

The case of ‘protective fever and chest signs’:

towards a better understanding of general practice databases

General practice databases provide researchers with information about the realities of primary care. Empirical data of GPs’ care of patients is a key resource in primary care research,¹ and it is a positive sign that this trend is increasing in the worldwide health science community. In a number of countries, robust general practice databases are emerging which serve as an infrastructure for studies of diagnosis, prognosis, and management of health problems. An inherent strength of the databases are their strong links to the community in which the GPs practice and, as a consequence, they enable direct research of important health problems.² Although GP databases are relevant to large sectors of the science and healthcare communities, they can be difficult to interpret.

This month’s edition of the *BJGP* presents five studies on diagnostic acuity of serious (respiratory tract) infections in the general practice population (Hay *et al*,³ Van den Bruel *et al*,⁴ Holm *et al*,^{5,6} and van Duijn *et al*⁷). The studies examining different populations report a generally low *a priori* probability of serious disease — well below 1%. All studies analyse infectious signs and symptoms, while the value of additional testing of inflammatory markers is reported by Holm *et al*, for C-reactive protein and procalcitonin.^{5,6} A general finding was that signs, symptoms, and additional testing were related to diagnostic outcome. However, construction of a robust algorithm was only possible for ruling out serious disease, not for identifying it.

A problem with ruling out rather than diagnosing serious disease is that signs and symptoms seldom enter the surgery in their own right: patients’ beliefs and expectations interact in their presentation of illness, and determine as van Duijn *et al*,⁷ report, diagnostic and therapeutic

outcome. Additionally, GPs not only observe but also interpret while observing. Van den Bruel *et al*,⁴ conclude that a GP’s inference that ‘something is wrong’ is the best predictor of serious disease.

The data from these studies are based on the rich and detailed information⁸ derived from general practice databases. However, some problems of interpretation are apparent, and it is interesting to review the findings of these studies in this context.

The study of Hay *et al*,³ illustrates a methodological approach to cope with these problems in an elegant way. Their study investigated diagnostic evidence by predicting complicated courses of cough in preschool children from observed ‘fever’ and ‘chest signs’. The methodology to generate state-of-the-art diagnostic evidence is still under development,⁹ despite the fact that making a timely diagnosis is the most important clinical intervention to pursue in general practice. Hay *et al*, apply a two-step method in which they identify predicting signs and symptoms in an initial study (the ‘derivation study’),¹⁰ followed by validation of these findings in a different population. This methodological approach highlights the need for research access to a rich variation of general practice cohorts and databases, and emphasises their importance as a research infrastructure. Access to different databases is enhanced by standardisation of data recording. Therefore, an internationally-recognised classification of primary care should be used when recording core data.¹¹ Signs and symptoms identified during diagnosis are seldom independent from each other, and the nature of their interactions is often poorly understood. This is also the case with infectious signs and symptoms. Van den Bruel *et al*,⁴ identified up to five

factors to predict severe infection. GPs’ inference that ‘something is wrong’, which is a composite of observations, is difficult to replicate. All findings presented in these five articles are likely to benefit from validation in similar general practice populations elsewhere.

The main strength of these studies is in refuting severe, life-threatening disease. The studies of Holm *et al*,^{5,6} demonstrate that this is not just due to subjectivity-prone signs and symptoms. C-reactive protein and procalcitonin have proven their value as markers of inflammation in selected, hospitalised patients but did not add to diagnosis in general practice. The main problem was that few patients had truly elevated values of these inflammation markers.

The low rate of abnormal levels of these inflammation markers presents problems in positively identifying severe or rare cases in general practice. Calculating probabilities of common symptoms for relevant diseases requires large datasets,¹² much larger than the numbers in these five smaller studies. It is important to refute spurious relationships and highlight uncommon, severe morbidity. Linking databases and pooling data in the foreseeable future is likely to become a necessity for general practice research. This linking should not be limited to national boards.

Development of diagnostic strategies to confirm major disease should be encouraged, incorporating diagnosis of rare, severe disease. Excluding the probability of serious illness and disease is a core part of general practice which can benefit from research evidence. Studies that help practitioners to exclude serious illness and disease should not be ruled out by editors, reviewers, funding bodies, and the science community, as they are an important part of research.

Diagnosis is often regarded as the most important aspect of clinical reasoning. Most guidelines for medical performance focus on improving practitioner's diagnostic accuracy. Standardisation of diagnosis, based on unequivocal criteria¹³ that can be applied in various settings are needed. However, the study of Hay *et al.*,³ suggests that GPs are directing their interventions at prognosis, rather than diagnosis, of respiratory infection. Prognostication and its research are different from studies of the diagnostic process in that observation over time and patient's medical life history are required, in addition to the time-course of a single episode.

Van Duijn *et al.*,⁷ found that patients' health beliefs were strong predictors of antibiotic prescribing, and were found to be more decisive than physical signs and symptoms. The possibility of tracking patients' medical histories over time using general practice databases is overlooked by primary care and undervalued by the science community. It would have been difficult to conduct studies like that of Hay *et al.*, outside general practice.

