PLIDA Training Course

Introduction to the Person-Level Integrated Data Asset (PLIDA):

Working with PLIDA to address policy-relevant questions for the Department of Health and Aged Care

Professor Rosemary Korda
Professor Nicholas Biddle
Dr Jennifer Welsh

Australian National University

Module A1: Introduction to PLIDA

Professor Rosemary Korda
Professor Nicholas Biddle
Dr Jennifer Welsh

Australian National University

Introducing the convenors

Jenny Welsh

Nicholas Biddle

Rosemary Korda

Introducing yourselves

- What is your name?
- What is your current role?
- What is your background using PLIDA or similar data?
- What are one or two analytical questions that you are interested in that might be answerable using PLIDA?

Introducing the course

- Course aim: To introduce APS staff to the Person-Level Linked Data Asset (PLIDA) and its role in public policy analysis
  - Focus on data and analysis of relevance to the Department of Health and Aged Care
- Target audience: Staff who have started working with, or intend to work with, PLIDA data to address policy-relevant questions
- Course structure: The course will comprise 12 modules, delivered over three intensive days
  - The eight ‘A’ modules focus on fundamental concepts when using PLIDA data to inform policy and practice
  - The four ‘B’ modules focus on working with specific datasets within PLIDA: Census, Death Registrations, Medical Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS)

Introducing Module A1

1. Overview of PLIDA and underlying datasets
2. How the data are linked
3. Data quality and data privacy
4. Examples of how these data are being used to answer policy-relevant research questions, including:
    1. Examining variation in uptake of COVID-19 vaccination by population characteristics to inform the roll-out [Key data: Census and AIR]
    2. Equity in out-of-pocket costs for Medicare-eligible out-of-hospital services [Key data: Census, PIT and SSRI; MBS and PBS]

1. Overview and datasets in PLIDA	2. Approaches to data linkage	3. Data quality and data privacy	4. Policy relevant research projects

1. Overview and datasets in PLIDA

Overview of PLIDA

- The Person Level Linked Data Asset (PLIDA) is a secure, person-based research data asset
- Partnership among government agencies
- Combines information on health, education, government payments, tax information, employment and population demographics to create a comprehensive picture of Australia over time
- Access to data managed with the 5 Safes Framework
- Designed to answer policy-relevant questions (cannot be used for compliance)
- Securely stored by the ABS, accessed via a secure environment and made available to researchers for approved purposes
- Can also be linked to Business Longitudinal Analysis Data Environment (BLADE)

Analytical content

Australian Bureau of Statistics

ABS **DATA SERVICES**

MMP Datasets - By Domains

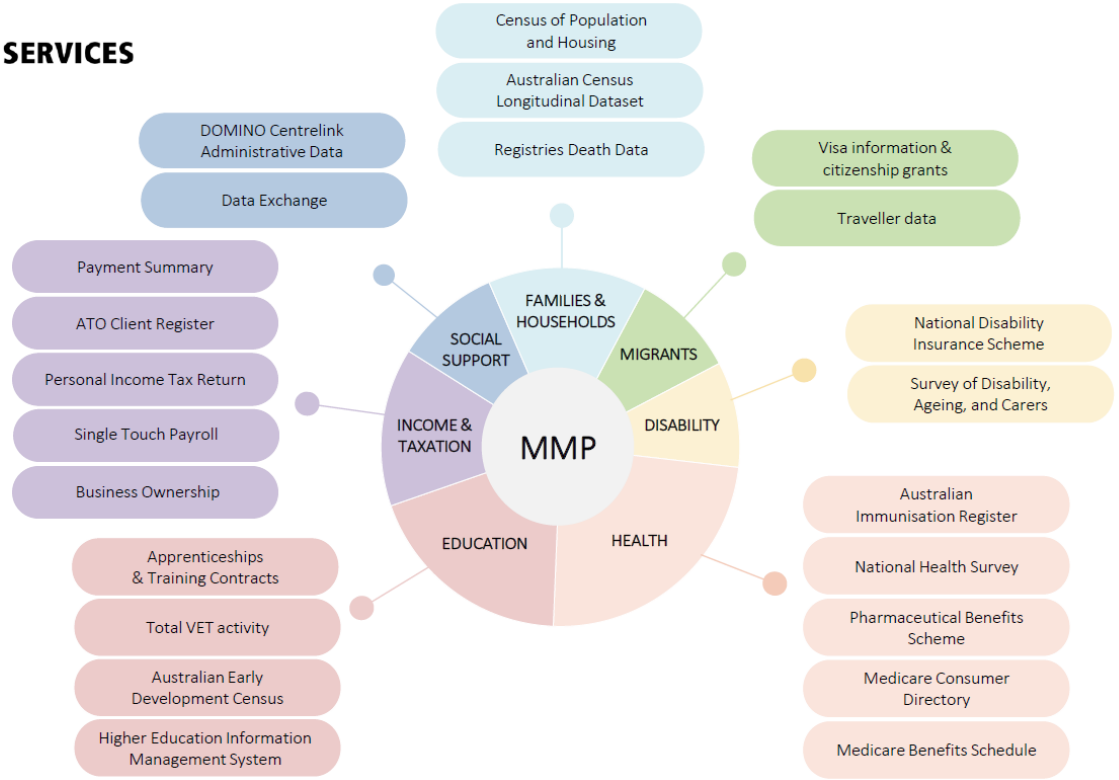


Chart from ABS User Information Pack, April 2023

1. Overview and datasets in PLIDA	2. Approaches to data linkage	3. Data quality and data privacy	4. Policy relevant research projects

2. Approaches to data linkage

Note: Thanks to Carrie Samuels, Director, Data Integration Assembly (ABS) who provided advice on this section

Person Linkage Spine (June 2022)

Spine contains 37.2 million elements

- The Person Linkage Spine is the base dataset to which all other person datasets are linked
- All datasets which are linked to the Spine are indirectly linked to each other in a secure and consistent way that protects the privacy of individuals
- Each element represents a single person resident in Australia at any point between 2006 – 2022
- Refreshed annually
- Further information in the 'PLIDA Data Availability and Access: User Information Pack' (April 2023)

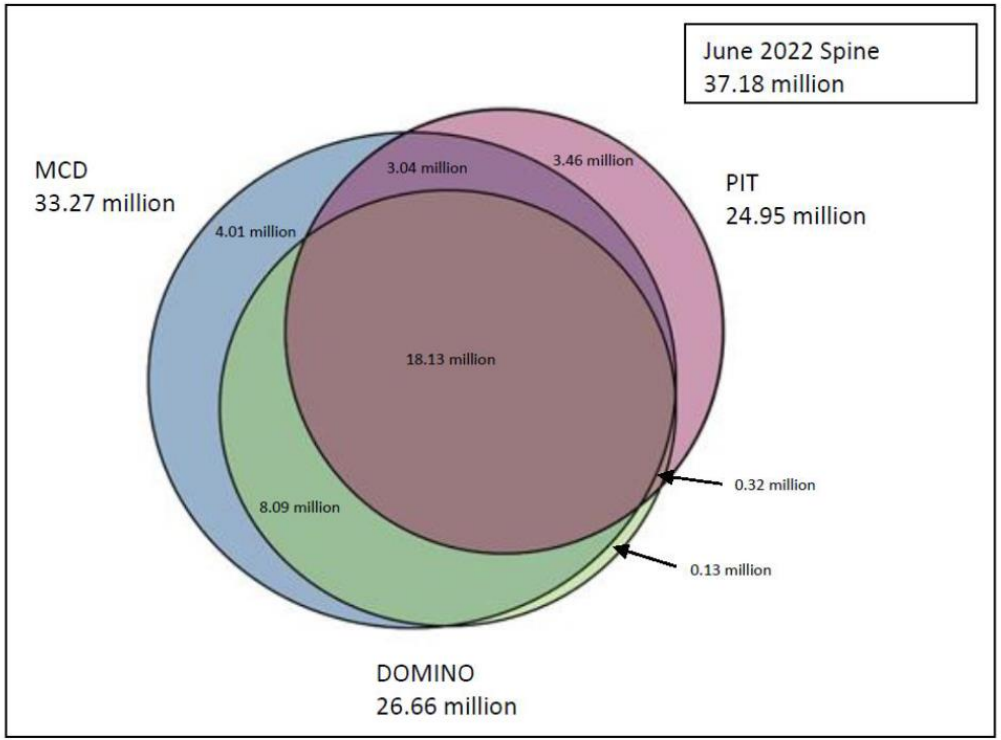


Chart from ABS User Information Pack, April 2023

Data linkage: Matching on key characteristics

- Criteria for matching variables: uniqueness, availability, accuracy and stability over time

