title

Reality check - reliable national data from general practice electronic health records

authors

Julie Gordon
Research Fellow
Family Medicine Research Centre, Sydney School of Public Health
University of Sydney
Email: julie.gordon@sydney.edu.au

Graeme Miller
Medical Director
Family Medicine Research Centre, Sydney School of Public Health
University of Sydney
Email: graeme.miller@sydney.edu.au

Helena Britt
Director
Family Medicine Research Centre, Sydney School of Public Health
University of Sydney
Email: helena.britt@sydney.edu.au
# Table of contents

## Contents

Key messages .................................................................................................................. 1  
Executive summary ....................................................................................................... 2  
1. Background ................................................................................................................ 3  
2. A perspective from the BEACH program ................................................................ 3  
3. A history of specifications for general practice EHRs .......................................... 4  
4. Possible reasons for lack of EHR standardisation ................................................. 5  
5. Data extraction from EHRs using data extraction tools ......................................... 6  
6. Questions to ask yourself when assessing data extracted from GP EHRs ............ 7  
7. Implications for data extraction .............................................................................. 7  
   7.1 Research and data analytics ............................................................................... 7  
   7.2 Integrated care programs ................................................................................... 8  
   7.3 Uploading clinical data to the MyHealth Record ............................................ 8  
8. Solutions .................................................................................................................. 9  
   8.1 The four essentials for high quality data from EHRs ...................................... 9  
   8.2 Implementing an EHR data model (standardising the structure of general practice EHRs) ................................................................. 9  
   8.3 Standardising data element labels and definitions ........................................ 10  
   8.4 Use of standardised clinical terminology sets for each data element, and standardised mapping of terminologies to classifications for data extraction and data analytics ....... 11  
   8.5 GP EHR software accreditation ......................................................................... 12  
9. Stakeholders .......................................................................................................... 13  
10. Conclusion .............................................................................................................. 14  
References ................................................................................................................... 15
Key messages

- General practice electronic health records (EHRs) are currently unregulated. Each system has been developed independently, with no common standards across all systems.
- In Australia, there are at least eight EHRs with inconsistent structures, data elements and use of clinical terminologies and classifications.
- The lack of standards across EHRs has made it difficult to: transfer clinical data between EHRs for clinical purposes; link individual health data for integration of care across different sectors of the health care system; and reliably extract patient data for research purposes.
- A national, cohesive approach is needed to develop and implement standards for general practice EHRs. There are four elements to be addressed:
  1. A defined EHR data model that links related data elements
  2. Consistent data element labels and definitions
  3. Use of standardised clinical terminologies and classifications
  4. Accreditation of general practice EHRs.
Executive summary

Since 1998, data about general practice activity in Australia has been collected, analysed and disseminated through the Bettering the Evaluation and Care of Health (BEACH) program. BEACH has provided valuable information about how general practice has changed over time, the impact of policy on practice and general practitioner (GP) professional development, and is the most reliable national source of data on GP activity. However, its cross-sectional design precludes comparison of outcomes of different approaches to care.

It is estimated that 96% of GPs currently use computers for clinical purposes. However, some GPs only use Electronic Health Records (EHR) for part of their clinical work, such as prescribing or ordering pathology tests. Others are paperless and only use EHRs, but even in these circumstances the EHRs themselves lack the structure to reliably link management actions to a patient problem. There are at least eight EHRs used in general practice, each developed independently and structured differently.

In short, there are no nationally agreed and implemented standards for EHRs in Australia, in three areas:

- EHR structure (including linkages)
- data element names and definitions
- use of clinical terminology and classifications.

Therefore it is not possible to reliably export standardised data from general practice EHRs of a sufficient quality to be used for clinical and research purposes.

With current policy focuses on data linkage, integration of care, improved use of the My Health Record (formerly the PCEHR) and attempts to use EHR data for research, the need for a reliable source of data from general practice EHRs has never been higher. Unfortunately there is no ‘quick fix’ solution, but the issues can be addressed with a targeted work program to address the three underlying problem areas.

