Author's response to reviews

Title: Addressing the deficiencies in the evidence-base for primary practice in regional Australia - Sentinel Practices Data Sourcing (SPDS) Project: A Pilot Study

Authors:

Abhijeet Ghosh Mr (aghosh@isml.org.au)
Karen E Charlton Ms (karenc@uow.edu.au)
Lisa Girdo Ms (lgirdo@isml.org.au)
Keith McDonald Mr (kmcdonald@isml.org.au)

Version: 3 Date: 2 July 2013

Author’s response to reviews: see over
RESUMBISSION: ORIGINAL RESEARCH PAPER FOR PUBLICATION IN BMC Public Health:
Addressing the deficiencies in the evidence-base for primary practice in regional Australia - Sentinel Practices Data Sourcing (SPDS) Project: A Pilot Study

I am pleased to submit the revised manuscript of an original research paper for consideration in BMC Family Practice. The paper proposes a novel way to monitor chronic disease prevalence on a population level through data collected routinely during patient interactions in General Practice. To date, information gathered from primary care interactions, using sentinel sites, has not been investigated as a potentially valuable surveillance system in Australia.

The study assessed the feasibility of accessing data obtained through a computerised clinical reporting tool that has been designed for desktop application (Pen Computer Systems (PCS) Clinical Audit Tool:™ PCS CAT). Collated patient data included information on chronic disease management and prevention, prevalence of health risks such as overweight and obesity, mental health indicators, medication profiling, as well as uptake of preventive health services (immunisation, cervical cancer screening and health checks). The study has shown that higher than national average estimates were found for the age-adjusted prevalence of chronic diseases such as hypertension, anxiety disorders and obesity/overweight within the cohort of the population that consult General Practice.

This study has demonstrated that the scope of data collected by patient visits to their General Practitioners, facilitated through the Medicare-funded primary health care system in Australia, offers a feasible opportunity for surveillance of chronic disease prevalence and its associated risk factors. It hence illustrates the potential of using routine general practice data to inform evidence based primary practice in Australia.

In Australia, as in most other countries, general practitioners are the first source of referral within the larger health care system, and thereby provide an optimal opportunity for effective monitoring of morbidity data of the population. The findings of the paper propose a way to implement population-level surveillance. The paper fits within the scope of the BMC Public Health journal in that the findings are applicable to improved planning for health service delivery at a local level.
As per the comments from the reviewers that were sent to us through an email from Ms Eloisa Nolasco on behalf of Prof Tony Dowell (email received on 14/06/2013) I have amended the manuscript to best incorporate the suggested changes. These are detailed as follows:

**History of Changes:**

**COMMENTS from REVIEWER 1:**

a) How valid are the case finding algorithms? There is some faith by the authors that the case finding algorithms used in the PCS CAT are valid but provide no published evidence that they are. How do they know that those identified with diabetes as an example actually have the disease? How many diabetics were missed because diabetes was not recorded in a part of the EMR that was accessible to extraction? How many patients with prediabetes were diagnosed by the program as having diabetes?

**Change:**

- The first paragraph within the METHODS section has been expanded to include:

  The Pen Computer Systems (PCS) Clinical Audit Tool™ (CAT) is a program that has been designed to collect and collate population-level health data for the purpose of chronic disease management. The program can be installed onto compatible desktop computer systems within General Practice and interfaces with the practice’s clinical software. It hence helps to simplify the review of clinical and practice management databases and facilitates the analysis of practice data through easy and accurate identification of chronic diseases and health risk prevalence. All practice clinical softwares utilise one of the several nationally validated health coding and medical classification systems such as SNOMED-CT, DOCLE, PYEFINCH and ICPC2+. These are commonly known as medical vocabularies. The data entered by clinicians within their practice softwares is hence as per these medical vocabularies which are used in their existing formats by PCS CAT to illustrate existing disease and health risk prevalence and facilitate further analysis.

