POLAR DIVERSION

OUTCOME HEALTH

POLAR Diversion: Using GP data to calculate risk of ED presentation.

Principal Investigator

A/Prof Christopher Pearce – Director of Research, Outcome Health

Co-Investigators

Adam McLeod – CEO, Outcome Health

Natalie Rinehart – Population Health and Research Manager, Outcome Health

Anna Fragkoudi – Research Officer, Outcome Health

Jason Ferriggi – CIO, Outcome Health

Jon Patrick- CEO, Health Language Analytics

Marianne Shearer- CEO, Gippsland PHN

Paula Eustace- Research Officer, Eastern Melbourne PHN

Acknowledgements

Outcome Health would like to acknowledge the HCF research foundation for its funding of this project.

Table of Contents

ETHICS APPROVALS

Monash University Project ID: CF14/218-2014001187 (old), now 5260 (new), Project Title: *A Hospital Presentation Risk Score using General Practice Data.*

Monash University Project ID: CF12/1057 – 2012000504, Project Title: *Melbourne East MonAsh GeNeral PracticE DaTabase (MAGNET).*

Eastern Health Project ID: LR09/2014 Data Linkage, Project Title: *Request for data linkage using a Statistical Linkage Key.*

Eastern Health Project ID: LR55/2014 Focus Group, Project Title: *Factors in avoidable presentations to the emergency department: Clinician perspectives.*

LAY SUMMARY

This study developed an algorithm to calculate a patient's risk of a hospital Emergency Department (ED) presentation. The algorithm was developed by a machine learning based on de-identified general practice patient clinical information linked to Emergency Department (ED) admissions data. The final algorithm was based solely on general practice data that is routinely collected and the tool developed is designed for use in Australian general practice.

To further gauge a General Practice perception of the risk scores, seven GPs evaluated the algorithm using their own patients. They were asked if they perceived the risk score to be accurate, what patient attributes were useful and if it would change their treatment for patients.

The aim was to highlight patients with higher risk of an ED presentations to the GP at the time of attendance, who can then potentially take action to avoid the ED presentation and reduce unnecessary visits to emergency departments nationally.

The algorithm results indicated that post the last GP visit 73.6% who presented to ED within 0-30 days were correctly identified. Patients who attended within 31-365 days fell to 36.8% and those who had no need to attend ED were correctly identified 82.3% of the time. GP perceptions of the algorithm when run over their own patients varied with higher agreement with the 31-365 time category and lower risk scores.

EXECUTIVE SUMMARY

Mapping research project and implementation

This project explored how an ED presentation risk identification process designed for use in general practice can identify patients' risk of ED attendance and the potential interventions that GPs may use to reduce this risk.

Approximately 70 per cent of the total burden of disease in Australia is attributable to disease groups that could be either prevented or managed outside the inpatient system (1). Reduction of avoidable hospital admissions is key to improving quality of life of patients and effectively managing expensive hospital resources. With the reduction of avoidable hospitalisations as one of the goals for Primary Health Network activities (2, 3), a clinically proven mechanism to highlight patients at risk is essential (4).

A retrospective, multisite cohort study using a mixed methods approach was conducted. General Practice data was extracted from 50 practices and linked with local health network data across three Emergency Departments (Box Hill Hospital, Angliss Hospital and Maroondah Hospital).

The final linked data sets identified GP patients with a linked ED presentation within 1 year, not injury based, with adequate data fields across 17,067 GP visits for 8,479 unique patients. This linked data was used to develop a machine learning algorithm to determine a range of variables, which in combination led to specific risk scores. Thirteen per cent of the data was kept separate to allow for blind testing of the final algorithm.

The results were then tested with active patients across seven GP Clinics (420 patients), where GPs evaluated the risk scores, to determine if the scores aligned with their clinical judgement. The patient results and attributes were presented in an electronic format, which included any changes they would conduct in their treatment based on their scores.

Significant work was undertaken mapping the diagnosis fields in GP clinical software to a standard, (SNOMED CT), the medications to the World Health Organisation ATC Drug Classification System, pathology to a new model relevant for General Practice and a series of severity mappings for key measures such as BP, BMI etc. This allowed inputs to the machine learning to be ranked, highlight specific high-risk groups of diagnosis, medications or pathology as key factors that may further contribute to the machine learning.

The model was built with this set of attributes, including diagnosis, measurements, pathology tests etc. The tool used 21 'group attributes' (e.g. Medication is one attribute with around 6000 items), including 52 'relational attributes' (e.g. Medication dose or frequency), from GPs' clinical software to inform the algorithm. This means many thousands of variables were considered to build the risk score, not just a handful of chronic diseases as is the case with most traditional risk algorithms. Due to GP data quality, many attribute values were left empty or have content equivalent to "unknown". The included attributes are explained in detail later in the report.

To account for different historical timeframes in the GP data, the data was divided into a 'Last Visit' highlighting what occurred at the most recent GP encounter and 'Previous Visits' which was a concatenation of previous visits to the GP. The historical value of each of the variables in the 'Previous Visits' was given a time value based on historical significance, e.g.: the Previous Visit Diagnosis used up to ten years, but for medications only two years of prescriptions were used.

Electronic online survey tools were also developed for the GP feedback section of the study, as well as semi-structured one-on-one telephone interviews with participating GPs and a focus group with ED staff.

Using a portion of the retrospective data that was not used to create the algorithm (i.e. it was deliberately kept separate), results showed:

Patients who did present to ED:

At 0-30 days post the last GP visit 73.6% of GP patients who subsequently presented to ED were correctly identified by the algorithm. At 31-365 days, this fell to 36.8% highlighting the difficulty in long-term identification.

Patients who did not present to ED:

Of patients who did not present to ED, 82.3% were correctly identified as not presenting to ED at any point in the year post the last GP visit recorded.

GP Perceptions:

GPs tended to agree with the results but felt that the data was more correct for the 31-365 days than at 0-30 day, which was opposite to what the 'factual' data showed. This highlights the tension between the tacit knowledge of clinicians versus the explicit knowledge of the data – and highlights the challenges integrating those two aspects.

INTRODUCTION

The hospital Emergency Department (ED) is an integral part of the health care system. With Primary care/General Practice, EDs form first contact in urgent situations. Worldwide, however, EDs face the common issue of patients presenting with issues that potentially could have been addressed in a timely fashion in the non-hospital setting (5, 6). Such issues may account for as many as up to 40% of ED presentations (7). The cited consequences of extending the ED to cater for these avoidable presentations fall into two main groups: the fiscal burden on the tertiary care sector (ED's are much more expensive than GP), and the presumed drain on its limited resources, affecting the quality of care (8-10).

Another policy issue is the notion of potentially preventable hospitalisations (PPH); those that could have been prevented through the use of appropriate primary health interventions in a non-hospital setting (i.e. general practice, community health, nurses, allied health, dental, etc.) (11). In other words, how could a hospital admission been prevented by better interventions and management within the primary care system? It has been suggested that rates of PPHs across different categories can be used as a measure of care in the primary health sector and are used as a key performance indicator (KPI) for Primary Health Networks (PHNs) across Australia(3).

In an effort to develop solutions, there has been considerable research directed at the drivers of presentations to the ED. Researchers often approach the task of identifying these drivers with different sets of assumptions. For instance, the issue of inappropriate presentations and preventable hospitalisations may be viewed through fiscal, policy, service delivery, access/equity or other lens'. There is a need for appropriate community supports and targeted care initiatives, such as case management, allied health or specialist intervention, wound care, home help, and hospital-at-home services to reduce the likelihood of a presentation. The early engagement of appropriate community supports to improve health and well-being and reduce the burden on acute services will make important inroads into the number of avoidable hospital presentations and admissions. In any exploration of the issue, it is instructive to consider perspectives of the clinicians at the coalface, who would be expected to have an overview of the unique and interacting clinical, socio-demographic and circumstantial factors affecting the person's reason/s for presentation.

Many people suffer from multiple risk factors especially increasing with age (12). Chronic conditions and their risk factors, often diagnosed and managed within general practice, are the primary source of Australian morbidity and mortality with some research suggesting it is responsible for up to 90% (13, 14). Therefore, attention to preventive measures, timely intervention and optimal use of support strategies and services to address chronic conditions and risk factors can reduce people's risk for hospital presentation. Addressing reversible health risk factors, managing chronic conditions and minimising their impacts are important primary health care priorities (15).

In addition, managing patients within primary care may lead to financial benefits for the community, government and health insurances. The comparison between the cost of activities within primary and tertiary care reveals a substantial difference. Although GPs set their own visit fee, the average cost to the patient (on top of the Medicare rebate) is approximately $30 (16). Thus, the combined cost of a GP visit is approximately $65, shared by government and the patient. On the other hand, when a patient presents to a hospital's ED, this cost can be $586 (17), to the public sector which does not include any subsequent hospital admissions. Thus, it is reasonable to conclude that in certain occasions, when it is possible to prevent an unnecessary ED presentation, which can be managed within primary health care, the financial benefit for the health system will be significant.

Primary care provides a low cost and low intensity approach to health care that ideally is suited to addressing both illness management and prevention.

Considerable research has gone into prediction of emergency presentation of patients (18-22), predominantly using linear regression models. Numerous predictive algorithms/models have been developed over the years internationally with the aim to identify high risk of emergency presentation, admission and re-admission among patients (22-32).

PARR (and the modified PARR2) were developed for the Department of Health in the United Kingdom (UK) and used hospital inpatient data to identify individuals at high risk of re-admission to hospital (27). PARR and PARR2 are algorithmic models (predictive models) that are run through software named PARR++. PARR developers linked the hospital episode statistics (HES) data and inpatient data to the HES data from community services (GP records, district nursing records and social services data) using criteria that are known to be risk factors in future admissions to hospital. The purpose was to develop a model that would be able to predict patients at risk of hospital presentations.

PARR's last update was in 2007 and included a range of variables such as socio-economic, diagnostic, prior hospital utilisation/cost, pharmacy data health status/functionality and clinical data. Although there was initial success of the tool, funding was withdrawn for further development and the models are no longer maintained (33). As a result, because PARR++ is based on older data it is likely to be ineffective at predicting the risk of Emergency Department readmission (34).

The QAdmissions tool, also developed in the UK, is a model involving 30 variables initially based on general practice data (28) such as demographics, lifestyle, chronic diseases, prescribed medication, clinical values and pathology test results. The algorithm attempted to estimate the risk of an emergency hospital admission for adults only, excluding children aged under 18. Initially, the tool raised some doubts concerning its reliability, as its predominant aim of reducing emergency admissions was not achieved.

In order to overcome its weaknesses, Hippisley-Cox *et al.*, 2013, re-developed a new calculation score, using GP data linked to HES (Hospital Admission) data. As a result, the tool was significantly improved presenting positive predictive values[1]. The tool demonstrated moderate predictive ability as the positive predictive value for the top 10% of patients was 42% with GP-HES linked data and 40% with the GP data

[1] Positive Predictive Value expresses the likelihood that a person with a positive test has a disease. The positive predictive value depends not only on the accuracy of the test, but also on the prevalence of the disease.

only, while sensitivity was a respective 39% and 37% (28). However, the updated QAdmissions tool includes recorded hospital emergency admissions that occurred during the previous year. In Australia, the healthcare system is not unified with data sets generally kept in isolation, which makes building models that utilise a combination of data types (i.e. general practice and hospital) difficult. As such, the QAdmissions tool is difficult to be automated in an Australian general practice context.

This difficulty was also evident in Australia when the Patient Admission Prediction Tool (PAPT) was developed by the CSIRO to manage hospital inpatient bed occupancy efficiently (22). It aimed to forecast Emergency Department presentations using hospital admission data. However, there was no link to patient diagnosis meaning the tool could forecast ED presentations, but failed to identify the reasons for presentations and provide a point of intervention.

Other models in Australia, specifically Victoria include a range of Hospital Admission Risk Program (HARP) algorithms. The HARP programs aim to assist people with multiple complex risk factors for health/social problems, who are at high risk of emergency department presentation or hospital admission (35). It suggests that the risk effect is magnified when multiple risk factors are combined or cluster together (36). However, a key limitation is that HARP models is they tend to only target patients with chronic diseases such as diabetes, cardiac failure, COPD, mental illnesses or people with complex needs rather than a broader audience and again, requires multiple data sources to complete and therefore cannot be automated on a single data set. This need to input data to make a tool work in real time for practitioners can reduce the real time applicability of a tool as it relates to a GP workflow (37).