If unique identifier variable is available, e.g. Medicare PIN, direct match can be made. Otherwise one or combination of matching methods can be used:

- Deterministic
- Probabilistic

Dataset A (e.g. Person Linkage Spine)

Record No	Name	DOB	Gender
A1	George Brown	01/02/1954	M
B2	Martha Smith	12/09/1991	F

Dataset B (e.g. Personal Income Tax)

Record No	Name	DOB	Gender
E5	James Grey	14/06/1946	M
F6	George Brown	01/02/1954	M
F7	George Brown	01/02/1954	M

Deterministic linkage (I)

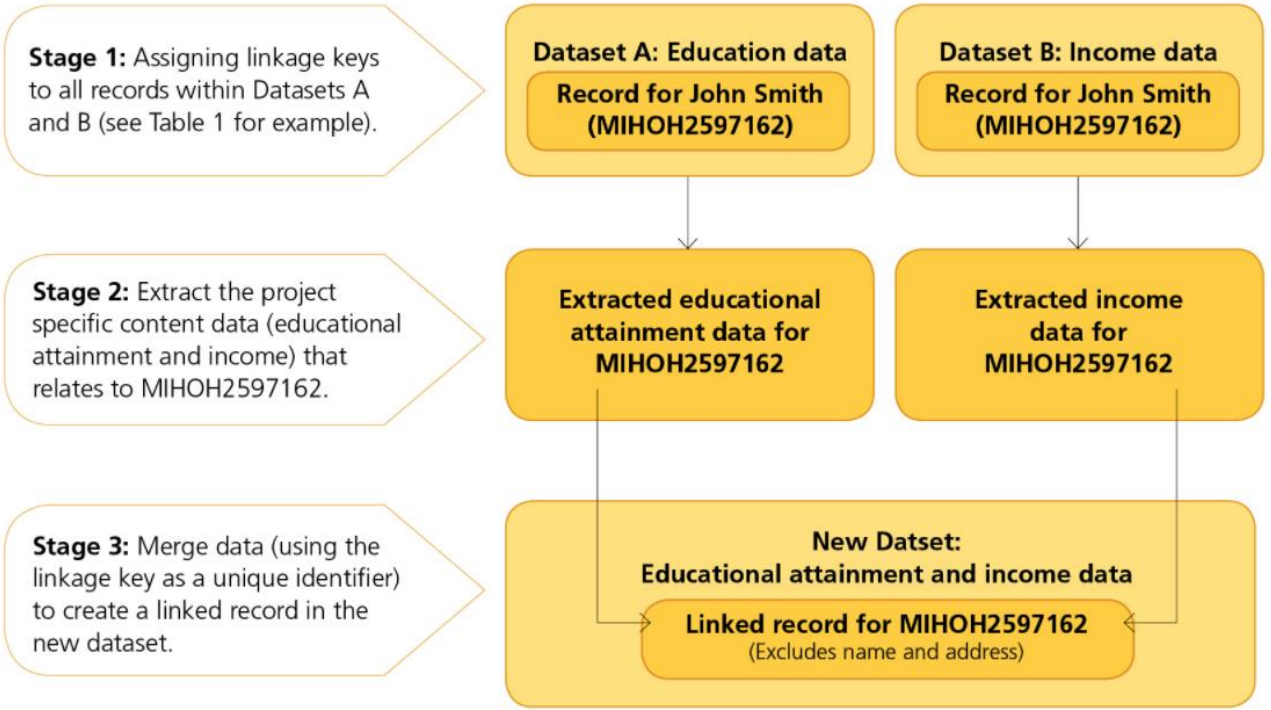
- This method currently used for PLIDA data linkage, where unique identifiers are not available
- All or nothing match
- The two datasets must agree exactly on every character of every matching variable to conclude that they correspond to the same entity
- But rules may be relaxed to allow for non-exact matches on some variables. Alternatively, different stages of matching can be used, slowly relaxing the number of variables that need to be exact matches

Name	Date of birth	Gender	Postcode
John Smith	12/05/1970	Male	5623
Linkage key = MIHOH2597162			

Deterministic linkage (II)

Name	Date of birth	Gender	Postcode
John Smith	12/05/1970	Male	5623

Linkage key = MIHOH2597162



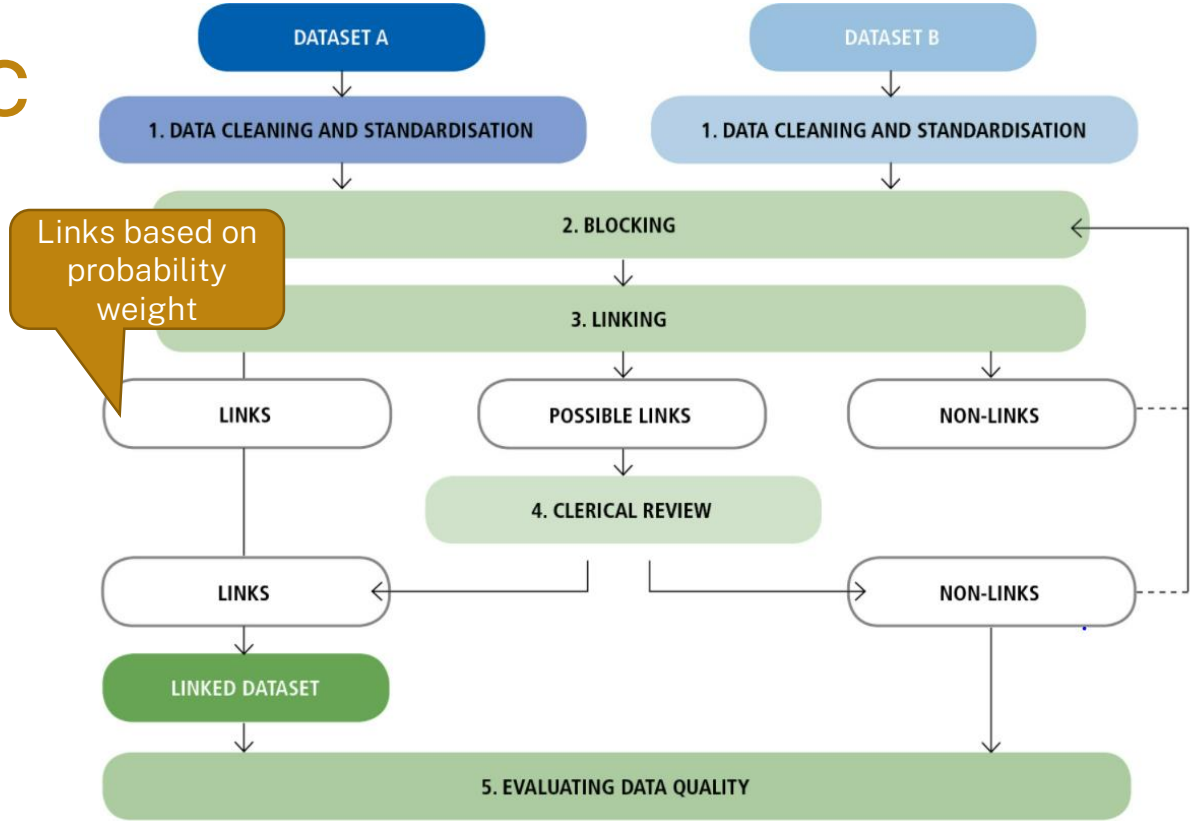
Source: https://toolkit.data.gov.au/data-integration/data-integration-projects/deterministic-linking.html#creating-a-linkage-key

Probabilistic linkage (I)

- Involves a number of steps, including calculating linkage weights for each pair of records
- The weights reflect the similarities of each pair of records (i.e. how likely they refer to the same entity)
- Cut-offs are then used to decide whether they are links or not
- In choosing cut-offs, trade-offs between missing true matches (false non-matches) and including too many false matches

Source: https://www.ausstats.abs.gov.au

Probabilistic linkage (II)



