

Hypertension Health Literacy in Australians

Data management plan and supplementary material

Joshua Brown, Owen Cho, Murray Keogh, Collin Sakal

Table of contents

Table of contents	1
1. Data management plan	2
1.1. Administrative.....	2
1.2. Data Collection and Management.....	2
1.3. Metadata	3
1.4. Ethics.....	3
1.5. Data Storage and Backup.....	3
1.6. Data Access and Retention	3
1.7. Responsibilities and Resources.....	4
2. Metadata document.....	5
3. Data dictionary for the survey data	6
3.1. Data dictionary for patient survey	6
3.2. Data dictionary for GP survey	8
3.3. Response data processing with data dictionary	8
4. Confidentiality, privacy and security	9
References	10

1. Data Management Plan

1.1. Administrative

Project name

Hypertension Health Literacy in Australians

Project Description

The purpose of this project is to examine the similarities and differences in hypertension health literacy by a range of patient characteristics. The project will also estimate the incidence and prevalence of hypertension in Australia and examine healthcare utilization by patient groups (hypertensive and non-hypertensive patients).

Investigator Details

Sanja Lujic, Centre for Big Data Research in Health (CBDRH), UNSW Sydney

Joe Stark, Data Analysis Section, Department of Health (DoH), Canberra

Contact Personnel

Sanja Lujic, CBDRH, s.lujic@unsw.edu.au

Funding Agency Details

This work is supported by a UNSW seeding grant (ID 1234/56)

DMP Version Dates

DMP Created: 16/6/2020

Last DMP Update: 30/6/2020

1.2. Data Collection and Management

The study will consist of three main data sources: Self-reported questionnaires filled by the patients and the GP, Pharmaceutical Benefit Scheme (PBS), and Medicare Benefit Schedule (MBS).

The self-reported questionnaires will include the patient's self-reported information consisting of demographic details and answers to questions relating to hypertension literacy. The GP specific questions incorporate patient's measured height, weight and blood pressure and prior history of hypertension diagnosis and treatment.

The PBS data will include information consisting of the patient's prescription medication dispensing history, including medication name and amount. The MBS data will include information consisting of the patient's health services received that were funded by Medicare, including the type and date of service. These data sources currently exist.

The study population includes patients aged 30 years and older who attend GP practices in Australia. A total of 3,000 patients will be recruited from 150 GP practices across Australia during May – September 2021. Once 150 GPs are recruited to take part in the study, 20 consecutive patients aged 30 and over will be recruited from each GP. The questionnaires will be administered to the GP and patients electronically, and then collected via mail or email (responder preference). When questionnaires are collected, the study participants will then be linked to PBS and MBS data sources

by the Australian Institute of Health and Welfare (AIHW) Data Integration Services Centre, an accredited authority.

Additionally, the Australian Bureau of Statistics (ABS) mid-year estimated residential population will be utilized to produce adjusted prevalence estimates.

Data cleaning, management, and analysis will be conducted using SAS 9.4 Software. The de-identified linked questionnaire and MBS/PBS data will be stored as flat files on a network drive. These flat files will be utilized as data sources in SAS. A project directory will be created on the network drive containing folders for code, data, and research documents. The data folder will include original and processed data sets and will follow the naming convention:

HHL_Data_TypeofEdit_Initials_Version. The code folder will include SAS scripts used in the cleaning and analyzing of the data and will follow the naming convention:

HHL_Code_TypeofAnalysis_Initials_Version. The research folder will include all other documents pertaining to the study such as presentation slides, journal write-ups, and project management documents.

Version control will be used in file naming conventions to keep track of changes across timelines and investigators. Where applicable, a version number and initials of editing person will be appended to the end of the file name. The initials will be removed with each new version to ensure readability.

1.3. Metadata

Descriptive and structural metadata will be provided in a text file format documenting information on data quality, context, use, and accessibility. Additionally, the project will document information about the authors. The Dublin Core Metadata Initiative (DCMI) format is followed throughout.

1.4. Ethics

The ethics clearance for this study will be sought from the UNSW Human Research Ethics Committee (HREC) as well as the AIHW Ethics Committee prior to the commencement of the study. The investigator will obtain written informed consent from both GPs and patients participating in the study after informing them of the study objectives, along with potential benefits and risks. The original signed consent will be retained in the UNSW records. Furthermore, the study will seek to maintain participant privacy by not storing the names and addresses of study participants with the questionnaires. Intellectual property and copyright conditions as states in the UNSW Intellectual Property Policy will be followed and copyright will be attributed to the Centre for Big Data Research in Health (CBDRH).

1.5. Data Storage and Backup

Data will be stored jointly at the University of New South Wales (UNSW) and the Australian Government Department of Health (DoH). At UNSW completed questionnaires and consent forms will be stored in locked filing cabinets at Ms. Lujic's office which is accessible through swipe card only. Electronic copies of the questionnaire and consent forms will be kept on the UNSW secure network drive. At the DoH de-identified linked questionnaire and Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS) supplied by AIHW Data Integration Services Centre will be stored on DoH network drive.

