## Affordability of Medicare-Supported Primary Healthcare for Australian Residents

## Background

### Medicare:

Medicare is Australia’s national public health insurance scheme. The scheme provides free or subsidized healthcare to all Australians and most permanent residents. It includes visits to doctors, specialists, optometrists and, in some cases, other healthcare professionals such as dentists. Medicare also covers the cost of treatment in public hospitals.

The government subsidizes more than 5,700 different health services and countless medications. Each service that is offered by the health professional is assigned a Medicare Benefits Schedule (MBS) item number and MBS Fee. The government decides what it believes to be a reasonable fee for the medical services and sets the MBS Schedule Fees accordingly. How much a patient is then reimbursed by Medicare is governed by the Schedule Fee for that service or treatment.

In some circumstances, Medicare reimburses 100 per cent of the Schedule Fee and at other times it is a smaller percentage. The amount and the percentage of reimbursement depends on which healthcare provider (public or private) is providing the treatment, and what the service or treatment is.

Medicare pays 85 per cent of the MBS fee for a specialist and 100 per cent for a general practitioner. Some doctors bulk-bill, which means they accept the MBS fee as full payment. However, doctors, specialists and other healthcare professionals are free to charge more than the MBS fee for services if they wish. In this case, the patient will be required to pay the difference between the MBS fee and the service fee (out of pocket costs).

### Primary Health Care

In Australia, primary care is often the first point of contact for individuals with the health system and most widely used part of the healthcare system. It relates to the treatment of non-hospitalised patients in the community that encompasses general practitioners, primary health care nurses and allied health. The services provided by these health professionals fall under the non-hospital Medicare-subsidised services, which will be the focus of the analysis.

## Objective:

Determine if non-hospitalised Medicare subsidised primary health care services are affordable to Australians in Australia. The services include:

1. GP
2. Nursing and Aboriginal health worker services
3. Allied health
4. Diagnostic imaging
5. Specialist Diagnostic

## Scope:

Focus is on non-hospitalized primary health care services subsidized by the government through the Medicare Benefit Schedule program.

## Dataset Sources & Limitations

### Census Datasets

**Datasets:**

1. 2011 Australian Census of Population and Housing,
2. 2016 Australian Census of Population and Housing
3. 2021 Australian Census of Population and Housing

**Date Range**: 2011, 2016, 2021

**Description**: The Australian Census of Population and Housing is the official count of population that collects details of age, sex and other characteristics of the population.

The Census measure the number and key characteristics of people in Australia on Census Nights:

1. 2011 - Tuesday 9th August 2011
2. 2016 - Tuesday 9th August 2016
3. 2021 - Tuesday 10th August 2021

Census data extracted for each year was population by sex, age and personal income for statistical area level 3 geographical area.

**Data Source**: The data is an external source from Australian Bureau of Statistics (ABS). ABS is Australia’s national statistical government agency that provides statistical services to Australian states and territory governments. As government data and source, the data source is considered trustworthy.

**Data Collection Method:**

Census Collection & Processing in 2011

1. **Inclusion**: All people in Australia, except foreign diplomats and their families, were included. Visitors were counted regardless of their length of stay, while Australian residents abroad were excluded.
2. **Distribution**: Census Management Units were tasked with providing every household with either eCensus access information or a paper Census form prior to Census Night.
3. **Collection**: Households could complete the Census online via eCensus or use the paper form. Assistance was available through the Census Collector or the Census Inquiry Service. In some urban areas, interviews were conducted to collect data. Collectors were responsible of collecting census forms after the Census Night.
4. **Follow-up**: If a household used the paper form or did not include all members in the eCensus, Collectors would visit up to five times within 20 days post-Census Night to retrieve the form.
5. **Processing**: Paper forms were checked and prepared for scanning, with damaged forms transcribed. Data capture and processing technologies such as intelligent forms processing and automatic coding were utilised.
6. **Data Repair and Coding**: Manual intervention was used to 'repair' unrecognizable characters. Manual coding was used when automatic determination was not possible.
7. **Data Derivation and Imputation**: Some data was derived from other responses on the form, like labor force status. Imputation was used for critical demographic data items when responses were not provided.
8. **Census Output Data Release**: Output data is released in stages over the course of 2 years.

Census Collection & Processing 2016:

2016 Census included the same population criteria as 2011 but the census form, distribution, data collection and follow-up methods were changed as follows. Due to the below changes, higher percentage of households completed the form online in 2016 than 2011.

Distribution & Collection**:**

1. Around 80% of Australian dwellings were mailed information containing a unique login number for the online form, instead of Census Field Officers visiting every dwelling in 2011.
2. The online form was redesigned for easy and secure use on various devices, from smartphones to desktop computers.
3. Households not responding online had option to request a paper form to complete and return via a prepaid envelope.
4. Reminder letters and subsequent visits by Census Field Officers were reserved for households that did not initially participate.
5. In some regions, Field Officers delivered materials and contacting residents. Follow-up visits were made to non-responding dwellings.

