BMJ open accessible medical research

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

To cite: Jorm LR, Leyland AH, Blyth FM, *et al*. Assessing Preventable Hospitalisation InDicators (APHID): protocol for a data-linkage study using cohort study and administrative data. *BMJ Open* 2012;**2**:e002344. doi:10.1136/bmjopen-2012-002344

▶ Prepublication history for this paper are available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2012-002344).

Received 14 November 2012
Revised 14 November 2012
Accepted 20 November 2012

For numbered affiliations see end of article.

Correspondence to
Professor Louisa R Jorm; l.jorm@uws.edu.au

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.

ARTICLE SUMMARY

Article focus

- 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 messages

- PPH have been adopted widely by international health systems as an indicator of the accessibility and overall effectiveness of primary care.
- However, much of the existing evidence is based on ecological (aggregate) analyses, and comes from the USA.
- 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 information about health risk factors, health status and health service use.
- Limitations include the use of administrative claims data containing only limited information about the quality of primary care services, and reliance on self-reported data for some predictor variables.

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,[1–4] but has been adopted widely by international health

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 1990s[1–3] 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 healthcare. 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 supply[13–18] and presence of community and rural health centres[19–21] 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

Vaccine-preventable conditions	Influenza and pneumonia
	Other vaccine-preventable conditions
Acute conditions	Dehydration and gastroenteritis
	Pyelonephritis
	Perforated/bleeding ulcer
	Cellulitis
	Pelvic inflammatory disease
	Ear, nose and throat infections
	Dental conditions
	Appendicitis with generalised peritonitis
Chronic conditions	Asthma
	Congestive cardiac failure
	Diabetes complications
	Chronic obstructive pulmonary disease
	Angina
	Iron deficiency anaemia
	Hypertension
	Nutritional deficiencies
	Rheumatic heart disease

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[25] 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/) 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),[26] height and weight, functional status (Medical Outcomes Study Physical Functioning scale),[27] psychological distress (Kessler 10 scale)[28] 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).[29] Audits have shown good-to-excellent coding of diagnoses and procedures in Australian hospital data.[30] 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

(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'[31] 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.[32] 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[33]), 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

Table 2 Projected* numbers of potentially preventable hospitalisation (PPH) episodes and persons experiencing PPH episodes, Assessing Preventable Hospitalisation InDicators (APHID) study

Condition	Episodes	Persons
Chronic	22440	13000
Asthma	650	480
Congestive cardiac failure	2500	1870
Diabetes complications	11400	7000
Chronic obstructive pulmonary disease	4100	2400
Angina	1850	1600
Iron deficiency anaemia	1400	1100
Other chronic	540	480
Vaccine-preventable	620	550
Influenza and pneumonia	500	450
Other vaccine-preventable	120	100
Acute	7350	7400
Dehydration and gastroenteritis	2300	2050
Pyelonephritis	2100	1750
Dental conditions	970	930
Other acute	1980	1592
All conditions	30 410	19 850

*Projections are based on linked data for the 45 and Up Study from study entry up to December 2010, and assume that similar age-specific event rates will apply for the period from study entry up to December 2011.

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+),[33] 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[35]); 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+)[33]; number of beds; peer group; number of admissions (emergency, planned and medical surgical); measures of casemix; patient experience survey data; measures of hospital workforce.

LHD-level predictor variables will include: number of hospitals and community-based services; measures of LHD workforce.

Statistical analysis will be performed using MLwiN[36] and SAS.[37] Cross-classified multilevel models will be used to account for the non-nested hierarchies of, for example, GP provider, the hospital attended and the SLA of residence[38] and will be estimated using Markov Chain Monte Carlo methods in MLwiN. Time to first PPH episode will be modelled using multilevel Cox (proportional hazards) regression analysis. All other outcomes will be modelled using multilevel Poisson regression, applying transformations to continuous variables as appropriate. All models will be corrected for possible overdispersion, either by using an adjustment parameter in the Poisson model or by fitting a negative binomial regression model. In the case of a highly right skewed distribution of the number of admissions beyond the first, zero-inflated Poisson or multinomial models will be fitted. For all models, squared and cubed terms of continuous predictor variables will be tested to improve the model fit. A method of fractional polynomials for continuous predictor variables will be used where appropriate.

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); 0.3 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

Table 3 Minimum detectable mean ratios* for analyses investigating the relationship between potentially preventable hospitalisations and health outcomes for people with chronic conditions, Assessing Preventable Hospitalisation InDicators (APHID) study

	Self-reported chronic condition							
Outcome	Hypertension		Asthma		Diabetes		Angina	
Mean length of stay	0.97	1.03	0.96	1.05	0.96	1.04	0.95	1.05
Mean hospital cost	0.98	1.02	0.97	1.03	0.97	1.03	0.95	1.04

*5% significance, 80% power.