In a significant policy shift, the Australian Government is undertaking a trial of 'Health Care Homes' – where people are allotted to their general practice who receives an annual payment based on the patients risk of ED presentation. The trial will use a combination of the 'Western HARP' and 'QAdmissions' algorithms but still requires data entry on behalf of practices and to date no ED presentation data is being made available further diluting the model.

Other models that attempted to predict patients at risk of admitting to hospitals within a period of 12 months are the Combined Predictive Model (CPM) (27) and Adults Veteran Association: Ambulatory Care Sensitive Conditions (AVA: ACSC) (38). The CPM is based on elements of the PARR model, meaning it supplemented the already obtained PARR data with additional one. This was mainly because the data used for PARR were limited and the identification of patients' at risk was more narrow (27). On the other hand, AVA concerned only patients registered with the Veterans Health Administrator (VHA) (USA) and tried to identify those at risk of hospitalisation. This model used logistic regression for prediction purposes, however, its predictive power was moderate and relied only on a specific population impacting on its not generalisability (38).

Some other models developed to predict ED presentations/admissions both in Australia and overseas (26, 31, 39, 40) have demonstrated good predictive values such the CSIRO Chronic Disease model (31) and DPM targeting patient 65+ years (40). The DPM used 89 variables across general practice and hospital data sets and is utilised in a considerable amount of practices in the UK. The CSIRO Chronic Disease model was developed by the Australian E-Health Research Centre and the Health and Hospital Service in Brisbane, aimed to predict patients with chronic diseases risk of re-hospitalisation (31) and also showed good predictive power. Nevertheless, many models do not present evidence of the reduction of hospital admissions (26) and others target only elderly patients most at risk of hospitalisation (39). Other models that target patients who previously have been admitted to hospital and are at risk of re-admission, such as EMR (32), LACE (41) and Queensland Hospital Statistical Algorithm (30) have shown to be poor predictors of hospital re-admission, especially when concerning the elderly population and as such, do not offer strong in primary care to identify high risk patients.

One recent Australian model, the Gold Coast Integrated Care Model (developed by Queensland Health, the Gold Coast HHS Board and the Gold Coast Primary Health Network) aims to reduce presentations to the EDs, decrease admission rates and enhancing the capacity of specialist outpatients (42) has shown some early positive results. The model utilises hospital data and GP clinical information for multiple diagnoses, high use of clinics or pharmacy to risk stratify patients risk of admission. It is assumed that 3% of patients have a high degree of complexity and will benefit from a shared care plan across multidisciplinary teams and 11% are diagnosed but stable patients who can self-manage with GP support. Although still in the early phase it has developed a shared patient record linked across data sets and is currently reviewing across 140,000 patients in the Primary Health Network for effective identification and management of higher risk patients. Although showing some early promise, the model requires substantial data linking across hospital and general practice data.

Other alert or decision support tools utilised in Australia have often been condition dependent. Tools such as the Absolute Risk Calculator (43) develops a risk of risk score of getting heart, stroke or blood vessel disease (cardiovascular disease) in the next five years based on patient demographics and GP stored data (i.e. age, gender, smoking, blood pressure, cholesterol, diabetes and ECG LVH). The AusDRisk (44) calculator aims to determine the risk of a type 2 diabetes diagnosis within the next 5 years through the review of age, gender, ethnicity, family history, High BP, smoking, diet, exercise and waist measurement. The QCancer risk tool was developed by the same British researchers that developed the QAdmissions tool (45). It was implemented within general practice trying to test its feasibility in the Australian context. The tool aims to improve the diagnostic assessment of patients with indications of cancer symptoms within primary health sector. These risk calculators aim to alert practitioners to at risk patients to intervene prior to a more serious diagnosis that may lead to long-term chronic conditions with a range of negative consequences for the patient and the health system. The study found that although GPs felt the risk tool could a valuable addition, there was considerable variation in their perceived accuracy of the tool, differences in how they viewed the cancer risk of standardised patients and concerns of how it would interrupt the flow of their consultations.

Despite some of the difficulties in this area, these tools aim to contribute to a reduction in patient conditions becoming more acute, and potentially an ED presentation occurs. Although clinicians recognise the need for these tools be a part of their practice, they also appear reluctant as they need to be assured about the quality and the accuracy of the produced risk scores provided of the respective tools (45, 46) in comparison to their own clinical judgement.

In addition, these tools generally do not offer effective ‘realtime’ information to inform GPs at the point in time care with patients. The POLAR Diversion project endeavored to develop an automated algorithm based on general practice clinical and billing data to identify potential ED presentation risk scores. The development of such a tool that automatically calculates patient alerts based on GP data can support practices to identify high-risk patients and where possible, intervene to reduce their risk and avoidable hospitalisations.

AIM

To develop a risk assessment tool for presentation to ED, based on general practice data alone, capable of being deployed in real time.

METHOD

The project essentially involved four phases (47):

- The linking of data for 5 years of ED presentation data with general practice data, to understand the general practice journey of those patients.
- Mapping, grouping and ranking of General Practice data to allow a consistent data set to be fed into the algorithms.
- Using that data to develop a machine learning application to predict risk of an ED presentation, and test reliability with a set of data.
- Develop a predictive tool to be deployed in general practice and receive feedback on tool performance.

The POLAR Diversion project is the first project to have developed and trialed an algorithm predicting risk of presentation to the Emergency Department, not only targeting patients with chronic diseases, but all conditions at risk for presentation for varying reasons.

In an effort to explore the ED presentations and the drivers of presentations from general practice data, the study was underpinned by the following assumptions.

1. That an accurate prediction of risk could be made using stand-alone GP data.
2. That the use of a machine learning algorithm would allow for a large number of variables to be taken into account.
3. That by exposing the developed tool with GPs the findings could be further validated

Ultimately, the results showed that 1 and 2 were possible, the findings from 3 showed that while overall most GPs agreed with the results they were more likely to view a short-term acute ED presentation risk with the long-term possibility.

Ethics

POLAR Diversion project was granted ethics approval from Monash University (MHREC 5260) along with the ability to collect and store general practice data (CF12/1057 – 2012000504). Following this, Eastern Health granted two additional ethics' approvals for linking general practice data with emergency data (Ethics LR09/2014) and for holding a focus group with ED clinicians to obtain views and perceptions on localised factors potentially contributing to avoidable presentations (Ethics LR55/2014).

Timeline

A range of steps took place throughout the four-year study, which can be broken into four different key activities as shown below.

Table 1 Timeline of activities

	Activity	2014	2015	2016	2017
Process preparation	ED Focus Group				
	Data Mapping and Linkage				
	Algorithm development				
Process implementation	GPs evaluation of risk algorithm				

Data Mapping and Linkage

Aim

This process aimed to highlight common patients across the general practice and ED datasets to indicate when and why patients presented to emergency departments. A profile of presentations, characteristics of patient demographics and nature and features of chronicity/complexity within this profile were used to develop the risk algorithm.

Data Collection Process

During 2014/15, Melbourne East GP Network, trading as Outcome Health (previously trading as Inner East Melbourne Medicare Local) extracted and collected de-identified general practice data using the GRHANITE extraction tool (48). The collected data was extracted regularly from 50 participating general practices for population health planning and research.

Data was extracted from general practice clinical and billing software at the practice site. The data collection process was divided into two stages:

1. Collecting identifiable data that does not leave the practice to generate an encrypted linkage key and
2. Sending de-identified data with the linkage key to Outcome Health.

This division of data allowed Outcome Health to give General Practices patient identifiable reports while generating aggregated de-identifiable data for the then Medicare Locals (not Primary Health Networks), giving each practices and the Medicare Locals relevant reports without compromising privacy (49, 50).

During extraction, an encrypted Statistical Linkage Key (SLK) is produced for each patient record through key demographic variables (i.e. age, gender, name, address, etc). The SLK is retrieved with the patient records, but none of the identifying patient material leaves the practice.

The same SLK system was also used during extraction of Eastern Health's ED dataset. The de-identified data from the Eastern Health Emergency dataset provided by Eastern Health (EH) for the years 2008-2012. Linking the SLKs from the two data sets (i.e. general practice and ED) allowed for patients who attended the ED to be identified and linked to their general practice data.

Data

The GP data collected represents the structured data held for that patient in the record. No narrative information was collected, the data was cleaned (i.e. free text diagnosis grouped to SNOMED-CT) and any potentially identifying information removed. So structured diagnoses are collated and converted into SNOMED-CT codes (51), and code grouping made (all forms of diabetes, for instance). Medications data was coded according to the WHO Anatomic and Therapeutic Classification (ATC) coding system, and grouped into levels, again allowing for therapeutic grouping to be used, rather than individual drugs. Numeric information (such as weight, blood pressure and/or individual test results) were also mapped to indicate the level of risk associated with these measurements.

Development of the Algorithm

Algorithm Sample

At the time of the general practice data collection there were approximately 16 million de-identified GP patient records across 744,477 unique patients over the five years (2010-2015). A total of 12,448 of these patients had a linked ED presentation within the last year. As shown below, the sample used for the final algorithm decreased during the project as essential vs optional data was identified, removal of injuries within the ED data, etc. Once the model's parameters had been finalised the final data count for the algorithm included 17,067 GP visits, across 8,479 unique patients. The algorithm was also crosschecked across 29,892 GP visits for 29,185 unique patients for the Non-ED sample.

Table 2 Algorithm sample

Criteria	GP Visits	Unique Patient Count
All data supplied	16,305,096	744,477
Those with a linked ED visit	37,789	21,376
Those with a linked ED within 1 year	26,691	12,448
Those with a linked ED within 1 year and no injury	20,213	10,610
Those with a linked ED within 1 year, no injury, with adequate data fields	17,067	8,479
Those without a linked ED visit used for the Non-ED sample	150,000	144,490
Those without a linked ED visit used for the Non-ED sample with adequate data fields	29,892	29,185

Algorithm Method

The model was developed through 'machine learning' – using advanced computer techniques. Machine learning has many variations, but they all share an important difference from the traditional statistical methods such as logistic regression or analysis of variance—the ability to make accurate predictions on unseen data. To optimize the prediction accuracy, often the methods do not attempt to produce interpretable models. This also allows them to handle a large number of variables common in most big data problems.

Key data utilised within the model can be viewed in the results with 21 relational attribute groups including 52 attributes across the GP data. There were a total of 157,330 individual attributes to train the model.

Accompanying the flexibility of emerging machine learning techniques, comes uncertainty and inconsistency in the use of such techniques. Machine learning, owing to its intrinsic mathematical and algorithmic complexity, can be difficult to untangle when a result does not make sense to what a human would consider logical or common knowledge. This large number of conflicting factors can mean for minor less relevant variables that share a common link across an 'admission flag' these can at times be misconstrued by the machine learning. Ultimately humans and disease progression is not predictable by its nature, so creating an algorithm that can take all factors into consideration is a near impossible task The machine learning creates a best guess based on what it has learned or gleaned from the data over time.

The major algorithm building tasks included:

- Building a coherent representation of the patient records suited to computing a predictive model;
- Testing a variety of combinations of attributes for the best results;
- Converting the many attributes available into domain ranges that were relevant to the task;
- Testing many class configurations around 30-day, 90-day, 180-day, 365-day and post-1-year attendances.
- Devising representations of the various time lapses between the GP visits of patients;
- Separating the analysis to use non-injury cases.
- Splitting the patient information into two groups: current visit information and historical patient information, so GPs could easily review the most pertinent patient information.

Thirteen percent of the GP data was held back from developing the predictive model, and then used 'blind' to test the reliability of the model.

GP Evaluation

The evaluation of the risk scores by GPs took place between March-July 2017.

Aim

The aim of this aspect of the project was to collect GP's perceptions of the risk algorithm scores, the patient attributes and their potential treatment changes based on their own active patients. Implementing the risk scores in a live general practice environment and consequently acquiring feedback on the accuracy and utility of the report was considered an essential step to validate the future use of the alert tool.

