Primary Care Respiratory Journal (2011); **20**(1): 71-74

Primary Care RESPIRATORY JOURNAL
www.thepcrj.org

ORIGINAL RESEARCH

Developing and testing search strategies to identify patients with active seasonal allergic rhinitis in general practice

***Vicky Hammersley[a], Roger Flint[a], Hilary Pinnock[a], Aziz Sheikh[a]**

[a] Allergy and Respiratory Research Group, Centre for Population Health Sciences: GP Section, The University of Edinburgh, Edinburgh, Scotland, United Kingdom

Originally submitted 3rd February 2010; resubmitted 11th May 2010; revised 4th August 2010; accepted 18th September 2010; online 28th December 2010

Abstract

Aim: We sought to assess the accuracy of different search terms to identify individuals with active seasonal allergic rhinitis (SAR) in general practice.

Methods: A reference search strategy was developed to identify patients with active SAR. This was applied through inspection of electronic health records of patients aged 15-45 years in a 10% random sample of a general practice database. Searches used Read codes and medication relating to SAR. Sensitivity, specificity, and positive and negative predictive values were calculated.

Results: Using the reference search strategy, 54/1092 (4.9%) of 15-45 year-old patients had current SAR. Searching for drugs used in nasal allergy had the highest sensitivity (85%) and good specificity (86%). Searching for a recorded history of SAR (H170) in the last two years was more specific (100%) but this approach only had limited sensitivity (17%).

Conclusions: Electronic searches can be used to identify patients with current SAR, but the accuracy varies widely. Larger numbers of sufferers can be identified using broader search parameters, but with increasing numbers of false positives. In contrast, more focused search strategies give a smaller yield needing less cleaning of data to identify true positives, but there is an associated increase in the number of false negatives.

V Hammersley *et al*. *Prim Care Respir J* 2011; **20**(1): 71-74
doi:10.4104/pcrj.2010.00086

Keywords seasonal allergic rhinitis, general practice, electronic search, Read codes, medication

See linked editorial by Levy on page 11

The full version of this paper, with online Appendix, is available at www.thepcrj.org

Introduction

Seasonal allergic rhinitis (SAR) is common in the UK and is responsible for considerable morbidity, impairment in quality of life and healthcare utilisation.[1] It has recently been shown to be associated with impaired educational attainment in adolescents.[2]

The Allergic Rhinitis in Asthma (ARIA) classification scheme was introduced in 2001 and reinforced in 2008; this subdivides allergic rhinitis into "intermittent" or "persistent" disease.[3,4] Previously, based on time of exposure, allergic rhinitis was subdivided into "seasonal" (more commonly known as hayfever), "perennial" and "occupational" forms. Intermittent allergic rhinitis (IAR) is defined as symptoms being present for less than four days per week for less than four weeks, and persistent allergic rhinitis refers to symptoms that are present for more than four days a week and for more than four weeks. Although this clinical definition is officially accepted in the UK as in other countries, the classification has not yet been widely adopted in UK primary care, and patients with IAR are thus still given a Read code[5] for seasonal allergic rhinitis in their computerised medical records. Hence, these terms and codes need to be searched when identifying local populations.

There is a need for increased research into SAR in order to improve understanding of its aetiology, changing epidemiology, disease trends and possible interventions to improve outcomes in people with SAR.[6,7] For example, one study currently being undertaken is a large cluster randomised trial of an educational intervention for adolescent hayfever sufferers, where it was

* **Corresponding author:** Ms. Vicky Hammersley, Centre for Population Health Sciences, The University of Edinburgh, 20 West Richmond Street, Edinburgh, EH8 9DX, United Kingdom. Tel: +44 (0)131 651 4151 E-mail: vicky.hammersley@ed.ac.uk

crucial to identify accurately participants for recruitment using electronic health records. Undertaking such work in primary care populations is important because of the opportunity to generate research findings that will be directly applicable to the general practice setting in which the majority of UK patients with SAR are currently managed.

In countries where the vast majority of primary care practices use electronic health records, novel primary care database searching methods using appropriate history and medication codes potentially provide an efficient means to identify such patients.[5] However, problems remain, such as the diversity of clinical computing systems and the inconsistent use of coding schemes by clinicians.[8]

Internationally, several medical coding systems exist.[8] However, for allergic rhinitis the codes are reasonably interchangeable between various systems. For example, the International Classification of Disease-10 code J30.1 (due to pollen) and the International Classification of Primary Care-2 R97 (hayfever/allergic rhinitis) will potentially identify the same patients as H170 (seasonal allergic rhinitis), the Read (5-byte version 2) coding scheme used in this pilot. Date ranges included in a search strategy can be adjusted for the relevant aeroallergens of a particular country. Previous work in other diseases[9,10] suggests that search strategies should be bespoke, reflecting differences in clinical and coding definitions nationally and internationally.

