MSOD Report: Spatial Mapping Medical Schools and Student Origins

Executive Summary
By Paul Konings, Ian McRae, Danielle Butler, Baldeep Kaur

The National Centre for Geographic & Resource Analysis in Primary Health Care (GRAPHC) at Australian Primary Health Care Research Institute was commissioned by Medical Deans to examine MSOD data from 2006 to 2010 and create an online mapping tool which would allow universities, policy makers and other stakeholders to reveal the regional source or footprint of medical school applicants and their intention to practice in urban/rural areas, analysed by area, university and background factors (such as rurality).

A collaboration between The Australian Primary Health Care Research Institute (APHCRI), The Medical Schools Outcomes Database (MSOD) and the Robert Graham Centre in the USA utilised geospatial analysis and mapping tools meet this commission.

The resulting interactive web based mapping tools offer the Medical Deans of Australia & New Zealand and invited associates accessible, visually engaging output to illustrate complex issues. These data, available via GRAPHC’s secure HealthLandscape Australia environment, may inform medical schools, community, potential students and policy makers about the nature of medical students progress from enrolment to practice, and may inform initiatives to address community objectives on workforce shortages.

The longer term aim of this project is to gather data and to monitor changes in 5yr cohorts of medical graduates to explore the influence of medical schools on workforce issues.

This report outlines the data available from the MSOD data base and how it was adapted for mapping purposes, a summary of the outcomes of some of the spatial analyses undertaken, and a summary of the modules now available in the web-based mapping facility; HealthLandscape Australia to explore the MSOD data. The report also discusses issues identified with data access and security, and recommended next steps for deployment of the mapping facility and for longer term development of the system.
# MSOD Report: Spatial Mapping Medical Schools and Student Origins

## Contents

**Executive Summary** ........................................................................................................... 1  
**Objective** ............................................................................................................................ 3  
**Background:** ....................................................................................................................... 3  
**Data:** .................................................................................................................................. 4  
  - Student Origin ..................................................................................................................... 4  
  - Spatial Reference ................................................................................................................. 4  
**Spatial Analyses:** ................................................................................................................. 5  
**Web Mapping Functionality:** ............................................................................................ 10  
  1. Medical School Categories ............................................................................................... 10  
  2. Medical School Catchment Postcodes .............................................................................. 12  
  3. Medical Students Data ...................................................................................................... 13  
  4. Med-School Catchment Radials ....................................................................................... 14  
**Next steps for consideration:** ........................................................................................... 15  
**Security** ............................................................................................................................. 15  
**Expansion of the data set** .................................................................................................. 15
MSOD Report: Spatial Mapping Medical Schools and Student Origins

By Paul Konings, Ian McRae, Danielle Butler, Baldeep Kaur

Objective: This report describes in summary the modules developed by the National Centre for Geographic & Resource Analysis in Primary Health Care (GRAPHC) within the Australian Primary Health Care Research Institute (APHCRI) and incorporated into the web-based mapping facility; HealthLandscapes Australia. This report also discusses issues identified with data access and recommended next steps.

Background: The Australian Primary Health Care Research Institute was commissioned by Medical Deans to examine MSOD data from 2006 to 2010 and create an online mapping tool which would allow universities, policy makers and other stakeholders to reveal the regional source or footprint of medical school applicants and their intention to practice in urban/rural areas, analysed by area, university and background factors (such as rurality).

The collaboration between The Australian Primary Health Care Research Institute (APHCRI), The Medical Schools Outcomes Database (MSOD) and the Robert Graham Centre in the USA is utilising geospatial analysis and mapping tools to help provide definitive answers on a national scale about how student origin and medical school culture affects subsequent workforce type and distribution.

The resulting interactive web based mapping tools offer the Medical Deans of Australia & New Zealand and invited associates accessible, visually engaging output to illustrate complex issues. As such they are a powerful tool in knowledge exchange and translation. These data, available via GRAPHC’s secure HealthLandscape environment, may inform medical schools, community, potential students and policy makers about initiatives to address community objectives on workforce shortages.

The longer term aim of this project is to gather data and to monitor changes in 5yr cohorts of medical graduates looking at influence of medical schools on workforce issues.

