Technology report

Author address for correspondence
Christopher Pearce
Department of General Practice, School of Primary Health Care, Monash University, Australia
Melbourne East General Practice Network
Nunawading, Victoria, Australia
Email: chris.pearce@monash.edu.

The Melbourne East Monash General Practice Database: Using data from computerised medical records for research

Danielle Mazza
Department of General Practice, School of Primary Health Care, Monash University, Australia

Christopher Pearce
Department of General Practice, School of Primary Health Care, Monash University, Australia, and Melbourne East General Practice Network, Australia

Lyle R. Turner
Department of General Practice, School of Primary Health Care, Monash University, Australia

Maria de Leon-Santiago
Department of General Practice, School of Primary Health Care, Monash University, Australia

Adam McLeod
Melbourne East General Practice Network, Australia

Jason Ferriggi
Melbourne East General Practice Network, Australia, and Department of Health and Human Services, Victorian Government, Australia

Marianne Shearer
Melbourne East General Practice Network, Australia, and Gippsland Primary Health Network, Australia

ABSTRACT

The Melbourne East Monash General Practice Database (MAGNET) research platform was launched in 2013 to provide a unique data source for primary care and health services research in Australia. MAGNET contains information from the computerised records of 50 participating general practices and includes data from the computerised medical records of more than 1,100,000 patients. Data extracted are patient-level episodic information and includes a variety of fields related to patient demographics and historical clinical information, along with the characteristics of the participating general practices. While there are limitations to the data that are currently available, the MAGNET research platform continues to investigate other avenues for improving the breadth and quality of data, with the aim of providing a more comprehensive picture of primary care in Australia.

Keywords: health services research, general practice, Melbourne East Monash General Practice Database (MAGNET), primary care data

INTRODUCTION

The availability of general practice data for analysis has the potential to benefit both health outcomes and health services research. The large sample sizes, breadth and uniquely longitudinal nature of the information that is available (e.g. medications used) are attributes not commonly found in other health data sets.[1–3] The U.K.'s General Practice Research Database (now part of the Clinical Practice Research Datalink)[4] is the largest and most comprehensive source of general practice data in the world. This database is considered the 'gold standard' for anonymised longitudinal medical records from primary care. Analysis of these data has resulted in the publication of a large number of studies that have contributed significantly to primary care policy and wider health research and practice.[5]

In contrast, primary health care in Australia has suffered from a particular lack of research capability to inform both policy and practice.[6–8] A fundamental barrier to developing this capability has been the limited access to patient and health services data that could underpin this research.[9] Approximately 75% of all medical consultations in Australia take place in general practice, with more than 85% of the population accessing a general practitioner (GP) every year.[10] For this reason, access to pooled data represents a significant potential resource.

However, access to general practice data in Australia has been limited. The segregated nature of information systems that are currently being used to collect and store patient records has led to incompatibility between software packages, coding regimes, specific fields collected and privacy policies. Also, general practices, by nature, exist as individual small businesses,[11] which makes recruiting individual practices for the purpose of data collection arduous and complex.

In Australia, there are currently only two significant sources of general practice data: Medicare Australia and the Bettering the Evaluation and Care of Health (BEACH) program.[12] Data held by Medicare Australia (on behalf of the Australian government) are used primarily for financial administration. Apart from some broad descriptors of the types of consultations being billed (e.g. health assessments or mental health plans), Medicare data do not contain any information about the clinical problems being managed. Also, it is structured around individual GPs, but not practices. In contrast, the BEACH program uses a cross-sectional, paper-based survey to collect data on the characteristics of GPs, GP-patient encounters, and the services and treatments provided. However, BEACH only captures information at the GP level (i.e. GPs describe the care that they have provided) and is limited by its paper-based and cross-sectional design. Again, it is not a whole of practice program. Consequently, there is a need to develop other general practice data sources in Australia that are capable of addressing the limitations inherent in the data sources that are currently available.

DEVELOPMENT

Setting

Inner East Melbourne Medicare Local (IEMML) was one of the 61 Medicare Locals (primary care organisations) in Australia that were responsible for coordinating and delivering health services at a regional level. IEMML served four municipalities in the South East of Melbourne (Figure 1). In July 2015, Medicare Locals were replaced by larger Primary Health Networks. Consequently, the data services are now held within the Melbourne East General Practice Network (MEGPN) and will be offered to all Primary Health Networks, potentially expanding the footprint.