Census Collection & Processing 2021:

2021 Census included the same population criteria as 2011 and 2016. Similar distribution and collection strategy was used as 2016 with 85% of Australian dwelling receiving mail containing instruction and login letter. Some received the form. 15% of householders had Census field staff deliver paper forms containing census form and login details. These householders were in rural areas where small towns that no street delivery service.

Census Digital Service was introduced in 2021 which allowed users to request the form, report they would not be home on census night and login without Census letter, improving the number of responses and accuracy of the data.

2021 Census was the first time that the ABS actively encouraged households to complete and submit their Census form as soon as they received their materials, meaning this could be done before Census Night on 10 August.

**Data Limitation:**

1. Census form is manually completed by respondents via paper or online form which is susceptible to manual errors or purposeful incorrect input of data.
2. Census is conducted every 5 years, resulting in missing data. Population data of missing years will need to be imputed.
3. Census data includes everyone present in Australia on the census night. This includes individuals who don’t receive Medicare benefits such as visitors, temporary residents, international students etc.
4. Census data does not consider Australian residents that were out of the country on the census night but used MBS services during the year.
5. Census data does not consider changes in population due to migration.
6. In 2011, approximately 1.7% of the population was missed from the census due to travels, unreturned forms or mistakenly not completing the census form. There could be instances some people are counted more than once.
7. Computer editing procedures are used to detect and correct obvious errors made by individuals in completing the form, but the procedures cannot detect and correct all householders' errors, and some remain in final output.
8. There were minor differences in census form questions in 2011, 2016 and 2021 that may result in respondents answering questions differently in each census for the same topic. Below is the list affecting the current project.
   1. Additional option for sex was added to enable people to report neither male nor female in 2016.
   2. Additional option was added to sex to allow respondents to identify themselves as non-binary in 2021
   3. Targeted supplementary questions and supporting text were added in 2016 and increased the usage in 2021 to improve accuracy of responses
   4. Households were asked questions that applied to them, based on their previous answers in 2016 and 2021. For example, if a respondent reported that they were not in the labour force, they were sequenced out of questions relating to labour force.
   5. In 2021 census, long term health conditions topic questions were added.
9. Inconsistent geographical statistical area level 3 between 2011, 2016 and 2021 whereby additional areas were added or removed based on population growth.
10. In 2021, there was declined in response rate of regional areas, possibly due to census staff recruiting and moving difficulties due to COVID-19 restrictions.

**Relevancy**: The dataset was collected by trustworthy data source; government agency. As the data was collected ABS and is used by other government organisations, it’s the most reliable and complete set of data available of the Australian population.

Census data enables the segmentation of the population by state, statistical area levels, gender, and age, allowing analysis of the impact of out-of-pocket costs of primary healthcare on these demographics from 2014 through 2022.

While absent census data requires estimation, the census dataset remains vital for performing the required analyses to evaluate the impact of primary healthcare across different demographics and income. This data will be used for the project.

### Medicare Subsidized Services Dataset (MBS)

Datasets:

1. Medicare-subsidised GP, allied health and specialist health care across locals 2021-22
2. Medicare-subsidised GP, allied health and specialist health care across local areas: 2019–21
3. Medicare-subsidised GP, allied health and specialist health care across local areas: 2013–14 to 2018–19

**Date Range**: 2013 to 2022 (financial year)

**Description**: Dataset contains data on Medicare benefits paid and total provider fees of residents in specified geographical areas (state, PHN, SAL3 areas) by different demographics (gender and age)

**Data source**:

The data is from the Australian Institute of Health and Welfare (AIHW), a government agency responsible for providing statistics that inform policy and service delivery. AIHQ collected the administrivia data from Australian Government Department of Health and Aged Care (DOHAC) and estimated resident population data from Australian Bureau of Statistics (ABS) to generate these datasets. Since all the source used are government source, AIHW data is considered a reliable.

**Data Collection:**

The MBS subsidised reports were created from Medicare Benefits Schedule claims data managed by Department of Health and Aged Care. The claims data was derived from administrative information on services that qualify for a Medicare benefit under the Health Insurance Act 1973 and for which a claim has been processed by Services Australia.

Administrative information is submitted by the healthcare partitioner when a clinically relevant service to a Medicare-eligible person is provided and claim for rebate or benefit is made to cover all or part of the cost of the service.

The claims data records persons Medicare enrolment postcode, not where the service was administered.

Data is reported by the financial year in which they are processed.