The utility of the various search strategies available is currently unknown. In order to investigate the relative accuracy of different search approaches, we sought to assess the sensitivity, specificity, and positive and negative predictive values associated with a range of search strategies in the context of identifying patients with current SAR.

Methods

Ethics and research management approval

Formal ethics and research management approval were not required as the research was undertaken using an anonymised dataset. [Dodds P. Personal communication, 4.6.09] Approval was obtained from the practice prior to commencing the fieldwork.

Setting and sampling

Our pilot study was carried out at a large three-site UK general practice with a list size of 32,436 patients served by 17 general practitioners (GPs). For the last four years, medical records – including all consultations and prescribing – have been recorded in an electronic health record ("Vision"™, In Practice Systems Ltd.) which is based on 5-byte Read codes (version 2). Read codes are organised in a hierarchical manner. For example, allergic rhinitis is coded as H17, and SAR is coded as H170; searches using H17 will thus include all people with SAR and other types of rhinitis, including perennial rhinitis, whereas searching H170 will only include those people with a specific diagnosis of SAR.

We selected all patients aged 15-45 years (n=10,920) and from this sample selected a random 10% sample of records for detailed manual interrogation (n=1,092) against the reference search strategy employed. This 10% sample was also used for all the test search strategies thereby allowing a direct comparison with the reference search strategy.

Reference search strategy

The reference definition used was:

- The presence of the SAR Read code (H170) as a history item or active problem recorded in the individual patient's electronic record or in the active problem field applied within the last two years, AND/OR
- A consultation with a clinician within the last two years diagnosing SAR (in free text or otherwise), AND/OR
- A diagnosis of SAR in the free text ("hay fever", "rhinoconjunctivitis", "SAR", "pollen allergy") OR H170 Read code applied previously (i.e. over two years ago) and a current prescription during the "hay fever" season for drugs used in treating SAR (nasal corticosteroids, sodium cromoglicate eye drops or antihistamines – checking that antihistamines and other drugs had not been given for any other condition).[11,12]
- The "hay fever season" was defined as "01.05.08 – 31.07.08 or 01.05.09 –31.07.09" inclusive.[13] The SAR season date ranges in searches were linked using the "OR" command.

Test search strategies approval

Three search strategies were tested:

1) based on Read codes suggestive of SAR
2) medication prescribed for treatments that can be used in SAR
3) a combination of the Read codes and medication prescribed.

Details of these test search strategies are reproduced in Appendix 1, available online at www.thepcrj.org.

Statistical techniques

We used the principles of the approach advocated in the

Box 1: Definitions of tests used

Sensitivity: The proportion of true positives that are correctly identified by the test

Specificity: The proportion of true negatives that are correctly identified by the test

Positive predictive value: The proportion of patients with positive test results who are correctly diagnosed with a search strategy

Negative predictive value: The proportion of patients with negative test results who are correctly diagnosed with a search strategy

STARD guidelines.[14] This involved calculating the following test parameters: test accuracy; sensitivity and specificity; and positive and negative predictive values (see Box 1 for details). Analysis was undertaken using Microsoft Excel software.

Results

Applying the reference search strategy

Applying the reference approach was a very labour-intensive process. Manually searching the electronic health records of the 1,092 records in the sampling frame took approximately 30 hours. These searches yielded 54 (4.9%) true positives, equating to approximately 33 minutes per case identified.

Accuracy of test search strategies

It was possible to execute successfully all the planned search strategies. The accuracy of the various test strategies employed is summarised in Table 1. The key findings were:

1) Searching on a specific SAR code recorded in the previous two years was 100% specific but the sensitivity was only 17%. The high positive and negative predictive values points to the efficiency of this approach. Such a search strategy would therefore be ideal for highlighting a small number of patients who are very likely to have SAR. However, it would not be useful for identifying the whole SAR population since 83% of patients were missed.
2) In contrast, searching for drugs used in nasal allergy had high specificity, sensitivity and negative predictive values, but low positive predictive values, indicating that this approach is particularly useful for identifying the largest number of potential SAR cases – although this would entail some cleaning of data to exclude false positives.
3) The combination of drugs used in nasal allergy and/or a SAR Read code (H170) during the previous two seasons had a 72% sensitivity and a 93% specificity, indicating that it may be useful for identifying large numbers of those with SAR whilst reducing the time requirement for data cleaning. There would be a trade-off between time required to identify the 7% of false positives versus the number of SAR positive individuals required for the study.

Table 1. Code and drug search results on 10% random sample of the 15-45 year old population.

