THESIS TITLE

Development of Tools for

the Design of Customizable Catchment areas

and

the Analysis of Geographic Variation in

Health Care

PHD Student: Ludovico Pinzari

TABLE OF CONTENTS

Contents

THESIS TITLE 1

TABLE OF CONTENTS 2

RESEARCH CASE 4

- MOTIVATION AND BACKGROUND 4
- RESEARCH STATEMENT 12
- GAPS IN THE CURRENT RESEARCH/KNOWLEDGE 13
- RESEARCH ARCHITECTURE 15

LITERATURE REVIEW 19

- ACCESS TO HEALTH SERVICES 19
- CATCHMENT AREA DESIGN 33
- HEURISTIC APPROACH 35

METHODOLOGY 37

- PEER GROUPING 38
  - HOMOGENEITY DIMENSIONS 41
  - HOMOGENEITY INDEX 47
  - DISSIMILARITY DISTANCE 74
- RATIONAL PRIMARY CARE SERVICE AREA DESIGN 78
  - LOCALIZATION INDEX 80
  - MULTI – OBJECTIVE SET PARTITIONING 84
- HEALTH WORKFORCE SUPPLY/DEMAND PLANNING 90

DATA SOURCES AND SECURE ACCESS 93

RESEARCH PLAN 94

FUTURE WORK 98

References 100

List of Figures 107

List of Tables 108

END 109

RESEARCH CASE

MOTIVATION AND BACKGROUND

There is much interest among health care stakeholders in studying the geographic variation of health related indicators, since it is commonly agreed that part of the unexplained variation is associated with inefficiencies and inequities in the provision and access of health care services. Such variation is associated with inefficiencies of health care where utilization rates vary across geographic regions [1], [2]. This variation is wasteful because higher utilization rates are not associated with better outcomes [3].

Practice variation varies from city to city, from state to state within Australia and from country to country worldwide. These variations are not fully explained by differences in patient need or preferences. Professor David S. Jones of the Culture of Medicine at Harvard Medical school noted that: ‘This isn’t variation that’s being caused by variations in the underlying burden of disease. If you control for the rates of the health of the population, you’ll still see this kind of practice variation. And because of that, researchers have identified this as the problem of *unwarranted* variation in medical practice – variation in practice that doesn’t make sense. Not that you can’t explain that, but that ought not to exist. And so since the 1960s, with the increased attention since the 1990s, since the cost of healthcare has increased dramatically, researchers have really tried to get to the bottom of this question of unwarranted variation to try to understand why is it that medical practice varies so much from place to place.’ [4].

‘*Unwarranted* variation raises questions about quality, equity and efficiency in health care. For instance, it may mean some people have less access to health care compared with others. It may suggest that factors other than patients’ needs or preferences are driving treatment decisions. It may indicate that some people are having unnecessary and potentially harmful tests or treatments, while others are missing out on necessary interventions. *Unwarranted* variation may also mean that scarce health resources are not being put to best use. As countries face increasing pressure on health budget, there is growing interest in reducing *unwarranted* variation in order to improve equity of access to appropriate services, the health outcomes of populations, and the value derived from investment in health care. Determining if variation is indeed *unwarranted* can be challenging, particularly without routine information on patient needs and preferences.' [5].

In order to study and explain the geographic variation of any indicator one has to select an appropriate geographic aggregation unit, such that spatial comparisons are meaningful and interpretable. However, geographic variation is usually studied using administrative boundaries that are not necessarily appropriate, since they were designed using criteria that were not related to most indicators of interest.

Countries select their preferred geographic unit for analysis, based on data availability and/or policy relevance [Table 1]. The population size of these geographic units varies widely [Figure 1].