**Data limitations:**

1. Does not include services where no Medicare benefit was claimed such as services subsidised by the Department of Veterans’ Affairs, compensation arrangements, or jurisdictional salaried GP services provided in remote outreach clinics.
2. Government expenditure associated with bulk billing incentives for non-hospital non-referred expenditures is not included.
3. There are instances where patient have used service outside their enrolled postcode but the claims data was recorded under the Medicare enrolment postcode
4. Records with missing postcode are included in the National total but not allocated a PHN or SA3
5. Medicare patients is higher than the Estimated Resident Population, the proportion of the population who had the service is reported as 100%
6. Patient’s Medicare enrolment postcode as recorded on the last service rendered (for any MBS service) in the reporting year was used. If a patient had more than one postcode listed on their last date of service in the year, then the postcode was taken from the last date of processing on that date of service.
7. Patients with an invalid age (less than zero or older than 117 years old) are excluded from the age demographic groupings but are included in the total patient’s category. These cases include a patient with an incorrect date of birth recorded in the Medicare Consumer Directory at Services Australia.
8. Over time MBS items are added, removed or amended which can affect the analysis done over time.
9. Providers sometimes bill a general item like 'GP Standard (Level B)' instead of a more specific service such as 'GP Health Assessment,' potentially leading to underreported usage of specific services.
10. The Australian population's official estimate, the ERP, is calculated by the ABS using Census data and records of births, deaths, and migration. This means population estimation does not consider population changes throughout the year.

**Relevancy**: The dataset was provided by trustworthy data source AIHW, a government agency. As the AIHW data is used by government organisations to advice on policies and service delivery, it’s the most reliable data available on Medicare subsidised primary health care services.

Medicare subsidised services datasets facilitate the analysis of provider charges, Medicare reimbursements, and the out-of-pocket expenses borne by patients, considering various demographics, geographic distributions, and temporal changes.

The dataset is crucial to assess the affordability of Medicare subsidised primary healthcare services in Australia and will be used for the project.

### Patient Experience

**Datasets:**

1. Experience of GP services 2022-2023
2. Experience of GP services 2021-2022
3. Experience of GP services 2020-2021
4. Experience of GP services 2019-2020
5. Experience of GP services 2018-2019
6. Experience of after-hours GP care 2022-2023
7. Experience of after-hours GP care 2021-2022
8. Experience of after-hours GP care 2020-2021
9. Experience of after-hours GP care 2019-2020
10. Experience of after-hours GP care 2018-2019

**Date Range**: 2019 to 2023 (financial year)

**Description**: The dataset includes yearly responses to a survey about the accessibility, obstacles, and patient experiences with healthcare services from general practitioners. It is organized by gender, age, and location.

**Data source**: The data is an external source from Australian Bureau of Statistics (ABS). ABS is Australia’s national statistical government agency that provides statistical services to Australian states and territory governments. As government data and source, the data source is considered trustworthy.

**Data Collection:**

Patient experience data is gathered via a survey, which is a topic on the Multipurpose Household Survey (MPHS) conducted by the ABS.

Information was collected from 25,934 fully responding persons in 2022-23, 23,949 in 2021-22, 28,386 in 2020-21, 29,793 in 2019-20 and 28,719 in 2018-19.

The data was collected monthly from a subset of the Labour Force Survey (LFS) sample. Each month, one-eighth of the LFS sample was chosen for the Multi-Purpose Household Survey (MPHS) through a random selection of a resident aged 15 or older using a computer algorithm.s

Interviews were conducted personally, with parental consent required for participants aged 15 to 17. In cases where the selected individual was incapacitated, a proxy could respond on their behalf.

The Computer Assisted Interviewing (CAI) method was utilized, with responses entered in an electronic questionnaire via notebook computers, primarily through telephone intervie

**Data Limitations:**

1. The patient survey was conducted on a sample, not on the population. Unknowingly there could be a sampling bias and sampling error.
2. Patient experience contains experience of patients of GP services and medical specialists, not all subsidised services that are part of the analysis.
3. There was a change in method of proxy interviews in 2023. Prior to 2022-23, proxy interviews were conducted for individuals aged 15 to 17 year and were not asked the below questions. But in 2022-23 all proxy interviews were not asked these questions.
   1. self-assessed health status
   2. experience with health professionals (listened to carefully, shown respect, enough time spent with person)
   3. whether waited longer than felt acceptable for a GP or medical specialist appointment
   4. whether would use telehealth again.
4. Since the method of collection was via telephone interviews, there is risk of data being entered incorrectly.
5. The tone, inflection, or mannerisms of the interviewer could have influenced the responses, which could have introduced bias into the data.
6. Survey answers reflect a population at a specific point in time. It does not consider any changes between the answers and the release of the results.

**Relevancy:** The data is an external source from Australian Bureau of Statistics (ABS) that provides data to other government agency and hence its considered a reliable source

The patient experience dataset will help analyse the perceived affordability of healthcare across different genders, ages, and geographic locations year on year.

Even though the dataset is a sample, it will be used for the analysis.

