

## 3 Lessons learnt

It is relatively rare for a research team to have the opportunity to conduct so many studies with the same basic methodology, among the same population, over such a long period. We have learnt many things over the last nine years in the SAND substudies and we have developed a number of rules of thumb when designing SAND research tools. We freely admit that the later SANDs are far better than the earliest SANDs, as we have moved through a continuous quality improvement program over the years.

In SAND substudies we are using the GP as an expert interviewer of his/her patient and are utilising his/her knowledge of the patient rather than relying purely on patient recall. The GP is an ideal person to ask patients about their health and health issues. Patients expect GPs to ask them questions about their health and their health behaviours and (we hypothesise) are likely to be more honest with their GP than with an interviewer (unknown to the respondent) approaching them for self-reported health information.

Some GPs have reported that some patients appear relieved when asked for specific information about their, for example, alcohol intake, as if this had concerned them but they had not raised it with their GP, or had not made their consumption levels clear to their GP on earlier occasions. Others report that they find completing the patient risk behaviours SAND provides them with an 'excuse' to raise these issues with their patients.

However, while some of the SAND substudies may assist the GP in initiating discussion with their patient about a specific topic, the GPs time must always be considered. If the additional time needed to ask the questions is too great they will simply drop out of the BEACH study altogether.

In an environment of increasing workforce shortages, many GPs are strapped for time. The tools we design for SAND must therefore be brief, simple to administer verbally and quick to complete. However, they must also provide valid and reliable results.

### 3.1 Rules of thumb in SAND tool design

#### **Remember your research impacts on a busy GP's day**

Completing the questions on the SAND form can sometimes be difficult for the GP. We tell GPs to leave the SAND section blank if they feel unable to ask the questions of a particular patient. This approach considers the nature of general practice, for example:

- the patient may not be seeing their regular GP and may not be comfortable answering specific health questions not related to the problems managed at the encounter
- the type of encounter (e.g. telephone encounter) and the parallel need for patient agreement for their data to be included in the BEACH program may make completion of the questions impossible
- the morbidity being managed (e.g. crisis situation such as suicide attempt, relationship breakdown, acute bereavement) may mean it is inappropriate for the GP to ask the SAND questions, which are not related to the crisis under management
- the patient (e.g. the cognitively impaired) may not be competent to give informed consent to the use of their data, or to answer the questions being asked.

## Impact on GP time

In any study using GPs to complete information about clinical activity or about patients, the participating GPs are being very generous in giving their time to the study. If you ask for too much it can result in GP overload and withdrawal from the study.

Each GP has three sets of SAND forms across their 100 BEACH forms. Therefore we have to consider the GP workload in terms of the combined effect of the three SAND topics. Some topics require the GP to ask every patient encountered all questions in that SAND. Examples include *Patient cultural background* (Abstract 95) and *Prevalence of chronic illnesses identified as National Health Priority Areas* (Abstract 61). Other SAND surveys have a filter question that limits the remainder of the questions to those who say 'Yes' to the first question. Examples are: *Prevalence and management of chronic pain* (Abstract 82) and *Lipid management in patients with high risk conditions* (Abstract 99).

In the 40 patient risk factor SAND forms in each GP's research pack (see Chapter 4), start and finish times for the consultations, and patient-reported height and weight are required for all patients and the alcohol and smoking questions are asked of all patients aged 18 years or more. We try to ensure that in a research pack we do not include two other SAND topics that require all questions to be asked of all patients, so that the GP will get at least one SAND topic that relies on a filter question, thus reducing the overall workload.

## Filter questions

When possible, for the topic under investigation we use a filter question to identify the patients of interest who meet certain criteria. For patients who have been identified (for example, patients with a specific condition) GPs are asked to complete additional questions. For patients who do not meet the criteria for inclusion, GPs are instructed to end the questions. Examples of filter questions are provided in Box 1.

### Box 1: Examples of filter questions used as the first question in a SAND survey

Does this patient have Type 2 diabetes? <input type="checkbox"/> Yes <input type="checkbox"/> No → end questions	Does this patient have any of the following risk factors? (Tick all that apply) <input type="checkbox"/> Existing CHD <input type="checkbox"/> Diabetes mellitus <input type="checkbox"/> Familial hypercholesterolaemia <input type="checkbox"/> Elevated cholesterol <input type="checkbox"/> Family history of CHD <input type="checkbox"/> Peripheral vascular disease <input type="checkbox"/> None of the above → end questions
--	--

## Enumerating and describing not judging or criticising

The BEACH study aims to describe what is really happening in general practice. We therefore try to avoid creating scenarios where the GP might feel that his/her management is being questioned or judged. The last thing we want (for both data reliability and continued participation) is for clinicians to feel they have to justify their management. For example, the box at the top of the GP instructions for *Prevalence, cause and severity of adverse pharmacological events* (Abstract 56), emphasises that this SAND is an investigation of the impact of pharmacological adverse events in the community, and that the GP may be unaware of the event, either because the patient did not inform them, or the event was related to a medication provided elsewhere. The study is not asking about clinician error or seeking to assign blame anywhere or to anyone for the adverse event.