Process

Participants were required to review 60 patient records each (n = 420) with a spread of scores to ensure adequate evaluation across all score spectrums. Patient information shown for review, when available, included:

- Patient Name
- Demographics: Age, gender, pensioner, DVA, Aboriginal or Torres Strait Islander status
- Risk Factors: Smoking, Alcohol, Allergies and BMI
- Diagnoses
  - Current Diagnoses: Marked as active in GP software or identified in the last visit
  - Historical Diagnoses: Up to 10 years
- Medications
  - Current Medications: Marked as current in GP software or used in past eight months
  - Historical Medications: used in the last 9-24 months
- Pathology Tests and Results
  - Current Pathology: Tests and results reviewed in the last visit
  - Historical Pathology: Tests and results reviewed in the last 12 months
- GP Measurements (i.e. BMI, BP, Temperature, weight, etc)
  - Current Measurements: Tests and results reviewed in the last visit
  - Historical Measurements: Tests and results reviewed in the last 5 years
- Patient risk scores for the 0-30 day and 31-365 day

GP participants answered a range of questions concerning the

- Accuracy of the risk score
- Accuracy of the attributes presented for each patient
- Any changes they would make in patient treatment given the score

Recruitment and Consent

GPs who participated in the study were employed by Practices who had pre-existing contract agreements, via three Victorian PHNs in both urban and rural settings (Gippsland, South Eastern Melbourne and Eastern Melbourne). All practitioners were required to work a minimum three sessions per week (0.6 FTE), in order to ensure that participants had sufficient active patient counts in the last 12 months to evaluate the risk scores.

The PHNs recruited GPs from their local area with 10 consenting to be involved in the study and data extracted from their practice. Due to one technical issue and two GPs not completing the process, the final sample consists of seven participants that completed both the patient record evaluation, final survey and the one-on-one telephone interview.

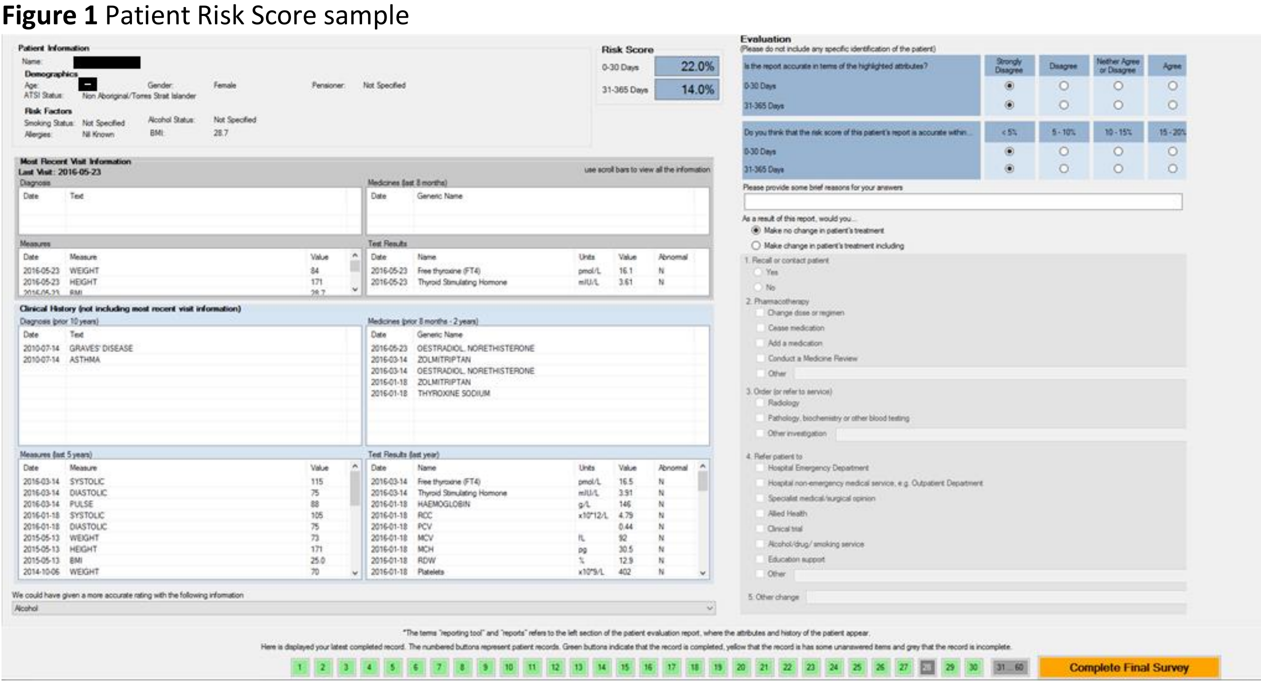
General Practitioners provided their consent prior their participation in the study, by signing a Consent Form. Outcome Health practice liaison staff approached the participants and provided an Explanatory Statement, a Consent Form and a Permission Letter to be signed by GPs' Practice. The participating GPs were also supplied with a Withdrawal Form, which allowed them to withdraw from the study at any stage.

Data Collection Tools

The patient record evaluation application was developed for study purposes and was hosted on a pre-existing software utilised by practices (POLAR GP). The survey presented demographic and clinical attributes of patients' medical history and allowed GPs to evaluate the efficacy of the risk scores based on the attributes included in each patient record.

The patient's evaluation tool comprises 15 items containing clinical information for each patient and nine questions, which allowed GPs to evaluate the accuracy of the risk score, the attributes presented and any treatment changes identified for the patients based on their clinical judgement, as seen on the example below.

Figure 1 Patient Risk Score sample



A final survey was developed as an adjunct to the patient's evaluation form. This form used Likert Scales over 11 questions, allowing GPs to evaluate the accuracy of the report overall, the usefulness and the intension of usage. The purpose was to acquire additional feedback, which would allow us to investigate whether the results from different approaches well class.

An interview guide was developed to support the one-on-one semi-structured telephone interviews that occurred during the week after the completion of the patient evaluation records. The topics discussed during the interviews concerned the experience of GPs using the tool.

Data Collection Process

Data extracted from the participating practices included patients that have visited the GPs during the past year of the extraction commencement. Approximately 1,000 patient records were extracted from each practice. The records were run through the algorithm in order to identify patients at-risk of an emergency presentation at hospital. A list of patients' de-identified IDs was created in order to perform a randomised application, which ensured a spread of scores across the two-time periods (0-30 days and 31-365 days). The randomly selected patients met the pre-determined alert criteria that have historically identified patients who are at risk of hospital presentation and provided to GPs for evaluation. The included attributes were specified in the report.

After reading the explanatory statement and signing the consent form, the participating GPs were asked to login to POLAR GP, already installed in the Practice's computing system. Each GP completed 60 patient reports totaling 420 unique patient records. Following the evaluation of the 60 patient records, practitioners completed a semi-structured survey with some free response items. GPs began reviewing patient records May-July 2017 and were given incentives for their time.

The final data collection took place during the semi-structured one-on-one telephone interview during the week following the completion of the patient evaluations. The interview was held at a day and time that suited participants, with an expectation that most GPs would want to be interviewed over the phone in between consultations.

Clinicians Perceptions towards Avoidable ED Presentation

Aim

This aspect of the project took place in 2015 and aimed to explore the perspectives of practicing ED clinicians on the localised factors potentially contributing to avoidable presentations.

A focus group interview was conducted involving a led discussion between six experienced ED clinicians (i.e. nurses, doctors). Participants were asked for their perceptions of the nature and drivers of avoidable presentations, following the general format of the interview guide. The interview lasted approximately 45 minutes.

Recruitment and Consent

The recruitment took place at Box Hill ED Services via a poster advertisement. Participants were provided a participant information statement and a consent form before interviews took place. Participants (nurses or doctors) were required to have minimum of two years post qualification experience in an Australian emergency department. Persons who did not meet the clinical requirements were excluded from the study.

RESULTS

This section presents the approach taken to analysis of data collected in the surveys and interviews.

Algorithm Development

Initial Data Review

Data attributes as part of the Patient information consisted of:

- Age
- Gender
- DVA Status
- Pension Status
- Aboriginal Status
- Alcohol (non-drinkers 28%)
- Care Plan Goals,
- Data attributes from the Clinical table includes:
    - Smoking status (12% smoker, 23% ex-smoker)
    - Alcohol status (9% drinker, 5% non-drinker)
    - Allergy status (known allergy 12%)

Data attributes as part of the GP Visits information consisted of:

- List of Diagnoses has about 30% (6478) of patients with reported diagnoses.
- About 24% of visits use 4500 unique diagnosis descriptions.
- The most frequent diagnoses by visit
    - Hypertension= 197
    - URTI= 174
    - Asthma= 174
    - Depression= 146
    - Bronchitis= 135
    - Tonsillitis= 115
    - UTI= 114
    - Otitis media= 97
    - Gastroenteritis= 95
    - Review= 87
- 86% of visits have some value for Diagnosis-Status-at-Visit with the most common including COPD, Bone Joint Disease, Diabetes, Cancer, CHD, Asthma, Gastroenteritis, Stroke, Influenza, Hypertension, Anxiety, Depression, Hepatitis

Model Development

After extensive data review, the data was split between a current visit view of the patient and a historical view. Its purpose was to deal with the relevance of certain types of information for a current ED attendance. The notion was to separate the more recent patient information to the GP from previous visits which were collapsed into the one "historical visit". The criteria for key attributes were specifically designed and are presented in the table below.

Table 3 Final variable listing in Current and Historical visits

Attributes	Time categories
	Clinical fields
Current Diagnosis	Current visit and ACTIVE diagnoses
Historical Diagnosis	Up to 10 years- Not including current diagnoses information
Current Immunisation	Current visit
Historical Immunisation	Within last 5 years- not including Current visit
Current Script	Within last 8 months
Historical Script	Between 9-24 months
Current Pathology test	Current Visit
Current Pathology result	Current Visit
Historical Pathology test	Within last 12 months- not including last visit
Historical Pathology result	Within last 12 months- not including last visit
Current Measurement	Current Visit
Historical Measurement	Within last 5 years- not including current visit
MBS	Any
	Other patient information
Alcohol Usage	Last recorded
BP recorded	Last recorded
Care goal	Last recorded
Reaction (allergies)	Last visit
Historical Reaction	All information apart from last visit
Tobacco	Last recorded
Age	Last visit
Gender	Last visit
DVA Status	Last visit
Pension Status	Last visit
Aboriginal Status	Last visit

In addition to this, extra mapping was laid over the data including risk groups for key measurements (i.e. BMI, BP, Blood Sugar Levels, cholesterol, falls and temperature) and a medications risk grouping where particular medications were given a higher risk score that others (i.e. medications for Cardiac issues or chemotherapy were scored at a higher risk than acne medication). These extra mappings allowed for another dimension of understanding of the data being worked into the model.

Time Categories

A range of models were explored including a 0-30 day, 30-90 day, 90-180 day, 180-365 day, over 365 day and a no ED attendance risk. The predictive capabilities of the 30-90 day, 90-180 day, 180-365 day were deemed too low and it was chosen to reduce a number of these time categories to two (i.e. 0-30 day and 301-365 day) as shown in **Table 4** and **Table 5**.

Table 4 Initial class accuracy for predictive model of ED attendances based on GP data.

Results	TP	FP	FN	Precision	Recall	F-Score
0-30 day	7999	3242	3456	71.16	69.83	70.49
30-90 day	111	377	2667	22.75	4.00	6.80
90-180 day	8	93	1465	7.92	0.54	1.02
180-365 day	5	53	1356	8.62	0.37	0.70
over 365 day	513	544	2890	48.53	15.07	23.00
Non-ED	28580	8837	1312	76.38	95.61	84.92
OVERALL	37216	13146	13146	73.90	73.90	73.90

Table 5 Final class accuracy for predictive model of ED attendances based on GP data.

Results	TP	FP	FN	Precision	Recall	F-Score
0-30 day	7732	2765	3723	73.66	67.50	70.44
31-365 day (90+180+365)	582	1000	5030	36.79	10.37	16.18
NON-ED	28708	6172	1184	82.31	96.04	88.64
OVERALL	37022	9937	9937	78.84	78.84	78.84

As shown below in the Confusion Matrix table, a significant number of 30-day and post-30-day records migrate into the non-ED attendance class but far fewer migrate in the other direction. This suggests that the model is not allowing a large number of non-ED attendees to be placed in a 0-30 day or 31-365 day grouping, thereby not alerting practitioners and patients unnecessarily.

