The quality of data on general practice

A discussion of BEACH reliability and validity

Helena Britt

BA, PhD, is Associate Professor and Director, Australian GP Statistics and Classification Centre, University of Sydney, New South Wales. helenab@med.usyd.edu.au

Graeme Miller

MBBS, PhD, FRACGP, is Associate Professor and Medical Director, Australian GP Statistics and Classification Centre, University of Sydney, New South Wales.

Clare Bayram

BAppSc(Hons), is Research Officer, Australian GP Statistics and Classification Centre, University of Sydney, New South Wales.

BACKGROUND
The quality of data used to measure quality is as important as quality of care itself. Any organisation undertaking studies within general practice should ensure the methods they adopt meet these 'best practice' standards, to ensure and demonstrate the reliability, representativeness and quality of their data. The BEACH (Bettering the Evaluation and Care of Health) program is a continuous national study of general practice activity in Australia that began in April 1998.

OBJECTIVE
This article describes the methods used to ensure and test BEACH data representativeness and reliability, and responds to frequently asked questions.

DISCUSSION
BEACH measures the process of care at general practitioner-patient encounters including problems managed and treatments provided. BEACH data are used by the profession, government, researchers and industry. We discuss how BEACH data differ from Medicare Benefits Schedule, the Pharmaceutical Benefits Scheme and National Health Survey data. We also consider challenges to achieving national electronic collection of health data in general practice in Australia.

The quality of data used to measure quality is as important as quality of care itself. The Quality Framework for General Practice being developed by The Royal Australian College of General Practitioners (RACGP) considers multiple aspects of quality. The background paper[1] quotes Donabedian's three components of quality:

- structures: material resources, facilities, equipment, range of services at practice
- processes: what is done in giving and receiving care
- outcomes: effects of care on health status of the patient and the community.[2]

Their measurement requires different tools. For example accreditation was designed around measurement of structures, the BEACH program was designed for the measurement of process, and in Australia we do not have a good national measure of health outcomes.

BEACH provides us with information about the content of general practitioner-patient encounters and the problems managed and treatments provided by GPs to the Australian community. It is the only continuous national randomised study of general practice activity in the world, and the only

one that provides direct linkage of management actions to the problem under management.

The BEACH data are used by a wide range of people: the profession and its organisations, government, researchers and industry. We are often asked questions about how reliable and representative BEACH data are, and we address these frequently asked questions in this article.

What is the BEACH method?

BEACH relies on the cooperation of randomly selected GPs across Australia. Each completes details for 100 consecutive patient encounters on structured encounter forms and provides information about themselves and their practice. About 1000 GPs participate each year and participants change each week. Participants gain points toward their quality assurance requirements for continued vocational registration.

The sample frame for the study is all vocationally registered (VR) GPs and registrars who claimed at least 375 A1 Medicare items of service in the most recent quarter. The Australian Government Department of Health and Ageing (DoHA) draws the samples from Medicare claims data.

What are the response rates?

In the 2005–2006 BEACH year we approached 3620 randomly selected GPs. Almost 10% could not be contacted, most having moved (untraceable). The final sample included 1017 GPs, 31.1% of those who were contactable. The inaccuracy of the contact details provided in the Medicare data makes GP recruiting difficult. Although we manage to locate most of those with inaccurate contact details, the proportion of all sampled GPs who had died/retired/moved (untraceable) doubled from 4.0% in 2003–2004 to 8.3% in 2005–2006, the untraceable proportion being highest among young GPs (27.5%).

The sample is 1000 GPs per year. How many individual GPs have participated to date?

Over the first 8 years of BEACH, 7991 GPs provided details for 799 100 encounters. General practitioners can only be selected once per quality assurance triennium but it is possible to be selected in multiple trienniums. All evidence indicates the 7991 participants represent 6463 individuals, more than one-third of all VR GPs and registrars (approximately 17 500).

If there is a 30% GP response rate, is this representative?

We test this question every year in BEACH. The DoHA provides Medicare data about all the GPs (as a group) in the sample frame (ie. all those who had a chance of being selected). We test whether final participants are representative of all in the sample frame. If there are no differences we have an extremely good representation of all the GPs. If they differ, then we have to apply statistical adjustment.

In 2005–2006 (and in all BEACH years) the BEACH GPs and all GPs in the sample frame did not differ in terms of gender, place of graduation, and distribution across RRMA classes.[3] However, BEACH GPs were significantly older and differed in their state distribution when compared with the total sample.

