Assessing Preventable Hospitalisation InDicators (APHID): protocol for a data-linkage study using cohort study and administrative data

Louisa R Jorm,1,2 Alastair H Leyland,3 Fiona M Blyth,4,2 Robert F Elliott,5 Kirsty M A Douglas,6 Sally Redman,2 on behalf of the APHID Investigators

ABSTRACT

Introduction: Potentially preventable hospitalisation (PPH) has been adopted widely by international health systems as an indicator of the accessibility and overall effectiveness of primary care. The Assessing Preventable Hospitalisation InDicators (APHID) study will validate PPH as a measure of health system performance in Australia and Scotland. APHID will be the first large-scale study internationally to explore longitudinal relationships between primary care and PPH using detailed person-level information about health risk factors, health status and health service use.

Methods and analysis: APHID will create a new longitudinal data resource by linking together data from a large-scale cohort study (the 45 and Up Study) and prospective administrative data relating to use of general practitioner (GP) services, dispensing of pharmaceuticals, emergency department presentations, hospital admissions and deaths. We will use these linked person-level data to explore relationships between frequency, volume, nature and costs of primary care services, hospital admissions for PPH diagnoses, and health outcomes, and factors that confound and mediate these relationships. Using multilevel modelling techniques, we will quantify the contributions of person-level, geographic-level and service-level factors to variation in PPH rates, including socioeconomic status, country of birth, geographic remoteness, physical and mental health status, availability of GP and other services, and hospital characteristics.

Ethics and dissemination: Participants have consented to use of their questionnaire data and to data linkage. Ethical approval has been obtained for the study. Dissemination mechanisms include engagement of policy stakeholders through a reference group and policy forum, and production of summary reports for policy audiences in parallel with the scientific papers from the study.

INTRODUCTION

Potentially preventable hospitalisations (PPH) (also termed as hospitalisations for ‘ambulatory care sensitive conditions’ (ACSC), ‘ambulatory sensitive hospitalisations’ and ‘preventable hospitalisations’) are those that could potentially be prevented by timely and effective provision of primary care.1 The concept of PPHs was originally developed in the USA,2-4 but has been adopted widely by international health systems as an indicator of the accessibility and overall effectiveness of primary care.5 Potentially preventable hospitalisations (PPH) (also termed as hospitalisations for ‘ambulatory care sensitive conditions’) have been adopted widely by international health systems as an indicator of the accessibility and overall effectiveness of primary care.6 However, much of the existing evidence is based on ecological (aggregate) analyses, and comes from the USA.7 Key questions about the validity and value of PPH measures, and their applicability in different settings, remain unanswered.

Strengths and limitations of this study

APHID will be the first large-scale study internationally to explore longitudinal relationships between primary care and PPH using detailed person-level data containing only limited information about health risk factors, health status and health service use.8 Limitations include the use of administrative data relating to the use of health services.

Key messages

- This article reports the protocol for the Assessing Preventable Hospitalisation InDicators (APHID) study.
- APHID will validate potentially preventable hospitalisations (PPH) as a measure of health system performance in Australia and Scotland.
- APHID will create a new longitudinal data resource by linking together detailed data from a large-scale cohort study and prospective administrative data relating to the use of health services.

Key questions about the validity and value of PPH measures, and their applicability in different settings, remain unanswered.
systems as an indicator of the accessibility and overall effectiveness of primary care.5–8 In Australia, rates of PPH for selected conditions are a key performance indicator specified in the National Healthcare Agreement (NHA),9 10 which is intended to track progress against the objective ‘ Australians receive appropriate high quality and affordable primary and community health services’. The conditions that are currently included in the NHA indicator are shown in table 1.10 These are based on the Victorian ACSC study,11 which in turn had its origins in work in the USA in the 1990s1–5 that used physician expert panels to identify and rank candidate conditions. However, there is considerable variation in the PPH condition sets that are used across countries, and for PPH measures are particularly attractive as indicators of health system performance because they can be generated from routine hospital data and yet focus attention on the outcomes of care, rather than process or throughput. However, a number of key questions about the validity and value of these measures remain unanswered.