This Issues Brief describes four steps required to produce high quality data from general practice EHRs:

1. A defined EHR data model that links related data elements
2. Consistent data element labels and definitions across EHRs
3. Use of standardised clinical terminology and classifications

This recommended program of work requires a national, cohesive approach, involving stakeholders from government, professional organisations, the EHR software industry and organisations that use data from general practice.
1. Background

Over the last 20 years, electronic health records (EHRs) have been used increasingly to capture clinical data in general practice. This has been done via a multitude of EHR products, each developed independently with little or no commonality between them.

Vast amounts of clinical data are now kept in EHRs. These data are being used for a variety of purposes beyond patient care—for audit and improvement activities, research, patient risk stratification, integrated care programs, and to populate shared health records such as the national My Health Record.

There is a widely held perception that because data are held electronically, they can be easily used for these secondary purposes. This brief describes the current issues when using data from EHRs for secondary purposes, commonly held misconceptions, and what needs to be done to produce data from general practice EHRs that are complete, representative and reliable.

2. A perspective from the BEACH program

The Bettering the Evaluation and Care of Health (BEACH) program was run continuously by the University of Sydney from April 1998 to June 2016. BEACH was a national, cross-sectional survey of the clinical activity of general practitioners (GPs).\(^1\) The program collected information from rolling random samples of GPs with the aim of gaining an understanding of the characteristics of the GPs themselves, the content of GP-patient encounters, and the services and treatments provided.

General practice has changed dramatically since BEACH began. Only 10% of GPs now practise solo\(^2\) (compared with 18% in 1998\(^3\)). The movement to larger practices, encouraged by government financial incentives, led to shared infrastructure costs and increased flexibility of working hours—on average, GPs now work 2.5 fewer face-to-face clinical hours per week than they did in 2005–06.\(^2\)

BEACH also measured changes in the management of specific conditions in response to new evidence, pharmacological and other products, screening and diagnostic tests, and government policies and incentives. This relied on the GP’s direct linkage of management to the patient problem, a linkage that is facilitated by the structure of the BEACH encounter form. In addition to extensive BEACH publications\(^4\), commissioned BEACH reports (over 100 per year) have been prepared for government, professional organisations, Primary Health Networks, non-government organisations, researchers and industry to guide policy and practice.

For 18 years BEACH relied on the contribution of GPs in completing details of encounters on structured paper forms. Data were collected on paper rather than automatically downloaded from EHRs for the following reasons:
1. Completeness of the data entered - based on BEACH data (2014–15), it is estimated that 96% of practising GPs used a computer at their desk for some clinical purpose. Of these 71% stated they were ‘paperless’ (used complete electronic medical records)², so there is still a considerable proportion using EHRs only for prescribing and/or ordering pathology tests.

2. There are no nationally agreed and implemented standards for the EHR – including data structure, systems of terminology and classification, and consistent data elements with definitions.

3. There is no compulsory minimum dataset to standardise the types of patient data that should be collected at a minimum for every patient at every consultation.

4. Most systems lack the structured problem orientation described by Dr L Weed⁵ which provides reliable linkage of management actions to a patient problem, even though this structure was introduced and widely adopted for paper records in Australia in the 1970s by the RACGP.

In the absence of these standards, data were collected on paper to adequately address the research questions and obtain the quality and reliability of data needed in a national research program such as BEACH.

3. A history of specifications for general practice EHRs

Theoretically, it should be possible to extract data from EHRs to examine outcomes of patient care, longitudinal patient pathways and the effectiveness of general practice care.

Specifications for general practice EHRs were developed by the Family Medicine Research Centre (FMRC) in the 1990s, and were found by independent reviewers to be excellent.⁶ A functional requirements specification for both clinical and administrative general practice computer systems was developed by IBM in 1997.⁷ In 2000 FMRC worked with Simsion Bowles and Associates, GPs and other stakeholders to develop a general practice data model and core data set.⁸ These projects were funded by the (then) Commonwealth Department of Health and Ageing, but their outcomes were not adopted, and the Government did not proceed with implementing data standards for electronic health records. This may have been because there was a lack of perceived cost-effectiveness and benefits by government, vendors and other stakeholders.

The International Classification of Primary Care Version 2 (ICPC-2)⁹ was recommended as the standard for classifying patient-reported and GP-recorded morbidity data in 2003 — another standard that was never mandated.