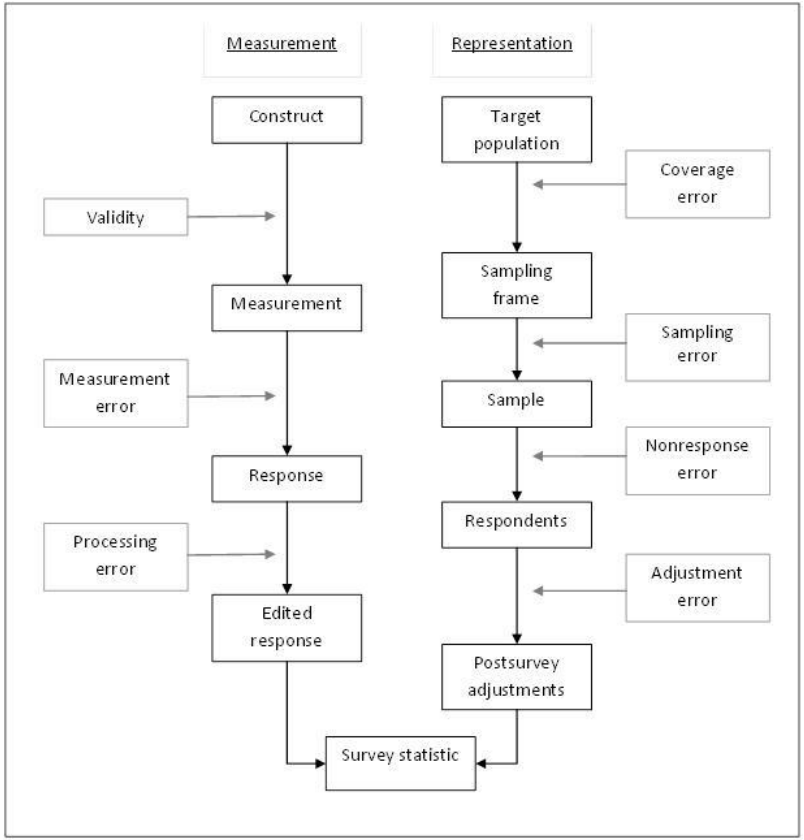
Source: https://toolkit.data.gov.au/data-integration/data-integration-projects/probabilistic-linking.html

1. Overview and datasets in PLIDA	2. Approaches to data linkage	3. Data quality and data privacy	4. Policy relevant research projects

3. Data quality and data privacy

The Total Survey Error Framework

"Total Survey Error refers to the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data. A survey error is defined as the deviation of a survey response from its underlying true value."



Ref: Public Opinion Quarterly, Volume 74, Issue 5, Special Issue, 2010

Quality frameworks for administrative and linked data

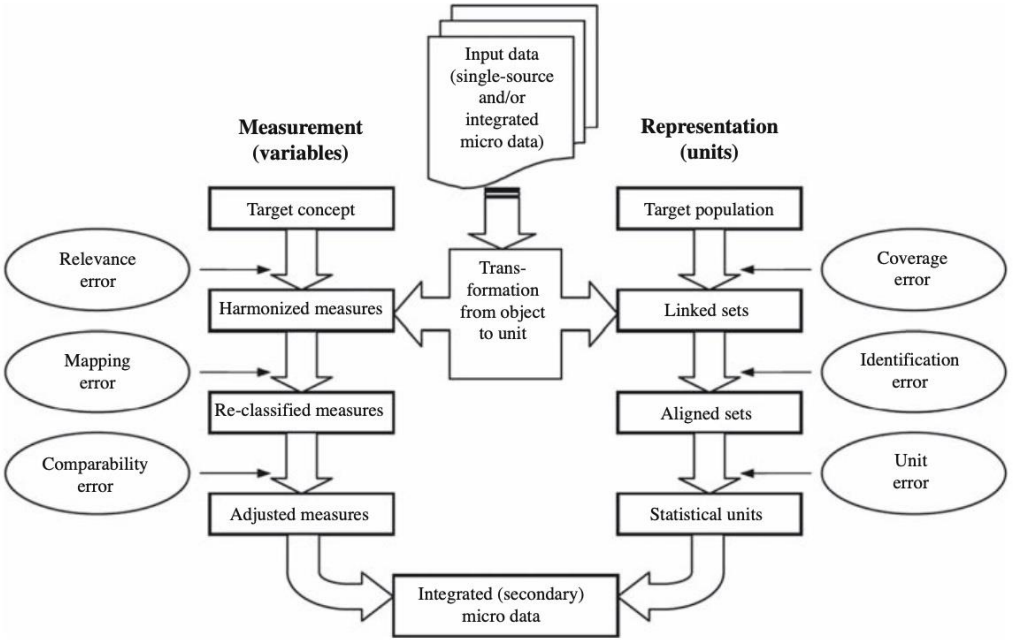


Fig. 2. Phase 2 of the quality assessment framework (Zhang 2012)

Reid, Zabala, and Holmberg (2017) 'Extending TSE to Administrative Data: A Quality Framework and Case Studies from Stats NZ'

- Phase 1 - how well a dataset meets its original, intended purpose
- Phase 2 - problems that can arise when integrating datasets from different sources
- Phase 3 - estimation, design, and evaluation

Data linkage quality

		True match status	
		Match	Non-match
Observed link status	Link	True match ✓	False link ✗
	Non-link	Missed links ✗	True non-match ✓

Data linkage errors

- Linkage error (missed & false links) results in missing data that can produce:
  - **Selection bias**: study sample that is not representative of target population
  - **Measurement error** and misclassification (exposure and/or outcome)

- Type of bias depends on how the linked data are used
- Linkage determines sample → selection bias
- Linkage defines exposure/outcome status → misclassification

- Extent of the bias depends on:

1. Linkage rates
2. Distribution of errors with respect to exposure and/or outcome

Privacy and the separation principle

- Record linkage staff doing the linkage use identifying variables (e.g. name, DOB), but do not have access to health information about the individuals
- Data custodians (e.g. state/territory health departments) only have access to data within their data collections
- Researchers receive data without identifying variables

The separation principle

No one working with the data can view both the linking (identifying) information (such as name, address, date of birth) together with the merged analysis (content) data (such as clinical information) in an integrated dataset

The 5 Safes approach

- Safe projects: Is this use of the data appropriate?
- Safe people: Can the users be trusted to use it in an appropriate manner?
- Safe settings: Does the access facility limit unauthorised use?
- Safe data: Is there a disclosure risk in the data itself?
- Safe outputs: Are the statistical results non-disclosive?

1. Overview and datasets in PLIDA	2. Approaches to data linkage	3. Data quality and data privacy	4. Policy relevant research projects

4. Policy-relevant research projects

Example 1. Using PLIDA to inform COVID-19 vaccine policies

AIR-MADIP project: Examining variation in uptake of vaccination by population characteristics to inform the roll-out

Acknowledgement: Work funded and supported by the Health Economics and Research Division in the Australian Department of Health and Aged Care

ANU Study team: Nicholas Biddle, Jennifer Welsh, Peter Butterworth, Ben Edwards, Rosemary Korda

See publications here:

- Socioeconomic determinants of vaccine uptake – July 2021 to January 2022 | Australian Government Department of Health and Aged Care
- Discretion in decision to receive COVID-19 vaccines and associated socio-economic inequalities in rates of uptake: a whole-of-population data linkage study from Australia - ScienceDirect

Sociodemographic variation in COVID-19 vaccinations

- As COVID-19 vaccination policy and immunisation program is rolled out, it is important to understand population groups where uptake is low and target communications accordingly
- Australian Immunisation Register (AIR) data linked to data sources within PLIDA
- Government interested in producing estimates of uptake in relation to a range of sociodemographic characteristics, including:

  - Age
  - Occupation
  - Indigenous status
  - Areas
  - People with health conditions
  - Socioeconomic position

Background

- The Australian COVID-19 vaccination program commended on 22 February 2021. Under the program, all people in Australia are eligible for vaccination (free of charge)
- As the rollout of the program continues, there is benefit in measuring and monitoring uptake in relation to sociodemographic and socioeconomic characteristics
  - Identify groups with low uptake → prioritise public health actions → increase population coverage

- This project examined socioeconomic (measured with education) variation in uptake of COVID-19 vaccines
  - Are there education-related gaps in uptake of 1st and 3rd dose?
  - If so, when did these gaps appear?
  - Have gaps narrowed or widened over time?
  - What are the current gaps?