Much of the research relating to PPHs has focused on socioeconomic, race and urban-rural differentials, which may reflect gradients in health status (disease prevalence and severity) as well as in access to or quality of health care. However, studies in the USA, mainly using ecological (correlational) approaches (ie, the unit of analysis was an aggregate of individuals, usually the population of a geographic area), have also focused on the association between PPHs and the nature of the healthcare system, reporting that self-reported access to medical care,12 increased physician supply15–18 and presence of community and rural health centres19–21 are inversely correlated with PPH rates, while high rates of emergency department (ED) attendances are positively associated with PPH rates.21

Evidence from other settings is sparse. A recent study in the UK, limited to chronic obstructive pulmonary disease (COPD) and using an ecological design, found that patient-reported access to consultation within 2 days, and primary care staffing, were protective for admission rates.22 The only Australian study to date,23 again using ecological methods, found that access to medical care and rate of general practitioner (GP) visits (both self-reported) and GP supply were negatively associated with PPH rates in Primary Care Partnership areas in the state of Victoria, independent of disease prevalence. However, these associations disappeared when rural residence was taken into account. The authors concluded that rural residence may be a greater risk factor for PPH than access to primary care, but did not identify the potential role of the hospital (eg, variability in admission practices, availability of beds) in driving urban–rural differences in PPH. Indeed, PPH rates are potentially an indicator of access to hospital care, as well as to primary care, an issue that to our knowledge has never been investigated, either in Australia or internationally.

The Assessing Preventable Hospitalisation InDicators (APHID) project will validate PPH as a measure of health system performance in Australia and Scotland. It involves three partner agencies with key roles in using these measures to drive change in the Australian health system: the Australian Commission on Safety and Quality in Health Care, the Agency for Clinical Innovation and the NSW Bureau of Health Information.

Using linked person-level data, the APHID project will explore relationships between frequency, volume, nature and costs of primary care services, hospital admissions for PPH diagnoses, and health outcomes, and factors that confound and mediate these relationships. Using multilevel modelling techniques, we will quantify the contributions of person-level, geographic-level and service-level factors to variation in PPH rates, including socioeconomic status, country of birth, geographic remoteness, physical and mental health status, availability of GP and other services, and hospital characteristics. The use of individual data and taking account of the different levels that may influence outcomes means that we will avoid the risk of the ecological fallacy common in ecological studies.24

Our specific objectives are:
1. To link questionnaire data from 267 000 participants in the 45 and Up Study to prospective data on use of primary care services, ED presentations, hospitalisations and deaths.
2. To analyse these linked data to establish the relationships between use of primary care services and

---

### Table 1 Conditions included in the Australian National Healthcare Agreement potentially preventable hospitalisations performance indicator

<table>
<thead>
<tr>
<th>Vaccine-preventable conditions</th>
<th>Influenza and pneumonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute conditions</td>
<td>Other vaccine-preventable conditions</td>
</tr>
<tr>
<td></td>
<td>Dehydration and gastroenteritis</td>
</tr>
<tr>
<td></td>
<td>Pyelonephritis</td>
</tr>
<tr>
<td></td>
<td>Perforated/bleeding ulcer</td>
</tr>
<tr>
<td></td>
<td>Cellulitis</td>
</tr>
<tr>
<td></td>
<td>Pelvic inflammatory disease</td>
</tr>
<tr>
<td></td>
<td>Ear, nose and throat infections</td>
</tr>
<tr>
<td></td>
<td>Dental conditions</td>
</tr>
<tr>
<td></td>
<td>Appendicitis with generalised peritonitis</td>
</tr>
<tr>
<td>Chronic conditions</td>
<td>Asthma</td>
</tr>
<tr>
<td></td>
<td>Congestive cardiac failure</td>
</tr>
<tr>
<td></td>
<td>Diabetes complications</td>
</tr>
<tr>
<td></td>
<td>Chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td></td>
<td>Angina</td>
</tr>
<tr>
<td></td>
<td>Iron deficiency anaemia</td>
</tr>
<tr>
<td></td>
<td>Hypertension</td>
</tr>
<tr>
<td></td>
<td>Nutritional deficiencies</td>
</tr>
<tr>
<td></td>
<td>Rheumatic heart disease</td>
</tr>
</tbody>
</table>
measures of PPH, and the contributions of person-level, geographic-level and service-level factors to these relationships.

