MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190232

# What a Comprehensive, Integrated Data Strategy Looks Like: The Population Level Analysis and Reporting (POLAR) Program

**Christopher Pearce[a,b], Adam McLeod[a], Natalie Rinehart[a], Jason Ferrigi[a], Marianne Shearer[c]**

*[a] Outcome Health, East Burwood, Victoria, Australia,*
*[b] Department of General Practice, Monash University, Victoria, Australia*
*[c] Gippsland Primary Health Network, Victoria, Australia*

## Abstract

*In Australia, general practice (GP) acts as the gatekeeper to the rest of the healthcare system, and therefore the vast majority of the population have an electronic medical record. It follows that the largest database of the population is therefore on the distributed GP computers. Informed by a comprehensive system-wide data strategy, the Population Level Analysis and Reporting program extracts data from the GP electronic medical records and repurposes it for multiple uses. The program requires the data to be coded and then structured for multiple uses clinical care, clinical governance, research, and policy.*

***Keywords:***

Primary Health Care, Electronic Health Records, Medical Record Linkage.

## Introduction

With 85% of the population seeing a general pratice (GP) at least once a year [2], GP currently contains the most comprehensive database of health data in Australia. Hospital systems contain largely episodic data, and national projects contain limited project-specific data. For instance, the national schemes Medicare (the national health insurance scheme) and Pharmaceutical Benefits Schedules (subsidised medicines), contain only data collected for administrative purposes. Some organisations rely on registries – purpose-built data collections around specific diseases – and often based on hospital settings. Others use small, practice-based research networks.

The described project occurs in the context of a worldwide move to use the data from electronic medical records for research [4; 5], although significant barriers apply, including data extraction [9], and data quality.

Australia has meso-level organisations designed to assist general practice to deliver population-based care. Formerly called GP networks [15], these Primary Health Networks (PHNs) fulfil a critical role as the chosen delivery platform for elements of key health reforms and initiatives and are now embedded in the primary health care landscape as drivers of quality, efficiency, coordination and improved access to health services. PHNs have adopted a comprehensive program to supply their general practices with a multi-faceted approach to help them provide patient centred care, improve their data quality, and take a population-based approach to analytics, based on common principles.

The Population Level Analysis and Reporting (POLAR) program and its research arm, POLAR Data Space, represent an attempt to bridge the power of the former with the needs of the latter, by creating a useful environment to understand the health needs of the population using Australian GP derived data. POLAR is a platform designed primarily to improve quality of care for the Australian population. It does so at several levels – acknowledging a data hierarchy [14] designed to take the use of health data beyond the term 'secondary use of data' – to a position where all uses of data are important. Just because a piece of information is recorded for the care of a patient, does not mean uses of it beyond that are 'secondary'. Indeed, the most significant impact of the data could be from the use of data beyond the individual clinical record.

The program is informed by the principles of an 'organisational wide data quality management program' [8] which expands the data hierarchy to be:

- Identification of the patient
- Clinical care
- Coordinate and integrate services (clinical governance)
- Care of populations
- Research, evaluation, and monitoring of safety and quality
- Policy and strategy
- Administration and logistics

The objective of the POLAR program is to construct a data quality platform that allows data to be 'fit for purpose' across all of these possible uses.

Lacking often in consideration of a data strategy is a theoretical basis of the interaction of data with the business of health. While society has been extensively altered by the arrival of the Internet and the digitization of society – creating the term 'Digital Revolution' to mirror the 'Industrial Revolution' – its effects have been delayed in health, in part because health remains heavily reliant on human-to-human interactions.

Relatively under-theorised, the patient-doctor relationship has been characterised using grounded theory [1], and complexity theory [10]. More recently, the patient-doctor-computer relationship has been described using the frameworks of both dramaturgy [20] and the work of Habermas [11]. These frameworks were created at a time when the potential of data to alter the interaction was not fully appreciated. Early work focused on the role of the computer as an agent, whereas increasingly the computer is a conduit for data to influence care. The focus now needs to be on data as the active agent. Data from multiple sources now influences the patient in their healthcare interactions. In developing this integrated strategy, we looked to extend the use of Habermas.

Habermas framed the world according to the actions of the *lifeworld* and the *system* [6; 7]. *Lifeworld* represents the individual and cumulative actions of individuals – through what is known as communicative action. As the actions of individuals, communicative action represents their expression as aspects of personality, culture and even society. Importantly in our context, it is the result of communication that creates this.