In 2005, the National E-Health Transition Authority (NEHTA) was established, and other dedicated government funding of IT development in general practice was stopped. NEHTA focused on developing the Personally Controlled Electronic Health Record (PCEHR, now My Health Record). The My Health Record is a repository of health care data uploaded from
EHRs, not an EHR itself. Due to the lack of standards within the EHRs it draws from, it does not link the patient problem with subsequent management and outcome. Yet a large government financial investment was made to provide incentives to vendors and GPs to roll out the PCEHR.

In 2016, there remains a variety of EHR systems with inconsistent structures, data elements and terminologies. The absence of compulsory basic standards has four negative effects:

1. The transfer of reliable, machine-readable patient data to other GP practices and to other health providers is extremely difficult.
2. Linking clinical data between general practice and other sectors in the health system (e.g. hospitals) is extremely difficult.
3. It is hard for practices to change to a different EHR system because transfer of historical patient data to a new system is unreliable, due to different data structures and coding systems.
4. Obtaining high quality data from passive data collection from GPs’ EHRs is impossible, restricting the ability to provide reliable national information about the care provided to individuals and the population. This is unsatisfactory given that in 2014-15 there were 139.4 million GP services provided at a cost to the Commonwealth of about $6.8 billion.

Much can be learned from BEACH about general practice and how it changed during the last 18 years. The program provided valuable input into primary care reform and GP professional development. More could be achieved if standards and minimum datasets were finalised and mandated — so that reliable, high quality, longitudinal patient-based data could be collected from GP EHRs.

4. Possible reasons for lack of EHR standardisation
There are many factors that may have contributed to the lack of basic standards in general practice EHRs. The following is a list of hypotheses but it is likely that a combination of these factors is responsible.

- Unwillingness from policy makers (governments, NEHTA) to invest in the changes.
- The desire to achieve consensus from those involved (vendors, professional bodies, users).
- Resistance from vendors who want users ‘locked in’ to their product and who would be potentially required to make enormous changes to their products to meet imposed standards. This may have a flow-on effect to the ‘look and feel’ of the product, and may create resistance from users.
- The problem is not well understood – there is limited understanding about the underlying design features required to obtain quality data from EHRs amongst those who are not familiar with general practice data.
5. Data extraction from EHRs using data extraction tools

Over the last 15 years, data extraction tools have been developed that interrogate and extract data from GP EHRs. Some of the tools involved are the CAT tool (from PEN Computing)\(^\text{12}\), GRHANITE (University of Melbourne)\(^\text{13}\) and the Canning Tool.\(^\text{14}\)

There are two significant issues with these tools.

1. The ability to extract meaningful EHR data is only as good as the structural capacity of the record to link related data types. The problem-oriented medical record structure (POMR) is the ideal structure to be used, so that all investigations and treatments are inherently linked to each problem. This linkage is the only way to build a complete picture of the patient’s health care, which does not rely on assumptions.

2. The tools cannot extract what is not in the record. If there is missing data in the EHR (and evidence suggests this is common), there is only a certain amount of work that can be done to curate the record to an acceptable level for research use. Data extraction tools often attempt to identify areas where data may be missing. For example, they may identify instances where a prescription for a particular medication is in the record but there is no indication present (e.g. a diabetic medication is recorded in the medications list without a diagnosis of diabetes). However, this relies on the validity of the assumed linkage between particular diagnoses and medications. For example, it is often incorrectly assumed that anti-depressants are only given for the management of depression. In fact, BEACH data (with its problem – treatment linkage) suggested that only 70% of prescriptions for anti-depressants were prescribed for depression. Another 16% were prescribed for another psychological problem (e.g. anxiety, sleep disorders) and a further 10% for musculoskeletal, neurological and premenstrual/menopausal problems, because of the medication’s muscle relaxant properties.\(^\text{15}\)

A 2013 Australian study assessed the consistency of results generated by three data extraction tools from two GP EHRs, to determine the patient prevalence of diabetes. The study authors concluded (page 822):

“...the DET (data extraction tool)/EHR (electronic health record) combinations did not extract similar counts of diabetics and indicators of diabetes care. This renders current DETs ineffective as tools for measuring the quality of care in a way that might be compared between systems. When we add the lack of transparency for proprietary reasons and a lack of technical and professional standards and safety regulations for medical software, this situation is unable to ensure that practice is safe, or able to support clinical governance.”\(^\text{16}\)
This study’s alarming findings indicate that the combination of proprietary EHR software and proprietary data extraction tools, both developed without needing to adhere to standard data models, means users have no way of knowing whether the data generated are valid and reliable.

