contribution of outdoor air pollution to the incidence, prevalence, mortality and hospital admission for chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Int J Environ Res Public Health* 2014;11:11822–32.
25. Li C, Fang D, Xu D, *et al*. Main air pollutants and diabetes-associated mortality: a systematic review and meta-analysis. *Eur J Endocrinol* 2014;171:R183–90.
26. Wang B, Xu D, Jing Z, *et al*. Effect of long-term exposure to air pollution on type 2 diabetes mellitus risk: a systemic review and meta-analysis of cohort studies. *Eur J Endocrinol* 2014;171:R173–82.
27. Clifford A, Lang L, Chen R, *et al*. Exposure to air pollution and cognitive functioning across the life course—A systematic literature review. *Environ Res* 2016;147:383–98.
28. Rollin L, Darmoni S, Caillard JF, *et al*. Searching for high-quality articles about intervention studies in occupational health—what is really missed when using only the Medline database? *Scand J Work Environ Health* 2010;36:484–7.
29. Gehanno JF, Paris C, Thirion B, *et al*. Assessment of bibliographic databases performance in information retrieval for occupational and environmental toxicology. *Occup Environ Med* 1998;55:562–6.
30. Vandenbroucke JP, von Elm E, Altman DG, *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007;4:e297.
31. National Center for Biotechnology Information. My NCBI. http://www.ncbi.nlm.nih.gov/account/ (accessed 9 May 2016).