By contrast, *system*, and the concomitant *strategic action* represents the actions of the created society on the individual. This creates for us a tension when considering a framework to examine data – for data is generated by the micro-interactions that represent communicative action, yet is also utilised by the system, and represents a means by which stratifications can be implemented. This is big data at work.

By example – a consultation (or other healthcare interactions) represents communicative action based on of the *lifeworld*. In past work, this would be understood as understanding the whole person in developing a plan [17]. Now, however, such an interaction may be influenced by the actions of the computer – which may be the conduit for the strategic action of the system – recommending specific actions based on national initiatives and allowing for the passage of information to inform the system [12]. The conduit for this activity is the data codes generated by the system.

So in describing the data hierarchy below, in effect it represents a ladder from communicative to strategic influence – from changing the communication between individuals though working between individuals and the system, and to the system itself. This creates the data as a mediator between the two worlds, and the organisation wide governance framework is a means of making that work.

## Methods

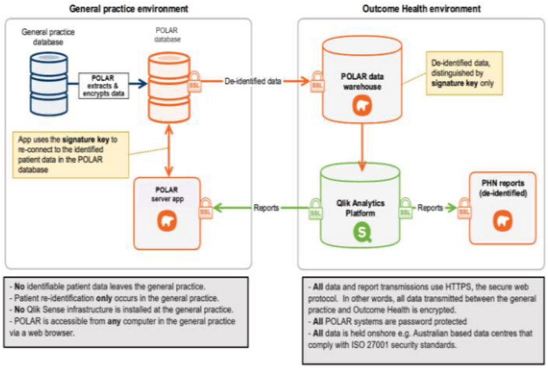
At the initial level, GPs (and practice nurses, other practice staff) collect and record data in their Clinical Information System (CIS) primarily for the care of an individual patient. Thus a blood pressure, test result or social history recording helps an individual GP with an individual patient.

Data in the systems is collected in a variety of formats. Within the record, there is both structured and unstructured data. Unstructured data can be found in the clinical notes section, in incoming and outgoing correspondence, and other documents such as discharge summaries. At the time of writing, because of the potential of identifiability, unstructured narrative data is not collected.

Next is structured data – which may or may not be coded. Structured data includes elements such as diagnoses, medications, and measurements. This data is extracted and processed, de-identified, and placed in a data repository (see figure 1).

The POLAR extraction tool (called Hummingbird) extracts data into the POLAR database and then de-identified data is sent outside the practice to the POLAR data warehouse, where it is framed and processed for the various uses – including feedback to the practice.

At POLAR we now have a multi-layered approach to cleaning and rationalising data to be used at multiple levels. Medication data is organized according to the World Health Organizationav's Anatomic and Therapeutic Coding (ATC) system. Diagnoses are extracted and undergo an automated process to apply a SNOMED-CT-AU code. Use of this process allows a code to be applied to 95% of data extracted from the diagnosis field. Most of the remnants are not diagnoses, but either administrative or other notes recorded in the diagnosis section. Once the diagnoses are coded, there are further overarching groups created – all diabetes codes into a single diabetes category. Key chronic disease groups are utilised as a qualifier as well.



- **No** identifiable patient data leaves the general practice.
- Patient re-identification **only** occurs in the general practice.
- **No** Qlik Sense infrastructure is installed at the general practice.
- POLAR is accessible from **any** computer in the general practice via a web browser.

- **All** data and report transmissions use HTTPS, the secure web protocol. In other words, all data transmitted between the general practice and Outcome Health is encrypted.
- **All** POLAR systems are password protected
- **All** data is held onshore e.g. Australian based data centres that comply with ISO 27001 security standards.

*Figure 1– Extraction Environment*

Pathology codes are recorded in the system generally using LOINC codes, but again there are many variations across the laboratories and inconsistent naming conventions. We have gone through the extractions and grouped them according to clinical utility (hepatitis testing, etc.) [18]. A similar process has been used for radiology investigations, where no current coding schema is available.

### Governance

Governance is just as important as technical ability. Outcome Health provides the POLAR program to the PHNs, and the PHNs provide the service to the practice. As part of accreditation requirements, practices must inform patients of the potential uses of their data.