6. Questions to ask when assessing data extracted from GP EHRs
   1. How are the participating practices recruited? Is the study drawn from a random sample of practices, or do participants volunteer to participate? Is the sample of practices and their GPs representative of Australian general practice as a whole, or of practices and GPs in the selected geographic area (e.g. such as primary health networks (PHNs))? 
   2. Are there any restrictions about participation? For example, are participants restricted to those using EHRs, or those using one or two particular brand(s) of EHR? 
   3. Is the methodology used to collect the data published? 
   4. Is the study population representative of GPs in Australia as a whole? Have any comparisons been published about the representativeness of the data, comparing the characteristics of the participants with the characteristics of the Australian population? If not, why not? 
   5. Are missing data reported?

7. Implications for data extraction
   Data is extracted from EHRs for many purposes.
   
   7.1 Research and data analytics
   Numerous research projects rely on EHR data extraction, on local, regional and national scales. One such project is MedicineInsight (run by NPS MedicineWise) which extracts data from general practice EHRs for quality improvement activities at the practice level, and to provide aggregated data to government to inform policy. Practices enrol themselves to participate.

   There are currently few published studies from data collected through the MedicineInsight program, but current publications note that a limitation of the data is the completeness and accuracy of the data recorded in the EHR, as well as data being available for extraction.

   Another limitation of many research projects relying on EHR data extraction, including MedicineInsight, is that due to the structural differences between general practice EHRs, they are often restricted to practices using particular EHRs. For example, MedicineInsight is currently restricted to practices using the latest versions of the Best Practice or Medical Director EHRs. Although these two systems accounted for 81.4% of the EHRs used by Australian GPs in 2014–15 (unpublished BEACH data), there are at least six other EHRs, each with a small, but important, market share. Best Practice and Medical Director are the two programs that are easiest to match, because they were designed by the same designer.
This demonstrates the difficulties involved with extracting data from general practice EHRs in the current ‘standard-free’ environment. Every researcher or data analyst who tries to extract EHR data is faced with these same issues, and at present, each comes up with their own solution, which often involves restricting their study to a level that is easier than attempting to extract data from all EHRs. While initially appearing to be a viable solution, this approach does not address the underlying problem, and in fact, perpetuates the situation by publishing data with significant limitations. This limits the suitability of data for research, policy planning and quality improvement.

7.2 Integrated care programs

These programs focus on care planning and co-ordination, for patients deemed at high risk of consuming health care resources. They aim to improve the care provided to high risk patients in the community, reducing preventable hospital admissions and improving patient outcomes.

In Australia these projects are in their infancy and at present are pilot projects focussing on selected groups of patients. Two high profile examples are the Western Sydney LinkedEHR (with a focus on diabetes management) and the Gold Coast Integrated Care Project, which is taking a broader approach focussing on frequent users of the area’s hospitals. The Commonwealth Department of Health has announced that trials of Health Care Homes will commence in 2017 in up to 7 regions across Australia.

One important aspect of care integration is the integration of data—to identify patients eligible for inclusion in the projects and to communicate clinical information about patient care to members of the integrated care team. True care integration cannot occur if the health care data about the patient is kept in silos and not shared.

7.3 Uploading clinical data to the My Health Record

The My Health Record is not an EHR, but a repository of clinical data that is uploaded from EHRs. As a result, the quality and reliability of these clinical data are almost totally reliant on the quality of data in the source EHR. If, as discussed earlier, there is missing or incomplete data in the EHR, this has a flow-on effect to the My Health Record. This poses a risk for patient safety when clinical decisions are made purely on data in the My Health Record. It is therefore imperative that some basic information standards are introduced to general practice EHRs.
8. Solutions

8.1 The four essentials for high quality data from EHRs

1. Consistent data element labels and definitions
2. A defined EHR data model that links related data elements
3. Use of standardised clinical terminology sets for each data element in the record, and standardised mapping of terminologies to clinical classifications for data extraction and analytics
4. Accreditation of GP electronic health records

Each of these is interdependent, and needs to be addressed as part of an overall solution to introduce and roll out standards for general practice EHRs.