1.6. Data Access and Retention

During the study access to data stored at UNSW or DoH will only be given to principal investigators and researchers specified in the metadata document. After project completion, all de-identified data

will be archived at UNSW and DoH. Linked questionnaire, MBS, and PBS data will be retained in accordance with relevant legislation and appropriate determination of its value in future research. Outside access to archived study data can be obtained pending ethics approval.

Aggregate study findings will be disseminated through peer review publications and presentations. Authorship on publications and presentations will be decided based on the International Committee of Medical Journal Editors (ICMJE) Uniform Requirements for Manuscripts Submitted to Biomedical Journals.

1.7. Responsibilities and Resources

Sanja Lujic will be responsible for implementing, reviewing, and revising the Data Management Plan. Responsibility for each of the activities outlined in the Data Management Plan, including but not limited to data capture, metadata production data quality and analysis, data storage and backup, data archiving and sharing, will be jointly shared by Sanja Lujic and Joe Stark in their roles as Principal Investigators. SAS 9.4 Software will be utilized in the analysis and management of the data, and no further expertise, hardware, or software resources will be needed.

2. Metadata Document

<i>Title</i>	Hypertension Health Literacy in Australians
<i>Identifier</i>	DOI: TBA
<i>Date of Data Collection</i>	Estimated: May – September 2021
<i>Publication Date</i>	Estimated: Paper 1 - Mid 2022, Paper 2 - Early 2023.
<i>Description</i>	The Health Literacy Project examines similarities and differences in hypertension by a range of patient characteristics. It will estimate the incidence and prevalence of hypertension in Australia and examine healthcare utilisation by patient group. The study population includes individuals over the age of 30, who will be sampled through a random selection of GP's through the Australasian Medical Publishing Company mailing list. Data will be collected through a questionnaire that will be completed by both the patient and the GP. The patient will answer questions on demographic details and hypertension literacy. The GP will answer questions on patient height, weight, blood pressure and prior history of hypertension diagnosis and treatment.
<i>Spatial Coverage</i>	Australia
<i>Subject</i>	Medicine, Biomedical Sciences
<i>Topic Classification</i>	Hypertension, Health Literacy, Healthcare Utilisation
<i>Type</i>	Dataset derived from questionnaire, Text (Research paper)
<i>Creator</i>	Centre for Big Data Research in Health & Department of Health
<i>Contributors</i>	Sanja Lujic, Centre for Big Data Research in Health (CBDRH), UNSW Sydney Joe Stark, Data Analysis Section, Department of Health (DoH), Canberra
<i>Funder</i>	The University of New South Wales - Seeding grant ID 1234/56
<i>Contact</i>	Sanja Lujic, CBDRH, s.lujic@unsw.edu.au
<i>Publisher</i>	Biomedical Journal: TBA
<i>Language</i>	English
<i>File/Data Format</i>	TBA
<i>Related Resource</i>	Hocking A, Laurence C, Lorimer M. Patients' knowledge of their chronic disease-the influence of socio-demographic characteristics. Australian Family Physician. 2013; 42:411-6
<i>Rights</i>	All rights held by the University of New South Wales(UNSW), The Centre for Big Data Research in Health(CBDRH) and the Department of Health(DoH).
<i>Software</i>	SAS Version 9.4
<i>Source</i>	N/A

[1-2]

3. Data dictionary for the survey data

This section contains detailed data dictionaries for the data containing responses of the surveys. It was assumed that the surveys are to be electronically distributed so that survey metadata, such as participant ID, can be automatically embedded. Each row represents a response about an individual participant and each column is for a question, resulting in each cell containing a response to a specific question. The variable names are the column names and were shortened and lowercased for the convenience of processing in coding interface. The responses will be validated at the time of the participants' input.

3.1. Data dictionary for patient survey

The table below is for response data from the patient survey.

Variable name	Variable type	Main question number	Main question narrative	Sub question number	Sub question narrative	Possible choices	Input validation
date_pt	Date		Date of survey:				ISO format (YYYY-MM-DD)
id	String		Participant ID:				Unique participant ID is embedded
q01	Categorical	1	Has a doctor or a health care practitioner (e.g. nurse) told you that you have high blood pressure?			Yes No Don't know	
q02	Categorical	2	In the past month have you been treated for high blood pressure?			Yes No Don't know	
q03	Categorical	3	What is meant by 'hypertension'? (please select one)			High blood pressure High blood sugar Overactivity High tension/stress Don't know	
q04	Categorical	4	How common do you think high blood pressure is in Australia? (please select one)			1 in 100 people (1%) 1 in 20 people (5%) 1 in 5 people (20%) 1 in 2 people (50%) Don't know	
q05	Categorical	5	What measures can help in lowering blood pressure? (please select one)			Only medication can help Only lifestyle (healthy diet and exercise) can help Medication and lifestyle can help Nothing can help Don't know	