Despite these positive aspects of the research, the results of Hay *et al.*, are, at face value, a little puzzling. In their derivation study they identify a number of factors that were related to a complicated course of illness, notably fever and chest signs. However, in their validation study neither is related to complicated outcome of respiratory tract infection. Particularly difficult to interpret is that in this part of their study 'fever' and 'chest signs' appear to be protective against adverse outcome of respiratory tract infection. Factors related to a certain outcome in one study cannot be confirmed in another, which is why the researchers applied a two-step derivation-validation study design. However, a shift from a 'risk' for complications in the initial study to a 'protective' factor in the second one is an entirely different story. This would suggest the inclusion of spurious variables for adverse outcome in the study design, or recruitment of highly-selected patients or practices. The comprehensive general practice databases makes the finding of 'spurious' relations a realistic one. Indiscriminate inclusion of patient and

practice characteristics in derivation studies may turn this strength into a weakness, as the likelihood of statistical significance would present a researchers' lottery with only jackpot. But uncritical selection of prognostic factors would be about the last criticism to make of the study by Hay *et al.*³

'Fever' and 'chest signs' were more than just significant factors in the derivation study. These factors are likely to be linked to 'infection' and, logically, the higher the fever, or the stronger the indications of involvement of the usually sterile lower airways, the more severe the infection. The validation study dismantled this common sense as too simplistic, but that makes it still hard to accept these factors as a protective of adverse outcome of respiratory tract infections in young children. Recruitment of highly-selective patients and practices is also unlikely. The participating practices in both studies represent, beyond reasonable doubt, regular primary care patient populations. No important exclusions were imposed on patient recruitment for the studies.

A possible explanation for the apparent anomaly in the Hay *et al.*, study could be 'diagnostic shift'. The ultimate purpose of the study was to identify criteria for antibiotic prescribing in respiratory tract infections; the complicated course of the infection was to be the marker of that. GPs are highly aware of the discrepancies in their prescription of antibiotics. A diagnostic shift could have occurred in relation to children with symptoms of a respiratory infection and additional markers of complicated infection including, but not exclusively, fever and chest signs. With these children labelled under another diagnosis — and excluded from the study — 'fever' in the children remaining in the validation study would have lost its prognostic meaning. An alternative explanation for the anomaly could have been that routine treatment of respiratory infections with antibiotics prevented the development of complications. The same mechanisms may have affected the other studies reviewed here.

Such a course of events is conceivable, but it is uncertain whether this applies to the study by Hay *et al.* It should signal,

though, that clinical processes and decisions influence general practice data with, at face value, unpredictable research effects. The simple knee-jerk reaction would be to refute practice-derived databases for research purposes. An alternative approach would be to invest in the critical appraisal of research that is based on general practice databases to improve understanding of, among others, diagnostic shifts.

The general practice research community has been apologetic and defensive towards one of the pearls of clinical research: longitudinal patient-centred databases of information collected under regular general practice care. A more pro-active approach is required. Raising the profile of general practice research should include presenting the values of general practice databases. General practice data are far from 'simple' data ready for scientific processing. The role of general practice researchers is to provide guidance and in-depth understanding for other scientists so that they can make use of general practice databases for clinical research. This starts with gaining and sharing this in-depth understanding ourselves. The intriguing case of 'protective fever and chest signs' is as good a starting point for this as it can get.

Chris van Weel

Professor of General Practice, Radboud University Nijmegen Medical Centre, the Netherlands

REFERENCES

1. Van Weel C, Rosser WW. Improving health care globally: a critical review of the necessity of family medicine research and recommendations to build research capacity. *Ann Fam Med* 2004; **2**(Suppl 2): S5–S16.
2. Nutting PA, Beasley JW, Werner JJ. Practice-based research answer primary care questions. *JAMA* 1999; **281**(8): 686–688.
3. Hay AD, Gorst C, Montgomery A, *et al.* Validation of a clinical rule to predict complications of acute cough in preschool children: a prospective study in primary care. *Br J Gen Pract* 2007; **57**(540): 530–537.
4. Van den Bruel A, Aertgeerts B, Bruyninckx R, *et al.* Signs and symptoms for diagnosis of serious infections in children: a prospective study in primary care. *Br J Gen Pract* 2007; **57**(540): 538–546.
5. Holm A, Nexoe J, Bistrup LA, *et al.* Aetiology and prediction of pneumonia in lower respiratory tract infection in primary care. *Br J Gen Pract* 2007; **57**(540): 547–554.
6. Holm A, Pederson SS, Nexoe J, *et al.* Procalcitonin versus C-reactive protein in predicting pneumonia in