The data warehouse underpinning Melbourne East Monash General Practice Database (MAGNET) contains information from the computerised medical records of patients attending general practices located within the region served by IEMML (50 practices have contributed data). Currently, the data warehouse contains data from the computerised medical records of more than 1,100,000 registered patients.

To provide an alternative source of primary health care data in Australia, IEMML collaborated with Monash University to create the MAGNET research platform. It was designed to replace (and improve on) pre-existing systems that provided audit loops for practices, but not a research program. Data were initially collected from the computerised medical records of each practice using one of the commercially available clinical audit tools.[13] These were designed for population health reporting focusing on chronic disease. The development of the database was informed by a theoretical framework on data governance at an organisational level.[14]

Data source and extraction

The extraction of data from general practices involves a number of steps (Figure 2). Initially, participating general practices are asked to sign a consent agreement informing them of the data extraction process and the use of deidentified data in the data warehouse for research purposes. As part of national practice accreditation requirements, practices are also required to inform patients that data collected by the practice may be utilised for quality improvement and research. Individual patients are able to opt out of data collection. Once a practice agrees to participate in the data extraction program, a practice liaison officer visits the practice to install the extraction tool. The extraction process produces an initial encrypted data file, which is sent by secure data transfer from the general practice to MEGPN. Subsequent incremental data extractions are then set to run at defined intervals from the practice.

Extraction is facilitated through the use of the GeneRic Health Network Information Technology for the Enterprise (GRHANITE) extraction technology.[15] This tool is designed to extract patient-level episodic information across the various medical software programs used by GPs in Australia. GRHANITE™ is an important component of the POpulation Level Analysis and Reporting (POLAR) tool, which primary health networks utilise to provide comprehensive feedback
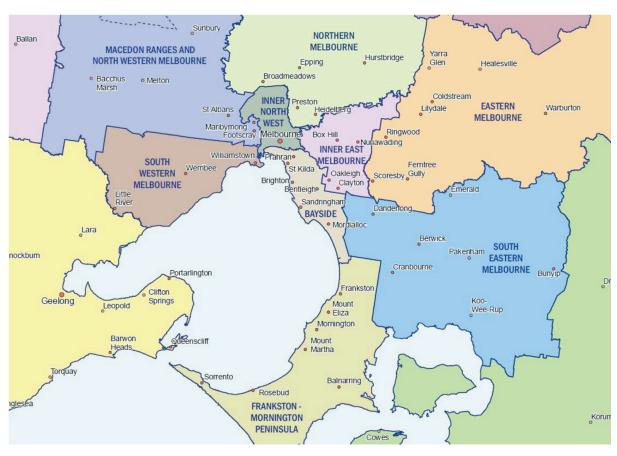
Figure 1 IEMML catchment area – location of general practices that contribute data to MAGNET (© 2011, IEMML)
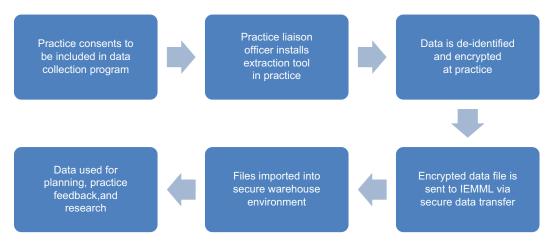

Figure 2 Procedure for MAGNET data collection

to general practices. This feedback, provided as a report to the practices, combines practice data with local emergency data and local government-level data. The web-based tool and feedback reports developed by MEGPN are offered free of charge to member practices as a quality improvement initiative to assist in monitoring and improving their own clinical governance and population health activities.[16] The data available for research are, therefore, a by-product of clinical governance processes within the practice, which is supported and encouraged by MEGPN. Consequently, data quality improvement is a practice-level activity rather than a research project-led activity. The useful feedback of practice-relevant information has been a critical component of uptake of the extraction program.

Creation of MAGNET data warehouse

The MAGNET data warehouse consists of multiple datasets that are aggregated and connected by common data elements. The connection of datasets is performed at a geographic level using postcodes, statistical local areas (SLAs) or local government areas. The data warehouse is gradually moving towards the Australian Bureau of Statistics (ABS) geo-mapping standards[17] as more datasets have the capability to map to SLA levels.

The data warehouse has a 'star structure' construction, with the 'fact' or 'raw data' being central to the structure. Dimensions of interest and metrics are then added to the fact tables to incorporate additional information or analysis potential. Data extracts that have been collected from the general

practices are then uploaded and processed onto a staging database. Validation and data cleansing are performed during this staging process and only clean, validated data are transferred to the data warehouse. Validation rules are constantly being reviewed to ensure that the data are relevant, and any datum that fails the validation process is quarantined for further analysis.