Aims of this study were:

1. To examine how rates of uptake of 1st and 3rd dose varied in relation to level of education
2. To document current education-related gaps in these outcomes

PLIDA datasets

Australian Bureau of Statistics

ABS **DATA SERVICES**

MMP Datasets - By Domains

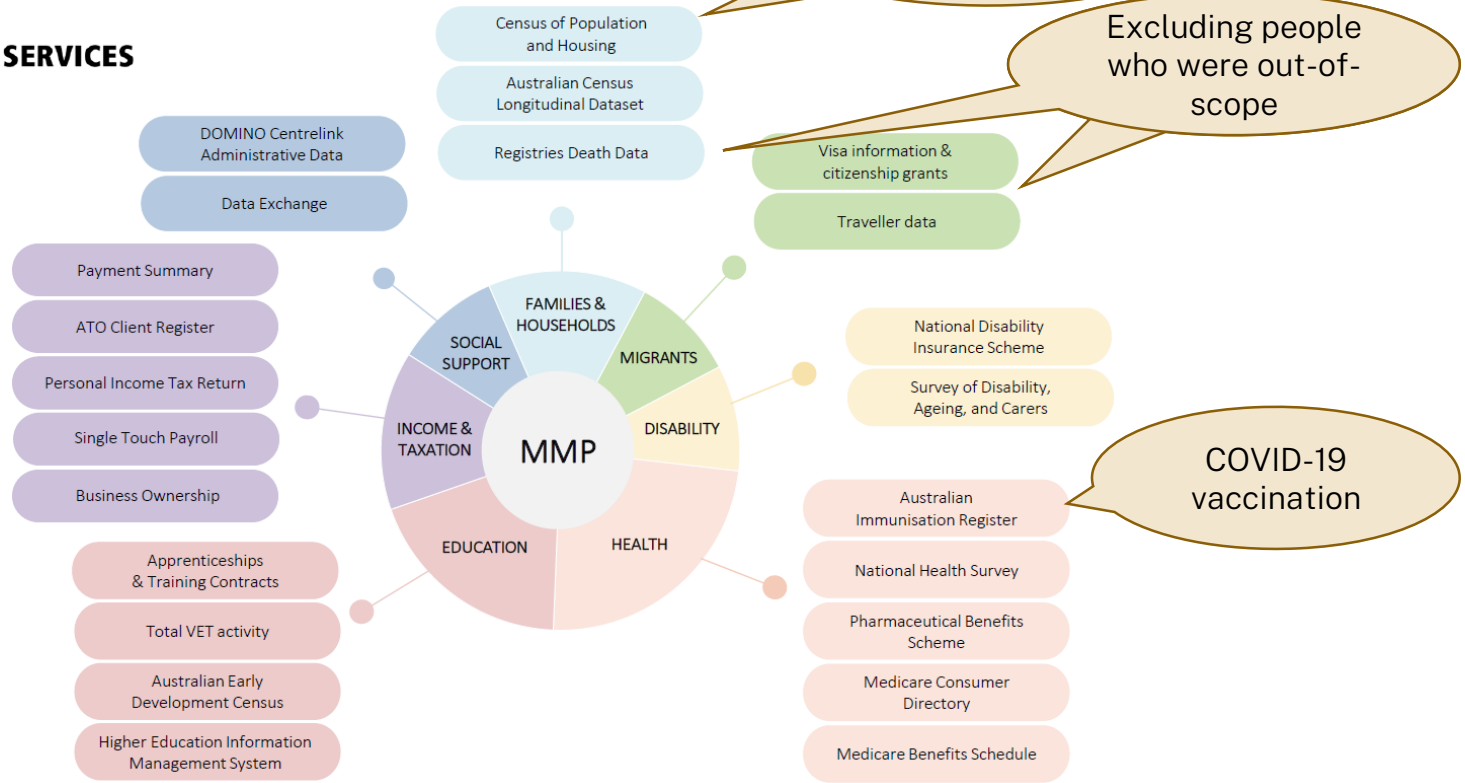


Chart from ABS User Information Pack, April 2023

Methods - data and study population

- Two outcomes, ascertained from AIR:
  1. 1st dose any COVID-19 vaccine
  2. 3rd dose of any COVID-19 vaccine
- Main exposure: education level (measured with Census 2016), comparing:
  1. Highest education: University degree
  2. Lowest education: Did not finish year 12, no other qualifications

  Other groups included in the analysis, but not shown here for brevity
- Study population:
  - 1st dose analysis: People aged 30-89 years (on 1 Jan 2021) with 2016 Census record linked to MADIP Spine, without death or outward migration record prior to 22 Feb 2021. Excluded people who had first dose <22 Feb 2021.
  - 3rd dose analysis: 1st dose pop (above), who had received two doses of COVID-19 vaccine. Excluded anyone who had third dose before 8 Nov 2021 (likely immunocompromised).

Methods – analysis strategy

1. Describe proportions who received 1st and 3rd dose every week for each 10-year age group and state
   - Accounted for deaths, outward migration, change in age groups at end of every week
   - Compared highest and lowest education groups
   - Report gaps as at 15 May 2022

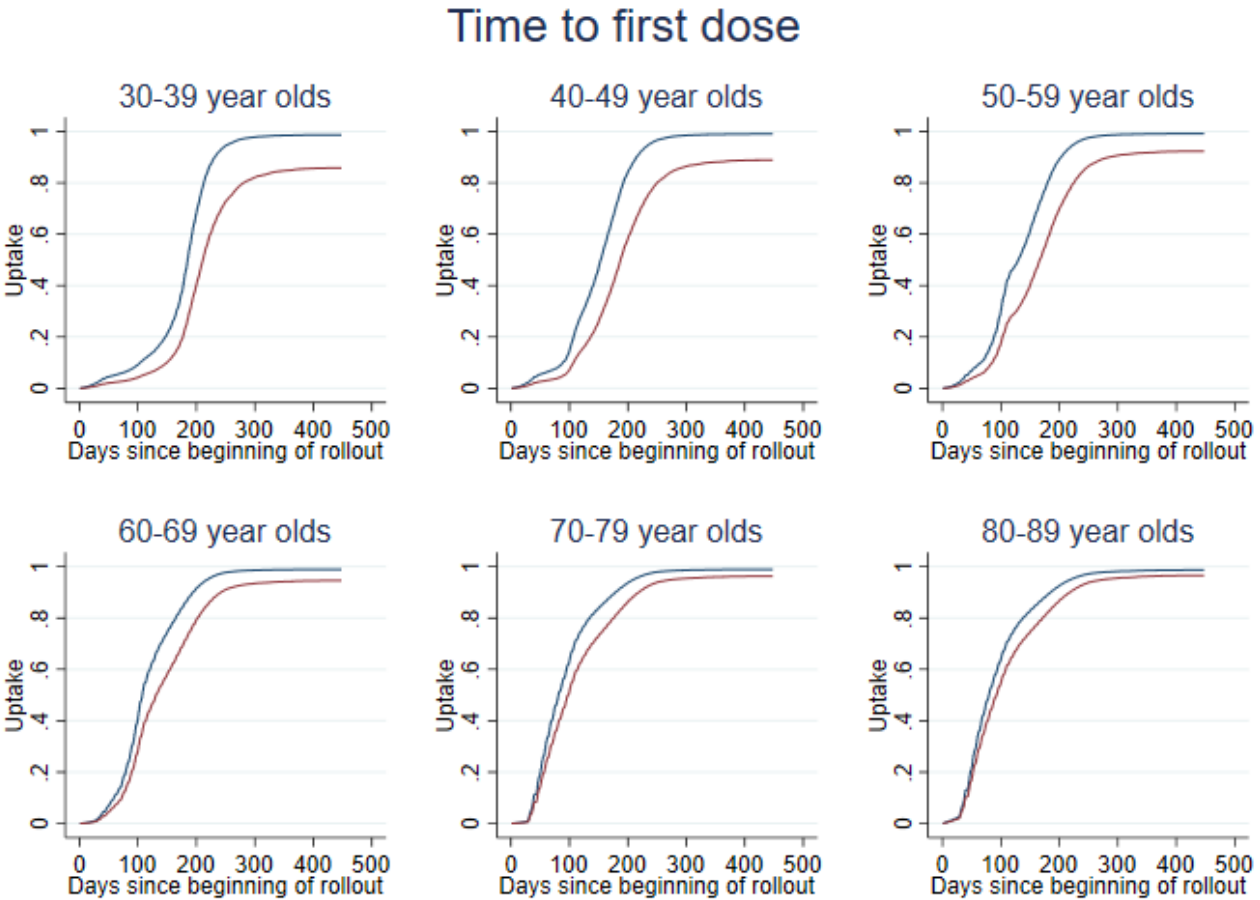
2. Use Cox regression estimating hazard ratios (HRs) for time to vaccination in relation to level of education, stratified by 10-year age group
   - Model 1: age- (single year) and sex- adjusted
   - Model 2: further adjusted for state (2021 geography), remoteness, country of birth, English language proficiency, carer, disability status, measured with Census