Codes	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Code only (last 2 years)				
H17 (Allergic rhinitis)	0.22	1.00	1.00	0.96
H170 (SAR)	0.17	1.00	1.00	0.96
Drug group only				
Antihistamines (3.4.1)	0.78	0.86	0.22	0.99
Corticosteroids (6.3)	0.26	0.92	0.15	0.96
Other anti-inflammatory preparations (cromoglicate)	0.43	0.96	0.38	0.97
Drugs used in nasal allergy	0.85	0.86	0.24	0.99
Drug combinations (2 SAR seasons)				
Sodium cromoglicate OR nasal allergy drugs	0.41	0.98	0.55	0.97
Sodium cromoglicate OR antihistamines	0.61	0.98	0.62	0.98
Sodium cromoglicate corticosteroids	0.35	0.96	0.31	0.97
Antihistamines or corticosteroids	0.59	0.94	0.35	0.98
Antihistamines or drugs used in nasal allergy	0.67	0.97	0.50	0.98
Corticosteroids or drugs used in nasal allergy	0.54	0.95	0.34	0.98
Any drug	0.67	0.94	0.36	0.98
All drugs and Read code H170 (last 2 years)	0.72	0.93	0.35	0.98

Discussion

This study has, for the first time, demonstrated the different yields of patients with current SAR identified by conducting different search strategies to interrogate GP records. We hope that our work will inform clinicians and researchers and enable them to reflect critically on the search strategies employed to identify patients with SAR and also to begin to understand the resource implications for cleaning data associated with employing different search strategies.

Strengths and limitations of this work

One of the main strengths is that all potentially eligible patients were equally likely to be studied; there was therefore no risk of selection biases.[15] Another strength of the design is that it was based on clinician-diagnosed SAR, not on self-reporting by patients with current or historical symptoms, although this may have had a significant impact on the estimated population prevalence in the practice database as a result.

The major limitations include the fact that the reference and test assessments were conducted by one researcher, although in order to mitigate against the possible risks of misclassification error, clear protocols and explicit criteria were consistently used. Also, this work was only undertaken in one general practice, albeit a very large practice, raising the possibility that practice idiosyncrasies in coding could limit the generalisability of this work. For example, sodium cromoglicate was included in the search strategy as in this particular practice this is the only drug from this drug group in the practice formulary prescribed for hayfever. However, other drugs within BNF group 11.4.2 (other anti-inflammatory preparations) may need to be included in other general practice searches. There is

therefore a need to replicate this work in other practices to assess the robustness of our test statistics. A few treatments (specifically montelukast) licensed for the treatment of SAR were excluded from our search strategy as they are not included in the relevant BNF groups. These drugs are not in this practice's formulary for the treatment of SAR, and are therefore rarely used for this indication – so exclusion from our searches will have had little effect on our findings. There is also the limitation that up to half of the patients with SAR may be self-diagnosed and hence neither known to, nor coded by, their GP.[16] This constraint, whilst important, cannot currently be overcome through interrogation of practice electronic health records.

Conclusions and recommendations for research

There is clearly no "ideal" search strategy, due to the trade-off between specificity and sensitivity. If the clinician or researcher needs to identify a small number of people definitely suffering from SAR, then a search for the H170 code will identify such people most efficiently. Time-limiting these searches to the last two years, for example, will allow the identification of those with current SAR.

To identify the largest possible population of SAR sufferers, the names of all drugs used to treat nasal allergy should be incorporated into the search strategy. Potentially, this strategy could be improved further by integrating general practice and pharmacy computer systems to take account of over-the-counter prescriptions. However, cleaning the results of such a search to remove false positives is a labour- and time-intensive process, and so this is only likely to be appropriate for large scale, well resourced projects.

More broadly, in the UK, management of some long-term conditions (e.g. asthma) now attracts incentive payments under the Quality and Outcomes Framework (QOF), which is dependent on consistent coding. If, in the future, SAR were to be adopted as an indicator within QOF, coding would very likely improve.

Finally, the Read code system has inherent limitations. Often, useful information is written in the free text within the medical record. A final recommendation, therefore, would be the ability to carry out free-text searches on individual patient medical records. A method by which one could search for that free text, such as "natural language searching", would facilitate investigation of a large number of records, more so with patients who have long and complex medical histories.[17]

Conflict of interest declarations

Aziz Sheikh is Joint Editor-in-Chief of the *PCRJ* but was not involved in the editorial review of, nor the decision to publish, this article.

Funding

This project was funded by the MSc in Public Health Research dissertation budget at The University of Edinburgh.