Data types and limitations

The data extracted from computerised medical records are related to patients and their episodes of care (e.g. patient demographics and historical clinical information). The characteristics of the participating general practices are also collected.

Patient demographics

Patient demographics that have been extracted include date of birth, gender, pension and Department of Veterans Affairs status and location of residence (Table 1). Identifying information related to both patients and the general practices has been excluded to meet privacy concerns.

Clinical information

GRAHNITE™ extracts all data related to the patient's episodes of care within the practice (Table 1). Specific information about each consultation is also collected (e.g. time and duration of the consultation). For most of these variables, the information is coded using the computerised medical record's internal coding system, which relies on the GP to select an appropriate option from a predefined list. Other data include diagnoses, reason for encounter, prescribed medications, structured observations, investigations ordered and received, immunisations and many other fields. Understanding the workflow, cognitive load and financial incentives is crucial to understanding the data.

A unique function of GRAHNITE™ is the ability of the software to build a comprehensive set of statistical linkage keys. These linkage keys enable the identification of the same patients at different practices (thus removing patient duplication), which makes the final dataset more representative of the catchment's population. The linkage keys can also be used to link MAGNET with other datasets when the appropriate ethical considerations have been met. Linking patient records creates a 'richer' dataset consisting of a timeline or sequence of health events for a particular cohort of patients. Such information is useful not only for research purposes but also for improving health services and measuring the outcomes of health care programs.

Privacy issues

Privacy of health data is considered paramount, and privacy is built into the data governance protocols. Australian general practices are required as part of their accreditation to give all patients information about their use of health data, including in practice use for quality purposes and potential use for research. MEGPN then enters into an agreement with the practice regarding the use of data. Crucially, data are not used outside the practice without

Table 1 Patient- and practice-level variables that are included in the MAGNET dataset

Characteristics	Fields
Patient	Year of birth
	Gender
	Marital status
	Pensioner status
	Postcode
	Aboriginal and torres strait islander (ATSI) status
Clinical	Consultation characteristics (date, duration, reason for visit, Medicare Benefits Schedule (MBS) items billed)
	Smoking and alcohol usage characteristics
	Allergies
	Blood group
	Pathology ordered characteristics and related results
	Clinical measurements (eg, blood pressure, height, weight, cholesterol)
	Immunisation details
	Current and past diagnoses
	Pregnancy-related history
	Prescribed medications
	Medication-related information (dosage, strength, frequency, duration, therapeutic class, generic equivalent)
	Medication review details
	Asthma plan details
	Care plan details
Practice	Opening hours
	Group/solo status
	Number of nurses & doctors in the practice
	Doctor/nurse characteristics (year of birth, gender)

the practice's consent, and only deidentified data are used outside the practice. Ethics approval has been granted by Monash University for the use of the data.

DISCUSSION

Principal findings

Data within MAGNET acknowledges a data hierarchy,[14] where the data are primarily used to assess quality improvement at both the practice level and the population level and to monitor the activities of primary care services within the region.[16,18] These data can also be used to support targeted and priority-driven strategic research in primary care.[19] For example, data within MAGNET can be used for research that focuses on improving the quality of care in general practice and optimising primary care practice organisation and health care systems.

Data within MAGNET are currently being analysed for a number of studies. One example is the REDIRECT project, which involves the analysis of general practice data to examine the 'primary care patient journey' of older patients

who present unnecessarily to hospital emergency departments. Another example is a study examining the utilisation of guidelines for managing overweight and obese patients in general practice.[20] Measurable components of the guidelines that are included in the MAGNET dataset (e.g. weight, height and waist circumference) have been analysed to determine the extent to which GPs are currently implementing guideline recommendations. The linkage capability within MAGNET is also currently being used to link local emergency data with patient-level data in general practice, allowing the mapping of the general practice journeys of all patients admitted to the regional emergency departments.

Limitations of MAGNET data

General practice software in Australia is only in its infancy in truly using contained data to inform clinical decisions.[21] Decision support in clinical systems not only requires quality data to inform the computational issues but also drives data quality. Decision support is either rudimentary or absent, and therefore remains dormant for both uses. Consequently, there are some limitations to the data within MAGNET. Firstly, the 'completeness' of certain variables (e.g. pregnancy information and aboriginality) is an issue. Because the primary use of the data is to inform the clinical care of a patient, only data that serve a clinical purpose have a high degree of validity and reliability. Therefore, items such as aboriginality are not necessarily well recorded.