3. To analyse these linked data to establish the relationship between PPH and health outcomes for people with chronic conditions, and the contributions of person-level, geographic-level and service-level factors to these relationships.

4. To conduct comparative analyses using data from the Scottish Morbidity Records.

5. To consider, synthesis and effectively communicate these findings in order to drive change.

**METHODS AND ANALYSIS**

**Data sources**

The 45 and Up Study is a cohort study of 267,000 men and women aged 45 years and over and resident in New South Wales (NSW), Australia’s largest state. Briefly, participants were randomly sampled from the Medicare Australia database and joined the Study by completing a mailed self-administered questionnaire (available at [http://www.45andup.org.au/](http://www.45andup.org.au/)) and providing consent for long-term follow-up, including linkage to health records. People resident in non-urban areas and those aged 80 and over were oversampled. Recruitment occurred from February 2006 to April 2009. The overall response rate is estimated at 18%, consistent with similar cohort studies.

Data captured in the 45 and Up Study baseline and follow-up questionnaires include the following self-reported chronic conditions that are relevant to PPH: (ever diagnosed) heart disease, high blood pressure, diabetes and asthma; and (treated in the last month) ‘other’ heart conditions, high blood pressure and asthma. Additional data coded from free text fields allow identification of participants who reported angina, congestive heart failure and COPD.

Questionnaire data also include information on key potential confounders and mediating factors, including age, sex, household income, level of education, smoking history, alcohol use, physical activity (Active Australia questionnaire), height and weight, functional status (Medical Outcomes Study Physical Functioning scale), psychological distress (Kessler 10 scale) and medical and surgical history.

The NSW Admitted Patient Data Collection (APDC) will be used to identify PPH admissions and to provide information for risk-adjustment (comorbidities, previous admissions). The APDC includes records for all separations (discharges, transfers and deaths) from all NSW public and private sector hospitals and day procedure centres. The information reported includes patient demographics, source of referral to the service, service referred to on separation and diagnoses, procedures and external causes of injury coded according to the Australian modification of the International Statistical Classification of Diseases and Related Problems, 10th revision (ICD-10-AM). Audits have shown good-to-excellent coding of diagnoses and procedures in Australian hospital data. Because the APDC contains information on episodes of care (ending with the discharge, transfer or death of a patient), rather than periods of stay in hospital, a continuous period of stay will be constructed by combining all contiguous episodes of care, including nested and non-nested transfers, for the same patient.

The NSW Emergency Department Data Collection (EDDC) will provide information about ED presentations. The EDDC covers 80 EDs, including all those in public hospitals in the Sydney metropolitan area, and captures about 75% of all presentations to NSW EDs. The information reported includes patient demographics, mode of arrival, triage category, mode of separation, service referred to on separation, diagnoses and procedures.

NSW Registry of Births, Death and Marriages (RBDM) death registration (fact-of-death) data will be used to ascertain deaths following hospitalisation and for censoring in person-time analyses. Most deaths are registered within 4 weeks of the date of death.

Most non-hospital medical care in Australia is provided on a fee for service basis, paid by the universal health insurance scheme, Medicare, according to the Medical Benefits Schedule (MBS) of payments. We will use MBS claims data to measure use of GP, specialist and other Medicare-funded services. For each claim for service processed, the MBS claims data include the date of the service, the item number for the service, patient age, gender and postcode, provider business code, the amount charged by the provider, the Medicare benefit for the service, the method of payment and de-identified information relating to the provider. Individual providers who practice in more than one location can have multiple Medicare provider numbers. GP claims for patient care under MBS distinguish brief, standard, long and prolonged consultations.

Under the Pharmaceutical Benefits Scheme (PBS), the Australian government subsidises the cost of medicine for most medical conditions for all Australian residents who hold a current Medicare card. We will use PBS claims data to measure use of medications and medication compliance. Only pharmaceuticals that qualify for subsidy are included in PBS claims data, coded to the Pharmaceutical Benefits Schedule. PBS data include PBS item number, date, type of prescription, PBS payment category, specialty of provider and pharmacist’s business postcode.