Variable name	Variable type	Main question number	Main question narrative	Sub question number	Sub question narrative	Possible choices	Input validation
q06a	Categorical	6	Do you agree or disagree with the following statements?	A	Older age is associated with high blood pressure	Strongly agree Agree Disagree Strongly disagree	
q06b	Categorical			B	Cigarette smoking increases the risk of high blood pressure	Strongly agree Agree Disagree Strongly disagree	
q06c	Categorical			C	High levels of physical and emotional stress increase the risk of high blood pressure	Strongly agree Agree Disagree Strongly disagree	
q06d	Categorical			D	Family history of high blood pressure increases the risk of developing high blood pressure	Strongly agree Agree Disagree Strongly disagree	
q07	Numeric	7	What is your age?				A positive integer under 150 (years)
q08	String	8	Which Australian state do you reside in?				Valid Australian state name
q09	Numeric	9	What is your residential postcode?				Valid Australian postcode
q10	Categorical	10	What is your gender			Male Female	
q11	Categorical	11	What is the your highest level of education?			Attended high school, did not complete Year 10 High School, Year 10 High School, Year 12 (HSC) Traineeship/apprenticeship TAFE University – Bachelors degree University – Postgraduate degree	
q12	Categorical (multiple selection is allowed)	12	Are you of Aboriginal or Torres Strait Islander Origin? [can select multiple]			No Yes, Aboriginal Yes, Torres Strait Islander	
q13	Categorical	13	Do you speak a language other than English at home?			No Yes	
q14	String	14	In which country were you born?				Valid country name

3.2. Data dictionary for GP survey

The table below is for response data from the GP survey, and this can be joined with the patient survey data using participant ID variable as a key.

Variable name	Variable type	Main question number	Main question narrative	Sub question number	Sub question narrative	Possible choices	Input validation
date_gp	Date		Date of survey:				ISO format (YYYY-MM-DD)
loc_gp	Numeric		GP location (postcode):				Valid Australian postcode
id	String		Participant ID:				Unique participant ID is embedded
q15	Numeric	15	Patient's height (meters)				
q16	Numeric	16	Patient's weight (kg)				
q17a	Numeric	17	Systolic blood pressure/ Diastolic blood pressure	a	Systolic blood pressure		
q17b	Numeric			b	Diastolic blood pressure		
q18	Categorical	18	Does the patient have hypertension?			No Yes	
q19	Categorical	19	Patient currently treated for hypertension?			No Yes	
q20a	String	20	List of medications for treatment of hypertension	a	medication #1		
q20b	String			b	medication #2		
q20c	String			c	medication #3		

3.3. Response data processing with data dictionary

Above two tables can become lookup tables for processing the response data. The original response data with each question in each column would be difficult to be summarised. On the other hand, if the data are reshaped into tidy data with columns of the variable name and response value, it can be joined with above data dictionaries.[3] With this format, summarising for analysis would be easier and more systematic.

4. Confidentiality, privacy and security

Before participating in the study, each participant will be informed on relevant privacy laws, what personal information will be collected, how this information will be utilised, to whom this information will be disclosed to, as well as the risks of security breaches. All participants will sign an informed waiver of consent to acknowledge this. All researchers and data custodians will strive to uphold their duty of confidence for each participant's privacy through the following approach. Firstly, names and addresses will not be stored within the questionnaires. Secondly, all physical material (questionnaires and informed consent forms) will be stored in Ms. Lujic's office, only accessible through a swipe card of authorised personnel. Thirdly, all digital material will be stored solely on encrypted networks, with digital copies of physical material stored on UNSW's secure network drive and de-identified linked questionnaire - Medicare Benefits Schedule - Pharmaceutical Benefits Scheme data being stored on the Department of Health's network drive. Both these network drives have strict identification and authentication protocols in place to ensure that only authorised personnel can access the data. Fourthly, when linkage of different data is carried out, only essential information with anonymised patient ID will be shared between the data custodians. This is to minimise the risk of identifying what medical condition a specific patient at a specific practice has from full information from both data.[4] Fifthly, data will be only be retained in accordance with relevant laws and specifications laid out to participants when signing the waiver of consent. Again, all retained information will be stored on encrypted networks requiring authentication to be accessed.

References

1. ADA Dataverse [Internet]. Australia: Australian Data Archive; 2020. The Australian Longitudinal Study on Male Health, Wave 2 (Release 2); 2017 Aug 17 [cited 2020 Jun 22]. Available from: <https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.4225/87/2KHTSV>
2. Dublin Core Metadata Initiative [Internet]. U.S.: Dublin Core Metadata Initiative; c1995-2020. Dublin Core Metadata Element Set, Version 1.1: Reference Description; 2012 Jun 14 [cited 2020 Jun 22]. Available from: <https://www.dublincore.org/specifications/dublin-core/dces/>
3. Wickham H. Tidy data. Journal of Statistical Software. 2014 Sep 12;59(10):1-23.
4. National Statistical Service [Internet]. A Guide for Data Integration Projects Involving Commonwealth Data for Statistical and Research Purposes; 2014 Jul 1 [cited 2020 Jun 24]. Available from: <https://statistical-data-integration.govspace.gov.au/topics/applying-the-separation-principle#>