- Possibilities of missing cases has been further discussed in a later paragraph in the METHODS section as follows:

  While incomplete history taking by the clinician will continue to decrease the comprehensiveness of case finding, data cleansing processes of systematically recording and allocating coded diagnosis allows the practice clinical software to draw on the coded information and hence assists in accurate and complete case identification. For example if a patient with Type 2 Diabetes Mellitus was not coded into the clinical system and Type 2 Diabetes was just inserted as a free text entry in his patient record by the GP, it would lead to this case of Type 2 Diabetes being missed in the case finding and data analysis. The data cleansing process helps to avoid this bias from occurring. The free text entry which has been inserted by the GP as per his clinical judgement, will then be replaced to the coded item of ‘Type 2 Diabetes Mellitus’ during the data cleansing process and hence the case will be illustrated as one of Type 2 Diabetes within the data extract and will be available as such during analysis. Such cleaned and correctly coded information is then easily extractable from the Best Practice clinical software using the PCS CAT tool.

b) How was the denominator estimated for calculation of prevalence?

**Change:**

- Table 2a shows the age specific counts of persons within the sample and the age specific counts of all chronic conditions in the sample. These age specific counts were used to construct age standardised prevalence figures for the sample as shown in Table 2b.

c) Figure 5 is difficult to read and the data could be displayed graphically in other ways that would be easier to interpret.

**Change:**
• Figure 5 has been redrawn as a column chart to easily illustrate the disease prevalence for all age groups

COMMENTS from REVIEWER 2:

a) Acknowledge and address bias arising from the definition of the denominator – the sample, and whether there is bias as a consequence of an ill population being more likely to visit the practice (implied by the difference between the sample age sex profile in Figure 4, and the ISML profile in Figure 3 – unsurprisingly 0-4 year olds are highly represented in the sample, whereas people in their twenties are low). At the very least there should be some estimate of the proportion of the local population who have consulted, and therefore the extent of such bias. If this correction was actually performed in the process of standardisation, then it should be reported fully. Otherwise it is hard to defend the results as a true estimate of prevalence, comparable with population prevalence from the Australian population overall.

Change:

• The second paragraph within the DISCUSSION section has been expanded to include:

This data obtained from patient interactions provides an accurate and timely picture of the major primary health care needs of patients that access general practitioner services within an identified smaller geographic region within a local government area. For example, it is clear that activities that target weight reduction and prevention of hypertension are required in the catchment area of this General Practice. Compared to estimates obtained from the ABS Australian Health Survey 2011-12, this pilot study identified higher than national figures for the prevalence for clinically diagnosed anxiety, hypertension, and overweight and obesity, but a lower prevalence of asthma, COPD, chronic heart disease, depression, osteoarthritis and osteoporosis. The prevalence of diabetes was similar to Australian national estimates. While these prevalence figures represent only the section of the population that goes to a GP and can thereby be argued to be a biased estimate of regional population disease prevalence; the figures do help primary care initiatives to target vulnerable groups that present to general practices with an identified need to address specific chronic conditions. They hence enable primary health planners to tailor healthcare services to meet local needs of the populations that actively visit general practices to receive primary healthcare services. It is not possible to determine from this single pilot practice data source whether the age-standardized prevalence figures are truly reflective of regional differences, or whether they may be a result of some disease classification and coding anomalies, nor is it possible to hypothesize why such differences may exist. This will be a topic of further larger studies.

b) The validity of the underlying disease coding is not addressed, and must be considered before a meaningful claim can be made that this approach produces “a valid and sensitive surveillance system on chronic diseases”. The authors should note the extensive validation conducted on the UK GPRD (see for example the 2010 review of validations by Herret in BJ Clin Pharmacology, which makes recommendations for assessing and reporting validity).

Change:

• The paper suggested by the reviewer has been utilised and added as a reference (no. 33) and the second last paragraph of the DISCUSSION section now includes the following:

One of the main benefits of using general practice databases for disease surveillance is the ability to access data from large patient populations across a wide population coverage. However, the data are collected primarily for clinical and routine use, rather than for surveillance or research purposes. Hence, data quality and reliability may be compromised [33] and additional data cleaning is required before being extracted to a usable format, as was performed in the present study. The SPDS study relies on the clinical judgement of GPs for accuracy of disease diagnosis and assumes inclusiveness of GP history-taking, clinical
inquiry and data recording skills. The validity of medical terminology and coding systems such as PYEFINCH, DOCLE and ICPC2+ also need further standardisation. With the introduction and implementation of new e-health requirements in Australia in February 2013, the National E-Health Transition Authority (NEHTA) has advised all General Practice clinical software vendors to start working towards mapping their local medical vocabulary against SNOMED-CT, that has been identified by NEHTA as the preferred nationally recognised disease classification or terminology system [34]. In due course, coded General Practice clinical data in Australia may be considered a reasonably valid and reliable source of information for mapping regional disease prevalence and conducting surveillance of chronic conditions. Such data is particularly useful for planning of primary care services to meet the needs of local populations.