**Data linkage**

Linkage of MBS and PBS data to 45 and Up Study data will be performed by the Sax Institute. All 45 and Up Study participants consented specifically to this linkage, and it is performed under approvals from the Australian Government Department of Human Services and the Australian Government Department of Health and Ageing ethics committee. This will be a direct
Assessing Preventable Hospitalisation Indicators (APHID)

(deterministic) linkage, using an encrypted version of the Medicare number. All other linkages will be undertaken by the Centre for Health Record Linkage (CHeReL), which links questionnaire data from the 45 and Up Study to the other databases that will be used in this study on an ongoing basis.

The CHeReL uses probabilistic record linkage techniques to link personal identifiers (including full name, date of birth, sex and address) from records in these datasets. Evaluation of the accuracy of the linkage is determined by clerical review of samples of matched records. Quality assurance data show false positive and negative rates of 0.4% and <0.1%, respectively. The CHeReL uses the ‘best practice protocol’ for preserving individual privacy. The data custodians will supply de-identified datasets for each of the data sources to the researchers, who will then merge the records together using a unique identifier.

The linked data will include all records available from July 2000 onwards for each dataset. This will provide retrospective data for risk-adjustment, as well as prospective data for ascertainment of incident events.

Analysis plan

To establish the relationships between use of Medicare-funded services and measures of PPH (objective 2), a series of multilevel models will be built using prospective data for all participants in the 45 and Up Study. Follow-up will be to the end of the period for which APDC data are available (December 2011), or death, whichever comes first, giving a duration of follow-up from 2 to 4 years depending on the date of recruitment into the 45 and Up Study. Model building will be an iterative process and models will have up to six levels: individuals (n=267,000), GP providers (n=14,500), Divisions of GP or Medicare Locals (n=20); Statistical Local Areas (SLAs, n=250), hospitals (n=400) and Local Health Districts (LHDs, n=18).

Analyses to explore the relationship between PPH and health outcomes for people with chronic conditions (objective 3) will be restricted to 45 and Up Study participants who self-reported chronic PPH conditions (asthma, angina, congestive heart failure, COPD, diabetes and hypertension) at baseline. They will compare outcomes for people with these conditions who have PPH episodes during follow-up with those who do not. Analyses will be performed for all the chronic conditions as a group and stratified according to individual condition, where numbers support this. Again, a series of multilevel models will be built, with up to six levels: individuals (n=126,000), GP providers (n=14,000), Medicare Locals (n=20); SLAs (n=250), hospitals (n=400) and LHDs (n=18).

Outcome measures related to PPH will include: time to first PPH episode; number of PPH episodes; total PPH inpatient bed days; average length of PPH hospital stay; estimated costs of PPH episodes to the health system and the individual. Hospital costs will be estimated using Australian Refined Diagnosis Related Group (AR-DRG) costs obtained from national public and private sector cost data. Costs to the individual will be estimated using information on labour force status and household income. Analyses will be performed for all causes and stratified according to vaccine-preventable, chronic and acute PPH, and individual PPH diagnoses where numbers support this.

Health outcomes will include: time to death; number of non-PPH hospital stays; total bed days for non-PPH hospital stays; number of ED presentations; number of MBS services; cost of hospital services; cost of MBS services (benefit paid and out-of-pocket costs); cost of PBS services; estimated costs to the health system and to individuals.

Individual-level predictor variables will include:

- Demographic and socioeconomic factors: age, sex, country of birth, language spoken at home, indigenous status (note small numbers will preclude Indigenous-specific reporting), accessibility and remoteness (using the Accessibility/Remoteness Index of Australia (ARIA+) which assigns scores to geographic areas based on the road distance to service towns of different sizes(38)), private health insurance and healthcare card holder status, household income, highest level of education, employment status and household composition.

- Use of Medicare-funded services: number of GP consultations (brief, standard, long and prolonged), proportion of consultations that are long or prolonged, MBS benefits and out-of-pocket costs and PBS benefits and out-of-pocket costs.

- Quality of GP care: number of GP providers providing care; usual provider continuity index (proportion of visits to most frequently seen GP provider); continuity of care index (derived from number of different GP providers seen and number of visits to each GP provider)(34); claims for GP Management Plan or Team Care Arrangement (chronic conditions only); completion of annual cycle of care (diabetes only); multidisciplinary diabetes care, that is, participants saw a GP and at least one of the following: specialist, practice nurse, diabetes educator, dietician and ophthalmologist/optometrist (diabetes only); medication compliance ( persistence of dispensing).