Table 6 Final model confusion matrix

	0-30 day	31-365 day	NON-ED
0-30 day	7732	724	2999
31-365 day	1857	582	3173
NON-ED	908	276	28708

The Final Model

The final model used the following variables when available:

Essential

1. Diagnosis Details (current visit and active only)
    a. diagnosis-name
    b. diagnosis-SCT category
2. Historical Diagnosis Details (all time)
    a. diagnosis-name
    b. diagnosis-SCT category
3. Patient Information
    a. visit-type
    b. gender
    c. DVA
    d. ATSI
    e. pension-status
    f. age-at-visit

Optional

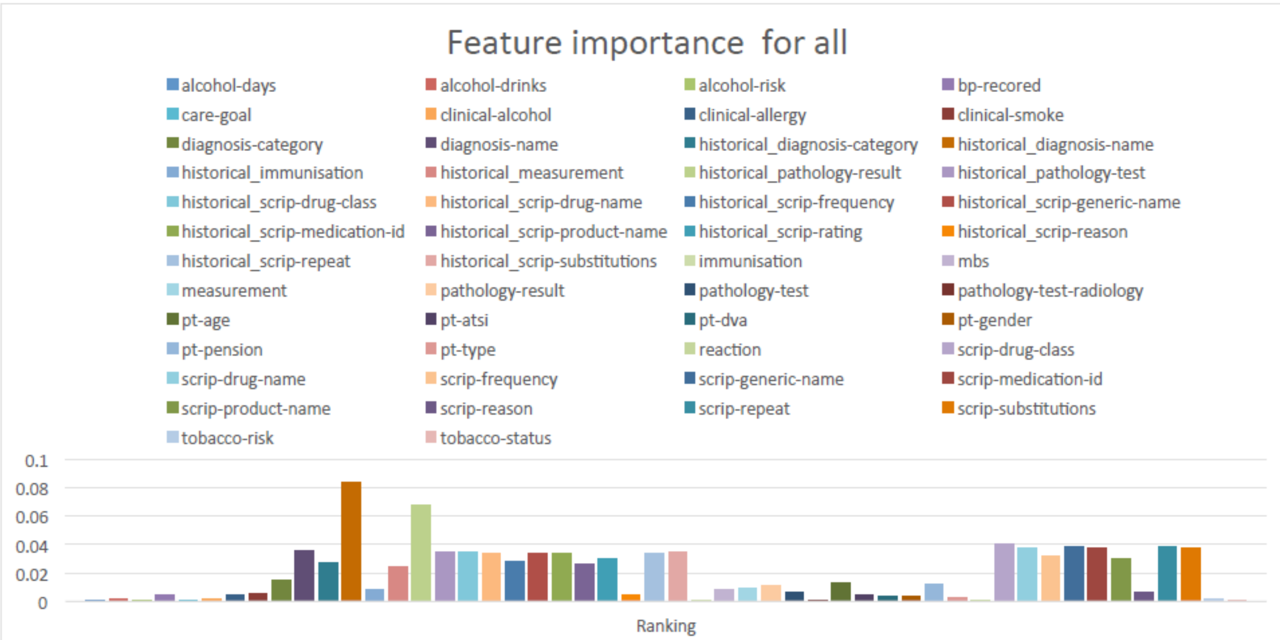
4. Alcohol Usage
    a. alcohol-days per week
    b. alcohol-drinks per day
    c. alcohol-risk factor
5. BP recorded
6. Care goal
7. Clinical fields
    a. clinical-smoke info
    b. clinical-alcohol info
    c. clinical-allergy info
8. Immunisation (current visit)
9. MBS
10. Measurement
11. Reaction types and values (current visit)
12. Pathology test (current visit)
    a. test-name
    b. radiology-test
13. Pathology result types and values (current visit)
14. Script Details (within 8 months)
    a. script-generic name
    b. script-drug name
    c. script-product name
    d. script-frequency
    e. script-repeat

   f. script-substitutions
   g. script-reason
   h. script-medication id
   i. script -rating
   j. script-drug-class
15. Tobacco Usage
   a. Tobacco-risk factor
   b. tobacco-quit status
16. Historical Immunisation (up to 5 years)
17. Historical Measurement
18. Historical Reaction types and values (up to 5 years)
19. Historical Pathology test (within 12 months)
   a. test-name
   b. radiology-test
20. Historical Pathology result types and values
21. Historical Script Details (between 9 months and 24 months)
   a. script-generic name
   b. script-drug name
   c. script-product name
   d. script-frequency
   e. script-repeat
   f. script-substitutions
   g. script-reason
   h. script-medication id
   i. script-rating
   j. script-drug-class

GP Data Item Importance

The importance of items were reviewed across the final model. As shown below, some of the key variables within the model included diagnoses and pathology results, with less but still apparent input from medications both current and historical. Individual patient demographics such as age, gender, pension status were less important than the more clinical patient information types.

Figure 2 Feature importance for final model



When passed though the untested data held back for validation purposes (13%), the results were very encouraging. Overall, the algorithm best predicted the 0-30 day time category with a 73.7% precision score. The 31-365 day time category was able to predict 36.8% and the no ED attendance category showing a precision score of 82.3%.

The next step was to deliver the model to the test practices. The purpose was to assess the acceptance and applicability of the tool in a general practice setting. Each participating general practitioner was required to evaluate 60 patient attribute reports and associate the risk scores. The tool also recorded the suggested actions taken in response to the reports. Following the interim assessment of the 60 patient records, GPs were asked to complete a semi-structured survey with some free response items. After the online evaluation, they were also asked to participate in a one-to-one telephone interview to detail the usefulness of the report in general.

GPs Evaluation of Risk Algorithm: Quantitative

Quantitative analysis of the online surveys was conducted with a sample of 420 patient records reviewed by the seven participating General Practitioners. All quantitative analysis was conducted on SPSS 20.0.

Algorithm Patient Capture

The risk score algorithm was used to develop risk scores for the participating GPs patients who had visited in the previous 12 months. The spread of risk scores (n = 6,671) were reviewed across all patients and showed that only a small proportion of patients fell into the High or Extreme Risk Groups indicating reasonable delineation across patients as seen in **Table 7**.

Table 7 Patient spread across risk groups for time categories

Risk Group	Score Range	0-30 Day		31-365 Day	
		Patient n	*Patient proportion*	*Patient n*	*Patient proportion*
Very Low	(<20%)	4,094	61.4%	6,322	94.8%
Low	(20<40%)	1,266	19.0%	249	3.7%
Medium	(40<60%)	695	10.4%	66	1.0%
High	(60<80%)	391	5.9%	24	0.4%
Extreme	(80+%)	225	3.4%	10	0.1%

Overall Results

The overall results suggested that for the 0-30 scores, 71.1% of records were within 20% of what the GPs thought was an accurate score for patients. The 31-365 day scores also showed a good level of agreement with 76.1% of records being within 20% of what the GPs thought was accurate.

Analyses suggested that participants found the accuracy of the scores better in the 31-365 day category ($M = 2.67$, $SD = 1.60$) than the 0-30 day category ($M = 2.94$, $SD = 1.61$) with lower scores indicating less discrepancy with their own clinical judgement. This difference was found to be significant with a paired samples t-test, $t(359) = 2.44$, $p = .015$.

Review of the accuracy of the highlighted attributes presented to GPs in the report showed that for the 0-30 category 61.7% of patient records were either strong agree/agree, 11.7% neither agree or disagree and 26.6% disagree/strongly disagree. This was similar to the 31-365 category 67.6% of patient records were either strong agree/agree, 10.7% neither agree nor disagree and 21.7% disagree/strongly disagree. A paired samples t-test found the GPs perceived the accuracy of the attributes used in the algorithm significantly better in the 31-365 day category ($M = 3.55$, $SD = 1.13$) than the 0-30 day grouping ($M = 3.40$, $SD = 1.22$) with higher scores indicating more agreement, $t(419) = -2.738$, $p = .006$.

The reasons given by practitioners for their answers varied considerably. Some identified though the risk score was accurate, too high or too low. Others focused more on the ‘stable’ nature of the patients’ status that may have conditions that could put them at a higher risk. Examples included “asthma well controlled”, “Stable chronic conditions”, and “Stable IDDM”. Other responses identified that the patient was ‘healthy’ with little reason for a high-risk score. Often this notion of health was linked to younger patients. Examples included “why the high short term risk in healthy 27 yo?”, “healthy six year old” and “healthy child. At risk of injury due to being an active boy”. Other comments related to the type of condition or characteristics a patient had that they felt was causing the risk score or if they felt the risk score was too low. Examples included “Frail elderly woman”, “She is an older lady with cancer and regular infections. Twelve-month score will be much higher", and 'recurrent injuries associated with drug and alcohol issues".

Risk Group Analysis

Analysis of GP perceptions were conducted over the risk groups through cross-tabulations and ANOVAs. As shown below, GPs perceived the risk scores to be more accurate in the lower risk groups for the 0-30 day category, but there was less of a decrease in agreement evident in the 31-365 day category (See Table 8). It should be noted that total patient counts are 360, as one GP identified after the evaluation that they did not answer this question properly. Therefore, this GPs results were removed from the analysis of this question.

Table 8 Risk group patient records and proportion of GPs perception of scores within accuracy

	Risk Group	**Within GP Perception of Accuracy**				
		<5%	*5-10%*	*10-15%*	*15-20%*	*20%+*
0-30 Day	Very Low	49.1%	11.4%	18.3%	6.9%	14.3%
	Low	17.3%	16.0%	21.3%	14.7%	30.7%
	Medium	12.5%	2.5%	20.0%	7.5%	57.5%
	High	14.9%	10.6%	19.1%	12.8%	42.6%
	Extreme	4.3%	8.7%	26.1%	4.3%	56.5%
	Total	31.1%	11.1%	19.7%	9.2%	28.9%
31-365 Day	Very Low	46.9%	12.1%	11.3%	8.8%	20.9%
	Low	21.0%	21.0%	27.4%	3.2%	27.4%
	Medium	19.0%	23.8%	19.0%	9.5%	28.6%
	High	14.8%	7.4%	33.3%	11.1%	33.3%
	Extreme	18.2%	0.0%	36.4%	9.1%	36.4%
	Total	37.5%	13.6%	16.9%	8.1%	23.9%

To explore these differences in the 0-30 day category, an ANOVA was conducted and found that participants rating of the accuracy of the very low scores ($M = 2.26$, $SD = 1.477$) were rated significantly better than the other risk groups (Low- $M = 3.25$, $SD = 1.48$; Medium- $M = 3.95$, $SD = 1.431$; High- $M = 3.57$, $SD = 1.5$; Extreme- $M = 4.00$, $SD = 1.279$) where Ms closer to zero were more accurate through a significant Tukey HSD post hoc comparison, $F(4, 355) = 20.35$, $p = .000$.

ANOVA results also showed a significant difference for the 31-365 day category with participants perceiving the lower risk categories more accurate, $F(4, 355) = 10.29$, $p = .003$. However the only significant difference was between Very Low and High categories (Very Low- $\underline{M} = 2.45$, $SD = 1.62$; Low- $\underline{M} = 2.95$, $SD = 1.49$; Medium- $\underline{M} = 3.05$, $SD = 1.53$; High- $\underline{M} = 3.41$, $SD = 1.42$; Extreme- $\underline{M} = 3.45$, $SD = 1.51$) through a Tukey HSD post hoc comparison.

The perceived accuracy of the attributes showed an interesting variation in agreement with the very low and high-risk groups being perceived better. The 31-365 day category showed less variation with scores ranging from 61.1%- 70.4% (See **Table 9**).

Table 9 Risk group patient records and proportion of GPs perception of accuracy of attributes presented

Risk Group	**0-30 Day**				**31-365 Day**			
	Patient n	*SD/D*	*NA/D*	*SA/A*	*Patient n*	*SD/D*	*NA/D*	*SA/A*
Very Low	199	18.6%	11.6%	69.9%	283	22.3%	8.5%	69.3%
Low	92	33.7%	10.9%	55.5%	72	22.2%	16.7%	61.1%
Medium	49	34.7%	16.3%	49.0%	25	28.0%	8.0%	64.0%
High	57	31.5%	8.8%	59.6%	27	18.5%	11.1%	70.4%
Extreme	23	39.1%	13.0%	47.8%	13	0.0%	30.8%	69.2%
Totals	420	26.6%	11.7%	61.7%	420	21.7%	10.7%	67.6%

SD/D = Strongly Disagree/Disagree, NA/D = Neither Agree or Disagree, SA/A = Strongly Agree/ Agree

This pattern was also present through the ANOVA calculations that only found significant differences in the 0-30 day category, $F (4, 415) = 3.62$, $p = .006$. The presented patient attributes was perceived as significantly different for the Very Low and Medium risk categories (Very low- $M = 3.62$, $SD =1.09$; Low- $M = 3.3$, $SD = 1.29$; Medium- $M = 3.08$, $SD = 1.29$; High- $M = 3.21$, $SD = 1.32$; Extreme- $M = 3.0$, $SD = 1.31$) through a significant Tukey HSD post hoc comparison. There were no significant differences for the 31-365 day category.