Note: HRs can be interpreted like a relative rate

Adjusted* proportions received 1st dose over time

High edu
Low edu

Day 0 = 22 Feb 2021
Day 447 = 15 May 2022



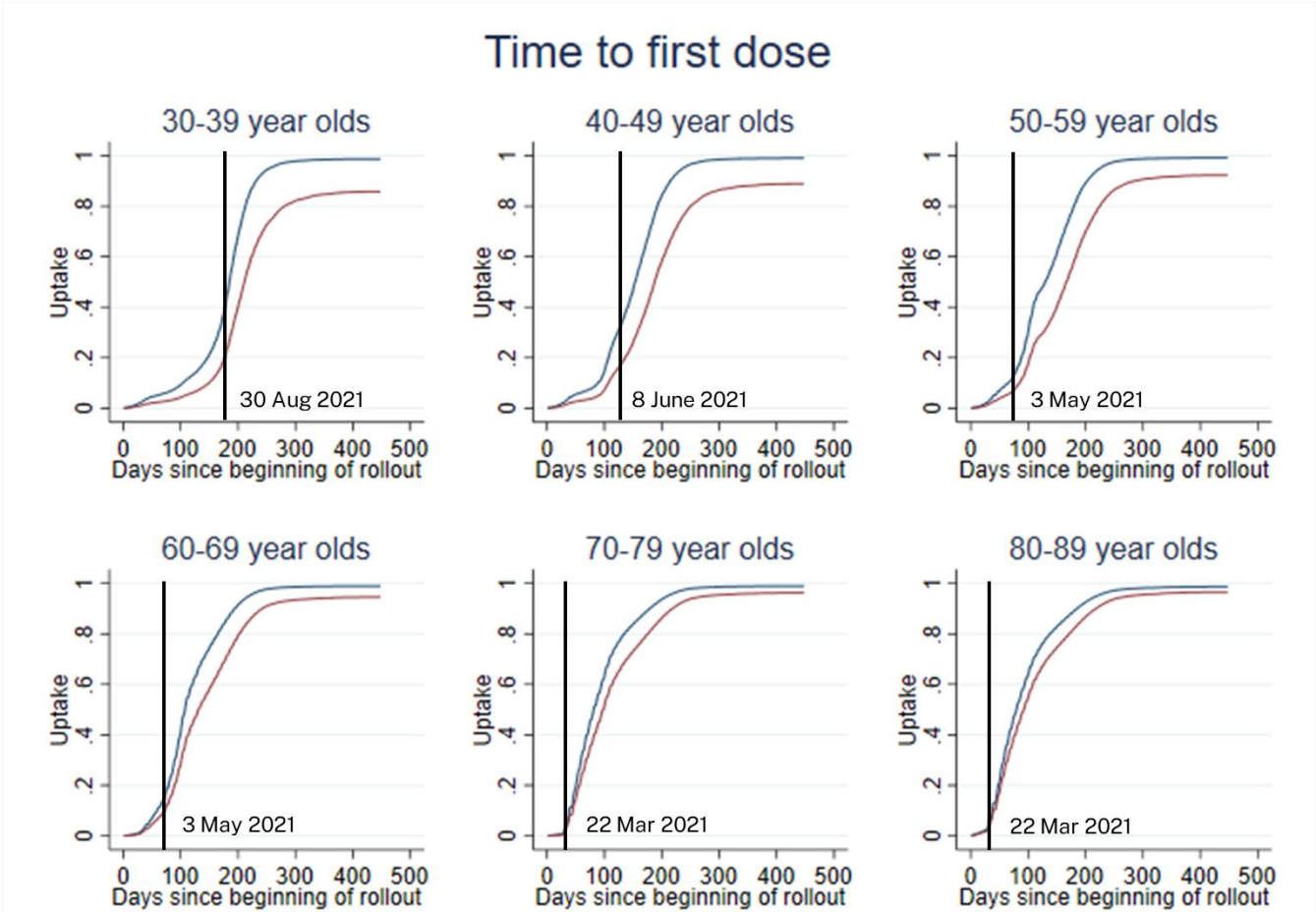
*Adjusted for age, sex, state, remoteness, country of birth, English proficiency, care giving status, disability status

Source: Welsh, J., Biddle, N., Butler, D. & Korda, RJ. (2023). Discretion in decision to receive COVID-19 vaccines and associated socioeconomic inequalities in rates of uptake: A whole-of-population data linkage study from Australia. *Public Health*, 224, 82-89, doi.org/10.1016/j.puhe.2023.08.020.

Adjusted* proportions received 1st dose over time

High edu
Low edu

Day 0 = 22 Feb 2021
Day 447 = 15 May 2022



*Adjusted for age, sex, state, remoteness, country of birth, English proficiency, care giving status, disability status. Note: Black line indicates date at which age group was eligible for all vaccines under Federal Government rollout program.

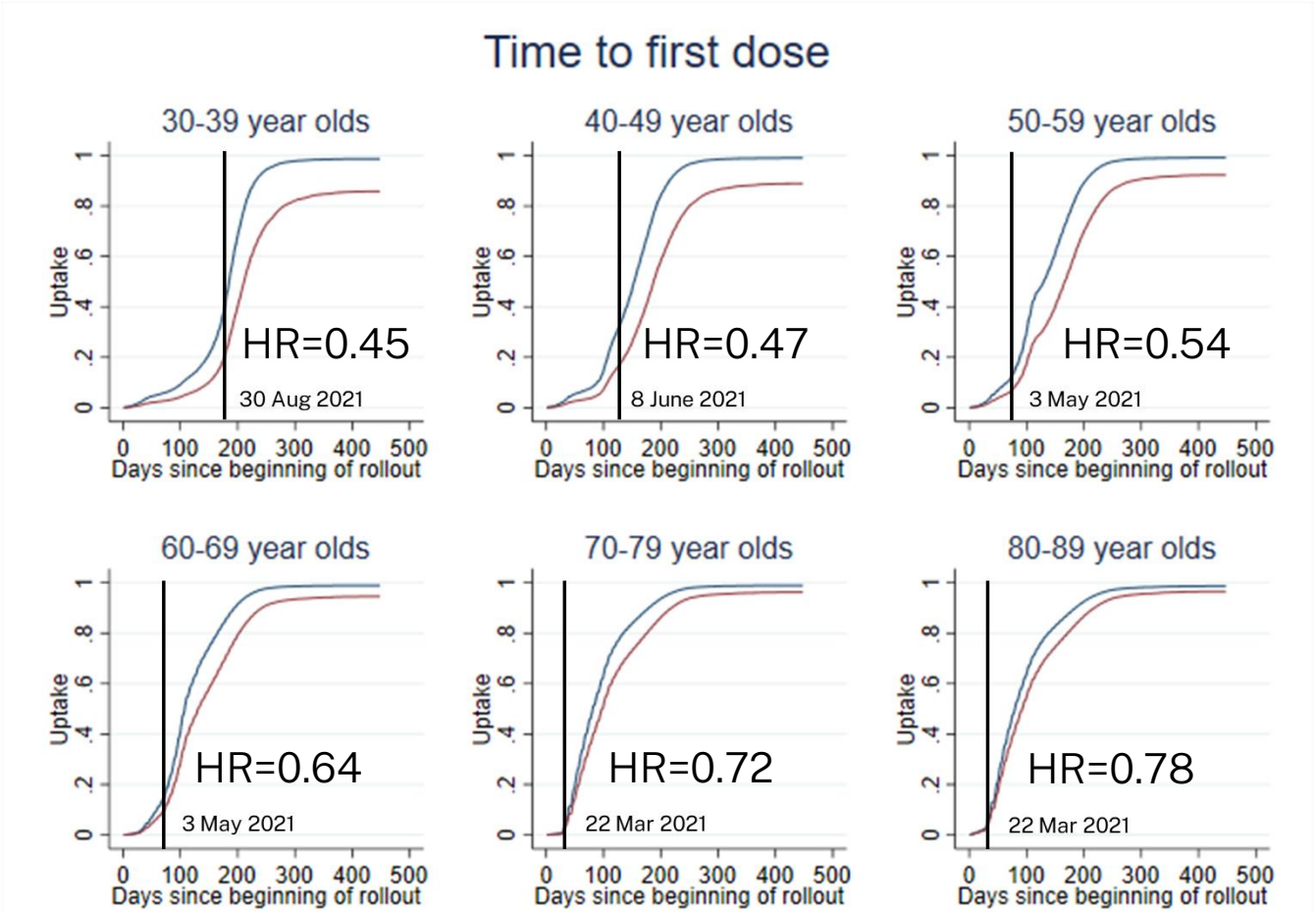
Source: Welsh, J., Biddle, N., Butler, D. & Korda, RJ. (2023). Discretion in decision to receive COVID-19 vaccines and associated socioeconomic inequalities in rates of uptake: A whole-of-population data linkage study from Australia. *Public Health*, 224, 82-89, doi.org/10.1016/j.puhe.2023.08.020.