We have demonstrated there are very few differences in clinical activity of GPs practising in different states and territories,[4] so differences in state distribution have little impact on the national picture.

The relatively consistent under representation of young GPs in BEACH is probably due to the fact that about one-quarter of the youngest age group is not traceable. We deal with this by giving more weight (statistically) to the encounters recorded by young GPs and less weight to those from the older age groups according to a formula derived from Medicare statistics.

There is another statistical step needed. In BEACH, each GP provides details for 100 encounters, but GPs differ in the number of services provided in a given year. For example, Dr X and Dr Y each complete 100 encounter records, but Dr X claimed 1500 Medicare A1 items in the previous quarter and Dr Y claimed 500. As we are aiming to represent all encounters, Dr X's BEACH forms must contribute to the BEACH database at three times the rate of Dr Y's. We therefore weight each GPs 100 records by a factor derived from the number of Medicare A1 items claimed in the previous year (data supplied by the DoHA).

Are GPs who participate in BEACH as busy as those who don't?

We address this question by comparing the participants with those who declined, in terms of the number of Medicare A1 items of service claimed in the previous quarter (data supplied by the DoHA). We found no significant difference ($p=0.75$) between the participants (mean number of A1 MBS items claimed 1300 claims) and those GPs who declined to participate (1309).

BEACH aims for a random sample of encounters, but randomisation is of GPs and then the encounters of each are recorded – is that representative of all encounters?

It is true that BEACH uses a 'cluster sample' design. This means that while we have a random sample of GPs, each GP records information about a 'cluster' of 100 encounters. Each 'cluster' has its own characteristics, being influenced by GPs characteristics and geographic location. Ideally we would like a true random sample of all the GP-patient encounters, but this is impossible in the current Australian health care system. So we are stuck with a cluster sample design – which must be used in any data collection in general practice – whether on paper or from electronic health records (EHRs). This is not a bad thing – it just means we need a very large sample of GPs with a cluster as small as possible and we have to 'adjust for the cluster' with statistical programs.

The sample size of 100 000 encounters from a random sample of 1000 GPs was scientifically determined. We tested different sample sizes and found the 1000 x 100 sample was the best balance between cost, statistical power and validity.[5] We use computer programs in a statistical package (SAS V9.1) to adjust for the effect of the cluster design. Adjusting for the cluster does not change the resulting estimates.

Are the encounters representative of real general practice activity?

This is a difficult question to test. There are no other national data that provide comparative information about all GP-patient encounters. However, the Medicare Benefits

Schedule (MBS) has the age-gender distribution of patients at all Medicare A1 items of service. It does not include the Department of Veterans' Affairs (DVA) paid encounters, nor those paid for by others (eg. state/territory governments, workers' compensation). To test the representativeness of the final encounter sample we compare the age-gender distribution of patients at BEACH A1 Medicare claimable encounters with all encounters claimed as Medicare A1 items in same period.

Even the raw unweighted BEACH data are in excellent agreement with the MBS in terms of the age-gender distribution of patients at encounter (*Figure 1*). BEACH varies from the MBS data by less than 10% (from total Medicare A1 items) in each age-gender group. The tight range of raw precision ratios (0.9–1.1) proves that the BEACH sample of encounters is an excellent representation of Australian general practice patient encounters claimed as A1 items of service.

What is the reliability and validity of other aspects of BEACH?

In the development of a database such as BEACH, data gathering moves through specific stages: GP sample selection (discussed above); cluster sampling around each GP (above); GP data recording; secondary coding and data entry. At each stage the data can be invalidated by the application of inappropriate methods. Although we have considered and tested these processes, others rarely question such issues.

Our earlier work demonstrated: the extent to which a random sample of GPs recording information about a cluster of patients represents all GPs and all patients attending GPs;[6] the degree to which GP reported patient reasons for encounter and problems managed accurately reflect those recalled by the patient;[7] and the reliability of secondary coding of reasons for encounter (RFE)[8]

and problems managed;[9] and the validity of ICPC as a tool with which to classify the data.[10] Quality control measures are applied regularly to all data entry by: use of health information management undergraduates trained in medical terminology and classification; CQI training methods; software locks to reduce error; one in 10 check of all coded forms against the paper original; and cleaning checks using both Access and SAS.

Are the problem labels GPs record on the BEACH form accurate?