Age Groups Analysis

Analysis of the patient age group was conducted over the two time categories. Both categories showed less agreement with those in a 65 and over compared to other age groups. However, ANOVA analyses did not show any significant differences across the age groups.

Table 10 Age group patient records and proportion of GPs perception of scores within accuracy

	Age Group	**Patient n**	**Within GP Perception of Accuracy**				
			<5%	*5-10%*	*10-15%*	*15-20%*	*20%+*
0-30 Day	<15	58	22.4%	12.1%	20.7%	15.5%	29.3%
	15-44	105	36.2%	10.5%	23.8%	6.7%	22.9%
	45-64	100	30.0%	13.0%	20.0%	9.0%	28.0%
	65 and over	97	32.0%	9.3%	14.4%	8.2%	36.1%
	Total	360	31.1%	11.1%	19.7%	9.2%	28.9%
31-365 Day	<15	58	46.6%	10.3%	19.0%	5.2%	19.0%
	15-44	105	39.0%	13.3%	20.0%	4.8%	22.9%
	45-64	100	35.0%	13.0%	15.0%	12.0%	25.0%
	65 and over	97	33.0%	16.5%	14.4%	9.3%	26.8%
	Total	360	37.5%	13.6%	16.9%	8.1%	23.9%

The accuracy of the attributes presented did show less agreement for patient records in the <15 years and 65 and over for the 0-30 day category. This trend was also visible for the 65 and over age group in the 31-365 day category as shown in **Table 11**.

Table 11 Age group patient records and proportion of GPs perception of accuracy of attributes presented

Age Group	Patient n	0-30 Day			31-365 Day		
		SD/D	*NA/D*	*SA/A*	*SD/D*	*NA/D*	*SA/A*
<15	83	27.7%	16.9%	55.4%	14.4%	15.7%	69.9%
15-44	131	19.1%	13.7%	67.2%	19.8%	9.9%	70.3%
45-64	106	24.5%	11.3%	64.2%	18.8%	8.5%	72.6%
65 and over	100	38.0%	5.0%	57.0%	33.0%	10.0%	57.0%
Totals	420	26.7%	11.7%	61.7%	21.7%	10.7%	67.6%

SD/D = Strongly Disagree/Disagree, NA/D = Neither Agree or Disagree, SA/A = Strongly Agree/ Agree

ANOVA was conducted and shows significant differences in how GPs perceived the accuracy of the presented attributes across age groups. In the 0-30 day time period, participants rating of the accuracy of the presented patient attributes was higher in the 15-44 age group ($\underline{M}$ = 3.61, SD = 1.12) compared to the 65+ age group ($\underline{M}$ = 3.14, SD = 1.28) through a significant Tukey HSD post hoc comparison, $F(3, 416) = 2.95$, $p = .033$.

Exploration of the 31-365 day time period also showed a significant difference between all the younger age groups (<15, $\underline{M}$ = 3.8, SD = 1.09; 15-44, $\underline{M}$ = 3.61, SD = 1.12; 45-64, $\underline{M}$ = 3.64, SD = 1.00) compared to the 65+ group ($\underline{M}$ = 3.55, SD = 1.13) where GPs perceived less score accuracy through a significant Tukey HSD post hoc comparison, $F(3, 416) = 6.61$, $p = .000$.

Missing Attributes Analysis

Missing attributes in patient records were explored through independent t-tests to compare the perception of accuracy for GPs. Overall this showed that GPs were more likely to perceive scores as accurate when they contained 'current' information for diagnoses ($t(117.4) = 3.60$, $p = .00$), pathology results ($t(358) = 2.15$, $p = .03$) and other measurements such as BMI, BP, cholesterol, weight, temperature, etc ($t(358) = 2.46$, $p = .01$). However, there did not seem to be an optimal number of attributes present that increased GPs perception of accuracy.

Treatment change

Participants were asked if they would make any changes in patients' treatment, testing, referrals or other activities based on the risk scores. These outcomes were analysed in terms of the five risk groups. As shown below in **Table 12**, a higher proportion of practitioners identifying that they would change the treatment of patients in the 31-365 day group, but this did not necessarily translate into many more patients. There were also less GPs identifying that they would change treatment in the 0-30 day 'extreme' group, which may be due to GPs perception that this risk group was less accurate. In the 31-365 day category, there is an increase in GPs wanting to change a patient's treatment, as the risk group increases.

Table 12 Participants making changes to patient treatment and recalls based on risk scores

Risk Group	0-30 Day			31-365 Day		
	N	*Change treatment*	*Recall patient (n)* #	*N*	*Change treatment*	*Recall patient (n)*#
Very low	199	14.6%	47.4% (9)	283	11.0%	72.2% (13)
Low	92	14.1%	87.5% (7)	72	6.9%	50.0% (1)
Medium	49	10.2%	100% (3)	25	24.0%	80.0% (4)
High	57	12.3%	100% (4)	27	29.6%	50.0% (3)
Extreme	23	0.0%	0% (0)	13	30.8%	66.7% (2)
Total	420	12.9%	67.6% (23)	420	12.9%	67.6% (34)

Recall patients is the percentage of the change treatment patients that GPs would recall back in.

The types of changes that a GP may want to make were reviewed as outlined below. The identified changes varied considerably; however, testing activities such as pathology or radiology were the most commonly cited activities (See **Table 13**).

Table 13 Participants patient change types based on risk scores

Risk Group	0-30 Day			31-365 Day		
	Pharmacotherapy (n)	*Testing (n)*	*Referrals (n)*	*Pharmacotherapy (n)*	*Testing (n)*	*Referrals (n)*
Very low	2.0% (4)	9.0% (18)	3.0% (6)	2.1% (6)	5.3% (15)	3.2% (9)
Low	1.1% (1)	6.5% (6)	5.4% (5)	0.0% (0)	4.2% (3)	0.0% (0)
Medium	2.0% (1)	2.0% (1)	2.0% (1)	0.0% (0)	12.0% (3)	12.0% (3)
High	3.5% (2)	1.8% (1)	5.3% (3)	7.4% (2)	7.4% (2)	11.1% (3)
Extreme	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	23.1% (3)	0.0% (0)
Total	1.9% (8)	6.2% (26)	3.6% (15)	1.9% (8)	6.2% (26)	3.6% (15)

GPs also identified 'other' activities they would conduct with patients they wished to make a treatment change with. This included measures such as BP, BMI and Spirometry. Other activities include reviewing for CVD risk and management of these conditions. Other items concerned checking on patient smoking status and immunisations. Other activities included more work or education with the patient on issues such as encouraging weight reduction, compliance with mental health treatment, setting up regular nurse contact and when to seek help on infections were also raised.

Final GP Survey

Respondents were asked to complete a final survey to gather their overall perceptions of the tool and suggestions for revisions. A 5 point Likert scale was used for a range of questions (i.e. 'strongly disagree' = 1 through to 'strongly agree' = 5).

Table 14 Final GP survey responses

Question	M (SD)
1. The format of the report is easy to use.	3.00 (0.58)
2. In general, the report is accurate in targeting patients at risk of ED admission.	2.29 (0.49)
3. The report is a useful addition to clinical support strategies.	2.57 (0.79)
4. If the report was offered real-time to my desktop, it would be useful.	2.71 (1.11)
5. I would likely use a real-time report as part of my practice.	2.86 (1.22)
6. I would be likely to use the report for high-risk patients to provide or access additional support (i.e. care co-ordinators, practice nurse etc).	3.14 (1.22)

As shown above, GPs generally found the format of the report easy to use and would likely use a report to provide additional support to higher risk patients. However, there was less agreement concerning the accuracy and if they would use it if it was delivered as a real-time report within their practice. GPs identified within their comments that some of these concerns related to those in the higher risk categories, which was also suggested within the individual patient records evaluations.

GPs Evaluation of Risk Algorithm: Qualitative

All interviews were audio recorded and transcribed in full by AF. The seven recordings were listened by NH while reading the transcribed text, in order to correct any potential mistakes and ensure accuracy. The interviews were coded using the NVivo 11.0 software.

A qualitative phenomenological approach was applied in conducting the semi-structured interviews and a thematic analysis to analyse the transcripts.

The experience of using the tool was perceived as positive from most of the participants. The accuracy of the tool was an overarching theme in GPs stories. Reflecting on their participation in the study and while evaluating the patient records, general practitioners said that a predictive tool constitutes an innovative idea that is useful and helpful to guide their thinking and decisions towards a patient's care. However, although the idea of using a predictive tool as part of their practice is beneficial, there were some suggestions that the tool needs further refinement in order to ensure better accuracy.

During the thematic analysis of the interview transcripts, three themes emerged:

First theme: Accuracy

Responses appeared to be mixed when GPs evaluated the tool's accuracy and whether the scores were realistic or not. Opinions seemed to be divided, expanding on the reasons that led them to form this decision.

"Yes, generally, I would say they did [align with my clinical experience]. I think generally it's a fairly good predictor"

"…The reports were reasonably realistic, but they are only as good as the data that the GP has put them into the program"

"I thought that the accuracy of the report was quite long way of, of what my assessment of the preference is…I think it overestimated the risk in the 30 months {meaning days] significantly and I think it underestimated the risk in the 12 months period"

Certain groups seemed to be more accurate than others, especially elderly people and the low risk patients.

"…the predictions …particularly I think for the lowest risk patients were really accurate…"

"…Probably the older ones were more accurate when they had multiple risk factors…"

Some participants stressed the need of having knowledge around the type of algorithm that was used, which calculated the risk score. Participants mentioned that this might have helped them to have a complete image and provide answers that are more accurate.

"…if I knew the algorithm I could then have suggested that the long term risk score for this patient would be higher than the short term risk…"

"Well, since I don't know how you do them [calculations], it is very difficult to kind of make a meaningful comment about…"

Keeping this in mind, it might be that GPs lack of knowledge concerning how the algorithm was calculating risk scores was frustrating for them and their evaluation of the tool.

Second theme: Usability

All of the GPs were in an agreement that a predictive tool consists an innovative idea that is very useful and helpful for GPs' daily practice. Having a predictive tool offers the opportunity to predict and potentially prevent a number of people presenting to emergency departments. Although some participants reported the tool needs further revising and improvement, most of them said that it constitutes a useful alerting tool. Reviewing a patient's record in a single scan and the time efficiency this represented was also highlighted as valuable.

"It might provide an early warning to people who I might not be considering as potentially risk to go into hospital."

"…that was easy to absorb and to read, it was quite well presented. It was a good summary."

"It got quicker as you went along, because I knew most of the patients, that was pretty quick, the ones that I didn't know or hadn't see for a long time, it took a bit longer."

Third theme: Suggestions

The third theme identified participants' suggestions for inclusion of further items, which might have provided better estimation around a patient's health. The views of the participants were mostly personal preferences, all underpinned by the same needs; the reason that a patient is flagged as high risk in the system. Suggestions offered by the GPs when applied to the tool might provide better accuracy and understanding, and as a result, create a more constructive tool.

“Well it would be helpful if the program flagged why it thought, that this risk was so high in that particular person…why was that person flagged as being as high risk…”

ED Focus Group

The focus group interview took place on 20-05-2015 in Box Hill Hospital ED premises. Six clinicians (five doctors and one registered nurse) consented to participate in the focus group interview, which was video recorded. The recorded focus group was subsequently transcribed and analysed using narrative techniques.

Data analysis revealed that responses fell under three themes: ‘Access’, GP skill set’ and ‘Frequent flyers”.

First theme: Access

Many comments were made by participants around access. As a publicly funded ED, with no charge, it has advantages over private EDs

“Privates will come here; I think it’s because of the fee?”

“Yeah, cost of 450 bucks, CTs…”

“Six weeks wait in the community – and we can get it done in 24 hours in short stay”

Access issue are particularly relevant when considering GP access as well. Several participants responded that patients presented either because they could not get an appointment to see a GP or that their regular GP was closed. This also applied to nursing home and residential aged care facilities (RACF) – where staff are not sufficiently trained to deal with emergencies, and getting GPs to visit was difficult

“Usual GP to dash out to the nursing home – it just doesn't happen”

Access also applied to the hospital – with note being made that often the ED was being used to arrange semi urgent appointments to outpatients or to inpatient units.

“Why not just make the call to gastro and organise a semi –urgent transfusion the next day”

Second theme: GP Skill Set

Many comments were made about the ED being used where GPs no longer had sufficient skills

“Small lacerations – do GPs still deal with them? I don’t know. We often glue little things”

“I think sometimes GPs just want us for a second opinion – they have 15 minutes to make a decision whereas we have time.”