Adjusted* proportions received 1st dose over time

High edu
Low edu

Day 0 = 22 Feb 2021
Day 447 = 15 May 2022

HR* compares rates of uptake in low compared to high education across entire time period



*Adjusted for age, sex, state, remoteness, country of birth, English proficiency, care giving status, disability status.
Note: Black line indicates date at which age group was eligible for all vaccines under Federal Government rollout program.

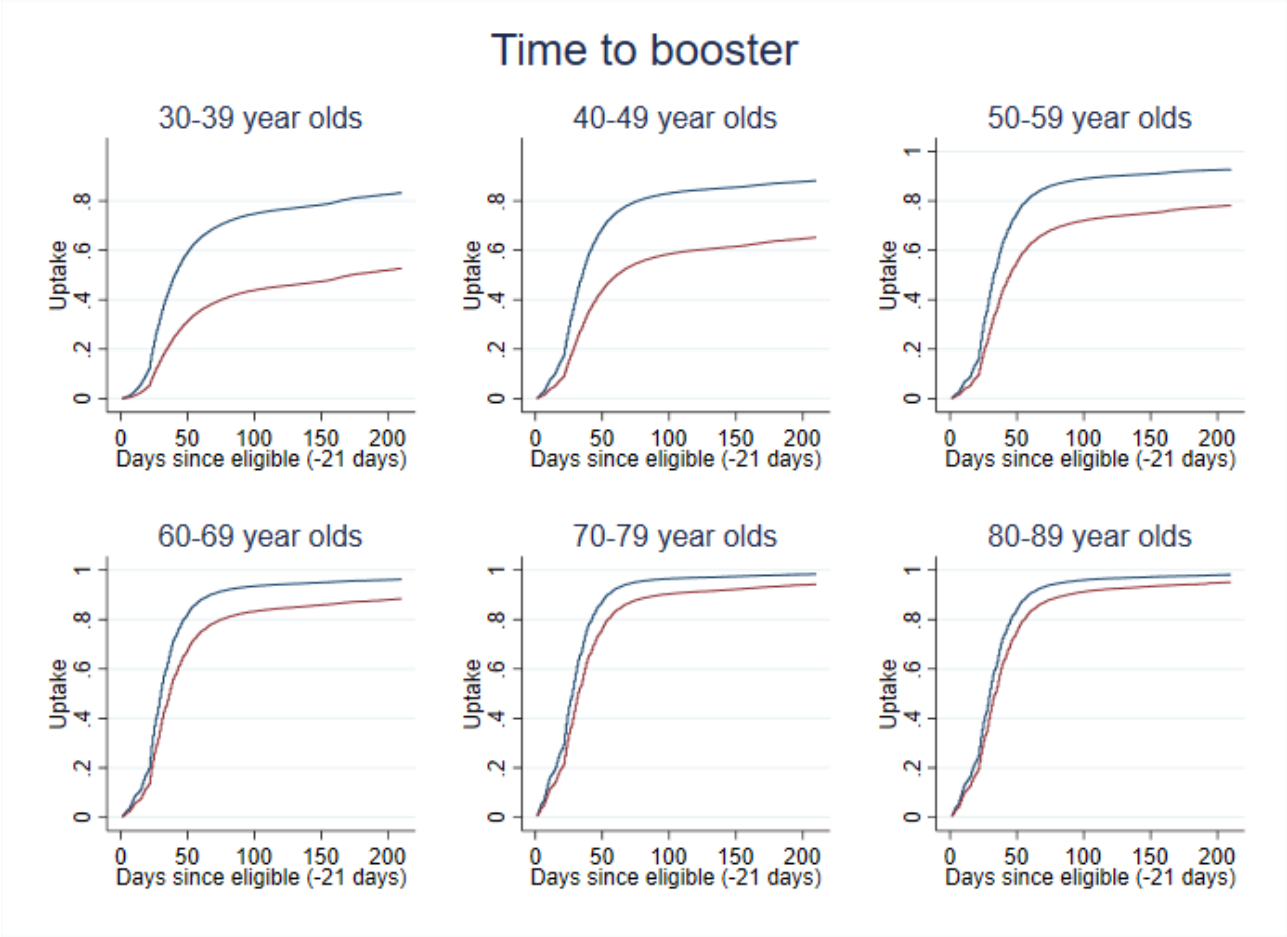
Source: Welsh, J., Biddle, N., Butler, D. & Korda, RJ. (2023). Discretion in decision to receive COVID-19 vaccines and associated socioeconomic inequalities in rates of uptake: A whole-of-population data linkage study from Australia. *Public Health*, 224, 82-89, doi.org/10.1016/j.puhe.2023.08.020.

Gaps in 1st dose
(at 15 May 2022)

	Unadjusted proportion received 1st dose		Gap
Age group	High education	Low education	% point
30-39 years	96.4	85.4	11.0
40-49 years	96.2	89.5	6.6
50-59 years	96.3	92.7	3.6
60-69 years	96.6	94.4	2.2
70-79 years	96.9	95.7	1.2

Adjusted* proportions received 3rd dose over time

High edu
Low edu

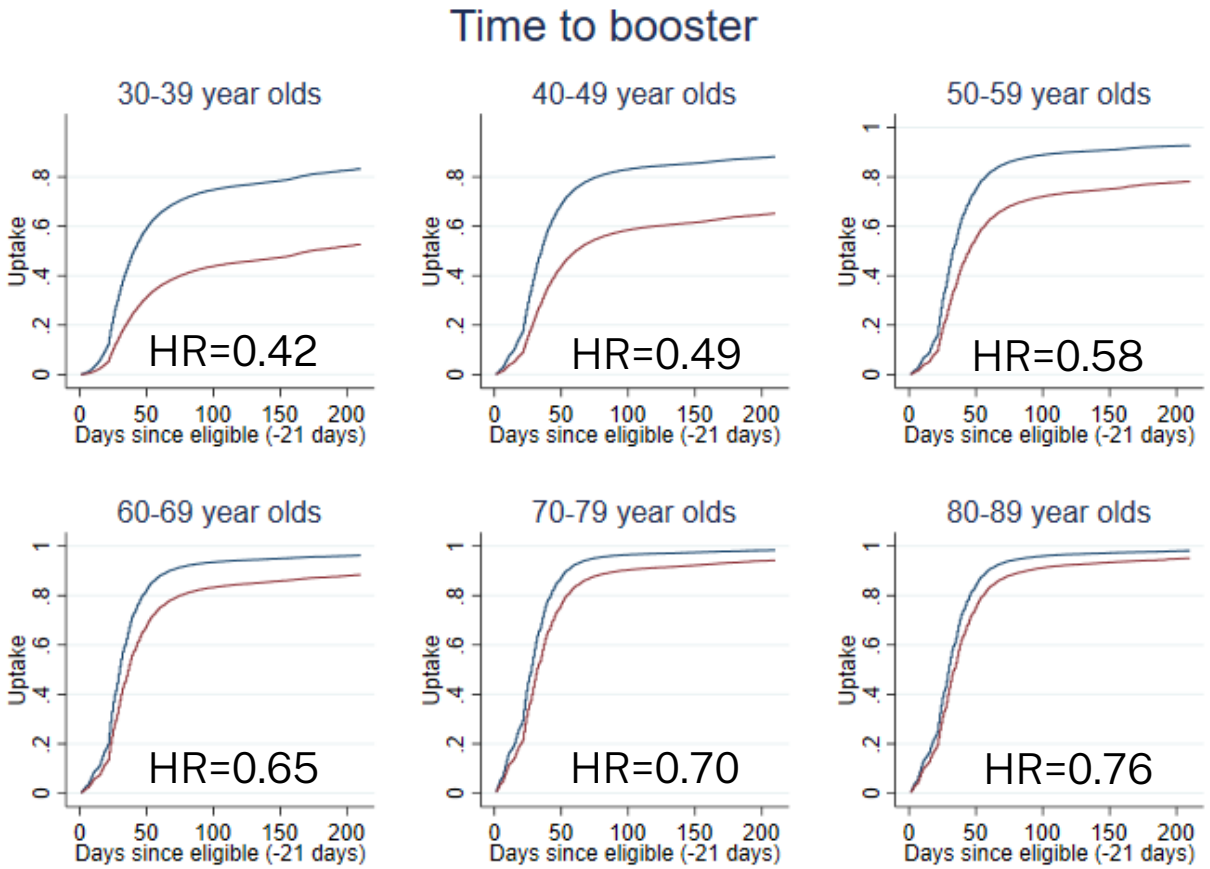


*Adjusted for age, sex, state, remoteness, country of birth, English proficiency, care giving status, disability status