References

1. Ghouri N, Hippisley-Cox J, Newton J, Sheikh A. Trends in the epidemiology and prescribing of medication for allergic rhinitis in England. *J Roy Soc Med* 2008;**101**(9):466-72. http://dx.doi.org/10.1258/jrsm.2008.080096
2. Walker S, Khan-Wasti S, Fletcher M, Cullinan P, Harris J, Sheikh A. Seasonal allergic rhinitis is associated with a detrimental effect on examination performance in United Kingdom teenagers: case-control study. *J Allergy Clin Immunol* 2007;**120**:381-7. http://dx.doi.org/10.1016/j.jaci.2007.03.034
3. World Health Organization Initiative, Bousquet J, van Cauwenberge P, Khaltaev NJ. Allergic Rhinitis and its Impact on Asthma (ARIA). *J Allergy Clin Immunol* 2001;**108**:S147-334.
4. Bousquet J, Khaltaev N, Cruz AA, *et al*. Allergic Rhinitis and its Impact on Asthma (ARIA) 2008. *Allergy* 2008;**63**(s86):8-160. http://dx.doi.org/10.1111/j.1398-9995.2007.01620.x
5. Anandan C, Simpson CR, Fischbacher C, Sheikh A. Exploiting the potential of routine data to better understand the disease burden posed by allergic disorders. *Clin Exp Allergy* 2006;**36**:866-71. http://dx.doi.org/10.1111/j.1365-2222.2006.02520.x
6. Gupta R, Sheikh A, Strachan DP, Anderson HR. Burden of allergic disease in the UK: secondary analyses of national databases. *Clin Exp Allergy* 2004; **34**(4):520-6. http://dx.doi.org/10.1111/j.1365-2222.2004.1935.x
7. Gupta R, Sheikh A, Strachan DP, Anderson HR. Time trends in allergic disorders in the UK. *Thorax* 2007;**62**(1):91-6. http://dx.doi.org/10.1136/thx.2004.038844
8. Simpson CR, Anandan C, Fischbacher C, Lefevre K, Sheikh A. Will Systematized Nomenclature of Medicine-Clinical Terms improve our understanding of the disease burden posed by allergic disorders? *Clin Exp Allergy* 2007;**37**(11):1586-93. http://dx.doi.org/10.1111/j.1365-2222.2007.02830.x
9. Thiru K, Donnan P, Weller P, Sullivan F. Identifying the optimal search strategy for coronary heart disease patients in primary care electronic patient record systems. *Informatics in Primary Care* 2009;**17**:215-24.
10. Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N, Wright L. Methodology for assessing the usefulness of general practice data for research in one research network. *Health Informatics J* 2004;**10**:91-109. http://dx.doi.org/10.1177/1460458204042230
11. Sheikh A, Singh Panesar S, Dhami S. Seasonal allergic rhinitis. *Clin Evid* 2005; **14**:684-95.
12. Fokkens WJ. Nasal corticosteroids, first choice in moderate to severe allergic rhinitis. What prevents general practitioners from using them? *Allergy* 2003; **58**(8):724-6. http://dx.doi.org/10.1034/j.1398-9995.2003.00256.x
13. Bremner SA, Carey IM, DeWilde S, *et al*. Vaccinations, infections and antibacterials in the first grass pollen season of life and risk of later hayfever. *Clin Exp Allergy* 2007;**37**(4):512-17. http://dx.doi.org/10.1111/j.1365-2222.2007.02697.x
14. Bossuyt PM, Reitsma JB, Bruns DE, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003; **326**:41-4.
15. Simpson C, Newton J, Hippisley-Cox J, Sheikh A. Incidence and prevalence of multiple allergic disorders recorded in a national primary care database. *J R Soc Med* 2008;**101**:558-63. http://dx.doi.org/10.1258/jrsm.2008.080196
16. Richards S, Thornhill D, Roberts H, Harries U. How many people think they have hay fever, and what they do about it. *Br J Gen Pract* 1992;**42**(360):284-6.
17. Jurfsky D, Martin JH. *Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Second edition. London: Pearson; 2009.

Available online at **http://www.thepcrj.org**

Appendix 1. Read codes and drugs used in searches.

Read Code:	Definition:
H17..	Allergic Rhinitis
H170.	Seasonal Allergic Rhinitis

Drug groups used in nasal allergy (BNF 57, 2009)

B.N.F. Drug Group:	Drug:
Nasal corticosteroids	Beclometasone Dipropionate
	Betamethasone Sodium Phosphate
	Budesonide
	Flunisolide
	Fluticasone Propionate
	Mometasone Furoate
	Triamcinolone Acetonide
Antihistamines	Alimemazine Tartrate
	Chlorphenamine Maleate
	Clemastine
	Cyproheptadine
	Hydrochloride
	Hydroxyzine Hydrochloride
	Ketotifen
	Promethazine Hydrochloride
	Cetirizine Hydrochloride
	Desloratadine
	Fexofenadine Hydrochloride
	Levocetirizine Hydrochloride
	Loratadine
	Mizolastine
	Azelastine hydrochloride
Cromoglicate	Sodium Cromoglicate 2% W/V Eye drops