Metadata

Data Source

Self-reported questionnaires filled by the patients and the GP

The primary data source will be the self-reported questionnaires completed by the patients and general practitioners (GPs) participating in the study. These questionnaires will capture information related to hypertension health literacy, demographics, medical history, and other relevant variables.

Australian Bureau of Statistics (ABS) mid-year estimated residential population

The secondary data source will be obtained from the ABS website. The ABS provides mid-year estimated residential population data, which will be used to produce adjusted prevalence estimates for hypertension health literacy among the Australian population.

Data Preservation and Archiving

Paper Questionnaires:

- Format: Paper-based
- Description: The questionnaires are printed on paper, and participants fill them out manually using pen.
- Data Capture Method: Responses from paper questionnaires will be manually entered into a digital format for data storage and analysis.

Digital Questionnaires:

- Format: Electronic
- Description: The questionnaires are displayed using electronic devices such as computers, tablets, or smartphones.
- Data Capture Method: Participant responses are directly captured in a digital format.

Data Storage Format:

The collected questionnaire data will be stored in CSV (Comma-Separated Values) format, which is a commonly used and widely supported format for structured data. This format has good compatibility with various data analysis tools and software. Each row in the CSV file will represent a participant, and each column will represent a variable or data element.

File Naming Conventions

- Be descriptive and consistent.
- Use standard abbreviations: For example, HHLA for Hypertension health literacy in Australia.
- Include timestamps (dd_mm_yyyy) and version number.
- Use alphanumeric characters and use underscores to separate words within the file names.

Data Elements and Variable Descriptions

1. Administrative:

- Date of survey: Date of when the questionnaire was completed.
- Participant ID: Unique identifier for each participant.

2. GP Survey:

- Patient's height: Height of the patient in meters, keep two decimal places.
- Patient's weight: Weight of the patient in kilograms, keep one decimal place.
- Systolic blood pressure: Systolic blood pressure measurement in mmHg.
- Diastolic blood pressure: Diastolic blood pressure measurement in mmHg.
- Is the patient currently treated for hypertension: Binary variable indicating whether the patient is currently receiving treatment for hypertension.
- List of current medications for treatment of hypertension: Free-text field to capture the names of the medications.

3. Demographics:

- Age: Age of the patient in years.

- Australian state of residence: State in Australia where the patient resides.
- Residential postcode: Postcode of the patient's residential area.
- Gender: Gender of the patient.
- Highest level of education: Education level attained by the patient.
- Aboriginal or Torres Strait Islander Origin: Binary variable indicating if the patient identifies as Aboriginal or Torres Strait Islander.
- Language spoken at home: Indicating if the patient speaks a language other than English at home.
- Country of birth: Country where the patient was born.

4. Patient Questionnaire- High Blood Pressure:

- Has a doctor or a health care practitioner told you that you have high blood pressure: Binary variable indicating if the patient has been diagnosed with high blood pressure.
- Treated for high blood pressure in the past month: Binary variable indicating if the patient has received treatment for high blood pressure in the past month.
- Knowledge-related questions: Multiple-choice questions capturing the patient's understanding and knowledge about hypertension.

Data Relationships

The data relationships exist between the different questionnaires completed by the patients and GPs. Each patient questionnaire is associated with a specific GP survey. The demographic information links to the analysed results as the complement and revise.

De-identify and Data Clean

Personally identifiable information of patient is not included in questionnaire. Instead, a unique Participant ID was assigned to each participant for identification purposes.
Anatomical Therapeutic Chemical (ATC) Classification System was used to coding hypertension medications in data clean. This will not allow easy identification of the actual medications and simplify subsequent analysis.
Numeric codes (integer) will replace the categorical data, also the specific codes for missing or implausible values. These will help to simplify subsequent analysis.
Limitations of Numeric data will be applied during data clean. If a data value falls outside of the predefined limitations but is still considered reasonable:

- Contact the GP: If you have access to the GP who collected the data, it can help to confirm whether the value was accurately recorded. Or the GP may provide additional insights to reach and verify the value.
- Use professional judgment: Depending on your expertise and knowledge of the subject matter, you can exercise professional judgment to assess the reasonableness of the value. If the value appears to be within a clinically acceptable range despite exceeding the predefined limitations, you may choose to retain it. PLEASE NOTE, document the reasons why you consider the value to be reasonable and any supporting information you have. This documentation will help ensure transparency and allow others to understand the decision-making process.
- Sensitivity analysis: If appropriate for your analysis, you can conduct sensitivity analyses by evaluating the impact of including or excluding values that fall outside the limitations. This can help assess the robustness of your results and provide insights into the potential influence of these data points.

Data Attribution

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