Adjusted* proportions received 3rd dose over time

High edu
Low edu

HR* compares rates of uptake in low compared to high education across entire time period



*Adjusted for age, sex, state, remoteness, country of birth, English proficiency, care giving status, disability status

Gaps in 3rd dose
(at 15 May 2022)

	Unadjusted proportion received 3rd dose		Gap
Age group	High education	Low education	% point
30-39 years	75.0	42.1	32.9
40-49 years	82.2	57.0	25.2
50-59 years	87.8	71.8	16.0
60-69 years	92.0	82.7	9.4
70-79 years	95.0	90.1	4.9
80-89 years	95.1	90.8	4.3

Limitations

- Census 2016 variables were 4.5 years out of date at the time the study period began; covariates potentially out of date
- No information on COVID-19 infections, which had implications for timing of 3rd dose
- Relied on provisional deaths data to exclude people who had died
- Migration data to exclude those who permanently left the country has relatively low linkage to PLIDA Spine
- Used education as main exposure and a study population of people aged ≥30 years at time of rollout (25 years at Census 2016)

Summary

- Uptake of 1st dose much higher than 3rd dose, but with that in mind, there are similarities between the two doses
  - At older ages, high uptake overall and relatively smaller amounts of education–related variation
    - » Small gaps in absolute terms (<5 percentage points)
  - For middle age groups (30-69 years), quite substantial gaps in uptake according to education, differences emerged early in the rollout and persist.
    - » With each increasing 10 year age group, gap between high and low education group decreases
    - » Little evidence that gaps are narrowing. Patterns in uptake of 1st dose suggest that gaps in 3rd dose unlikely to close without additional policy responses

- The PLIDA data asset was an incredibly powerful tool for measuring and monitoring uptake of COVID-19 vaccines
  - Enables flexibility to measure uptake according to sociodemographic and socioeconomic characteristics, including education.
  - These data can were used to inform the rollout, i.e. identify groups with low uptake, examine factors associated with uptake and prioritise public health measures to increase uptake of COVID-19 vaccines.

1. Overview of PLIDA and underlying datasets	2. Data linkage	3. Data quality and data privacy	4. Examples of how these data are being used

4. Policy-relevant research projects

Example 2. Using PLIDA to inform policies on out-of-pocket costs

See full publication: Progressivity of out-of-pocket costs under Australia's universal health care system: A national linked data study - ScienceDirect

Project: Equity in of out-of-pocket costs under Medicare.
How are out-of-pocket costs distributed in relation to ability to pay?

NHMRC partnership grant, Grant #1134707. Whole-of-population linked data: strengthening the evidence to drive improvement in health and health care in Australia

Research team: Hsei Di Law, Dinith Marasinghe, Danielle Butler, Emily Lancsar, Emily Banks, Nicholas Biddle, Jennifer Welsh, Rosemary Korda

Partners: Australian Bureau of Statistics, Australian Institute of Health and Welfare, National Heart Foundation of Australia

Out-of-pocket health care costs a growing concern

The Sydney Morning Herald

Out of pocket or out of care: Medical costs pile pressure on Australians March 6, 2022

GP patients' out-of-pocket costs outstrip Medicare rebate as bulk-billing falls to near-decade low January 14, 2023

9NEWS

A growing number of Aussies are choosing between crucial medication and food amid cost of living crisis Feb 22, 2023

Medicare: MBS and PBS

- Affordability and equity – important principles of universal healthcare
- Out-of-pocket (OOP) costs often incurred

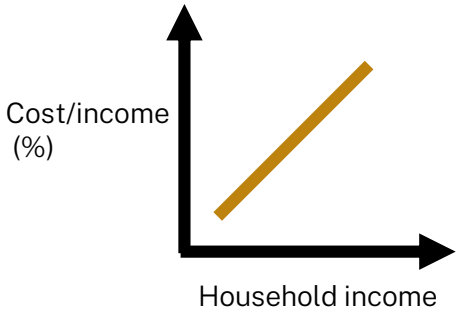
MBS	PBS
OOP cost = Fee charged – MBS benefit	OOP cost = Co-payment $42.50, or $6.80 if concession, or price of medicine if < co-payment*

- ➢ Measures to contain OOP costs
  - Bulk-billing (MBS)
  - Safety nets (MBS and PBS)
- ➢ Plus concessions to address equity

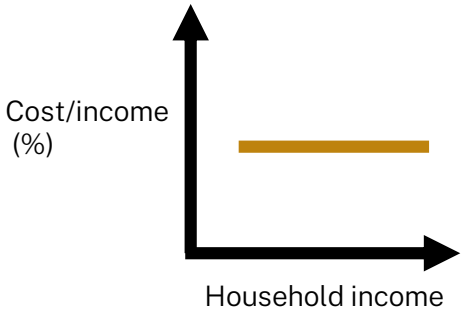
* Co-payments at time of the study

Research aim

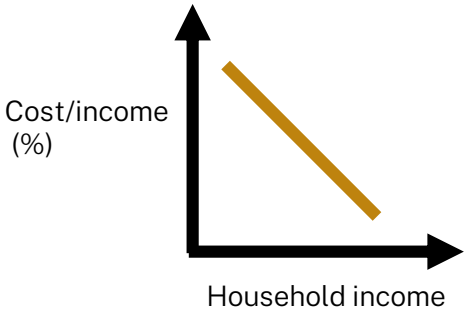
- To quantify the distribution of OOP costs according to disposable household income (“ability to pay”)