The participant commented that there were ‘good’ GPs who they probably never saw the patients of, but many GPs were operating below a standard skill set.

Third theme: Frequent Flyers

Participants drew out two specific types of frequent flyers:

Those with significant mental health issues – usually with extensive in-reach and outreach support in the community, but nevertheless needing constant attendances for extra support

Those with severe chronic and complex diseases – such as COPD or cardiac problems. Again, these people often have community supports and care plans etc., but the variability of their diseases means that they often need to attend ED for hospital support. This particular group was seen as not being amenable to preventing admissions – in that each attendance often required admission

The one exception to that rule was where the care plan or advanced directive specified not being sent to hospital, but community constraints meant that this happened anyway.

Summary

The overarching theme to emerge from this focus group interview is that to reduce avoidable admissions, more resources are required in the community – both in terms of GP training and standards, and access to investigative services.

DISCUSSION

Algorithm development

Through the use of a linked Hospital ED data set and 50 General Practices' data across the Eastern Melbourne Suburbs, an algorithm was developed that aimed to predict ED presentations based on General Practice patient data. The algorithm utilised machine learning in an effort to harmonise the complex data set and develop a predictive risk score for patients. This is one of the first instances of machine learning applied to predicting ED presentations in Australia using routinely collected data from general practices.

The algorithm was trained on 157,330 General Practice data points across 52 attributes within 21 relational groups and was tested across essential and optional variables.

Thirteen percent of the GP data was held back to 'blind' test the precision and sensitivity of the model and showed a high level of agreement/correlation.

Comparison with other models

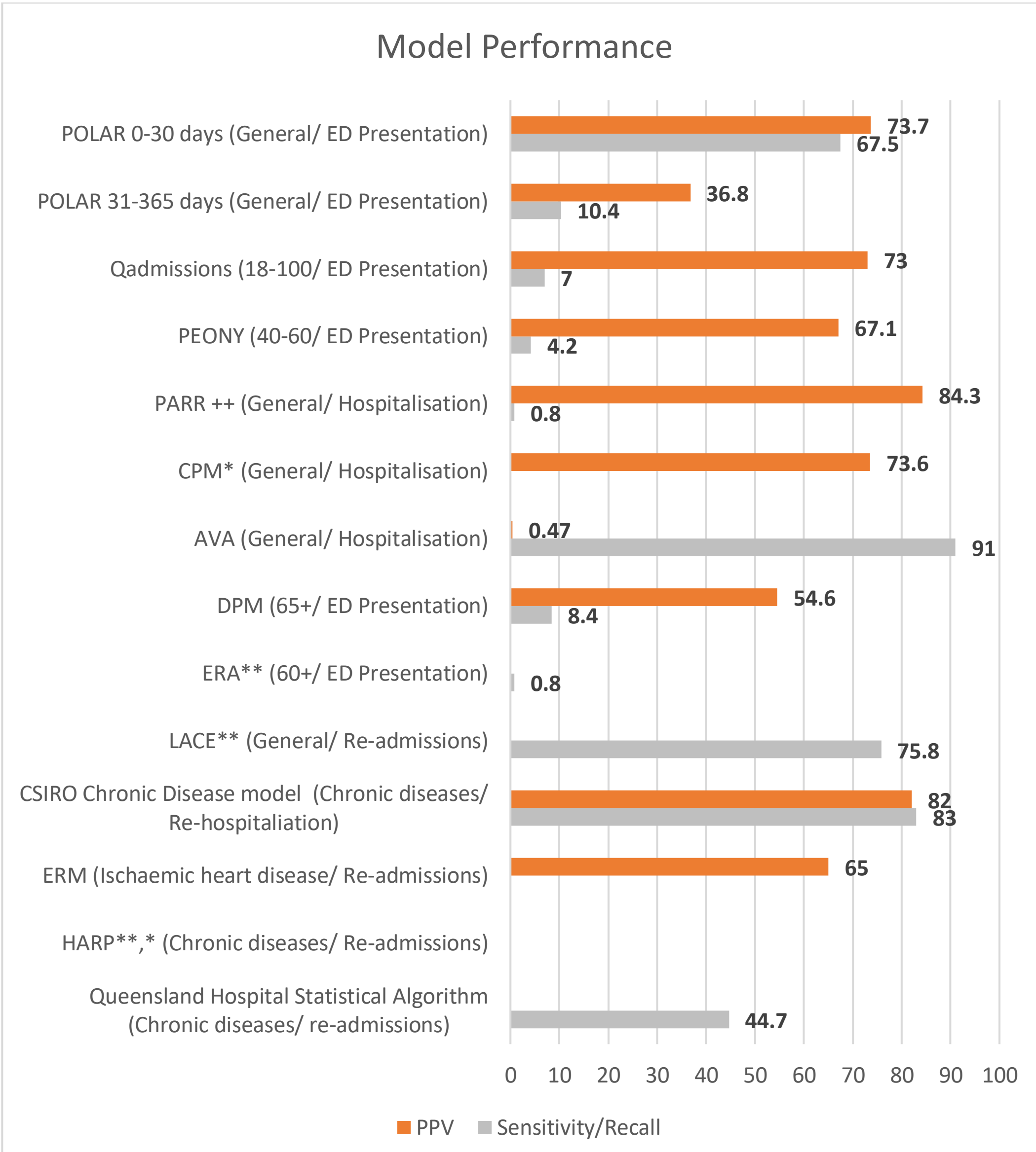
Numerous algorithms have been developed throughout the years in both Australia and overseas regarding hospital attendances, predominantly through logarithmic regression. However, the vast majority of the models target specific groups and populations with specific characteristics (i.e. particular chronic conditions and/or recent hospital attendances), narrowing the proportions of patients being considered at-risk of presentation to an ED. There are models targeting only ambulatory care sensitive conditions (ACSC), elderly people and hospital re-admissions (38-41). A range of these models are shown in **Table 15** below.

Table 15 Characteristics of the included models

Model	Population	Target type	Key Data Sources	Included variables	Prediction length
POLAR	General	ED presentation	GP Data	21 groups, 52 variables	0-30 days, 12 months
QAdmissions	18-100	ED presentation	GP and Hospital Data	30 variables	2 years
PARR	General	Hospitalisation	GP, Hospital, nursing and social services data	Not specific number	12 months
CPM	General	Hospitalisation	GP and Hospital	Not specific number	12 months
PEONY	40+	ED presentation	GP, Hospital, socio-demographic and pharmacy	35 variables	12 months
AVA	General	Hospitalisation	Hospital and Veterans National Patient Care Database	4 categories (not specific number of included)	90 days, 12 months
DPM	65+	ED presentation	GP and Hospital Data	89 variables	12 months
ERA	60+	ED presentation	GP, Hospital and socio-demographic	12 variables	30 days
LACE	General	Re-admission	Hospital	4 variables (later added more)	90 days
Queensland Hospital Statistical Algorithm	Chronic diseases	Re-admissions	Hospital	16 variables	12 months
HARP	Diabetes and chronic diseases (targets specific populations)	Re-admissions	GP, hospital and survey data	6 categories, 42 variables (calcularo version)	12 months
EMR	Ischaemic heart disease	Re-admissions	Hospital	22 variables	30 days
CSIRO Chronic Disease model	Chronic diseases	Re-hospitalisation	Hospital	19 variables	30 days

As discussed earlier, the various predictive models include a range of variables and forecast risk for a specific period of time, depending on target population and outcome. **Figure 3** demonstrates the positive predictive value (PPV) for different predictive risk models (PRMs) using similar population size and thresholds. The predictive performance of POLAR is clearly superior to the others with strong scores in both PPV[2] and Sensitivity (recall)[3], particularly in the 0-30 day category. This shows that the POLAR Diversion algorithm performs well in finding cases eligible for potential primary care intervention.

Figure 3 Comparison between predictive risk models



[2] Positive Predictive Value: expresses the likelihood that a person with a positive test has a disease. The positive predictive value depends not only on the accuracy of the test, but also on the prevalence of the disease.

[3] Refers to the accuracy of the test. Sensitivity/Recall measures the proportion of the false negatives, when a large number of positive and negative cases is tested.

*Does not provide sensitivity values in the final report
**Do not provide PPVs

QAdmissions: QAdmission's PPV value is almost the same as POLAR's, however, the sensitivity is extremely low, which means that the ability to recognise patients at risk for admission is hampered in comparison to the POLAR Diversion Risk Score. The QAdmissions used two validation cohorts and did not include children (<18). In addition, the clinical outcomes and cost-effectiveness of using the algorithm in primary care have not been well evaluated (28).

PEONY: POLAR performed better than PEONY, with a lower PPV and very low sensitivity. There is not much evidence of how the model reduces the ED admissions(26).

PARR: Additionally, POLAR demonstrated greater performance than PARR, as PARR proved unable to predict future admissions with no previous patient's admission. One of the reasons probably is that the tool is no longer maintained so the data is outdated (25).

CPM: CPM had similar results concerning PPV scores to POLAR, but in terms of sensitivity comparison cannot be applied as CPM's publications do not include this information. However, the project was de-funded and concern about a lack of accuracy raising questions about how the tool can identify new at-risk patients (27, 34).

AVA: Although the sensitivity of the tool is very high (91%), the PPV value is extremely low, which means that the predictability is controversial (38).

DPM: The PPV is in moderate levels (54.6%) where sensitivity is only 8.4%. Although DPM appeared to outperform some models (such as CPM) it did not demonstrate better performance than POLAR (40).

ERA: This model is targeting specifically elderly people at risk of re-admission to hospital. Findings from this tool do not demonstrate generalisability, as it narrows on targeting a specific population of elderly people, drawn from assisted-living community facilities in the US, for hospital and ED presentation. Compared to POLAR, this PRM demonstrated significantly lower sensitivity (39).

LACE: When the model was applied and tested in the US, it appeared to be a poor predictor for hospital re-admissions amongst elderly people. Although the model undergone some modification (altering the cut off values of prediction) its performance on predicting re-admissions did not improve significantly. However, its sensitivity increased, from 49% to 76% (41).

Queensland Hospital Statistical Algorithm: Although this is one of the few models that has included children, its performance is poor when compared to POLAR, as the sensitivity of this PRM was only 44.7% (30).

HARP: As HARP is not a predictive model per se, no PPV or sensitivity is provided by any reports, so practically its performance cannot be compared against POLAR's. However, this model is restricted only on targeting specific populations affected by specific chronic diseases, which means it cannot be applied to the general population (29, 52) nor automated within general practice as it uses previous hospital admission as a key indicator.

EMR: This model demonstrated extremely low sensitivity, 0.65% and focused only on targeting unplanned hospital re-admissions after patients experienced myocardial infarctions (32).

CSIRO: Although CSIRO has demonstrated strong performance predicting hospital re-admissions of patients with chronic disease values (31). However, it has been suggested that it's sensitivity results have are influenced by including dialysis patients for dialysis and requires further amendments before implementation (33).

As shown above, POLAR has performed better in estimating patients' presentations in an ED. This may be due to the use of machine learning, in place of the commonly used regression analysis in other PRMs. Machine learning allows for the inclusion of a large breadth of data and types of data into modelling and is becoming more common with health and biomedical research globally (53). Where many of the previous PRMs included key 'risk factors' for consideration such as chronic conditions, key specific medications or pathology outcomes and recent hospital attendances, the POLAR Diversion algorithm included a large range of General Practice data, thereby increasing its possible reach.

By specifying particular conditions, medications or pathology outcomes, there is the risk of not reviewing other clusters of influencing variables that could lead to a number of false negatives. In other words, unless a patient fits the profile, they will not be given a risk score even though they present to an ED. Machine learning in the POLAR Diversion project allows the data to identify which clusters of variables best fit together to develop multiple profiles across ages, disease states, medication usage, patient demographics, immunisations, MBS activities, pathology and so on. Therefore, the opportunity to enhance our understanding of what poses a risk grows exponentially.

In addition to this strength of machine learning, there was significant mapping of the General Practice data that was undertaken before its inclusion in the 'black box'. Clinical data is not always clean and straightforward (54, 55) given it is not developed for this type of research, but day-to-day patient care. Mapping of GP free-text and coded diagnoses to SNOMED (51) enhanced the usability of the diagnoses fields. Mapping of the medications to the World Health Organisations ATC Drug Classification system (56) also allowed for grouping up risky medications such as immunosuppressants, antipsychotics, insulins and analogues, etc. Measures such as Body Mass Index, Cholesterol, Blood Pressure recordings and Blood Sugar levels (amongst others) were mapped and identified as risky depending upon their outcome. This extensive mapping process ensured that the clinical understanding of the data was included into the algorithm alongside the actual data itself. As such, the possible profiles of risky patients was built on the data and a guideline about how some of the data fields should be interpreted from clinical best practice understanding.