## **Not intruding on the GP–patient relationship**

When designing a SAND the relationship between the GP and patient is always considered.

A simple example is our choice not to ask patients aged between 14 and 17 years the SAND questions on current smoking status and usual alcohol consumption. While it is certainly legal for the GP to ask such questions of patients in this age group we believe that they would often be intrusive on the GP–patient relationship. Teenagers show a massive decrease in attendances with GPs at about 15 years of age (boys more so than girls)<sup>21</sup> so this is a patient group that should be encouraged to build a relationship with their GP. Such questions (purely for research purposes) may well challenge the likelihood of further visits.

In other SANDs, such as the study on the prevalence of premature ejaculation (Abstract 93) we specifically tell the GP to stop the questions ‘if you feel at any stage that these questions intrude too greatly on your relationship with this patient’.

## **Sensitive topics**

Some topics covered in SAND can be sensitive. With practice we have learnt that in these cases we should provide the GPs with information about the issue and the reason we are investigating. For example, for the study on premature ejaculation (Abstract 93), we provided background information about the lack of data generally available on this topic.

Some questions may appear somewhat invasive if asked directly so in such circumstances we have found it best to ‘couch’ the sensitive issue as one part of multiple categorical options with tick boxes. For example, *Management of depression and anxiety* (Abstract 47) in the third question, the research interest really centred on the possible effect of the specified medications on sexual function. Rather than ask a direct question on this possible effect, we ‘buried’ it among a series of six possible effects, softening the question asked of the patient.

## **The question of patient honesty**

Sometimes we need to find out whether a patient is doing something that they are not supposed to be doing. The question might then arise as to the likelihood of an honest patient response. We therefore avoid using questions where honesty may be compromised.

Again the questions about smoking status and alcohol consumptions among patients of 14–17 years are good examples. We are uncertain of the likelihood of receiving an honest answer from these young people if they are accompanied by a parent/relative/carer or where the GP is the usual doctor seen by other members of their family.

## **The way you ask the question has a direct effect on the result**

### **The use of tick boxes versus free text**

Whenever possible we use categorical response variables (tick box options) rather than open ended questions requiring free text entry. This provides a more reliable response. For example in the first study on patient comorbidity (*Prevalence of common morbidities in patients encountered in general practice*, Abstract 37) we simply asked the GP to list in free text all (up to 12) ‘other significant diagnoses/problems... not managed at today’s encounter’. In later studies (e.g. *Prevalence of chronic illnesses identified as National Health Priority Areas among general practice patients*, Abstract 61) we asked ‘Does this patient have any of the following conditions which require ongoing management?’ and offered a series of tick boxes of morbidities related to the National Health Priority Areas (categorical responses). We believe the later method generated more reliable results, which demonstrated repeatability in a later study (*Estimates*

of the prevalence of chronic illnesses identified as Health Priority Areas among patients attending general practice, Abstract 89).

### Three different measures of the prevalence of asthma among general practice patients

The structure of the question (described above), combines with the independent influence of the SAND topic, on the final results. For example, we have undertaken 10 SAND studies that include a filter question on whether or not the patient has asthma, to generate estimates of asthma prevalence among patients attending general practice.

Sample sizes range from 2,500 to 11,300 respondents. These are summarised in Table 1.

- Between March 1999 and December 2004 there were six SAND studies that used a filter question. GPs were instructed to ask each patient if they *'currently suffer from asthma?'* The six studies produced remarkably consistent results, even without age-sex standardisation for minor difference in the age-sex distribution of the respondents in each sample. Prevalence of asthma was consistently estimated to be between 12.8% and 14.7%, with no significant differences between results, as judged by overlapping 95% confidence intervals around the estimates (see Abstracts 3, 22, 39, 48, 63, 70).
- In the midst of this 1999–2004 period, we ran a SAND which asked the GP to describe in free text the patient's major comorbidity not managed at the encounter, including: *'chronic illnesses or other health problems that requires continuing management or surveillance; past problems with may need consideration in future care; any significant health influencing social problems'*. Spaces were provided for up to 12 free text descriptors (Abstract 37). The estimated prevalence for asthma was significantly lower at 8.8% (95% CI: 8.1–9.5).
- In late 2003–early 2004 we again investigated the issue of comorbidity using categorical tick box responses for a list of selected national health priority areas (including asthma). The GP was asked: *'Does this patient have any of the following conditions which require ongoing management'*. This method generated an asthma prevalence estimate of 11.4% (95% CI: 10.5–12.3) (Abstract 61). Using the same method in 2005 we gained an estimate of 10.7% (95% CI: 9.8–11.6) (Abstract 89).