Ethics approval has been granted separately for the collection and storage of data for the POLAR program, and also for linkage activities. The ethics approval is not project-based, but standing ethics for the underlying processes.

Outcome Health then acts as the data custodian. Various groups then oversee the uses of the data. The research governance group consists of the representative PHN CEOs and reviews applications for research using specific de-identified data fields. The POLAR Data Governance Committee oversees the program as a whole and is made up of a range of internal and external experts. This group is guided by a Data Governance Framework through regular meetings to ensure that the data and processes are ethical, secure and beneficial for all stakeholders.

Other working groups that oversee the presentation of the data to the practices and the PHNs help ti ensure that the tools provided meet the needs of the end users.

## Results

Using the data hierarchy described above, the data is repurposed for use at multiple levels. Importantly, this is not secondary use of data – each use of the data is as valid as any other.

### Identification of the patient

At the core of the program is the necessary data to identify the individual concerned. At the practice level this includes basic demographics such as: date of birth, sex, address, government derived individual health identifier, etc. However, one of the principles of POLAR is that no identifying information leaves the practice, for privacy and security issues. Therefore, Outcome Health has created a linkage process, called ORCA. For any given patient record ORCA creates 3 Hash keys based upon a combination of identifiable data. ORCA uses both AES encryption and SHA256 hashing combined with a salting process to protect the integrity of the hash key.

The linkage process allows patient data to leave the practice, but if necessary be returned to the practice and be re-identified during analytics conducted by the general practice staff. For example, a practice may want to identify all their patients with an active diagnosis of diabetes to ensure they are all participating in a Diabetes Cycle of Care.

In addition, through the use of the HASH keys, de-identified patient data can then be linked across practices and with other sources of data. This mechanism allows for innovative research projects across other data sets such as hospital, mental health, etc. to trace and better understand patient journeys through the health system.

### Clinical care

The reason the data is collected is to provide a longitudinal view of the patient for all the treating clinicians in the practice. This EMR began as an electronic representation or reflection of the paper record – but has evolved beyond that. Digital health data has allowed the system to more efficiently manage clinical/administrative items such as recalls, tracing of results chronic disease management and many other items. [19] It also has made for more complete information on communications such as referrals [3].

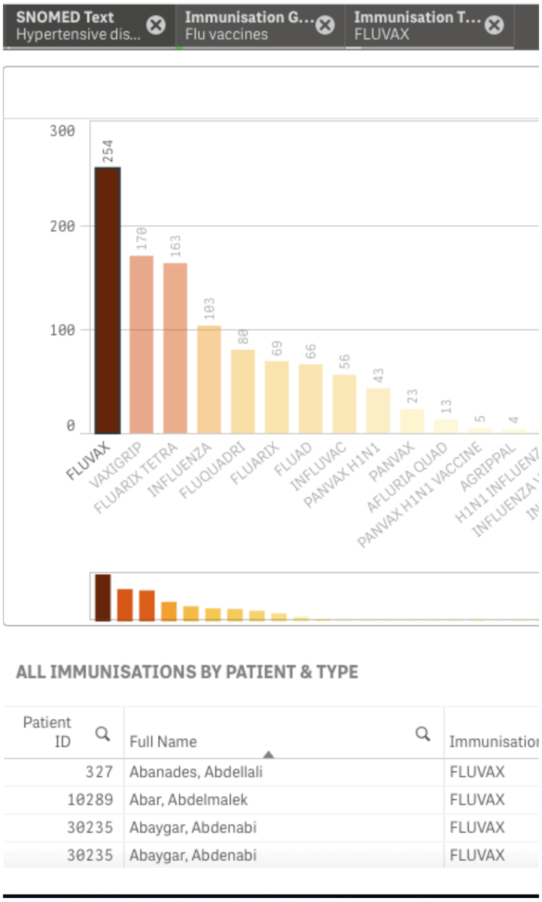
### Coordinate and integrate clinical services (Clinical Governance)

Within the practice, the coding and classification of data allow GPs to easily identify cohorts of patients. POLAR then extracts the data from the GP system for clinical governance – allowing GPs to look at their practice population – how many patients have untreated blood pressure, or abnormal test results. Alternatively, sophisticated calculations such as CHADS2-VASC scores. This allows individual GPs to monitor their populations and ensure consistency of care. System-wide recalls (such as adverse reactions to new drugs, or ensuring over 65s receive an influenza vaccination) are easily handled in batches.