- ED presentations: number of ED presentations. These analyses will be restricted to participants resident in SLAs serviced by EDs that are captured in the EDDC.

- Health risk factors: smoking status, alcohol intake, physical activity, body mass index, hypertension and high blood cholesterol.

- Health status: self-reported functional status, psychological distress, social support, number and type of self-reported chronic conditions, use of hospital and ED services in 5 years prior to baseline and pattern and volume of the use of Medicare-funded services in 5 years prior to baseline.

GP provider-level predictor variables will include: vocational registration; number of 45 and Up Study participants using this provider, annual number of
Consultations (brief, standard, long and prolonged); proportion of consultations that are long or prolonged.

Medicare Local-level predictor variables will include: presence of active Primary Care Collaboratives programme or active chronic disease management programme; proportion of local GPs who are participating in Medicare Locals.

SLA-level predictor variables will include: accessibility (ARIA+), socioeconomic status (using Socio Economic Status for Areas (SEIFA) indexes, four indexes that summarise different aspects of the socioeconomic conditions of people living in an area based upon sets of social and economic information from the Australian Census); full-time equivalent GPs; medical workers, nurses, pharmacists, Aboriginal health workers and community services workers per 10 000 population; rates of unemployment and labour force participation.

Hospital-level predictor variables will include: accessibility (ARIA+); number of beds; peer group; number of active Primary Care Collaboratives programme or active chronic disease management programme; presence of chronic conditions, acute conditions and low-prevalence chronic conditions.

Table 2 presents minimum detectable ratios for continuous predictor variables as well as appropriate models.

### Statistical power

Table 2 gives information on the projected numbers of PPH episodes, and persons experiencing PPH episodes, that will be available for analysis, based on linked data for the 45 and Up Study to date, and taking into account delays in the availability of data. Around 950 000 person-years of follow-up will be available for analysis. With the sample size available, we will be able to estimate average length of stay and inpatient costs, respectively, with a precision of 0.7 days and $650 for low-prevalence PPH conditions (eg, rheumatic heart disease, pelvic inflammatory disease); 0.5 days and $275 for mid-prevalence PPH (asthma, influenza and cellulitis) and 0.07 days and $65 for high-prevalence PPH (diabetes complications).

Table 3 presents minimum detectable ratios for comparisons of mean length of stay and hospital costs for people with self-reported chronic conditions who do and do not have a PPH episode during follow-up.
Assessing Preventable Hospitalisation InDicators (APHID)

The power of the study will be monitored during the model-building process, using sample size calculators specifically developed for multilevel models (PinT). These require estimates of variances and covariances for predictor variables and random effects that are not available from the literature.

Comparative analyses using Scottish hospital data

It is intended to undertake comparative analysis using Scottish data. This part of the project is at an earlier stage of development and the detailed research is still subject to funding and approval. Scotland offers an interesting comparator to Australia. Both countries contain large areas that are remote and inaccessible: 6% of the Scottish population live on areas classified as remote and rural while in Australia 10% live in areas classified as remote or very remote. Both countries face the challenges of providing healthcare in these areas. However, while in Australia health is poorer in rural areas than in urban areas, in Scotland the opposite holds and this is also true for deprivation. Comparative analysis offers the prospect of distinguishing the contributions of socioeconomic status, health status and rurality and remoteness to geographic variation in rates of PPHs.

GP supply has been reported to be negatively associated with PPH rates in Australia. Scotland has considerably fewer GPs per 100,000 population than does Australia: 81 and 110, respectively. Patients are registered with a specific practice in Scotland but not in Australia, and this may significantly alter organisational approaches to preventative healthcare. A further difference between the two countries is the funding of healthcare: in Australia 45% of the population holds private health insurance, compared with only 11% of the Scottish population.

Scottish morbidity records will be used to identify rates of PPHs across Scotland. Factors to be explored include quality of primary care, the availability of primary and secondary care, the rurality and remoteness of the population, the degree of deprivation and population characteristics including age, gender and ethnicity. The empirical strategy outlined above to analyse the Australian data will be employed and comparative analysis will be undertaken. Wherever possible, this will employ a common specification in order to distinguish the differential impact of the factors and the extent to which these might be explained by differences in underlying population behaviours and health delivery systems.