Country	Geographic units	Health decision making	Years
Australia	Medicare Locals (61)	No	2010/11
Belgium	Provinces (11)	No	2009
Canada	1. Provinces/territories (13) 2. Health regions (83)	Yes	2003/04 or 2006/07 and 2010/11
Czech Republic	1. Regions (14) 2. Districts (77)	Yes (Regions)	2007-10
Finland	Hospital districts (20)	Yes	2001-11
France	Administrative departments (95)	No	2005-11
Germany	1. Länder (16) 2. Spatial planning regions (96)	Yes (Länder)	2011
Israel	Districts (6)	No	2000-11
Italy	1. Regions (20) 2. Provinces (110)	Yes (Regions)	2007-11
Portugal	Grupos de municipios (28)	No	2002-09
Spain	1. Autonomous communities (17) 2. Provinces (50)	Yes (AC)	2000, 2005, 2010
Switzerland	Cantons (26)	Yes	2005-11
United Kingdom/England	Primary Care Trusts (PCTs) (151)	Yes	2010

Table 1 OECD (2014), Geographic Variations in Health Care What Do we know and What Can Be Done to Improve Health System Performance? OECD Health Policy Studies, OECD Publishing.

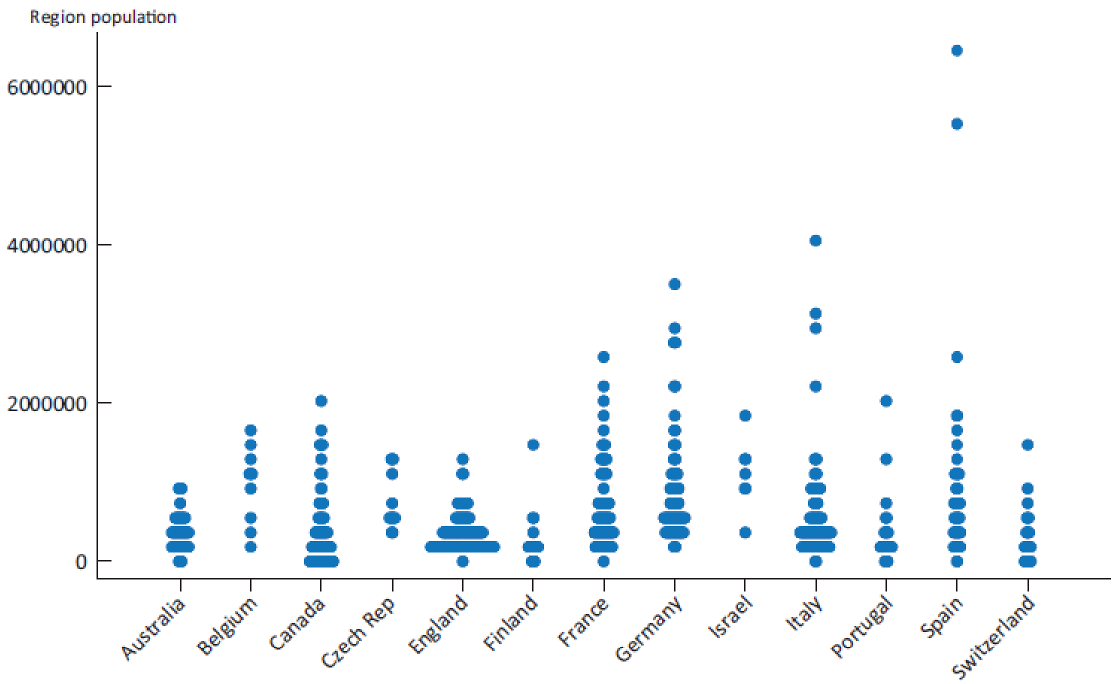


Figure 1 Population size of geographic units in participating OECD countries, 2011 or latest year. Source: National data submitted for the OECD project on Medical Practice Variations.

The size of the geographic unit *matters* for the analysis and interpretation of variations within and across countries. Health care utilisation rates observed in large territorial units will tend to be closer to the country's average while those in some less populated areas are more likely to deviate from this average for different reasons. This means that countries with smaller geographic areas are more likely, statistically speaking, to display higher variations across areas than countries with larger units [6].

The geographical unit used for analysis in Australia was the Medicare Locals (MLs) [Figure 2]. However, these administrative boundaries, like others such as Statistical Local Area, suffer of a limitation – Their aggregated results may be confused by large amount of heterogeneity, as they vary considerably in population size (40 000 to 800 000), health and socioeconomic status, geographic area, remoteness and proximity to territory hospitals.