This report outlines:

- the data available from the MSOD data base and how it was adapted for mapping purposes;
- a summary of some of the spatial analyses undertaken, including charts of the postcodes which comprise the largest sources of students of medical schools;
- a summary of the modules now available in the web-based mapping facility; HealthLandscape Australia to explore the MSOD data.
- issues identified with data access and security, and
- recommended next steps for deployment of the mapping facility and for longer term development of the system.
Data:
Data on medical schools and medical student numbers is from:

MSOD database, Commencing Medical Students Questionnaire (CMSQ) and the Medical Schools Data (MSD) (see http://www.medicaldeans.org.au/medical-schools-outcomes-database). Fields include in the data provided to GRAPHC are:

- Commencing Year
- School at commencement
- Residence indicator
- Secondary School Name
- Secondary School Postcode
- Rural Background
- Preferred Region of future medical practice
- Preferred Type of future medical practice

Student Origin: Student origins are based on the nominated Australian secondary school/college/senior high school for the final year at school. Student origins are based on the physical location of the high school NOT the nominated postcode from the MSODS dataset.

The school names were linked to several address databases including data from Department of Education, Employment and Workplace Relations (DEEWR). The location of the schools was determined from these disparate sources and then geocoded to a latitude and longitude.

Spatial Reference: Base map geographies (e.g. boundaries of Postal Areas or of Statistical Local Areas) used are based on the Australian Bureau of Statistics (ABS) Australian Standard Geographical Classification (ASGC). The ASGC provides a common framework of statistical geography. In this case the medical student origin (high school) data derived from the MSOD was geocoded and classified according to the Postal Area Geographies (POA2006) which is an approximation of Australian postcode boundaries. (see http://www.abs.gov.au/ausstats/abs@.nsf/0/736A18F4C81595CA2578D40012CF51?opendocument).

Because the high school locations are known, the students are also attributed with a range of other standard Australian administrative geographies that can be used for alternate visualisations in the future, including:

- ASGC¹:
  - Meshblocks,
  - Census Collection Districts, (CCD)
  - Statistical Local Areas, (SLA)
  - Local Govt Area Urban Centres and Localities (LGA)
  - State Suburbs (SSC)
  - Remoteness Areas (RA)
  - Indigenous Regions (IREG, IARE, ILOC)
  - Postal Areas (POA)

¹ ASGC structures will be replaced by ASGS structures in ABS data from the 2011 Census, and future MSOD analyses will need to consider whether to also follow the ASGS structure.
• Other
  o Electoral Divisions (CED, SED)
  o Medicare Locals (ML)

The numbers of students in the MSOD data base and numbers whose origins were able to be mapped were:

• Total students included in MSOD database 2006-2010 n = 14,251
• Mapped data n = 10,621
• Not Mapped data: international origin students = 2,117,
• Data missing or not locatable to postcode = 1,513

Spatial Analyses:
The nature of the data presented with two spatial components: the Medical School has location and the High School has location. The students can be attached and linked to both spatial attributes.

There are 2507 Postal Areas in Australia. Medical Students come from 963. The top 20 postcodes are listed here:

[Graph showing the top 20 postcodes]

As the students are located at their high school rather than their residence, this Figure shows that Brisbane City has the largest cluster of high performing schools in terms of providing medical students, but cannot be interpreted in any other way.

There were 28 unique references to medical schools in the original data. It was determined that 23 were spatially and functionally different and the Extended Rural Cohort (ERC) and International Medical University (IMU) classifications should be treated as attributes of medical schools instead of as separate medical schools. The undergraduate entry and graduate entry “schools” at the same university were however considered separate entities (and were frequently in different locations)
The universities where ERC or IMU streams were incorporated in the main student group for the purposes of this analysis were:

- Monash University (Undergraduate) 1238
- Monash University (Undergraduate) ERC 27
- University of Melbourne (Graduate) 292
- University of Melbourne (Graduate) ERC 22
- University of Melbourne (Undergraduate) 674
- University of Melbourne (Undergraduate) ERC 19
- University of Melbourne (Undergraduate) IMU 28
- University of Newcastle 572
- University of Newcastle IMU 1

![Med School Admissions 2006 - 2010 Chart](chart.png)
The 1st step in assigning spatial attributes to the student data involved cleaning the High School values. This data was presented as entered by the student and there are 2087 unique High School Name Values. Many of these however were the same school with a different name – either due to spelling variations, some schools called school by one student and college by another, some with longer names having short form rather than long form provided etc.