- adults with lower respiratory tract infection. *Br J Gen Pract* 2007; **57(540)**: 555–560.
7. Van Duijn HJ, Kuyvenhoven MM, Schellevis FG, Verheij TJM. Illness behaviour and antibiotic prescription in patients with respiratory tract symptoms. *Br J Gen Pract* 2007; **57(540)**: 561–568.
 8. Herbert CP. Future of research in family medicine: where to from here? *Ann Fam Med* 2004; **2(Suppl 2)**: S60–S64.
 9. Knottnerus JA (ed.). *The evidence base of clinical diagnosis*. London: BMJ Books, 2002.
 10. Hay AD, Fahey T, Peters TJ, Wilson AD. Predicting complications from acute cough in pre-school children in primary care: a prospective cohort study. *Br J Gen Pract* 2004; **54(498)**: 9–14.
 11. World Organization of Family Doctors. *ICPC-2: International Classification of Primary Care*. 2nd edn. Oxford: Oxford University Press, 1998.
 12. Okkes IM, Oskam SK, Lamberts H. The probability of specific diagnoses for patients presenting with common symptoms to Dutch family physicians. *J Fam Pract* 2002; **51(1)**: 31–36.
 13. Anonymous. ICHPPC-2 defined. *Inclusion criteria for the use of the rubrics of the International Classification of Health Problems in Primary Care*. Oxford: Oxford University Press, 1983.

ADDRESS FOR CORRESPONDENCE

Chris van Weel

Professor of General Practice,
Department of General Practice,
Radboud University Nijmegen Medical
Centre, 117-HAG, PO BOX 9101,
6500 HB Nijmegen, the Netherlands.
E-mail: c.vanweel@hag.umcn.nl

The Quality and Outcomes Framework: too early for a final verdict

An editorial in this Journal in 2002 suggested that the new GP contract could prove to be the requiem or renaissance for general practice.¹ What has been learned 5 years on?

One obvious impact of the Quality and Outcomes Framework (QOF) has been the high quality scores achieved in the first year and the widely-publicised financial rewards. In retrospect, it is easy to see how GPs were able to score so well. Care was already improving rapidly in the years leading up to the new contract. So, for heart disease, for example, the percentage of patients with controlled blood pressure rose from 47% to 72% between 1998 and 2003, and the percentage of patients with cholesterol within recommended levels increased from 18 to 61% in the same period.² The roots for these improvements go back a decade or more. Audit was introduced as a compulsory part of the 1990 GP contract and seemed to have a modest impact at the time.^{3,4} But what happened during that decade was that GPs gradually started using electronic records, they got used to comparing their care with others, and many GPs employed nurses to improve the care of chronic illness. So, when the QOF came along, much of the infrastructure for quality improvement was already in place, and GPs were able to respond rapidly to the new incentives.

Since the QOF was introduced, quality of care shows further improvement. For asthma and diabetes, care is now improving more rapidly than before the contract. For coronary heart disease, where care was

already showing major change, the improvement has continued at the same rate.⁵ Care in relation to these three diseases has undergone definite if modest improvements over and above what was already being achieved. These trends should have some important impacts on health.⁶ Several commentators have doubted whether the improvements so far represent value for money in terms of £1 billion annual investment, but what has been achieved is a mechanism for an ongoing programme of quality improvement in new areas that is unique among national healthcare systems.

Mangin and Toop⁷ are unable to find evidence for many of the indicators in the QOF. This is hardly surprising. When a rigorous process was used to develop quality indicators for asthma, angina, and diabetes, only a quarter of indicators that GPs rated as 'necessary to do and record' were strongly evidence based.⁸ Quality indicators are always going to contain a large element of professional judgement, and that should be applauded, not derided. What is needed is a robust and transparent process for incorporating professional judgement in new indicators.

Any scheme which includes large financial losses and gains is potentially open to cheating. Cheating is hard to detect, but one aspect of the QOF that still concerns government is exception reporting. The rationale for exception reporting is that evidence-based guidelines were never intended to apply to every patient who sits down in front of his or her GP. Allowing the

GP to say: 'This indicator doesn't apply to my patient', makes it easier to align managerial with professional incentives, and to avoid inappropriate distortions of care. So have GPs abused the ability to use exception reporting? On the whole, they have not. The median exception reporting rate was 6% in the first year of the contract,⁹ and 5.3% in the second year.¹⁰ One practice exception reported 86% of its patients in the first year, but this top figure for exception reporting has come down to 28% in the second year. Primary care trusts obviously have an inspection role for practices with high rates of exception reporting, but generally, there is little evidence of widespread abuse. Other forms of gaming are hard to detect. The suggestion that GPs recoded patients to diagnoses other than 'coronary heart disease' in the run up to the contract¹¹ is cause for concern, although this could be legitimate cleaning of disease registers. However, it is a serious problem that the current payment system systematically penalises practices serving deprived populations with high morbidity.¹² The payment formula needs to encourage casefinding in areas of high morbidity, not discourage it.

It is often suggested that incentives will widen health inequalities, because doctors will concentrate on patients who are easier to treat. When incentives were introduced for cervical cytology and immunisation in 1990, inequalities widened initially but over 6 or 7 years the gap narrowed so that there was an overall halving of inequalities between