

## Outline of process used to link MSOD to AHPRA data

### Background

MSOD survey data provides useful metrics on medical student graduates and their career intentions, but how those intentions subsequently unfold requires additional data. One such data source is provided by the Australian Health Practitioner Regulation Agency (Ahpra). The Ahpra public register of health practitioners contains details of all medical professionals currently registered to practice in Australia and includes practice location and medical specialty of all registrants. To create a linked dataset, each respondent's MSOD survey record must be 'matched' with the correct Ahpra registration record.

Data common to both datasets were used in the matching process:

- Last Name
- Middle Name
- First Name
- Gender
- Medical School
  - MSOD: Current Medical School
  - Ahpra: Awarding Institution of primary medical qualification
- Graduating Year
  - MSOD: year survey completed
  - Ahpra: year primary qualification obtained

### Automated matching

Fourteen sets of queries (steps) were applied to the datasets comparing these common fields, each getting progressively less specific. These were as follows:

Where...	Record marked as matched by comparison step...													
MSOD data = AHPRA data	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	10	11	20	21	30	31	40	41	50	51	60	61	70	71
Medical school = Awarding Institution	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Last Name = Last Name	✓	✓	✓	✓	✓	✓			✓	✓			✓	✓
First Name = First Name	✓	✓	✓	✓	✓	✓					✓	✓		
Gender = Gender	✓	✓	✓	✓										
Middle Name = Middle Name	✓	✓												
Year survey completed = Year qualification obtained	✓		✓		✓		✓		✓		✓		✓	
Year survey completed + 1 = Year qualification obtained		✓		✓		✓		✓		✓		✓		✓
Last Name = First Name							✓	✓						
First Name = Last Name							✓	✓						

Middle Name = First Name										✓	✓				
--------------------------	--	--	--	--	--	--	--	--	--	---	---	--	--	--	--

Note: The automated matching process only checks the data relevant to that step, meaning for example that a match could be made in step 9 where the medical school/awarding institution, last name, and year of qualification matches, where the MSOD middle name matches the AHPRA first name, however the MSOD First Name is different to the both the AHPRA First and Middle Names.

The process records the 'step' number against those records with a 1:1 match, so that the user of the data can be informed of the basis for that match.

### **Manual review of matched records**

Following these automated 14 steps, a manual review was conducted on:

- i) Records where an MSOD record was matched with multiple AHPRA records
- ii) Records where an AHPRA record was matched with multiple MSOD records

For these, the relevant records were displayed on-screen for review, and the correct match confirmed or all possible matches rejected.

An additional flag was recorded identifying these 'manually confirmed' matches.

### **Manual matching of unmatched records**

Unmatched MSOD records were then extracted to a separate file and a manual search of the AHPRA records was undertaken to enable further consideration as to whether there might be an associated AHPRA match.

This enabled additional matching of cases where;

- the Ahpra dataset lacked institution details for a qualification
- the Ahpra dataset lacked year of qualification details
- name spellings differed (e.g., Kandahapanny vs Kondanapanny; Smith vs Smiht)
- names were abbreviated (e.g., Rob vs Robert; Sam vs Samantha; Sarah vs Sarah-Jane)
- name order varied e.g., MSOD middle name = Ahpra surname
- surname differed and the respondent was female (i.e., name change due to marriage etc.)

An additional flag was recorded identifying these 'manually confirmed' matches.