Addresses for all schools in Australia were derived by combining a number of sources including but not limited to State Government departments, Department of Education Employment and Workplace Relations DEEWR, and a commercial database resellers. Schools reported by the students were then matched to this reference list, although given the variation between open text entries for the same school, considerable effort and some fuzzy logic was required to identify the correct schools. The reference datasets included physical addresses so once the MSOD data was paired to the appropriate school the addresses were geocoded. From an estimated 5,000 secondary schools in Australia, approximately 1,500 are linked to Medical Student enrolments.

Once geocoded, the high schools can be located within a variety of potentially useful geographies including POA and SLA. The spatially enhanced data was linked back to individual students (so that, for example, the student record now included the postcode in which they last went to high school). This means that students can be filtered according to a range of attributes other than those provided in the MSOD data.

The frequencies of medical students associated with each of the high schools and the defined areas and administrative boundaries was then calculated and joined back into the data. This enables us to symbolise areas (POAs) and points (high schools, or medical schools) according to the counts that have been spatially determined.

As the 23 Medical schools were also geo-located and incorporated into spatial layers, the resulting spatially attributed data could be visualised via:

- Medical Students origins (14251 points located at the high schools)
- High Schools’ attributes including Sector, Gender, Student population med student counts, by year
- Medical Schools attributes including a frequency counts.
Linking spatial attributes allows regional frequency analyses such as:

**Greater Capital City Statistical Areas:**

<table>
<thead>
<tr>
<th>Region</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater Capital City Areas</td>
<td></td>
</tr>
<tr>
<td>Rest of NT</td>
<td>17</td>
</tr>
<tr>
<td>Greater Darwin</td>
<td>61</td>
</tr>
<tr>
<td>Rest of Tas.</td>
<td>145</td>
</tr>
<tr>
<td>Greater Hobart</td>
<td>219</td>
</tr>
<tr>
<td>Rest of NSW</td>
<td></td>
</tr>
<tr>
<td>Greater Sydney</td>
<td>925</td>
</tr>
<tr>
<td>Rest of Vic.</td>
<td></td>
</tr>
<tr>
<td>Greater Melbourne</td>
<td>473</td>
</tr>
<tr>
<td>Rest of WA</td>
<td>134</td>
</tr>
<tr>
<td>Greater Perth</td>
<td>897</td>
</tr>
<tr>
<td>Rest of Qld</td>
<td>1047</td>
</tr>
<tr>
<td>Greater Brisbane</td>
<td>1392</td>
</tr>
<tr>
<td>Rest of SA</td>
<td>56</td>
</tr>
<tr>
<td>Greater Adelaide</td>
<td>771</td>
</tr>
<tr>
<td>Australian Capital Territory</td>
<td>276</td>
</tr>
</tbody>
</table>

Students whose high school origins are not known & international students are excluded.

**Remoteness Areas:**

<table>
<thead>
<tr>
<th>Region</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Remote Australia</td>
<td>32</td>
</tr>
<tr>
<td>Remote Australia</td>
<td>50</td>
</tr>
<tr>
<td>Outer Regional Australia</td>
<td>676</td>
</tr>
<tr>
<td>Inner Regional Australia</td>
<td>1682</td>
</tr>
<tr>
<td>Major Cities of Australia</td>
<td>8583</td>
</tr>
</tbody>
</table>

Students whose high school origins are not known & international students are excluded.
While this is not a geographic analysis, the attribution of students to high schools also enables some analysis of the nature of the schools the students attended:

GENDER:

Proportion of Medical Students who attended gender based High Schools

- Boys' School: 12%
- Girls' School: 13%
- Co-Ed School: 34%
- School type not known: 41%

Or SECTOR:

Proportion of Medical Students who attended High Schools in Public & Private sectors

- Non-government: 35%
- Government: 24%
- School type not known: 41%
Web Mapping Functionality:
APHCRI have a longstanding relationship with The Robert Graham Centre in the USA who are experts in the use and analysis of large datasets using Geographic Information Systems (GIS), and have a strong and team-based approach to GIS-related work and application development.

Together, GRAPHC and the Robert Graham Centre support the Australian web-based mapping environment known as HealthLandscape Australia.