Studies have found wide variance in the way GPs perceive and describe the problem under management. Label selection may be influenced by clinical experience,[11] by the individual's perception and inclination to 'bundle' (under one problem label) or 'split' the stimuli,[12] and by the management decisions already made.[13,14] Alderson[15] suggests that to many practitioners 'diagnostic accuracy is only important to the extent that it will assist them in helping the patient'. Morbidity statistics from family practice should therefore be seen as 'a reflection of the physician's diagnostic opinions about the problems that patients bring to them rather than an unarguable statement of the problems managed'.[14]

These issues should be borne in mind, as they apply equally to all general practice data – whether actively collected from encounter forms or paper medical records, or passively collected from EHRs.

What information does BEACH give us that is not available via MBS, PBS and NHS data?

There are other national sources of data about some aspects of general practice but they have substantial limitations in measuring activity and quality. The Pharmaceutical Benefits Scheme (PBS) counts only prescribed medications that are a cost to the Australian government, not those supplied by GPs and advised for over-the-counter purchase. It counts the number of times a particular drug is filled if it is subsidised by the PBS for that patient. Changes in the minimum subsidy level render measurement of changes in prescribing and its quality extremely difficult.[16] The PBS holds no data on the problem being managed with the medication, and assumptions of indication are unreliable.

The MBS collects data on those GP services that are billed to Medicare and does not usually include services funded through the DVA. The MBS holds the item number and some patient demographics but no information about content of the consultation.

The MBS also includes data on pathology and imaging tests paid for by the government. These do not reflect the
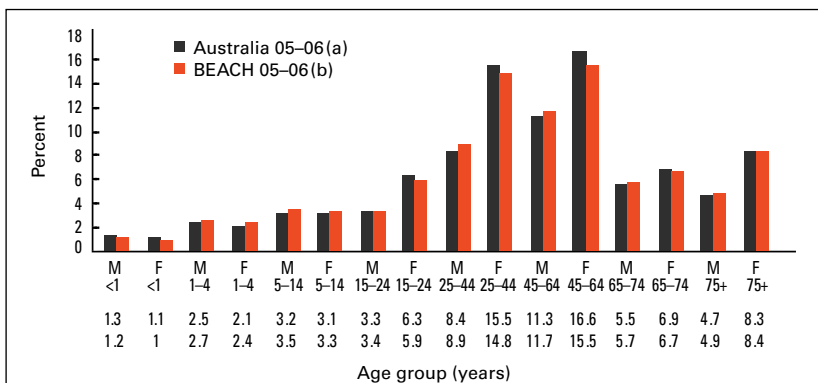

Figure 1. Raw unweighted BEACH data vs. MBS data
(a) Unweighted data, A1 items only
(b) Data provided by the Australian Government Department of Health and Ageing
Note: Only encounters with a valid age and gender are included in the comparison

	M <1	F <1	M 1–4	F 1–4	M 5–14	F 5–14	M 15–24	F 15–24	M 25–44	F 25–44	M 45–64	F 45–64	M 65–74	F 65–74	M 75+	F 75+
	1.3	1.1	2.5	2.1	3.2	3.1	3.3	6.3	8.4	15.5	11.3	16.6	5.5	6.9	4.7	8.3
	1.2	1	2.7	2.4	3.5	3.3	3.4	5.9	8.9	14.8	11.7	15.5	5.7	6.7	4.9	8.4

tests ordered by the GP as each pathology company can respond differently to a GP order; the pathology companies can only charge the MBS for the three most expensive tests undertaken. The MBS data include only those charged for and groups them on the basis of cost, therefore specific test rates are often not assessable.

For MBS imaging data coning is not an issue, but radiologists can decide whether the test is ordered by the GP, and can add other tests of their choosing. The MBS data therefore reflect tests undertaken by the radiologists (grouped according to cost) – not what the GP orders.

The National Health Survey (NHS) provides estimates of population prevalence of specific diseases and a measure of the problems taken to the GP by people in the previous 2 weeks. Data are based on self reported morbidity from a representative sample of the Australian population using a structured interview to elicit health related information from participants.[17] The survey gives little information about GPs' management of these morbidities.

Why isn't BEACH electronic?

The BEACH program, being paper based, is labour intensive for GPs and for secondary data entry. We recognise the need to move to national electronic data collection. The BEACH instrument and methodology provide an excellent starting point for developing any future electronic data collection from general practice. However, national electronic data collection will require:

• all GPs to use EHRs to have equal chance of selection (currently about 90% use them and only about 20% use all functions[18])

• the adoption of standardised data collection in all GP clinical software systems (this has already been developed for the General Practice Computing Group [GPCG][19])

• adoption of standard coding and classification systems and uniform application of these in all GP clinical software systems, and

• resolution of privacy and confidentiality issues.