**Table 1: Summary of methods, samples and results for asthma prevalence across multiple SANDs**

Method	Abstract no.	Page	Number of respondents	Data collection period	Estimated prevalence (95% CI)
Filter question	3	32	4,285	03/99–06/99	14.7 (13.3–16.1)
Filter question	22	75	5,495	11/00–01/01	12.8 (11.4–14.3)
<i>Free text—Comorbidity</i>	37	109	11,342	08/01–03/02	8.0 (95% CI: NA)
Filter question	39	113	3,070	04/02–05/02	13.9 (12.0–15.7)
Filter question	48	132	2,686	09/02–10/02	14.5 (12.7–16.2)
<i>Categorical (NHPAs)</i>	61	160	8,911	08/03–01/04	11.4 (10.4–12.3)
Filter question	63	164	2,527	09/03–10/03	14.5 (12.6–16.1)
Filter question	70	181	7,919	09/04–12/04	13.0 (11.9–14.0)
<i>Categorical (NHPAs)</i>	89	223	9,156	07/05–11/05	10.7 (9.8–11.6)
Filter question	96	238	5,911	02/06–06/06	11.6 (10.6–12.7)

Note: NHPAs—National Health Priority Areas; NA—not available.

- Most recently, in mid-2006 we repeated the asthma filter question (Abstract 96) and gained a similar result 11.6 (10.6–12.7) to previous asthma SAND substudies with filter

questions but one that is suggesting a trend towards lower prevalence of asthma among patients attending general practice. It will be interesting to see if this trend continues.

We conclude that the use of a filter question centred on the topic of asthma provides a slightly higher estimate of prevalence than the inclusion of asthma as one of many morbidities with tick boxes provided, and that free text recording of comorbidity present in the patient gives the lowest prevalence estimate. Considering the relative consistency of the results, we would regard the free text recording of comorbidity as less reliable than the other two approaches.

### **Structures to assist in counting the true level of missing data**

The structure of the questions should ensure that a response is possible from all who should answer it, even if the response is 'Don't know'. For example: if you only offer 'Yes' and 'No', where the information is not available the GP will leave it blank. If you offer 'Don't know' in addition to the other options, where the information is not available the GP can tick this option, thus reducing the size of the missing data. Such methods allow a better estimate of the true size of missing data. Keep in mind that where missing data for a question are greater than 5% the reliability of the result is in question.

### **Questions on clinical opinion**

In SAND we sometimes ask GPs to make a judgement based on their clinical opinion rather than based on, for example, formal guideline objectives. Judging a patient's progress based purely on guidelines does not consider the rest of the patient's 'health picture'. The patient may have other complex morbidities or personal history which means that they may not meet the guideline targets for management but in the GPs clinical opinion are meeting targets that are as good as could be expected in that patient. This is a wholistic approach rather than a clinical trial approach. An example can be found in *Diabetes Types 1 and 2 and coronary heart disease* (Abstract 86) where the question on the adequacy of the current control of the patient's cholesterol relies on clinical opinion.

### **The need for definitions**

In many cases the diagnostic label applied to a problem is clinical opinion rather than proven by tests, or by strict application of definitions (many of which have been designed in secondary and tertiary care). If asking whether the patient has been managed for depression in the previous 12 months, there is no point in providing a DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th edition) definition of depression – the patient either has, or has not, been managed for depression in the past 12 months, however it was defined at the time by the GP, a psychologist or a psychiatrist.

At other times providing a definition is essential to gain internal consistency among GP responses. For example in *Prevalence and management of chronic pain* (Abstract 42) we relied purely on the GP's clinical judgement in deciding whether or not the patient suffered from 'chronic pain'. When trying to publish the results internationally we were criticised for not selecting and providing the GPs with any one of a myriad of available definitions of 'chronic pain'. In Abstract 82 we repeated this topic and included a definition of 'chronic pain' to gain more reliable results.

We also provide definitions for disease attributes when there are standard definitions available. Abstract 48 is an example. We provided definitions for asthma severity among children and among adults, using the National Asthma Council's severity classification.