(objective 3). The power of the study will be monitored during the model-building process, using sample size calculators specifically developed for multilevel models (PinT).[39] 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 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.[40 41] 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[42] and this is also true for deprivation.[41 43 44] 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.[11] Scotland has considerably fewer GPs per 100 000 population than does Australia: 81[45] and 110,[46] 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%[47] of the population holds private health insurance, compared with only 11%[48] 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[49] 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[50] 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.[51]

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

1. Billings J, Zeitel L, Lukomnik J, *et al*. Impact of socioeconomic status on hospital use in New York City. *Health Aff (Millwood)* 1993;12:162–73.
2. Billings J, Teicholz N. Uninsured patients in District of Columbia hospitals. *Health Aff (Millwood)* 1990;9:158–65.
3. Weissman JS, Gatsonis C, Epstein AM. Rates of avoidable hospitalization by insurance status in Massachusetts and Maryland. *JAMA* 1992;268:2388–94.
4. Agency for Healthcare Research and Quality. *AHRQ Quality Indicators. Guide to Prevention Quality Indicators*. Version 3.1 (12 March 2007). http://www.qualityindicators.ahrq.gov/downloads/pqi/pqi_guide_v31.pdf (accessed Mar 2011).
5. Roos LL, Walld R, Uhanova J, *et al*. Physician visits, hospitalizations, and socioeconomic status: ambulatory care sensitive conditions in a Canadian setting. *Health Serv Res* 2005;40:1167–85.
6. Jackson G, Tobias M. Potentially avoidable hospitalisations in New Zealand, 1989–98. *Aust N Z J Public Health* 2001;25:212–21.
7. Purdy S, Griffin T, Salisbury C, *et al*. Ambulatory care sensitive conditions: terminology and disease coding need to be more specific to aid policy makers and clinicians. *Public Health* 2009;123:169–73.
8. Niti M, Ng TP. Avoidable hospitalisation rates in Singapore, 1991–1998: assessing trends and inequities of quality in primary care. *J Epidemiol Commun Health* 2003;57:17–22.
9. *Intergovernmental Agreement (IGA) on Federal Financial Relations: Schedule F National Healthcare Agreement November 2008*. Council of Australian Governments. http://www.coag.gov.au/intergov_agreements/federal_financial_relations/index.cfm (accessed Mar 2011).
10. Steering Committee for the Review of Government Service Provision. *National Healthcare Agreement. National Agreement performance information 2008–09*. Melbourne: Productivity Commission, 2009. http://www.pc.gov.au/__data/assets/pdf_file/0009/98757/healthcare-agreement.pdf (accessed Mar 2011).
11. Ansari Z, Carson N, Serraglio A, *et al*. The Victorian Ambulatory Care Sensitive Conditions study: reducing demand on hospital services in Victoria. *Aust Health Rev* 2002;25:71–7.
12. Bindman AB, Grumbach K, Osmond D, *et al*. Preventable hospitalizations and access to health-care. *JAMA* 1995;274:305–11.
13. Krakauer H, Jacoby I, Millman M, *et al*. Physician impact on hospital admission and on mortality rates in the Medicare population. *Health Serv Res* 1996;31:191–211.
14. Friedman B, Basu J. Health insurance, primary care, and preventable hospitalization of children in a large state. *Am J Manag Care* 2001;7:473–81.
15. Ricketts TC, Randolph R, Howard HA, *et al*. Hospitalization rates as indicators of access to primary care. *Health Place* 2001;7:27–38.
16. Backus L, Moron M, Bacchetti P, *et al*. Effect of managed care on preventable hospitalization rates in California. *Med Care* 2002;40:315–24.
17. Basu J, Friedman B, Burstin H. Primary care, HMO enrollment, and hospitalization for ambulatory care sensitive conditions: a new approach. *Med Care* 2002;40:1260–9.
18. Laditka JN, Laditka SB, Probst JC. More may be better: evidence of a negative relationship between physician supply and hospitalization for ambulatory care sensitive conditions. *Health Serv Res* 2005;40:1148–66.
19. Probst JC, Laditka JN, Laditka SB. Association between community health center and rural health clinic presence and county-level hospitalization rates for ambulatory care sensitive conditions: an analysis across eight US states. *BMC Health Serv Res* 2009;9:134.
20. Epstein AJ. The role of public clinics in preventable hospitalizations among vulnerable populations. *Health Serv Res* 2001;36:405–20.
21. Hossain M, Laditka JN. Using hospitalization for ambulatory care sensitive conditions to measure access to primary health care: an application of spatial structural equation modeling. *Int J Health Geogr* 2009;8:51.
22. Calderon-Larranaga A, Carney L, Soljak M, *et al*. Association of population and primary healthcare factors with hospital admission rates for chronic obstructive pulmonary disease in England: national cross-sectional study. *Thorax* 2011;66:191–6.