Another strength of the project was the removal of the 'injury' based ED attendances from the modelling. In most cases, injuries cannot be predicted unless it is due to a medical condition (i.e. falls in the aged) (57, 58). However, there are a large proportion of unexpected injuries (i.e. sports, transport collisions, burns, etc.) that occur to otherwise healthy individuals. The removal of these events from the model ensured it was trained on more predictable ED presentations with real and trackable health concerns. Finally, the ability of the POLAR tool to be delivered in general practice in real time represents a significant advance.

The POLAR Diversion Algorithm in General Practice

The algorithm was run across all patients who had visited their GP in the last 12 months for the seven participating GPs. The algorithm showed a reasonable delineation across the patients with most falling into the lower risk categories. The algorithm suggested that the 0-30 day time category showed 80.4% of patients were in the very low to low risk groups and 98.5% for the 31-365 day time category. The AIHW My Healthy Communities Report 2015/16 showed that approximately 13.5% of Australians attended an Emergency Department for their own health in the last 12 months (59). This was reduced to 11.1% in the Eastern Melbourne Primary Health Network, which is the local region where the initial data was extracted. Although this comparison (i.e. recent GP attenders attending an ED compared to the total population) is not exactly alike, it still suggests that the algorithm is not capturing unrealistic numbers of patients across a practice.

Machine Learning vs GPs Perceptions

Results from the validation of the algorithm demonstrated that the POLAR algorithm indicates high performance in accurately predicting patients' at risk of hospital presentation. However, during the 'live' phase of the project, GPs provided mixed opinions, which were often in direct conflict with the findings of the algorithm.

The algorithm's ability to predict patient presentation was higher within the 0-30 day time category than the 31-365 day. Although most GPs evaluated patient records within +/-20% of their own clinical judgement (0-30 day = 71.1% and 31-365 day = 76.1% of patient records), there was a significant difference between GPs perception between the two time category scores. Overall, GP perceptions suggested a higher approval of the 31-365 day time category scores. In addition, GPs also considered the patient attributes presented as more accurate within the 31-365 day time category.

GPs also seemed less comfortable the higher patient risk scores became. This was true of both time categories, but significantly so within the 0-30 day category. Patient attributes presented were considered significantly more accurate in the very low and high categories but less so in the middle risk groupings. This is despite the fact that the tool demonstrably performed.

The qualitative findings showed key themes across the GPs perceptions: accuracy, usability and suggestions. Opinions were divided with some across different GPs with some feeling it was accurate across all groups, others feeling the lower risk scores or risk scores for older patients were more accurate. For other GPs, there was clear frustration about not understanding about how the score was calculated. The 'stability' of the patient was also raised on a number of occasions. Even though GPs could see in some cases that there were a number of risk factors at play, they believed their patient was stable based on their activities with them and therefore, not at risk.

The variance in findings is not uncommon for GPs, who are often wary regarding tools designed to alert possible patients at risk within their practices (60, 61). Although practitioners often identify that a tool could be helpful to them (62), there is often concerns about perceived clinical accuracy. For example, the QCancer research, identified earlier, showed that GPs interpretation of patient risk (for cancer) has been shown to vary considerably across standardised patients (45) suggesting individual differences across practitioners. It also found that that younger practitioners were more likely to agree that the tool could aid them in identifying patients at risk compared to more experienced practitioners who were comfortable relying on their clinical intuition (45). This notion of a potential challenge to one's professional knowledge has also been found in other studies (62). It highlights the dilemma between clinical knowledge (or tacit understanding) versus the information derived from these large data sets.

Crilly *et al.* 2015, conducted a study involving the development of a tool to predict the patient flow in a hospital environment for better management of bed allocation (43). Prior to the implementation of the tool, the researchers of this project conducted a focus group in order to inform its development and evaluation qualitatively (63). The study found that although the expectations of the implementation of such tools are high, so is the anticipation as well. Results also indicated that once the tool had been implemented that participants had mixed perceptions towards the accuracy of the tool. Studies such as these indicate that what is perceived as accurate across practitioners can vary considerably. When coupled with high expectations about what a tool can achieve there are more often than not a spectrum of responses from users.

Some GPs identified frustration with not knowing how the score was calculated. This cultivation of trust or avoiding advice from the 'black box' has been raised in the past (60). Alerts need to be specific and allow the practitioner to subjectively evaluate their level of agreement with the alert. Regular alerts that are deemed as irrelevant reduce use of the tools. The interface of the POLAR alert allowed practitioners to see multiple levels of current and historic patient information in a single site, which many practitioners said was beneficial once they were acclimatised to the system. However, when the practitioner did not see patient information that allowed them to subjectively agree with the risk score, they called into question how the algorithm was constructing scores.

Data quality within the GPs system was raised by some GPs. "*The reports were reasonably realistic, but they are only as good as the data that the GP has put them into the program*". According to Horsky *et al.*, 2012, clinicians need to be persuaded that the advice is precise and appropriate within tools for best use (60). Data quality issues within general practice are not new in Australia and need to be an ongoing focus for accredited practices according to the RACGP (64). For tools such as the POLAR Diversion Algorithm to work optimally, practices must store data in the appropriate places within their clinical or billing systems. Information stored in notes, which are not extracted, will ensure that important patient information is lost and not taken into consideration by decision support tools.

The notion of patient 'stability' was raised on occasions by GPs. Some GPs recognised that there was a cluster of patient records that would put a patient at risk but felt that they were stable and therefore did not agree with the risk score calculated. However, it was noted by some that POLAR Diversion Risk score did give them an opportunity to review and ensure they were doing everything they could to keep the patient stable. This may imply that the use of the tool in the future may require an 'alert' label rather than a 'predictive' label to aid GPs comfortability with using the tool. If GPs are trained to see it as an alert that allows them to easily review the patient record as a whole, they can then subjectively choose if they feel that some change in treatment or referral is required.

The behaviour change associated with using the tool was identified for 12.9% of patient records, this reduced for the Extreme risk category, which is thought to be related to GPs not generally agreeing with extreme risk scores. Changes in treatment were more often related to testing (i.e. pathology and imaging) rather than referrals or pharmacotherapy activities.

Benefits

The POLAR Diversion project constitutes a substantial contribution to the evaluation of general practice and primary healthcare in Australia and hence, the potential for better health policy. As shown by its performance, POLAR can provide high predictability in the 30 day period and alert GPs when a patient is at higher risk for an ED presentation. Having a complete and active machine learning algorithm, which can alert clinicians for patients at risk for ED presentation, offers the advantage to the healthcare system to use it in its benefit. Therefore, not only a tool like this offers the potential of increasing the quality of life of many patients, but can also benefit the government and private health insurances financially by reducing costly ED attendances.

Strengths and Limitations

Although the study has added considerable value, data quality of general practice will be a limitation to the development and use of the algorithm. Data quality is only as good as how a clinician enters it (e.g. GPs neglecting to include all information in a patient's record or include the information in the wrong part of the software). However, the general principle is that if GPs can derive the benefit from better data, they will be driven to improve it.

Although the Australian health system is not currently unified across data sets, the POLAR Diversion Project has managed to link key data sets to develop an algorithm that can be solely run on routinely collected general practice data. This means that the algorithm is applicable in an Australian context and able to function right away in any practice using common clinical and billing software.

Outcome Health is the first in Australia to develop and trial a machine learning algorithm from General Practice data examining risk presentation to ED. However, the project also went a step further and actively involved GPs in evaluation of the tool across their own patients. The resulting perceptions of GP allow the project to determine how to best utilise the algorithm within a general practice context and any further work that may be needed to improve the tool for GPs best engagement.

The POLAR Diversion project focuses general practice data to develop risk scores, which allows GPs to mobilise available resources to patients in a timely manner. This could result in keeping patients out of hospital, which benefits the healthcare system and the community.

The way forward

Future iterations of the algorithm should involve mapping of chronic diseases, which will allow to better highlight acute versus chronic conditions and how this may relate to the 31-365 time category. The algorithm should be geared towards point of time risk stratification at a GP's computer during a patient visit to further optimise GPs ability to intervene.

Based on the results, the tool could be used as an alert based on risk categories rather than the per cent predictive risk score. A score such as 42% can be difficult to conceptualise in many cases, what does the 42% really mean when time and other human variables are taken into account. Rather moving forward a score based on groupings is proposed (i.e. Very Low <20%, Low 20-40%, Medium 40-60%, High 60-80%

and Extreme 80-100%) to allow GPs to determine if they have a complex but stable patient or if they need to review interventions such as diagnostics, management plans, referrals, medicine reviews, etc. However, in order for GPs to perceive the most value, the essential algorithm variables could be reviewed to ensure patient information in the 'current' or last visit tab (i.e. diagnoses, medications or pathology).

The algorithm can serve as a clinical decision aid and working with General Practices and PHNs regarding the design of the interface will be essential moving forward. This will facilitate the effective use of the alert in general practice and ensure that the algorithm remains responsive and relevant.

Subsequent studies should:

- Review the best way to improve the tool, develop its 'real time' capacity and interface and run a longer-term study across multiple general practices with access linked to multiple diverse ED data set, in order to review how the tool works across a broader spectrum of patient and practice groups.
- Explore the interface between the clinical knowledge (tacit) used by clinicians and the explicit knowledge contained in the tool. The tool should work with the clinicians, not displace them.
- Explore the utility of combining more data sets (pharmacy for compliance for instance) into the model
- Explore the models of care and/or resource needs to turn the GP contact into an admission reduction outcome.

CONCLUSIONS

This study has demonstrated the value of using linked data and modern computing tools to generate a machine learning model that has a high rate of predicting the risk of admission based on GP electronic health record data alone. This allows it to be deployed at the point of care, to maximise its effectiveness in influencing care provision.

The findings of the GP testing however highlight the ongoing dilemma of the application of such tools. The roll out of electronic record solutions is never about the technical capabilities, but the social environment in which they arrive – the socio-technical approach. In this study, this means that the next challenge is using and developing the tool in such a way that clinicians will trust it, and that it be used to influence change in the models of care for these patients.

ACKNOWLEDGMENTS

HCF Research Foundation

Eastern Melbourne Primary Health Network

Gippsland Primary Health Network

South East Melbourne Primary Health Network

Health Language Analytics

INDEX

ACSC= Ambulatory Care Services Conditions

AF= Anna Fragkoudi

AMc= Adam Mcleod

AVA= Adults Veteran Association

CP= Chris Pearce

CPM= Combined Predictive Model

CSIRO= Commonwealth Scientific and Industrial Research Organisation

DPM= Devon Predictive Model

ED= Emergency Department

ERA= Elders Risk Assessment

EMR= Electronic Medical Record

GP= General Practitioner

HARP= Hospital Admission Risk Program

HES= Hospital Episode Statistics

KPI= Key Performance Indicator

LACE= Length of stay, Acuity of admission, patient Comorbidity, and number of visits to the Emergency room