A strength of the Scottish analysis is that relatively robust measures of the quality of primary care can be constructed. In 2004, the Quality and Outcome Framework (QOF) was introduced to incentivise the quality of care provided by general practices. Practices could accumulate points on 146 indicators and receive payment according to the number of points achieved. For our analyses, measures of the quality of chronic disease management will be constructed using the practices’ performance on the clinical indicators within QOF, such as for example percentage of patients with asthma who had an asthma review in the preceding 15 months and investigate whether these measures are associated with PPH rates.

Ethics and dissemination

Participants have consented to use of their questionnaire data and to data linkage. Ethical approval has been obtained from the NSW Population and Health Services Ethics Committee, the Aboriginal Health and Medical Research Council of NSW Ethics Committee and the University of Western Sydney Ethics Committee.

Dissemination mechanisms include engagement of policy stakeholders through a reference group and policy forum, which will use deliberative dialogue approaches, with the goal being to identify potential actions and key implementation considerations.

We will write summary reports in formats designed for policy audiences in parallel with the scientific papers from the study. Our experience indicates that doing this enhances the clarity and quality of the scientific papers as well as facilitating quick uptake of the research findings. The summaries for policy will focus on incorporating the findings with other relevant evidence into compelling accounts that highlight potential actions.

Author affiliations

1Centre for Health Research, School of Medicine, University of Western Sydney, Campbelltown, New South Wales, Australia
2The Sax Institute, Sydney, New South Wales, Australia
3MRC/CSO Social and Public Health Sciences Unit, Glasgow, UK
4Concord Clinical School, University of Sydney, Sydney, New South Wales, Australia
5Health Economics Research Unit, University of Aberdeen, Aberdeen, UK
6Australian National University Medical School, Australian National University, Canberra, Australian Capital Territory, Australia

Acknowledgements We thank the men and women participating in the 45 and Up Study, which is owned and managed by the Sax Institute, Sydney, NSW (www.saxinstitute.org.au). We thank the Sax Institute, the NSW Ministry of Health, the NSW Register of Births, Deaths and Marriages and Medicare Australia for allowing access to the data, and the Centre for Health Record Linkage for conducting the probabilistic linkage of records.

Collaborators The APHID investigator team comprises Louisa Jorm, Alastair Leyland, Fiona Blyth, Robert Elliot, Kirsty Douglas, Sally Redman, Marjon van der Pol, Michael Falster, Neville Board, Danielle Butler, Douglas Lincoln, Sanja Lujic, Kate Needham, Damiola Olajide, Deborah Randall, Kim Sutherland, Diane Watson and Hunter Watt.

Contributors LJ had overall responsibility for the conception of this study and drafting this paper. LJ, AL, FB, RE, KD and SR all contributed to the design of the study and the writing of this paper and approved the final draft.

Funding The APHID study is funded by a National Health and Medical Research Council Partnership Project Grant (#1036858) and by partner agencies the Australian Commission on Safety and Quality in Health Care, the Agency for Clinical Innovation and the NSW Bureau of Health Information. The 45 and Up Study is funded by the Sax Institute with support from major partner Cancer Council NSW and other partners which, at the time of writing, include: Heart Foundation (NSW Division), NSW Ministry of Health, beyondblue: the national depression initiative, Ageing, Disability and Home Care, NSW Family and Community Services and Australian Red Cross Blood Service.
Competing interests None.

Ethics approval NSW Population and Health Services Ethics Committee, Aboriginal Health and Medical Research Council of NSW Ethics Committee, University of Western Sydney Ethics Committee.

Provenance and peer review Not commissioned; internally peer reviewed.

Data sharing statement The APHID study dataset has been constructed with the permission of each of the custodians of the respective source datasets and with specific ethical approval. The dataset could potentially be made available to other researchers if they obtain the necessary approvals. More information about these approvals is available from the authors on request.

REFERENCES

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
</tr>
</thead>
</table>
Assessing Preventable Hospitalisation InDicators (APHID): protocol for a data-linkage study using cohort study and administrative data

Louisa R Jorm, Alastair H Leyland, Fiona M Blyth, et al.

*BMJ Open* 2012 2:
doi: 10.1136/bmjopen-2012-002344