### Care of Populations (Population Health)

De-identified data is presented at the Primary Health Network level to allow the PHN to look at it at the population level – both in clinical and geographical measures. This becomes crucial to allow both population planning and also benchmarking back to the practices. Such programs have been in place for many years [15], but the level of sophistication has improved with the creation of the organised and coded values. It is essential for PHNs to understand what issues are crucial and in what areas to best leverage and administer their health program funding to improve the health system and the health outcomes for their communities.

Figure 2, for instance, can be replicated at the PHN level, giving valuable information on individual practice performance (for feedback and development initiatives) or for the PHN to design population-based interventions.


SNOMED Text
Hypertensive dis...
Immunisation G...
Flu vaccines
Immunisation T...
FLUVAX

ALL IMMUNISATIONS BY PATIENT & TYPE

| Patient ID | Full Name | Immunisatio |
|---|---|---|
| 327 | Abanades, Abdellali | FLUVAX |
| 10289 | Abar, Abdelmalek | FLUVAX |
| 30235 | Abaygar, Abdenabi | FLUVAX |
| 30235 | Abaygar, Abdenabi | FLUVAX |

*Figure 2– All hypertensive patients immunisation status*

### Research, evaluation, and monitoring of safety and quality

One feature of the program is POLAR Data Space, essentially a governance framework involving the PHNs that allows for the data to be used for research. With the focus of the PHNs on making a difference in primary care – the emphasis is on collaborative research with an intent to create practical outcomes through better primary care but also better-integrated care. Quality data can inform many different modes of research, from traditional descriptive methods through to advanced analytics.

Descriptive studies currently being undertaken or completed include:

- Analysis of after hours-presentations [22]
- Prescribing and antibiotic patterns [23]
- Cardiovascular screening in musculoskeletal disorders. [21]

More advanced research involves using the GP data for real-time monitoring of immunisation adverse events, with a trial project underway. Even mora advanced is using machine learning/artificial intelligence to provide advice at the GP consultation on the risk of emergency deparment (ED) attendance in the next 30 days. Using both linked data (mapping the GP journeys of 5 years worth of patients who attended local ED's) and using machine learning on the GP data [16], this tool can accurately predict the risk 75% of the time [13].

### Policy and Strategy

Good data is what should inform strategy and policy. It is impossible to plan if you do not know the landscape in which you are planning. The completeness of the data, its longitudinal nature, and the ever-increasing degree of linkage provide the PHNs significant opportunities to plan service delivery in their local areas, as well as influence policy at a state and national level.

Uniquely, the involvement of the PHNs as change agents allows the finding of the research groups to be rapidly implemented into practice, where appropriate, therefore closing the circle of learning. Examples of these are shown in figures 3, 4 and 5

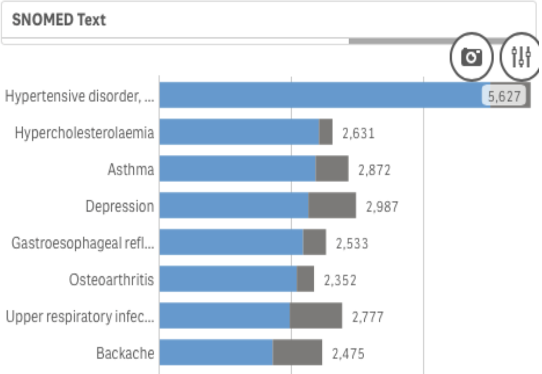


*Figure 3– Top 8 diagnoses of active patients who have attended more than 12 times in a year.*

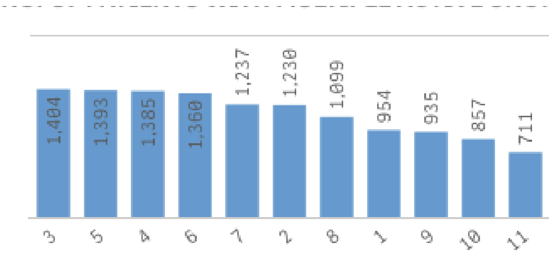


*Figure 4– Same patient cohort as figure 3, but counting the number of active diagnoses per patient.*