## Data Wrangling & Profile

### Census Datasets

**Rows**: 4,296

**Columns**: 27

**Date Range**: 2011-2022

#### Pre-processed & Combined Datasets

1. Created 3 datasets of 2011, 2016 and 2021:
   1. Combined census 2011 population by age, sex and personal income datasets.
   2. Combined census 2016 population by age, sex and personal income datasets.
   3. Combined census 2021 population by age, sex and personal income datasets.
2. Standardized 2011, 2016 and 2021 census datasets to ensure column consistency.
   1. In 2011, 2016, 2011 census datasets, *Year* column with corresponding year value was added.
   2. In 2011 census, SA3 area codes were updated to the 2016 edition list using the mapping below provided by ABS.

|  |  |  |  |
| --- | --- | --- | --- |
| SA3\_CODE\_2011 | SA3\_NAME\_2011 | SA3\_CODE\_2016 | SA3\_NAME\_2016 |
| 10101 | Goulburn - Yass | 10105 | Goulburn - Mulwaree |
| 21702 | Warrnambool - Otway Ranges | 21703 | Colac - Corangamite |
| 30802 | Gladstone - Biloela | 30804 | Biloela |
| 31604 | Nambour - Pomona | 31607 | Nambour |
| 50801 | Esperance | 51101 | Esperance |
| 50802 | Gascoyne | 51102 | Gascoyne |
| 50803 | Goldfields | 51103 | Goldfields |
| 50804 | Kimberley | 51001 | Kimberley |
| 50805 | Mid West | 51104 | Mid West |
| 50806 | Pilbara | 51002 | East Pilbara |
| 80102 | Cotter - Namadgi | 80101 | Belconn |

* 1. Income columns were renamed as below to ensure there was consistency across all 3 census datasets.

|  |  |  |
| --- | --- | --- |
| **Income Brackets** | **Average Income** | **New Income Column Name** |
| $1-$199 ($1-$10,399) | 5200 | average\_income\_$5200 |
| $200-$299 ($10,400-$15,599) | 12999.5 | average\_income\_$13000 |
| $300-$399 ($15,600-$20,799) | 18199.5 | average\_income\_$18200 |
| $400-$599 ($20,800-$31,199) | 25999.5 | average\_income\_$26000 |
| $600-$799 ($31,200-$41,599) | 36399.5 | average\_income\_$36400 |
| $800-$999 ($41,600-$51,999) | 46799.5 | average\_income\_$46800 |
| $1,000-$1,249 ($52,000-$64,999) | 58499.5 | average\_income\_$58500 |
| $1,250-$1,499 ($65,000-$77,999) | 71499.5 | average\_income\_$71500 |
| $1,500-$1,999 ($78,000-$103,999) | 90999.5 | average\_income\_$91000 |
| $2,000-$2,999 ($104,000-$155,999) | 129999.5 | average\_income\_$130000 |
| $3,000-$3,499 ($156,000-$181,999) | 168999.5 | average\_income\_$169000 |
| $3,500 or more ($182,000 or more) |  | average\_income\_$200000+ |
| Negative Income |  | negative\_income |
| Not Stated |  | not\_stated |
| Not Applicable |  | not\_applicable |

* 1. In 2011 census, income brackets *$3,000-$3,499 ($156,000-$181,999*) and *$3,500 or more ($182,000 or more)* was added and assigned blank values.
  2. In 2016 census:
     1. Column *average\_income\_$26000* was derived by adding column $*400-$499 ($20,800-$25,999* and *$500-$649 ($26,000-$33,799)*, The original 2 columns were dropped.
     2. Column *average\_income\_$9100s* was derived by adding the values from *1,500-$1,749 ($78,000-$90,999)* and 1*,750-$1,999 ($91,000-$103,999)* The original 2 columns were dropped.
     3. Income bracket *$3,500 or more ($182,000 or more)* was added to census 2016 and assigned blank values.
  3. In 2021 census
     1. Column *average\_income\_$26000* was derived by adding column values *400-$499 ($20,800-$25,999)* and *500-$649 ($26,000-$33,799)*. The original columns were dropped.
     2. Column *average\_income\_$91000* was derived by adding the values from 1*,500-$1,749 ($78,000-$90,999)* and *1,750-$1,999 ($91,000-$103,999).* The original 2 columns were dropped.

1. Combined (vertically stacked) census 2011, 2016 and 2011 datasets to create 1 dataset.

#### Imputation of missing years and population

1. Add missing years 2012-2015, 2017-2020 and 2023 to the combined census datasets
2. Used linear interpolation to add the population data of each year.