This solution is focussed on improving the quality of data contained in, and therefore extractable from, general practice EHRs. There are other areas that could be addressed to improve standardisation across the health system, but these are beyond the scope of this paper.

The proposed solution draws on the review or expansion of pre-existing work wherever possible.

8.2 Implementing an EHR data model (standardising the structure of general practice EHRs)

The GP core data set and data model, created in 2000 by Simsion Bowles Consulting, is now 15 years old but its content still forms the core standards required for GP EHRs.

Recommendation 1: That the GP core data set and data model (Simsion Bowles Consulting) is used as the starting point for the development of a minimum data set for general practice.

This review must involve all relevant stakeholders (identified in Section 8). Given the number of parties involved, it will be difficult to obtain consensus about the data model, but this step is imperative to improve the quality of data contained in general practice EHRs.

Recommendation 2: Subsequent to the data model review, measures must be introduced to enforce the implementation of the data model. This could include accreditation of GP EHRs (see Section 7.5).

It would also be useful to monitor the Fast Health Interoperability Resources (FHIR) specification, currently under development by HL7. FHIR is designed as a common standard for electronically communicating health information, but is not a complete data model for an electronic health record so cannot be used in place of the work outlined
above. Considerable resources are being spent on FHIR development, and in future the GP EHR data model work may need to be integrated with FHIR.

8.3 Standardising data element labels and definitions
The independent development of EHRs in general practice means there is no consistency in the label used to designate each data element, and how each data element is defined (in terms of how it should be populated). For example, the data element for why a patient has sought health care may be called ‘reason for visit’, ‘reason for encounter’, ‘presenting symptoms’ or ‘presenting problem’.

The data element that describes the problem managed at the completion of an encounter may be called ‘diagnosis’ or ‘problem managed’. A number of EHRs contain a data element called ‘reason for prescription’ – this could be interpreted to mean a diagnosis/problem or the reason for giving a prescription (e.g. need for repeat prescription). This field also implies that medication is given at each general practice encounter, or that the only reason a clinician would need to record a diagnosis is if they were prescribing. Both of these are incorrect assumptions.1

Each of these listed examples demonstrates that data element labels and definitions differ. Even though some may appear to be synonymous, they are often not. Therefore, what is used to populate one data element in one EHR may differ considerably from a (supposedly) equivalent data element in another EHR.

Another issue is whether recording in each data element is enforced in the EHR. We have heard anecdotally that users do not enter acute diagnoses (e.g. upper respiratory tract infections) into a ‘diagnosis’ field, because data in this field is used to automatically populate the problem list, and many GPs do not want acute problems in the problem list.

Minimum data sets have already been created for general practice, for example the GP core data set and data model8 and the General Practice Computing Group (GPCG) GP EHR and data query minimum data set23 for reporting from general practice. These can form the basis for the development work required.

Recommendation 3: Review all data element labels and their definitions, across all GP EHRs, and where possible standardise them across the system, to ensure that comparable data can be extracted from EHRs. This work must be done in conjunction with the work outlined in Section 7.2.

Recommendation 4: That previous standards work undertaken in this area is used as the basis for the development of a minimum data set specification.
Recommendation 5: Once the EHR data model and the data element labels and definitions are finalised, relevant aspects of this work should be entered into METeOR, the Australian online registry for health metadata.

8.4 Use of standardised clinical terminology sets for each data element, and standardised mapping of terminologies to classifications for data extraction and data analytics

This part of the solution involves creating, testing and implementing a standardised suite of classification and terminology tools, involving maps from each of the termsets currently used in general practice EHRs to SNOMED CT-AU, then to the ICPC-2 and ICD-10-AM classifications. The ‘create once, use many’ approach is required; ensuring that all data aggregated or communicated is accurate and reproducible.