**HINT:** Use the active diagnoses to see the top 10. Try using the diagnosis time filte if it changes over time.

TOP 10 CHRONIC DISEASE SNOMED DIAGNOSES PER AGE GROUP

| Age Band / SNOMED Text | Patient Count |
|---|---|
| **60 - 64** | |
| Anxiety | |
| Arthritis | |
| Asthma | |
| Backache | |
| Depression | |
| Diabetes mellitus (unknown type) | |
| Diabetes mellitus type 2 | |
| Hypertensive disorder, systemic arterial | |
| Ischaemic heart disease | |
| Malignant tumour of breast | |
| **65 - 69** | |
| Anxiety | |
| Arthritis | |
| Asthma | |
| Backache | |
| Depression | |
| Diabetes mellitus (unknown type) | |
| Diabetes mellitus type 2 | |
| Hypertensive disorder, systemic arterial | |
| Ischaemic heart disease | |
| Osteoporosis | |

*Figure 5– Example of PHN level data visualisation: Top 10 Active disease groups in specific age groups*

### Administration and Logistics

Within the practice, the data can be used for business and strategic planning to monitor appropriate billing for future service availability, or ensure clinician rostering is appropriate to meet the needs of their patient cohort, address waiting times or anticipate peaks in demand.

For the PHN, the de-identified data enables regional and local level planning to achieve better whole of system care. PHNs need data on health issues in order to understand the needs of communities, target and invest in services to address those needs, prioritise health system improvements, or evaluate performance and outcomes. With their emphasis on making a measurable difference, logistical analytics informs the journey of care so that fragmentation can be reduced by measuring integration and coordination of health services.

## Discussion

That which is readable by humans is not by a computer, and the challenge here is to make the data easily exchangeable by the system. Data in an electronic medical record exists in many forms. Free text is the free-flowing writing that exists in many areas such as narrative notes, and many individual data types may be mixed. Advanced computing techniques such as Natural Language Processing (NLP) are required to interpret these notes. Next comes structured text – in which the free text is at least placed in a recognisable area. Writing in a designated ‘diagnosis’ field fits into this category. Finally comes coded text – in which a specific term has been used to link to a computer readable code.

In order to make the data useable beyond the GP environment, POLAR has chosen to concentrate on the structured text that exists in the system. Narrative data contains much identifying information, and the commitment is that no identifying information shall leave the practice. Focusing on the structured data, we first concentrated on the diagnosis fields. Using a combination of automated and manual processing, we have been able to apply SNOMED-CT-AU codes to the vast majority of diagnoses. Once coded, we have been able to group relevant diagnoses into clinically meaningful areas. So all instances of diabetes, for instance – or a broader classification of chronic diseases. This makes manipulation of data for clinical governance and population health all the more manageable through the business intelligence analytics dashboards provided to general practice and PHNs for analysis.

Medications have been mapped to the Anatomic and Therapeutic Classifications (ATC) system, which applies five different levels, from 'all cardiovascular' down to individual drugs. This grouping, again, makes manipulation and interpretation of medications much simpler. A similar process is underway for pathology and radiology testing.

## Conclusions

The key to making digital data understandable by the system, and useful for multiple purposes, is a multi-pronged strategy of coding, grouping and iterative analysis, underpinned by robust governance structure and a clear strategy.

## Acknowledgements

We acknowledge the support and activity of the PHN's involved: Eastern Melbourne, Gippsland, South East Melbourne, Central East Sydney, South West Sydney.

## References

[1] A. Als, The desk-top computer as a magic box: patterns of behaviour connected with the desk-top computer; GPs' and patients' perceptions, *Fam Pract* **14** (1997), 17-23.

[2] J. Beilby and J. Furler, General Practitioner Services In Australia, in: *General Practice In Australia: 2004,* Australian Department Of Health and Ageing, ed., Australian Government Publishing Service, Canberra, 2005.

[3] Crampton N, Reis S, and Shachak A, Computers in the clinical encounter: a scoping review and thematic analysis, *J Am Med Inform Assoc* **23** (2016), 654-665.

[4] S. de Lusignan and C. Pearce, What are the informatics barriers to conducting international research using routinely collected primary care data?, in: *Medinfo 2010*, Cape Town, South Africa, 2010.

[5] S. de Lusignan and C. van Weel, The use of routinely collected computer data for research in primary care: opportunities and challenges, *Fam Pract* **23** (2006), 253-263.