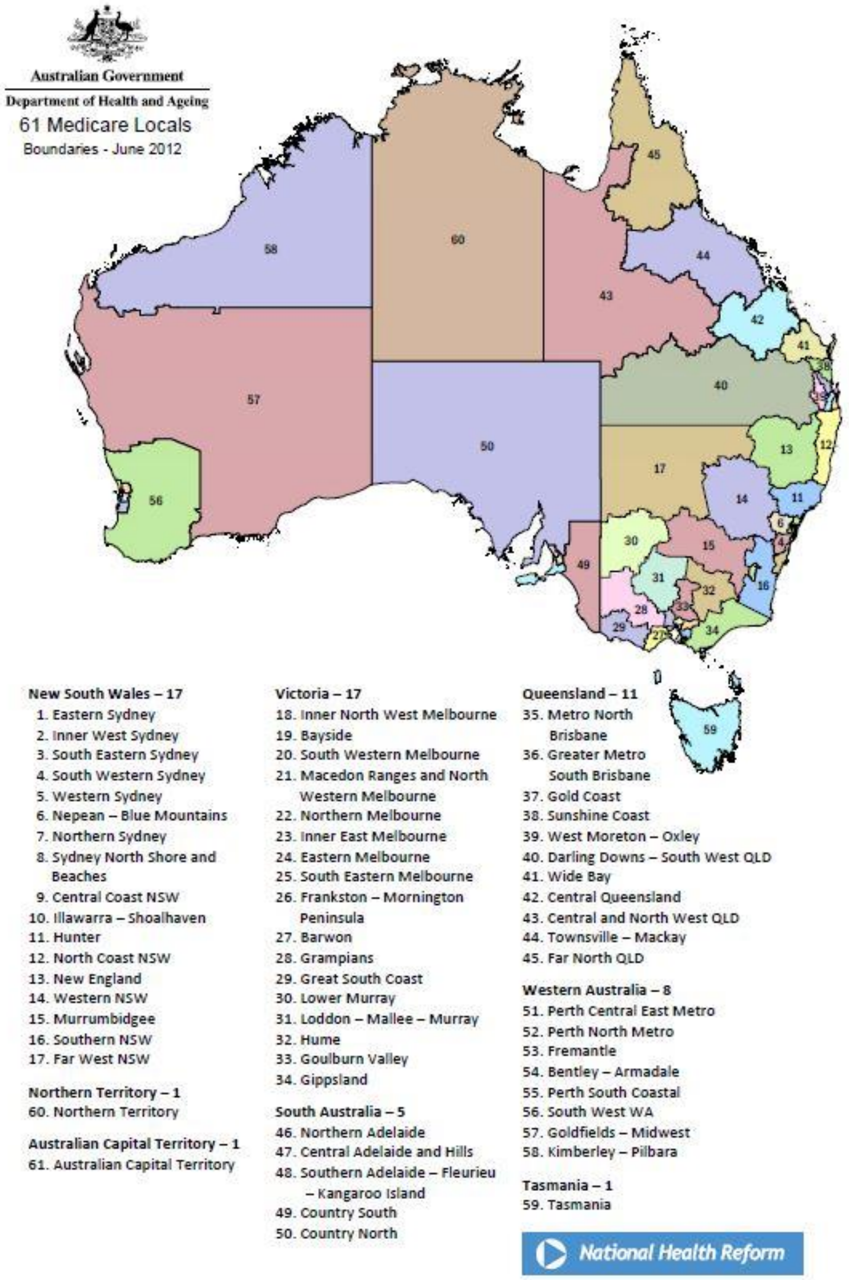


Figure 2 Medicare Local catchment

Therefore, the National Health Performance Authority (NHPA) have documented the variance between the 61 MLs in terms of the affordability, availability and accessibility of general practitioner and acute care by allocating each ML to one of seven peer groups ([7]Figure 3).