#### Data Types & Formats

|  |  |  |  |
| --- | --- | --- | --- |
| **Column name** | **Type** | **Qualitative/Quantitative** | **Qualitative: Nominal/Ordinal  Quantitative: Discrete/Continuous** |
| SA3 | String | Qualitative | Nominal |
| Year | Date Time | Quantitative | Discrete |
| age\_0-14 | Numeric | Quantitative | Discrete |
| age\_15-24 | Numeric | Quantitative | Discrete |
| age\_25-44 | Numeric | Quantitative | Discrete |
| age\_45-64 | Numeric | Quantitative | Discrete |
| age\_65-79 | Numeric | Quantitative | Discrete |
| age\_80+ | Numeric | Quantitative | Discrete |
| negative\_income | Numeric | Quantitative | Discrete |
| no\_income | Numeric | Quantitative | Discrete |
| average\_income\_$5200 | Numeric | Quantitative | Discrete |
| average\_income\_$13000 | Numeric | Quantitative | Discrete |
| average\_income\_$18200 | Numeric | Quantitative | Discrete |
| average\_income\_$26000 | Numeric | Quantitative | Discrete |
| average\_income\_$36400 | Numeric | Quantitative | Discrete |
| average\_income\_$46800 | Numeric | Quantitative | Discrete |
| average\_income\_$58500 | Numeric | Quantitative | Discrete |
| average\_income\_$71500 | Numeric | Quantitative | Discrete |
| average\_income\_$91000 | Numeric | Quantitative | Discrete |
| average\_income\_$130000 | Numeric | Quantitative | Discrete |
| not\_stated | Numeric | Quantitative | Discrete |
| not\_applicable | Numeric | Quantitative | Discrete |
| male\_pop | Numeric | Quantitative | Discrete |
| female\_pop | Numeric | Quantitative | Discrete |
| total\_population | Numeric | Quantitative | Discrete |
| average\_income\_$200000+ | Numeric | Quantitative | Discrete |
| average\_income\_$169000 | Numeric | Quantitative | Discrete |

#### Statistical Summary

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Column Name | count | mean | std | min | 25% | 50% | 75% | max |
| Year | 4296 | 2017 | 3 | 2011 | 2014 | 2017 | 2019 | 2022 |
| age\_0-14 | 4256 | 12,387 | 9,446 | 0 | 6,433 | 9,839 | 16,432 | 74,874 |
| age\_15-24 | 4256 | 8,375 | 6,540 | 0 | 3,934 | 6,647 | 11,319 | 40,357 |
| age\_25-44 | 4256 | 18,560 | 15,213 | 0 | 8,199 | 14,480 | 25,566 | 110,560 |
| age\_45-64 | 4256 | 16,582 | 11,168 | 0 | 8,873 | 14,429 | 22,512 | 55,602 |
| age\_65-79 | 4256 | 7,811 | 5,432 | 0 | 3,854 | 6,998 | 10,780 | 33,680 |
| age\_80+ | 4256 | 2,733 | 2,124 | 0 | 1,204 | 2,389 | 3,779 | 12,294 |
| negative\_income | 4256 | 326 | 280 | 0 | 150 | 260 | 411 | 2,204 |
| no\_income | 4256 | 4,496 | 4,098 | 0 | 1,861 | 3,226 | 5,780 | 26,787 |
| average\_income\_$5200 | 4256 | 2,483 | 1,957 | 0 | 1,162 | 2,013 | 3,310 | 13,276 |
| average\_income\_$13000 | 4256 | 3,786 | 2,953 | 0 | 1,770 | 3,132 | 5,072 | 23,608 |
| average\_income\_$18200 | 4256 | 4,585 | 3,336 | 0 | 2,119 | 4,018 | 6,364 | 20,746 |
| average\_income\_$26000 | 4256 | 7,809 | 5,599 | 0 | 3,646 | 6,818 | 10,568 | 29,570 |
| average\_income\_$36400 | 4256 | 4,291 | 3,091 | 0 | 2,050 | 3,709 | 5,859 | 16,934 |
| average\_income\_$46800 | 4256 | 4,413 | 3,258 | 0 | 2,115 | 3,806 | 5,804 | 20,376 |
| average\_income\_$58500 | 4256 | 4,582 | 3,447 | 0 | 2,144 | 3,830 | 6,314 | 22,909 |
| average\_income\_$71500 | 4256 | 3,271 | 2,546 | 0 | 1,436 | 2,645 | 4,569 | 16,555 |
| average\_income\_$91000 | 4256 | 4,627 | 3,785 | 0 | 1,910 | 3,657 | 6,421 | 25,960 |
| average\_income\_$130000 | 4256 | 3,366 | 3,249 | 0 | 1,082 | 2,365 | 4,613 | 26,688 |
| not\_stated | 4256 | 4,361 | 3,045 | 0 | 2,306 | 3,742 | 5,652 | 27,238 |
| not\_applicable | 4256 | 12,387 | 9,446 | 0 | 6,433 | 9,839 | 16,432 | 74,874 |
| male\_pop | 4256 | 32,800 | 23,315 | 0 | 17,246 | 27,508 | 44,115 | 149,962 |
| female\_pop | 4256 | 33,648 | 24,022 | 0 | 17,257 | 28,314 | 45,980 | 146,783 |
| total\_population | 4256 | 66,448 | 47,311 | 0 | 34,573 | 55,916 | 89,912 | 296,748 |
| average\_income\_$200000+ | 4216 | 1,015 | 1,732 | 0 | 121 | 436 | 1,143 | 17,716 |
| average\_income\_$169000 | 4256 | 1,041 | 1,503 | 0 | 188 | 560 | 1,276 | 15,561 |

**Observations**:

1. 8 SA3 geographical areas were introduced in 2016, so there is no population data between 2011 to 2015, resulting in 40 blank values per column.
2. average\_income\_$200000+ income bracket was introduced in 2021 census, hence there are blank value for years between 2011-2020.