Aboriginality is a good example of how data quality could drive better outcomes using computerised processes. The Aboriginal population has poorer health outcomes compared to the populace and there are different financial incentives to promote care. Thus, there is a different immunisation schedule for aboriginal children, for instance. If aboriginality is recorded in the system, the Electronic Medical Record can promote correct care and encourage use of financial incentives. As an example, the activities of MEGPN (and IEMML) during 10 years of practice improvement with the clinical governance program (prior to the formation of MAGNET) have increased the reliability around examples such as diabetes.

Secondly, data quality may be reduced because computerised medical records are often stored in many different computer systems, some of which use different coding terminologies and structures including non-standards-based terminologies. Each clinical information system has its own database structure and can record the same category of information in different parts of the software (e.g. diagnoses can be in a summary section, part of the consultation notes, or entered as a 'reason for encounter'). This structure and appearance can influence data collection.[22,23] Additionally, the encounter or consultation details that have been extracted are limited because of the interaction between different clinical and billing systems in some practices. Even the data extraction process itself may be fallible.[24]

Comparisons

Kaiser Permanente, a U.S.-based healthcare provider, is widely regarded to have one of the best clinical governance programs,[25] but it is based on using a single clinical system and a fee structure reliant on clinical coding. This is not replicable at any significant level in Australia. Other programs rely on extractions without necessarily providing the data quality program. The closest parallel to MAGNET is probably the UK's PRIMIS project,[26] which is a data quality program with similar foundations. However, none necessarily combines the data quality program with a comprehensive research outlook.

IMPLICATIONS AND FUTURE DIRECTIONS

The development of the MAGNET research platform contains many lessons for others considering similar programs. In developing expertise in combining disparate clinical systems while sitting at the juncture of clinical governance and clinical research, MAGNET bridges many of the issues experienced in other platforms. Its increased use of data linkage across primary care and secondary care will continue to breakdown the silo problems that bedevil other similar platforms.

Similarly, research involving data requires extensive work within a project to ensure data consistency and quality; otherwise, it is difficult to ensure that the results are both reliable and valid.[27] A key point of difference with the MAGNET research platform is in combining all the above issues.

The POLAR program allows flexible, tailored clinical governance programs to be delivered with the 'unintended' consequence of improving data quality across a range of systems and with due understanding of both the social and structural sources of data. By targeting practices within the area of a general practice support organisation, the program can be effectively replicated across the country, maintaining the population concentration whilst widening the geographical spread. This then represents a practical opportunity to unlock the potential of routine data for quality research.

CONCLUSION

The MAGNET research platform is the only source of routinely collected GP data in Australia with a regionally representative focus. Its availability for use in research is unique, which in conjunction with a program of tailored clinical governance programs, contributes to improving data quality across a range of systems. The nature of the practice recruitment, based around a central general practice support organisation, implies that MAGNET can be effectively expanded across the country. This represents a practical opportunity to unlock the potential of routine data for quality research.

Declaration of conflicting interests
The authors declare that they have no conflicts of interest.

REFERENCES

1. de Lusignan S, Hague N, van Vlymen J and Kumarapeli P. Routinely-collected general practice data are complex, but with systematic processing can be used for quality improvement and research. *Informatics in Primary Care*. 2006;14(1):59–66. http://dx.doi.org/10.14236/jhi.v14i1.615

2. de Lusignan S and van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice* 2006;23(2):253–63. http://dx.doi.org/10.1093/fampra/cmi106. PMid:16368704.

3. Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N and Wright L. Methodology for Assessing the Usefulness of General Practice Data for Research in One Research Network. *Health Informatics Journal* 2004;10(2):91–109. http://dx.doi.org/10.1177/1460458204042230

4. Williams T, Van Staa T, Puri S and Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Therapeutic Advances in Drug Safety*. 2012.

5. Chen YC, Wu JC, Haschler I, Majeed A, Chen TJ and Wetter T. Academic impact of a public electronic health database: bibliometric analysis of studies using the general practice research database. *Public Library of Science one* 2011;6(6):e21404.

6. Del Mar C. Is primary care research a lost cause? *British Medical Journal* 2009;339:b4810.

7. Howie J. Primary care's lost cause? Not lost if it combines medical and social science approaches. *British Medical Journal* 2010;340:c462.

8. Zwar NA, Weller DP, McCloughan L and Traynor VJ. Supporting research in primary care: are practice-based research networks the missing link? *Medical Journal of Australia* 2006;185(2):110–3. PMid:16842071.

9. Australian Institute of Health and Welfare (AIHW). *Review and Evaluation of Australian Information About Primary Health Care*. Canberra: AIHW, 2008.