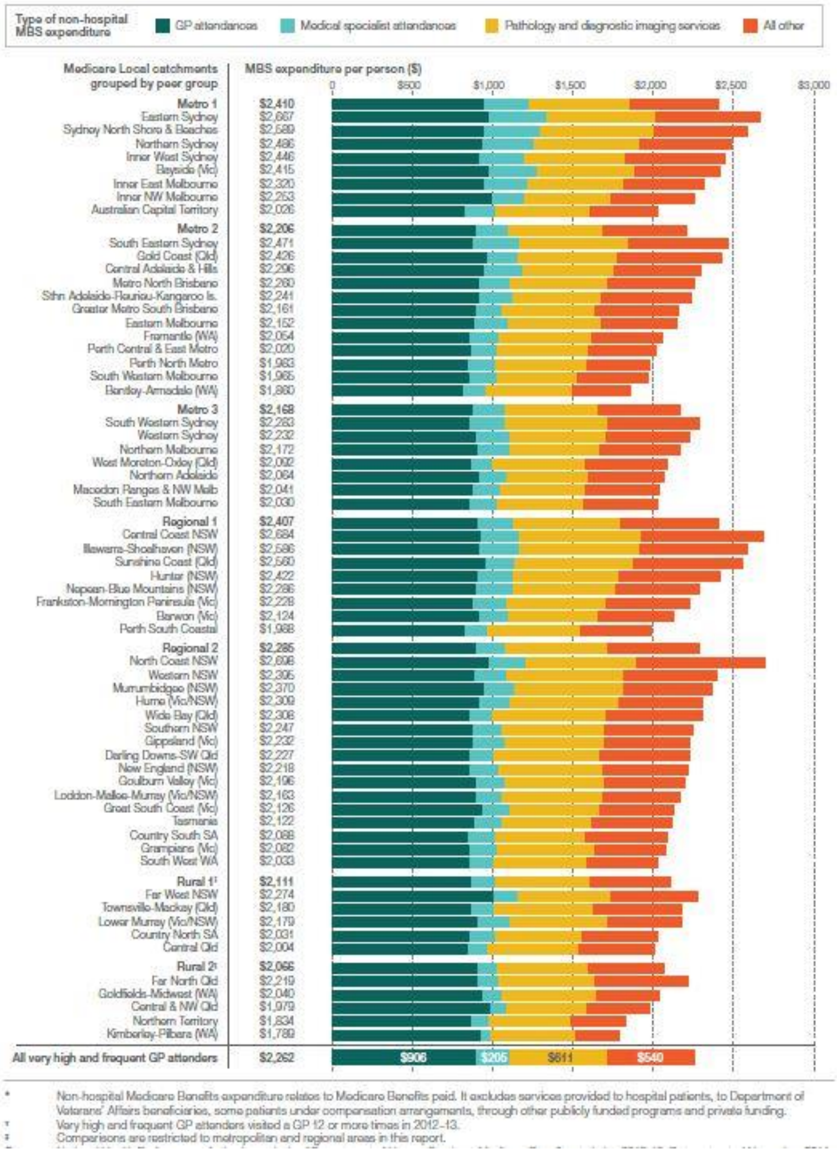


Figure 3 Average non-hospital MBS expenditure for very high and frequent GP attendance, by Medicare Local catchment, 2012 - 2013. NHPA analysis of Department of Human Services, Medicare Benefits statistics 2012 - 2013 Data extracted November 2014

However, from 1 July 2015 MLs have been replaced by a smaller number (31) of Primary Health Network (PHNs) Figure 4. The 31 PHNs areas are much larger in size and have greater diversity in their populations' characteristics and have therefore not been grouped. Instead, the Authority reported the variation using smaller local areas (Statistical Local Area Level 3- SA3) [8].