At present, there are three termsets or terminologies used in general practice EHRs. These are all ‘interface terminologies’, which incorporate local language preferences. The three termsets/terminologies are:

- PYEFINCH (used in Best Practice)
- The Medical Director termset (used in Medical Director)
- ICPC-2 PLUS (used in Zedmed, Genie, Practix, Profile and others).

Preparing a validated map from each of these to SNOMED CT will enable end users to continue to use the termset/terminology they are used to, while being able to transfer data to other parts of the health system using SNOMED CT. This is a common implementation scenario internationally, with many successful implementations of SNOMED CT using a local interface terminology mapped to SNOMED CT. For example, the United States Health Maintenance Organisation Kaiser Permanente has created a Convergent Medical Terminology, which is an interface terminology that incorporates content from SNOMED CT.

Recommendation 6: Each of the termsets/terminologies currently used in general practice EHRs should be mapped to SNOMED CT-AU, the Australian national clinical terminology.

Recommendation 7: SNOMED CT-AU should be used for communication to and from general practice (e.g. for referrals and discharge summaries, for transmitting information to the My Health Record).
Reporting of data from the health care system must still be in terms of the standard international classifications used in Australia, namely ICD-10-AM\textsuperscript{27} (used for hospital and mortality statistics) and ICPC-2\textsuperscript{9} (used in general practice).

**Recommendation 8:** Relevant concepts from SNOMED CT-AU should be classified to ICD-10-AM and ICPC-2.

Creating a standardised suite of classification and terminology tools will allow the transfer and reporting of data across the health care system. An overview of the process needed is shown in Figure 1.

---

**8.5 GP EHR software accreditation**

Historically, GP EHR vendors have not been required to adhere to standards in terms of EHR structure, data element labels and definitions, and classifications and terminologies. If the
solutions presented above are created, there must be a mechanism to ensure GP EHR vendors adhere to the resulting standards. It is proposed that accreditation of GP EHR software could be considered as an appropriate mechanism.

Recommendation 9: All GP EHR software should be required to meet accreditation standards.

Accreditation of EHRs is common overseas. For example, in the United Kingdom EHR accreditation has been in existence for many years, and currently occurs through a program called ‘GP Systems of Choice’, which is overseen by the Health & Social Care Information Centre. In the US, the Office of the National Coordinator for Health Information Technology coordinates a program of voluntary certification of health IT standards.

In Australia, some medical software is already regulated through the Therapeutic Goods Administration, but ‘medical records management system(s)’ are currently excluded from the types of medical software regulated. If GP EHRs were to be accredited or regulated, this could occur directly through a variation to the eHealth Practice Incentives Program (PIP), or indirectly through the Royal Australian College of General Practitioners practice accreditation standards.

9. Stakeholders
Solving this problem will need to involve multiple stakeholders, representing different interests. These include:

- professional associations, including the Royal Australian College of General Practitioners (RACGP), Australian Medical Association (AMA), the Australian College of Rural and Remote Medicine (ACRRM) and the Rural Doctors Association of Australia (RDAA).
- software vendors who develop EHRs for general practice, and the Medical Software Industry Association (MSIA)
- developers of data extraction tools
- Australian Government Department of Health, state and territory health departments
- other government instrumentalities, including the Australian Digital Health Agency, the Australian Commission on Safety and Quality in Health Care, and the Australian Institute of Health and Welfare.
- organisations currently attempting to extract data from EHRs, including university researchers, NPS MedicineWise, the Improvement Foundation, Primary Health Networks and developers of data extraction tools (GRHANITE, PEN, Canning tool).
10. Conclusion

Governments, PHNs, Local Health Districts and many other agencies are collectively spending millions of dollars to develop ways to extract data from GP EHRs. Largely, these developments are independent and the resulting data are not comparable.

In 2016, the lack of a cohesive, national approach to overcoming the issue of data extraction from GP EHRs is unacceptable. The longer it takes to address this problem, the harder it will be to develop and implement the necessary standards. Therefore this work must be done now, or we risk an ever-increasing cost— in money, quality of care and patient safety.
References


Contact
Deeble Institute for Health Policy Research
Australian Healthcare and Hospitals Association
T: 02 6162 0780
E: deeble@ahha.asn.au