10. Britt HC, Miller GC, Charles J, Henderson J, Bayram C, and Valenti L et al. *General Practice Activity in Australia 1999–00 to 2008–09: 10 Year Data Tables*. Canberra: Australian Institute of Health and Welfare, 2009.

11. Pearce C. Electronic medical records--where to from here? *Australian Family Physician* 2009;38(7):537–40. PMid:19575073.

12. Britt HC and Miller GC. The Bettering the Evaluation and Care of Health (BEACH) program: where to from here? *Medical Journal of Australia* 2013;198(3):125–6. http://dx.doi.org/10.5694/mja13.10012 PMid:23418681.

13. PenCS. The PCS Clinical Audit Tool (CAT). Available from: http://install.pencs.com.au/ClickOnce/ClinicalAudit/publish.htm. Accessed 16 April 2013.

14. Liaw ST, Pearce C, Liyanage H, Liaw GS and de Lusignan S. An integrated organisation-wide data quality management and information governance framework: theoretical underpinnings. *Informatics in Primary Care* 2014;21(4):199–206. http://dx.doi.org/10.14236/jhi.v21i4.87.

15. The University of Melbourne. GRHANITE Health Informatics Unit. Available from: http://www.grhanite.com/. Accessed 2 October 2014.

16. Pearce C, Shearer M, Gardner K and Kelly J. A division's worth of data. *Australian Family Physician* 2011;40(3):167–70. PMid:21597524.

17. Australian Bureau of Statistics. Australian Statistical Geography Standard (ASGS) 2014. Available from: http://www.abs.gov.au/websitedbs/d3310114.nsf/home/australian+statistical+geography+standard+%28asgs%29. Accessed 2 October 2014.

18. Pearce C, Shearer M, Gardner K, Kelly J and Xu TB. GP Networks as enablers of quality of care: implementing a practice engagement framework in a General Practice Network. *Australian Journal of Primary Health* 2012;18(2):101–4. http://dx.doi.org/10.1071/PY11121. PMid:22551830.

19. Department of Health and Ageing. Building a 21st century primary health care system: Australia's first national primary health care strategy. In: Department of Health and Ageing, ed. Canberra, Commonwealth of Australia, 2010.

20. Turner LR, Harris MF and Mazza D. Obesity management in general practice: does current practice match guideline recommendations? *Medical Journal of Australia* 2015;202(7):370–2. http://dx.doi.org/10.5694/mja14.00998. PMid:25877119.

21. Pearce CM, de Lusignan S, Phillips C, Hall S and Travaglia J. The computerized medical record as a tool for clinical governance in Australian primary care. *Journal of Medical Internet Research Research Protocols* 2013;2(2):e26. http://dx.doi.org/10.2196/ijmr.2700.

22. Kumarapeli P and de Lusignan S. Using the computer in the clinical consultation; setting the stage, reviewing, recording, and taking actions: multi-channel video study. *Journal of the American Medical Informatics Association*. 2013;20(e1):e67–75. http://dx.doi.org/10.1136/amiajnl-2012-001081. PMid:23242763; PMCid:PMC3715353.

23. Pearce C, Arnold M, Phillips CB, Trumble S and Dwan K. The many faces of the computer: an analysis of clinical software in the primary care consultation. *International Journal of Medical Informatics* 2012;81(7):475–84. http://dx.doi.org/10.1016/j.ijmedinf.2012.01.004. PMid:22325592.

24. Liaw S-T, Taggart J, Yu H and de Lusignan S. Data extraction from electronic health records – existing tools may be unreliable and potentially unsafe. *Australian Family Physician* 2013;42(11):820–3. PMid:24217107.

25. Shapiro J and Smith S. Lessons for the NHS from Kaiser Permanente. *British Medical Journal* 2003;327(7426):1241–2. http://dx.doi.org/10.1136/bmj.327.7426.1241. PMid:14644937; PMCid:PMC286235.

26. Horsfield P. Trends in data recording by general practice teams: an analysis of data extracted from clinical computer systems by the PRIMIS project. *Informatics in Primary Care* 2002;10(4):227–34. http://dx.doi.org/10.14236/jhi.v10i4.264.

27. de Lusignan S, Sun B, Pearce C, Farmer C, Steven P and Jones S. Coding errors in an analysis of the impact of pay-for-performance on the care for long-term cardiovascular disease: a case study. *Informatics in Primary Care* 2014;21(2):92–101. http://dx.doi.org/10.14236/jhi.v21i2.62.