## Questions that are not answered well

Some questions we have tried are answered poorly compared with others. These include:

- Questions that require the GP to look back over their clinical notes to find the answer. For example, lipid levels prior to starting medication – the patient may have started medication for lipid treatment more than 5 years ago and locating the record of the lipid level at the start of medication treatment may be difficult. Further, the patient may have started on treatment when at another practice and records may not be available to the current GP. Reliability of patient recall of such a result would probably be poor.
- Medication regimen information (i.e. complete dosage information). This is well recorded in the encounter form but not in the SAND form (unless the medication was prescribed today). Though the medication name appears to be easily recalled by the GP or patient, strength and dosage information are poorly reported. We hypothesise that this is because the GP may not be the prescribing doctor and the patient does not recall the details. Alternatively the GP may need to look through past notes to find the regimen data.
- Questions on duration of condition or duration of a treatment or medication, for example, *How long ago was the patients diagnosed with condition x? Specify number and circle weeks/months/years.* Depending on the time that has lapsed, if the data are not in the medical record the recall of the patient or the GP may be limited. If the research question can be answered with multiple choice, then use this option (e.g. <1 year, 1–3 years, > 3 years) as reliability will be far greater.

Always remember that in the current Australian health care system a single patient may have multiple complex problems, may see multiple GPs and may be managed simultaneously by many health professionals. Do not assume that a complete health record for the patient is available to the GP.

## Use of a suitable timeline for patient/GP recall

Length of accurate recall by patients and GPs can be limited. 'Within the last 6 months', 'during the last 12 months' or 'in the past' will provide results with varying degrees of accuracy. However, the most suitable recall time period will depend on the subject under investigation. For example: a question on whether the patient had been hospitalised could be asked for the previous 12 months, as it is a reasonably major event in anyone's life and therefore likely to be recalled. It also has a high chance of being in the patient's medical record.

However, when asking about the grade of pain suffered by people with chronic pain (see Abstract 82), we limited the time period to the last week. Using a 12 month recall period here would require the patient to try and average all his/her ups and downs in pain levels over a 1 year period and would be highly unreliable.

## Inadequate planning with insufficient space on the form

In a few SANDs we have asked about current medication for the morbidity of interest and allowed space for only one medication to be recorded. On receipt of the completed forms we found a number where the GP had recorded two medications by squeezing them into the allotted space. This means that at least two spaces should have been provided. However, having limited the GPs to one, we cannot assume that those who recorded only one, did so

even in cases where multiple were involved. Such errors lead us to question the validity of the final result about medications for that SAND.

### **Is general practice an appropriate setting for the subject?**

Not every subject of patient health and lifestyle is appropriate to investigate in a GP-based study of this type. There is one SAND study not reported in this book because it failed. The subject of the SAND was the amount and type of physical activity performed by the patient in the week prior to this encounter. So many patients reported that they had no exercise in the previous week, due to ill health, that we rejected the study as inappropriate. A question on average exercise levels over the previous month may have been more productive.

## **3.2 Methodological issues**

### **Test–re-test reliability of the SAND studies**

Where a SAND question has been repeated on a number of occasions we have an opportunity to investigate test–re-test reliability of the measure among different groups of GPs asking the questions of different sets of patients. If the method is to be regarded as reliable, prevalence estimates of a morbidity should remain constant, except where there has (in truth) been a change in the prevalence of the morbidity in the population at large.

Henderson et al. compared estimates of asthma prevalence generated from four SAND studies conducted in four different 5-week periods between March 1999 and October 2002. They demonstrated no significant difference in the estimated prevalence of asthma and no significant differences in severity levels over time for either adults or children.<sup>22</sup>

Similarly, Miller et al. demonstrated consistent results among three SAND studies that investigated adverse drug events experienced by responding patients over the previous 6 months. The subject was repeated twice because we were surprised by the result in the first SAND block – that 10% of patients seen by the recording GPs had experienced an adverse pharmacological event during the previous 6 months. The repeats produced the same result, demonstrating the reliability of the method.<sup>23</sup>

### **Recording effect**

It is possible that asking SAND questions about a specific topic will affect the frequency of management of that problem at the encounter, and/or the management actions at the encounter for the morbidity studied in the SAND. This has been investigated for the patient risk behaviours: weight (BMI), alcohol and smoking, to test whether asking these questions results in an increased likelihood of them being managed at the encounter. This investigation found no significant effect.<sup>7</sup> However, through our observation of the data we believe it may influence GPs' encounter behaviour in certain circumstances, though probably only for the encounters including that SAND topic. For example, we hypothesise that asking for the most recent HbA1c result for a patient with diabetes, and the time since that last test was done, may increase the chance that the GP will order a HbA1c at that encounter for the patients who have not had this tested for more than 6 months. We plan to undertake further analysis to determine if a range of SAND substudies influence the content of that encounter.