[6] Habermas, *The theory of communicative action, v 1: Reason and the rationalisation of society*, Beacon Press, Boston, 1984.

[7] Habermas, *The Theory of Communicative Action v2; Lifeworld and System: A Critique of Functionalist Reason.*, Polity Press, Cambridge, 1987.

[8] S.T. Liaw, C. Pearce, H. Liyanage, G.S. Liaw, and S. de Lusignan, An integrated organisation-wide data quality management and information governance framework: theoretical underpinnings, *Inform Prim Care* **21** (2014), 199-206.

[9] S.T. Liaw, J. Taggart, H. Yu, and S. de Lusignan, Data extraction from electronic health records - existing tools may be unreliable and potentially unsafe, *Aust Fam Physician* **42** (2013), 820-823.

[10] C.M. Martin and J.P. Sturmberg, General practice--chaos, complexity and innovation, *Med J Aust* **183** (2005), 106-109.

[11] C. Pearce, Computers, Patients and Doctors - Theoretical and Practical Perspectives, in: *Health Professionals Education in the Time of Clinical Information Systems,* A. Shachak, E. Borycki, and S. Reis, eds., Academoic Press, London, 2017, pp. 5-22.

[12] C. Pearce, M. Arnold, C. Phillips, S. Trumble, and K. Dwan, The patient and the computer in the primary care consultation, *Journal of the American Medical Informatics Association : JAMIA* **18** (2011), 138-142.

[13] C. Pearce, A. McLeod, N. Rinehart, J. Patrick, A. Fragkoudi, J. Ferrigi, E. Deveny, R. Whyte, and M. Shearer, POLAR Diversion: Using General Practice Data to Calculate Risk of Emergency Department Presentation at the Time of Consultation, *Appl Clin Inform* **10** (2019), 151-157.

[14] C. Pearce, M. Shearer, K. Gardner, and J. Kelly, A division's worth of data, *Aust Fam Physician* **40** (2011), 167-170.

[15] C. Pearce, M. Shearer, K. Gardner, J. Kelly, and T.B. Xu, GP Networks as enablers of quality of care: implementing a practice engagement framework in a General Practice Network, *Australian Journal of Primary HEalth* **18** (2012), 101-104.

[16] C.M. Pearce, A. McLeod, J. Patrick, D. Boyle, M. Shearer, P. Eustace, and M.C. Pearce, Using Patient Flow Information to Determine Risk of Hospital Presentation: Protocol for a Proof-of-Concept Study, *JMIR Res Protoc* **5** (2016), e241.

[17] S.J. Potter and J.B. McKinlay, From a relationship to encounter: an examination of longitudinal and lateral dimensions in the doctor-patient relationship, *Soc Sci Med* **61** (2005), 465-479.

[18] G. Sezgin, A. Georgiou, R.A. Hardie, L. Li, L.G. Pont, T. Badrick, G.S. Franco, J.I. Westbrook, N. Rinehart, A. McLeod, C. Pearce, M. Shearer, R. Whyte, and E. Deveny, Compliance with pathology testing guidelines in Australian general practice: protocol for a secondary analysis of electronic health record data, *BMJ Open* **8** (2018), e024223.

[19] D. Swinglehurst, T. Greenhalgh, and C. Roberts, Computer templates in chronic disease management: ethnographic case study in general practice, *BMJ Open* **2** (2012).

[20] D. Swinglehurst, C. Roberts, S. Li, O. Weber, and P. Singy, Beyond the 'dyad': a qualitative re-evaluation of the changing clinical consultation, *BMJ Open* **4** (2014), e006017.

[21] L.R. Turner, F. Cicuttini, C. Pearce, and D. Mazza, Cardiovascular disease screening in general practice: General practitioner recording of common risk factors, *Prev Med* **99** (2017), 282-285.

[22] L.R. Turner, C. Pearce, M. Borg, A. McLeod, M. Shearer, and D. Mazza, Characteristics of patients presenting to an after-hours clinic: results of a MAGNET analysis, *Aust J Prim Health* **23** (2017), 294-299.

[23] J. Yan, L. Hawes, L. Turner, D. Mazza, C. Pearce, and J. Buttery, Antimicrobial prescribing for children in primary care, *J Paediatr Child Health* (2018).

### Address for correspondence

Dr. Christopher Pearce, drchrispearce@mac.com.