### MBS Datasets

**Rows**: 258,535

**Columns**: 16

**Date Range**: 2014 – 2022

#### Pre-processed & Combined Datasets

1. Standardized and transformed 3 MBS datasets 2022-21, 2019-2021 and 2013-2019:
   1. Extracted SA3 geographical area dataset
   2. Populated blank State Territory values with ‘National-SA3Group’
   3. Populated blank Measure Values with n.p (not publish) as temporary filler
   4. Pivoted the dataset to have a unique row for each year, SA3 area, service and demographic.
   5. Renamed the columns for consistency across all datasets.
2. Combined the 3 MBS dataset to create 1 single dataset ranging from 2013-14 to 2021-2022
3. Cleaned the combined dataset as follows:
   1. Updated the Year values to be calendar years as per the below mapping.

|  |  |
| --- | --- |
| Original Year Value | New Year Value |
| 2021-22 | 2022 |
| 2020-21 | 2021 |
| 2019-20 | 2020 |
| 2018-19 | 2019 |
| 2017-18 | 2018 |
| 2016-17 | 2017 |
| 2015-16 | 2016 |
| 2014-15 | 2015 |
| 2013-14 | 2014 |

#### All measure of values had n.p or n.p. representing not published or missing values. These were set to blank. Measure of value types were converted to numeric from string.

* 1. *GeographicUnit* was dropped as it was not required for analysis.
  2. Gender values were standardized from *F* to *Females* and *M* to *Males*

#### Data Types & Formats

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Type** | **Qualitative/Quantitative** | **Qualitative: Nominal/Ordinal  Quantitative: Discrete/Continuous** |
| Year | Date Time | Quantitative | Discrete |
| StateTerritory | String | Qualitative | Nominal |
| GeographicCode | String | Qualitative | Nominal |
| GeographicAreaName | String | Qualitative | Nominal |
| GeographicGroup | String | Qualitative | Nominal |
| ServiceLevel | String | Qualitative | Nominal |
| Service | String | Qualitative | Nominal |
| DemographicGroup | String | Qualitative | Nominal |
| Medicare benefits per 100 people ($) | Numeric | Quantitative | Continuous |
| No. of patients | Numeric | Quantitative | Discrete |
| No. of services | Numeric | Quantitative | Discrete |
| Percentage of people who had the service (%) | Numeric | Quantitative | Discrete |
| Services per 100 people | Numeric | Quantitative | Discrete |
| Total Medicare benefits paid ($) | Numeric | Quantitative | Continuous |
| Total provider fees ($) | Numeric | Quantitative | Continuous |
| Estimated resident population | Numeric | Quantitative | Discrete |

#### Summary Statistics

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Column Names | count | mean | median | std | min | 25% | 50% | 75% | max |
| Year | 258,535 | 2018 | 2018 | 3 | 2014 | 2016 | 2018 | 2020 | 2022 |
| Medicare benefits per 100 people ($) | 234,301 | $ 6,546.57 | $ 1,146.11 | $ 11,229.59 | - | $ 97.00 | $ 1,146.11 | $ 7,662.81 | $ 110,280.55 |
| No. of patients | 234,301 | 21,725 | 3,290 | 228,847 | - | 521 | 3,290 | 11,101 | 23,099,650 |
| No. of services | 234,301 | 93,416 | 7,721 | 1,353,283 | 0 | 1,063 | 7,721 | 30,794 | 188,694,030 |
| Percentage of people who had the service (%) | 234,301 | 23.3 | 8.9 | 28 | 0.0 | 1.1 | 9.0 | 35.7 | 100.0 |
| Services per 100 people | 234,301 | 104.5 | 19.4 | 208.2 | 0 | 2.2 | 19.4 | 91.1 | 2041.2 |
| Total Medicare benefits paid ($) | 234,301 | $5,642,180.7 | $ 537,654.0 | $69,630,716.0 | - | $47,439.0 | $ 537,654.0 | $ 2,486,454.0 | $ 9,082,284,366.0 |
| Total provider fees ($) | 234,301 | 6,530,087 | 621,628 | 78,485,830 | - | 53,897 | 621,628 | 3,009,783 | 10,005,618,036 |
| Estimated resident population | 258,505 | 126,692 | 43,450 | 835,321 | - | 20,080 | 43,450 | 77,425 | 25,697,298 |

**Observations**:

1. The data shows that the median measure of value is significantly lower than the mean across all categories. This suggests a rightward skew in the distribution. A possible reason for this skewness is the inclusion of aggregated totals for "All persons" alongside detailed breakdowns by gender, age groups, as well as overall service totals and their subcategories.
2. There is 10% of data missing for every measure value column. This is due to data being unavailable, suppressed or SA3 levels being introduced later.