NR= Natalie Rinehart

OH= Outcome Health

PAPT= Patient Admission Prediction Tool

PARR= Patients At Risk of Re-hospitalisation

PHN= Primary Health Network

POLAR= POpulation Level Analysis and Reporting Tool

PPH= Potentially Preventable Hospitalisations

PPV= Positive Predictive Value

PRM= Predictive Risk Models

SLK= Statistical Linkage Key

UK= United Kingdom

USA= United States of America

VHA= Veteran Health Administration

REFERENCES

1. Watkins C, Harvey I, Langley C, Faulkner A, Gray S. General practitioners' use of computers during the consultation. Br J Gen Pract. 1999;49(442):381-3.
2. Adelaide PHN [Internet]. SA: Adelaide PHN; 2016 [cited 2017 September 04]. Available from: http://www.adelaidephn.com.au/publications-resources/activity-work-plans/
3. Australian Department of Health. PHN Performance Framework. 2017. [cited 2017 September 04]. Available from: http://www.health.gov.au/internet/main/publishing.nsf/Content/PHN-Performance_Framework
4. Billings J, Georghiou T, Blunt I, Bardsley M. Choosing a model to predict hospital admission: an observational study of new variants of predictive models for case finding. *BMJ Open*. 2013;3(8).
5. Ismail SA, Gibbons DC, Gnani S. Reducing inappropriate accident and emergency department attendances. *Br J Gen Pract*. 2013;63(617):e813-e20.
6. Roland M, Abel G. Reducing emergency admissions: Are we on the right track? *BMJ*. 2012;345:23.
7. Martin A, Martin C, Martin PB, Martin PA, Green G, Eldridge S. 'Inappropriate'attendance at an accident and emergency department by adults registered in local general practices: how is it related to their use of primary care? *J Health Serv Res Policy*. 2002;7(3):160-5.
8. Bernstein SL, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, et al. The effect of emergency department crowding on clinically oriented outcomes. *Acad Emerg Med*. 2009;16(1):1-10.
9. Collis J. Adverse effects of overcrowding on patient experience and care: John Collis presents the findings of a systematic literature review of how the number of people in emergency departments affects service delivery. *Emerg Nurse*. 2010;18(8):34-9.
10. Harrison MJ, Dusheiko M, Sutton M, Gravelle H, Doran T, Roland M. Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study. *BMJ : Br MedJ*. 2014;349.
11. Healthy Communities [Internet] Australia: Australian Institute of Health and Welfare (AIHW); 2015 [cited 2017 September 04]. Available from: https://www.myhealthycommunities.gov.au/our-reports/potentially-preventable-hospitalisations-update/july-2017/technical-note
12. Mazza D, Pearce C, Joe A, Turner LR, Brijnath B, Browning C, et al. Emergency department utilisation by older people in metropolitan Melbourne, 2008-12. Aust Health Rev. 2017.
13. Whitstock MT, Eckermann EJ, Marjoribanks TK, Pearce CM. Pharmaceutical economics and politics vs. patient safety: Lumiracoxib in Australia. International Journal of Risk & Safety in Medicine. 2008;20(3):161-7.
14. Kvamme OJ, Olesen F, Samuelsson M. Improving the interface between primary and secondary care: a statement from the European Working Party on Quality in Family Practice (EQuiP). Quality in Health Care. 2001;10(1):33.
15. Beaglehole R, Epping-Jordan J, Patel V, Chopra M, Ebrahim S, Kidd M, et al. Improving the prevention and management of chronic disease in low-income and middle-income countries: a priority for primary health care. The Lancet. 2008;372(9642):940-9.
16. Quarterly Medicare Statistics [Internet] Australia: Medicare; 2017 [cited 2017 September 04]. Available from: http://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/BudgetReview201314/Medicare

17. Strategic Review of the National Hospital Cost Data Collection [Internet] Australia: Independent Hospital Pricing Authority (IHPA); 2014 [cited 2017 August 02]. Available from: https://www.ihpa.gov.au/publications/strategic-review-national-hospital-cost-data-collection
18. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Medical Informatics and Decision Making. 2003;3(1):2.
19. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting Daily Patient Volumes in the Emergency Department. Academic Emergency Medicine. 2008;15(2):159-70.
20. Schweigler LM, Desmond JS, McCarthy ML, Bukowski KJ, Ionides EL, Younger JG. Forecasting models of emergency department crowding. *Acad Emerg Med*. 2009;16(4):301-8.
21. Wargon M, Casalino E, Guidet B. From Model to Forecasting: A Multicenter Study in Emergency Departments. Academic Emergency Medicine. 2010;17(9):970-8.
22. Crilly JL, Boyle J, Jessup M, Wallis M, Lind J, Green D, et al. The implementation and evaluation of the patient admission prediction tool: Assessing its impact on decision-making strategies and patient flow outcomes in 2 Australian hospitals. Quality Management in Healthcare. 2015;24(4):169-76.
23. Reuben DB, Keeler E, Seeman TE, Sewall A, Hirsch SH, Guralnik JM. Development of a method to identify seniors at high risk for high hospital utilization. *Med Care*. 2002;40(9):782-93.
24. Asplin BR, Flottemesch TJ, Gordon BD. Developing Models for Patient Flow and Daily Surge Capacity Research. *Acad Emerg Med*. 2006;13(11):1109-13.
25. Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *BMJ: Br Med J*. 2006;333(7563):327-30.
26. Donnan PT, Dorward DW, Mutch B, Morris AD. Development and validation of a model for predicting emergency admissions over the next year (PEONY): a UK historical cohort study. *Arch Intern Med*. 2008;168(13):1416-22.
27. Essex Strategic Health Authority Combined Predictive Model: Final Report. 2006.
28. Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions score. BMJ open. 2013;3(8):e003482.
29. HARP W. Summary: HARP risk calculator [Internet] Australia2009 [cited 2017 September 11]. Available from: https://www.adma.org.au/
30. Howell S, Coory M, Martin J, Duckett S. Using routine inpatient data to identify patients at risk of hospital readmission. *BMC Health Serv Res*. 2009;9(1):96.
31. Khanna S, Boyle J, Good N, editors. Precise prediction for managing chronic disease readmissions. Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE; 2014: IEEE.
32. Rana S, Tran T, Luo W, Phung D, Kennedy RL, Venkatesh S. Predicting unplanned readmission after myocardial infarction from routinely collected administrative hospital data. *Aust Health Rev*. 2014;38(4):377-82.
33. Oliver-Baxter J, Bywood P, Erny-Albrecht K. Predictive risk models to identify people with chronic conditions at risk of hospitalisation. 2015. Report No.: 0994187459.
34. Risk Stratification and next steps with DH Risk Prediction tools – Patients at Risk of Re-hospitalisation and the Combined Predictive Model [Internet] UK: UK Government 2011 [cited 2017 September 04]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/215489/dh_129005.pdf
35. Boyle J, Jessup M, Crilly J, Green D, Lind J, Wallis M, et al. Predicting emergency department admissions. Emergency medicine journal : EMJ. 2012;29(5):358-65.

36. Martens JD, van der Weijden T, Severens JL, de Clercq PA, de Bruijn DP, Kester AD, et al. The effect of computer reminders on GPs' prescribing behaviour: A cluster-randomised trial. International Journal of Medical Informatics. 2007.
37. Wang SJ, Middleton B, Prosser LA, Bardon CG, Spurr CD, Carchidi PJ, et al. A cost-benefit analysis of electronic medical records in primary care. The American Journal of Medicine. 2003;114(5):397-403.
38. Gao J, Moran E, Li Y-F, Almenoff PL. Predicting potentially avoidable hospitalizations. *Med Care*. 2014;52(2):164-71.
39. Crane SJ, Tung EE, Hanson GJ, Cha S, Chaudhry R, Takahashi PY. Use of an electronic administrative database to identify older community dwelling adults at high-risk for hospitalization or emergency department visits: the elders risk assessment index. *BMC Health Serv Res*. 2010;10(1):338.
40. Chenore T, Pereira Gray D, Forrer J, Wright C, Evans P. Emergency hospital admissions for the elderly: insights from the Devon Predictive Model. *J Public Health*. 2013;35(4):616-23.
41. van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Can Med Assoc J*. 2010;182(6):551-7.
42. Cooner MC, H and McMurray, A. The Gold Coast Integrated Care Model. International Journal of Intergrated Care. 2016;16(3):1-8.
43. Crilly JL, Boyle J, Jessup M, Wallis M, Lind J, Green D, et al. The implementation and evaluation of the patient admission prediction tool: Assessing its impact on decision-making strategies and patient flow outcomes in 2 Australian hospitals. *Qual Manag Health Care*. 2015;24(4):169-76.
44. Crilly JL, O'Dwyer JA, O'Dwyer MA, Lind JF, Peters JA, Tippett VC, et al. Linking ambulance, emergency department and hospital admissions data: understanding the emergency journey. *Med J Aust*. 2011;194(4):S34.
45. Chiang PP, Glance D, Walker J, Walter F, Emery J. Implementing a QCancer risk tool into general practice consultations: an exploratory study using simulated consultations with Australian general practitioners. *Br J Cancer*. 2015;112(Suppl 1):S77.
46. Jessup M, Crilly J, Boyle J, Wallis M, Lind J, Green D, et al. Users' experiences of an emergency department patient admission predictive tool: A qualitative evaluation. *Health Inform J*. 2016;22(3):618-32.
47. Pearce CM, McLeod A, Patrick J, Boyle D, Shearer M, Eustace P, et al. Using Patient Flow Information to Determine Risk of Hospital Presentation: Protocol for a Proof-of-Concept Study. JMIR research protocols. 2016;5(4):e241.
48. Mazza D, Pearce C, Turner LR, De Leon-Santiago M, McLeod A, Ferriggi J, et al. The Melbourne East Monash General Practice Database (MAGNET): Using data from computerised medical records to create a platform for primary care and health services research. J Innov Health Inform. 2016;23(2):181.
49. Pearce C, Shearer M, Gardner K, Kelly J, 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.
50. Pearce C, Shearer M, Gardner K, Kelly J. A division's worth of data. Aust Fam Physician. 2011;40(3):167-70.
51. SNOMED [Internet] Australia: Australian Government 2015 [cited 2017 September 18]. Available from: https://www.digitalhealth.gov.au/implementation-resources/ehealth-foundations/EP-2066-2015
52. Hospital Admission Risk Program (HARP) [Internet]. VIC: Victorian Government; 2011 [cited 2017 August 02]. Available from: https://www2.health.vic.gov.au/Api/downloadmedia/%7B7EBC12EB-385B-4395-A542-B7C9DB8399E5%7D

53. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. Journal of medical Internet research. 2016;18(12).
54. Abdelrahman W, Abdelmageed A. Medical record keeping: clarity, accuracy, and timeliness are essential. *BMJ Careers*. 2014.
55. PSO EI. Wrong-record, wrong-data errors with health IT systems. USA; 2015.
56. Thakurdas P, Coster G, Gurr E, Arroll B. New Zealand general practice computerisation; attitudes and reported behaviour. N Z Med J. 1996;109(1033):419-22.
57. Sherrington C, Tiedemann A. Physiotherapy in the prevention of falls in older people. Journal of Physiotherapy. 2015;61(2):54-60.
58. Ageing WHO, Unit LC. WHO global report on falls prevention in older age. World Health Organization; 2008. Report No.: 9241563532.
59. Healthy Communities [Internet] 2009 [cited 2017 September 20]. Available from: http://myhealthycommunities.gov.au/national/abs0009
60. Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. *J Biomed Inform*. 2012;45(6):1202-16.
61. Shibl R, Lawley M, Debuse J. Factors influencing decision support system acceptance. *Decis Support Syst*. 2013;54(2):953-61.
62. Porat T, Delaney B, Kostopoulou O. The impact of a diagnostic decision support system on the consultation: perceptions of GPs and patients. *BMC Med Inform Decis Mak*. 2017;17(1):79.
63. Jessup M, Wallis M, Boyle J, Crilly J, Lind J, Green D, et al. Implementing an emergency department patient admission predictive tool: insights from practice. *J Health Organ Manag*. 2010;24(3):306-18.
64. Royal Australian College of General Practitioners 2017 [cited 2017 September 20]. Available from: http://www.racgp.org.au/standards/311

APPENDICES

Project team governance

The Project comprised three groups:

a) Project Steering Group

Purpose:

Oversaw and guided the project management, risk management, reporting and progress of the project. This group had fortnightly meetings and used Basecamp™, an online project forum platform, for communication purposes.

b) Academic Consultants Group

Purpose:

To provide additional academic expertise to guide the Project Advisory Group on an as-needed basis. The Project Advisory Committee considered advice from the Academic Consultant Groups. The group consists of A/Prof Christopher Pearce and three academics with expertise relevant to the project.

c) Project Advisory Committee

Purpose:

- provided strategic, technical, clinical and operational advice to the researchers
- constituted the lead group providing advice and direction on the project deliverables
- monitored progress of recruitment and participation of General Practices
- provided feedback to the Outcome Health Board

Role:

The role of the Project Advisory Committee was to provide project advice to the Principal Investigator (CP) and Outcome Health CEO (AMc) around practitioner needs, clinical assumptions, and project and product applicability and intellectual property considerations. The Committee also advised on the specific alert criteria of the risk identification algorithm through a range of local best practice clinical guidelines and international learnings.

The Committee met as required.

Membership:

Assoc. Prof Chris Pearce (Monash University) (Chair)

Adam McLeod (Outcome Health)

Marianne Shearer (GPHN)

Dr Pat Crowe (GP/Outcome Health)

Lisa McGlynn (AIHW) –DoHA.

Andrew Howard (CIO-DoHVic)

Anna Burgess (DoHVic)

Paul Madden (CIO-DoHA)

Dr Andrew Maclean (ED Director, Eastern Health)

Dr Anthony Hobbs (DoHA)

Reporting:

The Project Advisory Committee reports to the Outcome Health Board.