The methodological studies leading up to BEACH, and the BEACH program itself, have demonstrated that it is not necessary to collect all data for all patients all of the time, to gain a reliable national picture of general practice activity. Electronic data collection (PC or web based), in which randomly sampled GPs record data for all the necessary BEACH data elements for a sample of patients could be introduced as an integrated process with the GP's desktop EHR software. The relevant data already recorded in the EHR could be transferred to a 'plug in' data collection tool. At the end of the encounter any BEACH data fields remaining empty could be highlighted for the GP's manual addition of information where required. A GP would only need to provide complete data for a sample of encounters, as in the current BEACH program.

When such a system proved reliable (as tested against parallel BEACH paper based data), and random sampling was possible (when all GPs are using EHRs) paper based data collection could be phased out. A move to passive data download could be made once all GPs used complete EHRs and standards were implemented and rigorously applied in all clinical systems. However, the same methodological rigour is needed as that undertaken for the development of BEACH.

Conclusion

Quality of data is important in any measurement of the activity of GPs. If you are measuring the impact of policy on practice, bad data may lead to the wrong conclusion and to inappropriate policy change. BEACH demonstrates that you can get quality data from a sample – you don't need the whole. We look forward to applying sampling techniques in future electronic data collection programs. The data will never be perfect in a clinical setting, no matter the method you use to collect it, but as BEACH shows, we can apply methodological rigour to make it as good as possible. Any study relying on GPs or practices for clinical data should consider all the issues discussed in this article if their data are to be representative. This is 'best practice' and should be followed by all organisations collecting clinical data.

References
1. Booth B, Snowdon T, Lees C. A quality framework for Australian general practice: Background paper. Melbourne: The Royal Australian College of General Practitioners, 2005.
2. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260:1743–8.
3. Britt H, Miller GC, Charles J, et al. General practice activity in Australia 2005–06. Canberra: Australian Institute of Health and Welfare, 2007.

4. Britt H, Miller GC, Knox S, et al. General practice activity in the states and territories of Australia 1998–2003. Available at www.aihw.gov.au/publications/index.cfm/title/10045.

5. Meza RA, Angelis M, Britt H, Miles DA, Seneta E, Bridges-Webb C. Development of sample size models for national general practice surveys. Aust J Pub Health 1995;19:34–40.

6. Driver B, Britt H, O'Toole B, Harris M, Bridges-Webb C, Neary S. How representative are patients in general practice morbidity surveys? Fam Pract 1991;8:261–8.

7. Britt H, Harris M, Driver B, Bridges-Webb C, O'Toole B, Neary S. Reasons for encounter and diagnosed health problems: convergence between doctors and patients. Fam Pract 1992;9:191–4.

8. Britt H. Reliability of central coding of patient reasons for encounter in general practice, using the International Classification of Primary Care. Informatics 1998;May:3–7.

9. Bridges-Webb C, Britt H, Miles DA, Neary S, Charles J, Traynor V. Morbidity and treatment in general practice in Australia 1990–1991. Med J Aust 1992;157(Suppl):S1–56.

10. Britt H. A measure of the validity of the ICPC in the classification of reasons for encounter. Informatics 1997;November:8–12.

11. Bentsen BG. The accuracy of recording patient problems in family practice. J Med Educ 1976;51:311–6.

12. Bensing J. The use of the RFE classification system in observation studies – some preliminary results. Proceedings of the Tenth WONCA Conference on Family Medicine, 1983. Singapore: WONCA, 1983.

13. Howie JGR. Diagnosis: the achilles heel? J R Coll Gen Pract 1972;22:310–5.

14. Anderson JE. Reliability of morbidity data in family practice. J Fam Pract 1980;10:677–83.

15. Alderson M. Morbidity and health statistics. 1st ed. Southampton: Stickton Press, 1988.

16. Robertson J, Fryer JL, O'Connell DL, Smith AJ, Henry DA. Limitations of Health Insurance Commission (HIC) data for deriving prescribing indicators. Med J Aust 2002;176:149–424.

17. Donner A, Birkett N, Buck C. Randomisation by cluster. Sample size requirements and analysis. Am J Epidemiol 1981;114:906–14.

18. Henderson J, Britt H, Miller G. Extent and utilisation of computerisation in Australian general practice. Med J Aust 2006;185:84–7.

19. Miller G, Britt H, O'Halloran J. General Practice EHR and data query minimum data set development, 2005. Phase 3 Development Grants. General Practice Computer Group.

afp **CORRESPONDENCE** email: afp@racgp.org.au