Progressive



Proportional



Regressive

Data

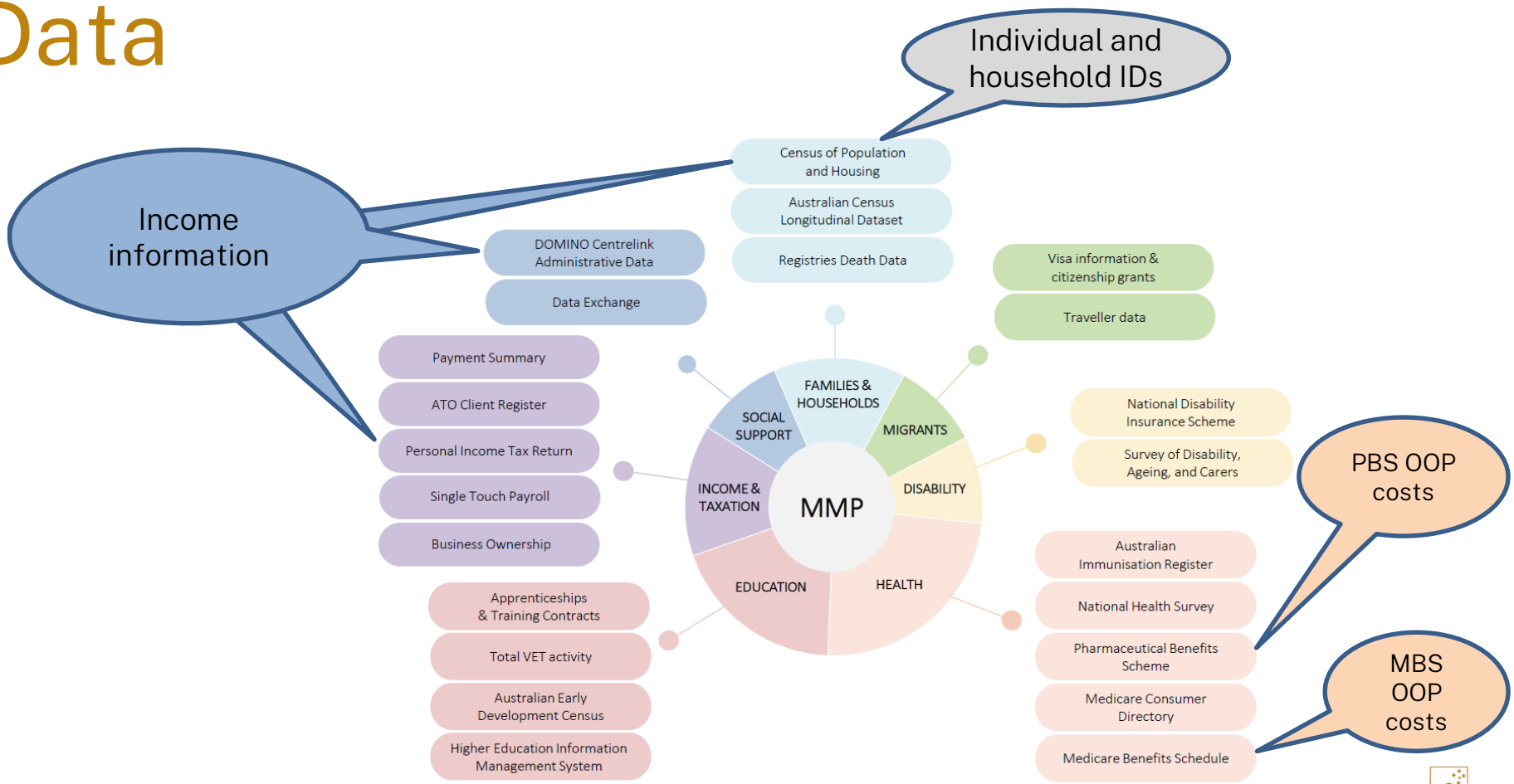
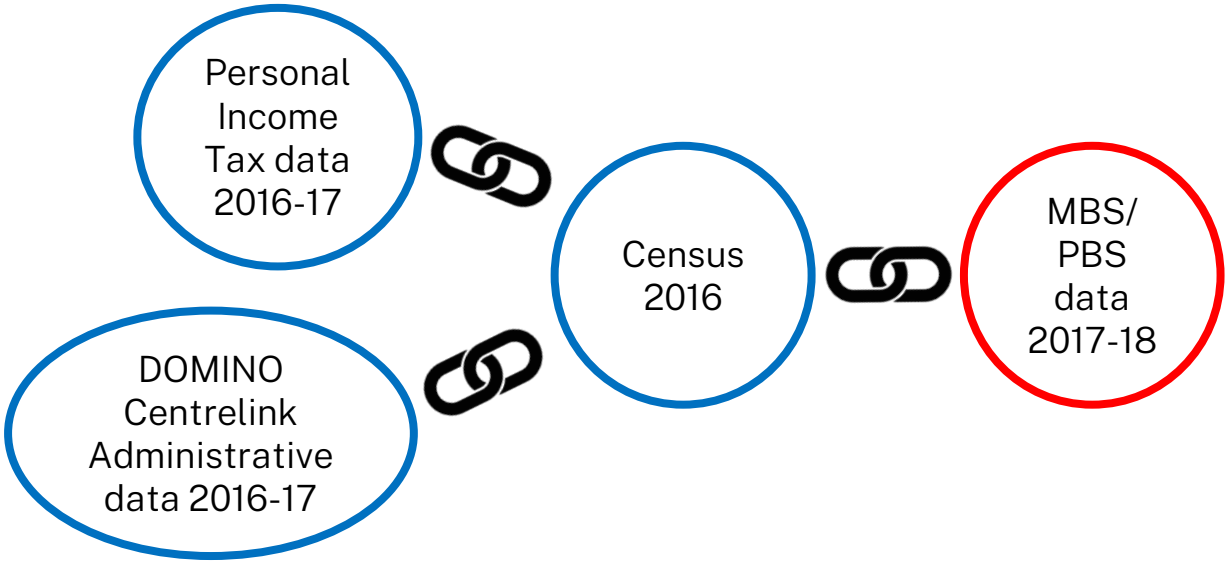


Chart from ABS User Information Pack, April 2023

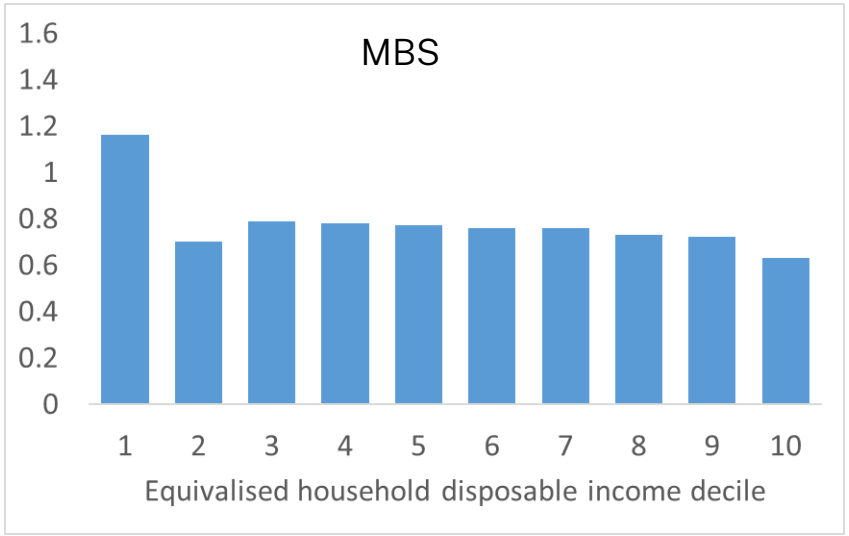
Data

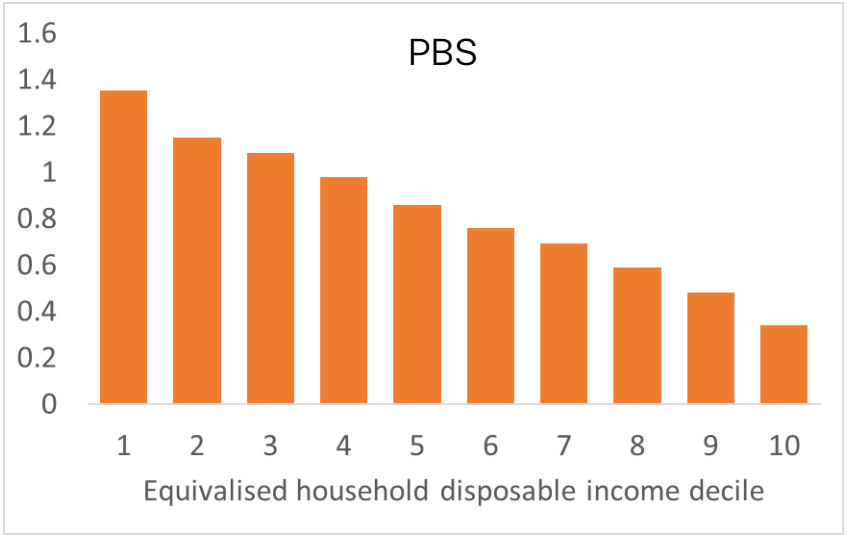


- Annual out-of-pocket costs MBS and PBS (A)
- Annual disposable income (B)
  (equivalised, i.e. adjusted for household size and composition)
- Outcome: Costs (A) / Income (B) (%)

Results

Annual OOP cost as a percentage of household income





Kakwani index MBS: -0.06

Kakwani index PBS: -0.24

Note: the range of K is -2 (most regressive) to 1 (most progressive)

Policy changes

Australian Government
Department of Health

Budget 2022–23

Improving access to medicines – Pharmaceutical Benefits Scheme Safety Net

The Australian Government is investing $525.3 million over four years to reduce the out of pocket costs for Australians who rely on Pharmaceutical Benefits Scheme (PBS) subsidised medicines, by reducing the eligibility threshold to meet the PBS Safety Net.

Ministers
Department of Health and Aged Care

Home | Media centre | Mark Butler | Anika Wells | Ged Kearney | Emma McBride | Malarndirri McCarthy

Home › The Hon Mark Butler MP › Minister Butler's media

3.2 million cheaper medicines in 2023

Australians are being encouraged to fill their prescriptions for medicines, especially as the maximum out of pocket cost for most medicines on the Pharmaceutical Benefits Scheme is now $12.50 lower.

The
Minister

Home › The Hon Mark Butler MP › Minister Butler's media

Cheaper medicines to ease cost of living

The Australian Government is easing cost of living pressures and making hundreds of common medicines cheaper, by allowing millions of Australians to buy two months' worth of medicine for the price of a single prescription.

The Hon Mark Butler MP
Minister for Health and Aged Care