HealthLandscape Australia is based on the Robert Graham Centre’s HealthLandscape which has been operating in America for over 10 years. HealthLandscape is an interactive web-based mapping tool that allows health professionals, policy makers, academic researchers, and planners to combine, analyse and display information in ways that promote better understanding and improvement of health and healthcare. The tool brings together various sources of health, socio-economic and environmental information in a convenient, central location to help answer questions about and improve health and healthcare. HealthLandscape can be used to create maps from publicly available data sets including regional criminal justice, education, healthcare, and demographic data, allowing users to discover community characteristics and share information with health professionals, policy makers, and legislators. ([http://www.healthlandscape.org/About-HealthLandscape.cfm](http://www.healthlandscape.org/About-HealthLandscape.cfm))

HealthLandscape Australia has all the capacities of the American system, and is gradually being upgraded and populated with data, mainly at this stage primary health care data as it is being managed within APHCRI.

The MSOD data has been deployed on HealthLandscape Australia in four separate modules.

- Med-School Categories
- Med-School Catchment Postcodes
- Medical Students Data
- Med-School Catchment Radials

Each of the modules can be visualised independently or in concert with other MSOD data modules or other HealthLandscape Australia layers such as Healthcare Workforce or Disadvantage.

1. **Medical School Categories**
   This module allows visualisation of the medical schools according to a number of criteria including:
   - Entry Level (graduate/postgraduate)
   - International Background:
Private School Background:

Rural Background

Student Count:
2. Medical School Catchment Postcodes
This module combines two tools “Tool Tips”, where by hovering the mouse pointer over a school a substantial amount of information is presented for each Medical School:

And the catchment depiction, where by selecting a school the postcodes from which students come can be highlighted:
Postcodes that are related to the students enrolled in each of the Medical Schools are depicted when a Medical school is clicked or when added from the menu.

The tool allows us to select a number of Medical Schools and to build combined catchments. The intensity of the colour in this case indicates the number of medical schools that are deriving students from this postcode.

The tool also offers the capacity to adjust the ‘threshold for visual inclusion. The “SAX tool” ranks postcode areas providing students to a university from the most to least. Those that are predominant catchments will remain visible as the threshold is lowered. The postcodes with the fewest students for the selected medical schools will be the first to be visually excluded.

3. Medical Students Data
This module presents the 14251 students according to their secondary school (as a proxy for Medical Student origins). The module also includes the hover Tool Tip which in this case provides information about the high school including Sector, Gender, Total School Enrolment and medical students that have enrolled from this school. That is to say that the number of medical students reported in the tool tip are always the total number of students from that high school – this cannot be filtered to different medical schools.

The information for the 14251 students can however be filtered, so that for example if a single medical school is selected the map will only show the schools which students enrolling in that medical school attended.
Each of the known attributes including those that have been linked to the student are available to build a subset of students using this filtering capacity.

The criteria which can be used are:
- Medical school
- Year
- Preferred region (ie whether wish to work in urban/rural area)
- Preferred field of practice
- State of origin
- Rural background
- Secondary school sector
- Secondary school type (gender)

The filter works cumulatively across the selection criteria and within each selection criteria. So the system allows multiple selection within the list of values and also to add extra selection criteria. For example:

Step 1: Select ANU school (405 enrolments mapped)

Step 2: select years 2006 + 2007 + 2008: (219 enrolments mapped)
NB this is (76 + 70 + 73)

Step 3: to identify how many came from local high schools, select State of origin - ACT (44 enrolments mapped)

4. **Med-School Catchment Radials**

This presentation style represents an alternate way of looking at the medical student origins by medical school. In this case the medical schools selected are UWA graduate and undergraduate schools and the ANU medical school. This provides a way of depicting the unit record data other than by points.
Next steps for consideration:

Security
Currently the MSOD is accessed via secure access. This is managed via a “white list” of accredited users which is provided by the MSOD administration. Secure access is necessary for the Medical Students Data module as it is possible with the filtering process to visualise and access information about individual medical students since the module is based on individual student data.

However, the Postcode Catchments and the Medical School Categories modules, do not offer access to unit record information and could reasonably be presented to the public via HealthLandscape Australia should the Deans wish to make the data more widely available.

Expansion of the data set
While the development of the current HealthLandscape Australia system to display data from the MSOD data base has been slower than was hoped for a range of reasons, including the complications that arose as a result of the process required to geo-locate the high schools, these problems have now been addressed and a system is in place which could readily incorporate data from later years, and in time data describing where the medical students practice. The Deans may wish to consider whether to add further years of data to the HealthLandscape system as they become available.