### Patient Experience GP Services Datasets

**Rows**: 40

**Columns**: 15

**Date Range**: 2023-2019

#### Pre-processed & Combined Datasets

1. Updated the year values from financial to calendar year as per the mapping below:

|  |  |
| --- | --- |
| Original Year Value | New Year Value |
| 2022-23 | 2023 |
| 2021-22 | 2022 |
| 2020-21 | 2021 |
| 2019-20 | 2020 |
| 2018-19 | 2019 |

1. Removed # from numeric values.
2. Renamed the region columns by shortening the names for easy reference for the analysis.

#### Data Types & Formats

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Type** | **Qualitative/Quantitative** | **Qualitative: Nominal/Ordinal  Quantitative: Discrete/Continuous** |
| year | Date time | Quantitative | Discrete |
| survey option | String | Qualitative | Nominal |
| 15–24 | Numeric | Quantitative | Continuous |
| 25–34 | Numeric | Quantitative | Continuous |
| 35–44 | Numeric | Quantitative | Continuous |
| 45–54 | Numeric | Quantitative | Continuous |
| 55–64 | Numeric | Quantitative | Continuous |
| 65–74 | Numeric | Quantitative | Continuous |
| 75–84 | Numeric | Quantitative | Continuous |
| 85+ | Numeric | Quantitative | Continuous |
| males | Numeric | Quantitative | Continuous |
| females | Numeric | Quantitative | Continuous |
| major\_cities | Numeric | Quantitative | Continuous |
| inner\_regional | Numeric | Quantitative | Continuous |
| outer\_regional\_remote | Numeric | Quantitative | Continuous |

#### Summary Statistics

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Column Name** | **count** | **mean** | **std** | **min** | **25%** | 50% | 75% | max |
| year | 40 | 2021 | 1.4 | 2019 | 2020 | 2021 | 2022 | 2023 |
| 15–24 | 40 | 40.8 | 34.0 | 0.8 | 17.3 | 27.6 | 72.2 | 99 |
| 25–34 | 40 | 41.4 | 34.3 | 0.7 | 17.8 | 27.0 | 74.6 | 99.3 |
| 35–44 | 40 | 41.5 | 34.6 | 0.7 | 15.3 | 28.9 | 73.5 | 99.4 |
| 45–54 | 40 | 41.2 | 35.8 | 0.6 | 13.4 | 27.0 | 76.5 | 99.4 |
| 55–64 | 40 | 40.6 | 38.3 | 0.5 | 9.5 | 22.0 | 80.5 | 99.7 |
| 65–74 | 40 | 39.8 | 41.5 | 0.2 | 5.2 | 17.1 | 87.7 | 99.9 |
| 75–84 | 40 | 39.2 | 43.9 | 0.1 | 2.3 | 12.1 | 91.8 | 100 |
| 85+ | 40 | 39.2 | 43.7 | 0 | 1.9 | 12.5 | 90.4 | 99.9 |
| males | 40 | 40.0 | 38.2 | 0 | 4.8 | 20.9 | 79.1 | 100 |
| females | 40 | 40.8 | 38.4 | 0 | 7.5 | 24.5 | 84.3 | 99.6 |
| major\_cities | 40 | 40.7 | 36.9 | 0.6 | 13.2 | 22.5 | 79.1 | 99.4 |
| inner\_regional | 40 | 41.0 | 36.2 | 0.9 | 13.4 | 24.7 | 77.5 | 99.2 |
| outer\_regional\_remote | 40 | 40.9 | 35.8 | 0.9 | 14.7 | 25.2 | 79.1 | 99 |

**Observations**:

1. Summary statistics shows there are no values above 100 or below 0 as expected.
2. There are no missing values
3. The distribution of data cannot be interpreted as these represent percentage of respondents to specific answers.
4. Since the # was removed for analysis, values with high margin of error will be part of the analysis.

### Patient Experience GP After Hour Service Datasets

**Rows**: 40

**Columns**: 15

**Date Range**: 2023-2019

#### Pre-processed & Combined Datasets

1. Updated the year values from financial to calendar year as per the mapping below:

|  |  |
| --- | --- |
| Original Year Value | New Year Value |
| 2022-23 | 2023 |
| 2021-22 | 2022 |
| 2020-21 | 2021 |
| 2019-20 | 2020 |
| 2018-19 | 2019 |

1. Removed # from numeric values.
2. Renamed the region columns by shortening the names for easy reference for the analysis.

#### Data Types & Formats

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Type** | **Qualitative/Quantitative** | **Qualitative: Nominal/Ordinal  Quantitative: Discrete/Continuous** |
| year | Date time | Quantitative | Discrete |
| survey option | String | Qualitative | Nominal |
| 15–24 | Numeric | Quantitative | Continuous |
| 25–34 | Numeric | Quantitative | Continuous |
| 35–44 | Numeric | Quantitative | Continuous |
| 45–54 | Numeric | Quantitative | Continuous |
| 55–64 | Numeric | Quantitative | Continuous |
| 65–74 | Numeric | Quantitative | Continuous |
| 75–84 | Numeric | Quantitative | Continuous |
| 85+ | Numeric | Quantitative | Continuous |
| males | Numeric | Quantitative | Continuous |
| females | Numeric | Quantitative | Continuous |
| major\_cities | Numeric | Quantitative | Continuous |
| inner\_regional | Numeric | Quantitative | Continuous |
| outer\_regional\_remote | Numeric | Quantitative | Continuous |