23. Ansari Z, Laditka JN, Laditka SB. Access to health care and hospitalization for ambulatory care sensitive conditions. *Med Care Res Rev* 2006;63:719–41.
24. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;88:216–22.
25. Banks E, Redman S, Jorm L, *et al*. Cohort profile: the 45 and up study. *Int J Epidemiol* 2008;37:941–7.
26. Australian Institute of Health and Welfare. *The Active Australia Survey: a guide and manual for implementation, analysis and reporting*. Canberra: AIHW, 2003.
27. Stewart A, Kamberg CJ. Physical functioning measures. In: Stewart A, Ware J, eds. *Measuring functioning and well-being: the medical outcomes study approach*. Durham, NC: Duke University Press, 1992:86–101.
28. Kessler R. *Final version of our non-specific psychological distress scale (memo dated 3/10/94)*. Ann Arbor: Survey Research Center of the Institute for Social Research, University of Michigan, 1994.
29. National Centre for Classification in Health. *The International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian modification (ICD-10-AM). Volume 3. ICD-10-AM Tabular list of procedures*. Sydney: Faculty of Health Sciences, University of Sydney, 2006.
30. Henderson T, Shepheard J, Sundararajan V. Quality of diagnosis and procedure coding in ICD-10 administrative data. *Med Care* 2006;44:1011–19.
31. Kelman CW, Bass AJ. Research use of linked health data—a best practice protocol. *Aust N Z J Public Health* 2002;26:251–5.
32. *National Hospital Cost Data Collection cost reports for Rounds 11–12*. Australian Government Department of Health and Ageing. http://www.health.gov.au/internet/main/publishing.nsf/Content/Round_12-cost-reports (accessed Mar 2011).
33. Australian Institute of Health and Welfare. *Rural regional and remote health: a guide to remoteness classifications*. Canberra: AIHW, 2004.
34. Cheng SH, Chen CC, Hou YF. A longitudinal examination of continuity of care and avoidable hospitalization evidence from a Universal Coverage Health Care System. *Arch Intern Med* 2010;170:1671–7.
35. Adhikari P. *Socio-economic indexes for areas: introduction, use and future directions*. Canberra: Australian Bureau of Statistics, 2006.
36. Rasbash J, Browne W, Healy M, *et al*. *MLwiN v2.25*. Bristol: Centre for Multilevel Modelling, University of Bristol, 2012.
37. *SAS Version 9.2*. Cary, NC: SAS Institute, 2007.
38. Hill PW, Goldstein H. Multilevel modeling of educational data with cross-classification and missing identification for units. *J Educ Behav Stat* 1998;23:117–28.
39. Snijders TAB, Bosker EJ. *PinT (Power in Two-level designs)*. Groningen, The Netherlands: University of Groningen, 2007.
40. People and Communities. National Statistics—Rural Scotland Key Facts 2011. http://www.scotland.gov.uk/Publications/2011/09/29133747/2 (accessed Sep 2012).
41. Demography and population. Services for Australian Rural and Remote Allied Health. http://www.sarrahtraining.com.au/site/index.cfm?display=143626 (accessed Sep 2012).
42. Godden D. Providing health services to rural and remote communities. *J R Coll Physicians Edinb* 2005;35:294–5.
43. Household weekly income by remoteness, latest year. Health Statistics New South Wales. http://www.healthstats.nsw.gov.au/Indicator/soc_income_aria (accessed Sep 2012).
44. Economy and Enterprise. National Statistics—Rural Scotland Key Facts 2011. http://www.scotland.gov.uk/Publications/2011/09/29133747/4 (accessed Sep 2012).
45. Connolly S, Bevan G, Mays N. *Funding and performance of health care systems in the four countries of the UK before and after devolution*. London: Nuffield Trust, 2011. www.nuffieldtrust.org.uk/publications/funding-and-performance-healthcare-systems. (accessed Sep 2012).
46. OECD Health Data 2009. OECD. http://www.oecd.org/health/oecdhealthdata2009comparinghealthstatisticsacrossoecdcountries.htm (accessed Sep 2012).

47. Health care delivery and financing. Australian Bureau of Statistics. http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1301.0~2012~Main%20Features~Health%20care%20delivery%20and%20financing~235 (accessed Sep 2012).
48. Office of Fair Trading. *Private healthcare market study*. London: Office of Fair Trading, 2011. http://www.oft.gov.uk/shared_oft/market-studies/OFT1396_Private_healthcare.pdf (accessed Sep 2012).
49. Dusheiko M, Gravelle H, Martin S, *et al. Does better disease management in primary care reduce hospital costs?* York: Centre for Health Economics, University of York, 2011.
50. Lavis JN. How can we support the use of systematic reviews in policymaking? *PLoS Med* 2009;6:e1000141.
51. Jewell CJ, Bero LA. 'Developing good taste in evidence': facilitators of and hindrances to evidence-informed health policymaking in state government. *Milbank Q* 2008;86:177–208.

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