c) The interpretation of key results is idiosyncratic. In the abstract and discussion the authors draw attention to anxiety, hypertension and obesity, where they observe a higher than national rate, but their results actually found much lower than national rates for a number of diseases, including heart disease, osteoarthritis, osteoporosis and COPD. COPD in particular, is less than 1/3rd of the national rate. Is this real (in which case it would be interesting to have some discussion about why it is the case)? Or is this an artefact of coding problems, as noted above (in which case it rather undermines the claim to validity)? These results should be discussed more fully, and their relevance to the overall validity of the dataset considered.

Change: -
- The second paragraph within the DISCUSSION section has been expanded to include: -

This data obtained from patient interactions provides an accurate and timely picture of the major primary health care needs of patients that access general practitioner services within an identified smaller geographic region within a local government area. For example, it is clear that activities that target weight reduction and prevention of hypertension are required in the catchment area of this General Practice. Compared to estimates obtained from the ABS Australian Health Survey 2011-12, this pilot study identified higher than national figures for the prevalence for clinically diagnosed anxiety, hypertension, and overweight and obesity, but a lower prevalence of asthma, COPD, chronic heart disease, depression, osteoarthritis and osteoporosis. The prevalence of diabetes was similar to Australian national estimates. While these prevalence figures represent only the section of the population that goes to a GP and can thereby be argued to be a biased estimate of regional population disease prevalence; the figures do help primary care initiatives to target vulnerable groups that present to general practices with an identified need to address specific chronic conditions. They hence enable primary health planners to tailor healthcare services to meet local needs of the populations that actively visit general practices to receive primary healthcare services. It is not possible to determine from this single pilot practice data source whether the age-standardized prevalence figures are truly reflective of regional differences, or whether they may be a result of some disease classification and coding anomalies, nor is it possible to hypothesize why such differences may exist. This will be a topic of further larger studies.

d) Fully report data manipulation. The method section refers to data cleansing (as does the author contribution statement). But the nature of the data cleansing is not made clear. What were the issues in the raw data which had to be corrected? Data cleansing is to be expected, but transparent reporting of method would require the nature of such data pre-processing to be explicit.

-----and------

e) Fully report the coding system. The underlying coding system for morbidity is not clearly stated. Was it ICPC, SNOMED or ICD-10 (ICD-10 is implied, but not stated clearly, in the
discussion section, but this would be an unusual coding system for routine data in general practice)?

Changes: -

- The METHODS section now includes the following:

  The selected pilot site for the project was the newly formed GP Super Clinic - Shell Cove Family Health (SCFH) located in the Shellharbour LGA of the Illawarra-Shoalhaven catchment of New South Wales, Australia. The practice clinical software was Best Practice which uses the PYEFINCH medical coding vocabulary. Practice staff were informed about the aims and objectives of the study assisted the researchers in undertaking the data cleansing process on their clinical software system. The data cleansing phase of the study was conducting using the data maintenance utility tool which is available within all GP clinical software. Within Best Practice this is called the Cleanup history tool. Data cleansing included:

- encouraging all practice staff to use the ‘drop down box functionality’ of Best Practice to define all medical diagnoses and other sections of the patient record;
- strictly avoiding free text entries in all sections of the patient record;
- finding all identifiable free text non-coded past medical history items, and either linking them to appropriate coded items or replacing them with the correct coded item; and
- coding all inactive patients as ‘Inactive’. An ‘active patient’ is one who has attended the practice three or more times in the past two years as defined in the RACGP Standards for general practices [16].

f) Amend Figure 5. A polar plot of prevalence by age is an unconventional and unduly complex presentation, which actually obscures the patterns in the data. I strongly recommend a simple x-y line plot for this diagram, which is likely to be much more accessible to the reader, and will display the patterns by age much more clearly to the eye.

Change: -

- Figure 5 has been redrawn as a column chart to easily illustrate the disease prevalence for all age groups

I look forward to further correspondence regarding this article.

Yours sincerely

Abhijeet Ghosh

Abhijeet Ghosh
Illawarra Shoalhaven Medicare Local
Email: aghosh@isml.org.au
Phone: +61 (0)2 4220 7600