#### Summary Statistics

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Column Name** | **count** | **mean** | **std** | **min** | **25%** | 50% | 75% | max |
| year | 40 | 2021 | 1.4 | 2019 | 2020 | 2021 | 2022 | 2023 |
| males | 40 | 41.5 | 32.4 | 1.2 | 15.2 | 29.5 | 73.5 | 95.0 |
| females | 40 | 42.1 | 29.5 | 2.2 | 17.3 | 37.3 | 69.1 | 92.1 |
| 15–24 | 40 | 41.7 | 31.1 | 1.3 | 14.0 | 33.8 | 69.5 | 93.4 |
| 25–34 | 40 | 41.7 | 30.7 | 0.6 | 14.1 | 34.3 | 71.2 | 91.7 |
| 35–44 | 40 | 41.9 | 30.4 | 1.7 | 17.5 | 32.0 | 69.4 | 91.4 |
| 45–54 | 40 | 42.1 | 30.0 | 1.7 | 17.2 | 36.6 | 68.4 | 93.1 |
| 55–64 | 40 | 41.9 | 30.9 | 0.6 | 15.6 | 35.3 | 70.4 | 95.1 |
| 65–74 | 40 | 41.9 | 30.8 | 0.0 | 20.5 | 35.2 | 69.0 | 96.4 |
| 75–84 | 40 | 41.3 | 31.9 | 0.0 | 15.4 | 34.5 | 69.7 | 96.4 |
| 85+ | 40 | 42.6 | 31.8 | 0.0 | 13.5 | 39.7 | 68.0 | 95.8 |
| major\_cities | 40 | 41.4 | 32.1 | 1.4 | 14.2 | 30.2 | 74.3 | 93.2 |
| inner\_regional | 40 | 43.4 | 27.7 | 3.1 | 24.9 | 43.9 | 58.5 | 94.6 |
| outer\_regional\_remote | 40 | 43.8 | 27.5 | 1.2 | 28.8 | 45.6 | 56.0 | 94.3 |

**Observations**:

1. Summary statistics show there are no values below 0 or above 100. This is as expected since the data shows percentage of respondents of survey answers.
2. There are no missing values.
3. Since # values indicating high margin error was removed, the analysis will consider the values as normal. They are not treated differently.

## Questions:

1. What percentage of the population use MBS primary health care?
2. Have out of pocket costs of primary health care Medicare services increased in Australia?
3. How much is an individual paying out of pocket per year (by state, gender, age, SAL3)
4. What percentage is that of their overall earnings? Has that increased or decreased over time?
5. What demographic is impated the most by the cost of primary health care? Has this changed over time? Reduced or increased?
   1. Gender?
   2. Age?
   3. Geographic?
   4. Income segment?
6. Why are this demographic group impacted the most?

## Definitions:

**General practice:** A general practitioner (GP) is likely the first point of contact for personal health and is important in the coordination of care of patients and referral to other health care services. A GP cares for patients in a whole of person approach, in the context of their work, family and community.

**Primary health care nurses:** Primary health care nurses play a key role in keeping people healthy by providing proactive care and health promotion. They work in a range of settings, including community health, general practice, aged care and schools.

**Allied health**: The allied health sector represents a broad range of health professionals who are not doctors, dentists, nurses or midwives, and includes psychologists, optometrists and physiotherapists. Allied health professionals use evidence-based practices to prevent, diagnose and treat a range of conditions and illnesses.

## References

**Medicare information**: <https://www.betterhealth.vic.gov.au/health/servicesandsupport/understanding-medicare>

**Primary health care**: <https://www.aihw.gov.au/reports-data/health-welfare-services/primary-health-care/overview>

**Dataset Information**:

2011 Fact Sheet: <https://www.abs.gov.au/websitedbs/censushome.nsf/home/factsheets?opendocument&navpos=450>

2011 Collection Method Information**:** <https://www.abs.gov.au/ausstats/abs@.nsf/mf/2903.0>

2016 Collection Method Information: <https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2008.0~2016~Main%20Features~Collection%20operations~93>

Datasets Links for Medicare Subsidised Data

2021-2022 [: https://www.aihw.gov.au/reports/primary-health-care/medicare-subsidised-gp-allied-health-and-specialis/data](file:///Users/patel/Documents/CF-Data%20Anaylst%20Course/portfolio_projects/mbs_analysis/project_management/https/www.aihw.gov.au/reports/primary-health-care/medicare-subsidised-gp-allied-health-and-specialis/data)

2019-2021 : <https://www.aihw.gov.au/reports/primary-health-care/medicare-subsidised-health-local-areas-2021-22/data>

2013 – 2019 - <https://www.aihw.gov.au/reports/primary-health-care/medicare-subsidised-health-local-areas-2019/data>