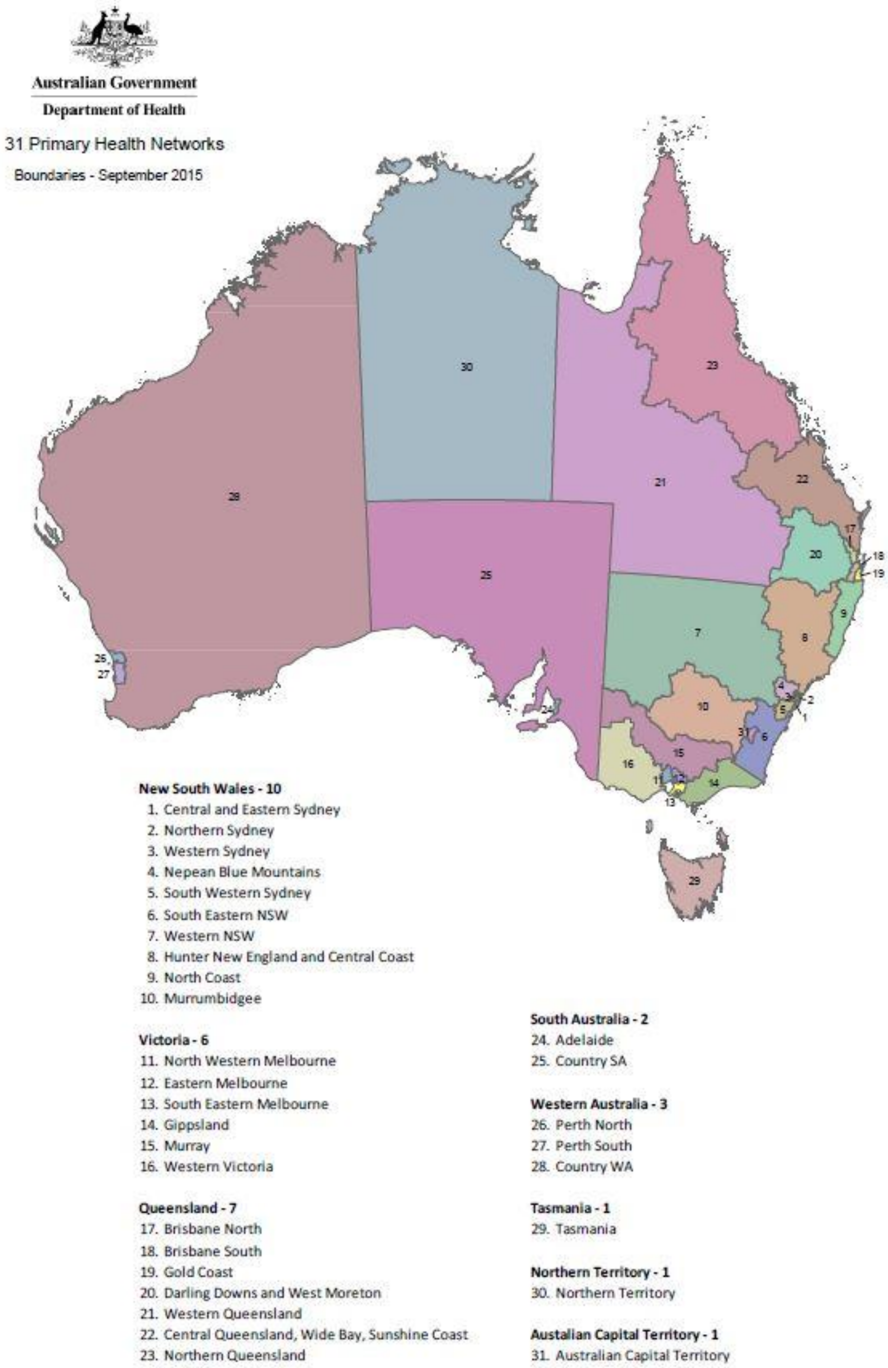


Figure 4 Primary Health Network - Australian Department of Health

However, it was soon realized that an alternative geographical unit should be selected for the study of the Geographic Variation. More precisely, the new geography should be able to incorporate **people preferences** toward provider location and must be **homogeneous** along a number of dimensions, and therefore they allow meaningful geographical comparisons. From this scenario also emerges that another important property is that they have to be **stable**: small changes in the data (as typically observed from one year to the next) should lead to small changes in the definition of their boundaries, so that the same set of areas can be meaningful used for a long period of time (say 5 years) without degradation in the results. This property allows any Agency to report the Geographic Variation without bothering of any changes in the organization and allocation of resources in the national health system. The geographic areas will also be **peer grouped** for fair comparison.

Hence the indicators under study are the **utilization of services**, such as GP visits or hospital admissions, and the appropriate geographies are usually called **catchment areas**. An example of catchment area is the Primary Care Service Areas (**PCSAs**), which has the defining property that most of the people in the same PCSA receive primary care within that PCSA. Other catchment areas may be designed with other objective in mind, that will depend on the specific indicator(s) that needs to be aggregated.

The design of catchment areas is notoriously difficult and there are few published sets of PCSAs [9], [10]. There is no general agreement on how catchment areas should be designed and there is no available software tool that could be used in Australia to design a set of catchment areas with specific properties. While there is agreement that it would be useful to be able to design catchment areas which are tailored to specific indicators, there is no general formulation of what that means.

In this research a new mathematical formulation to the definition of Rational PCSAs (**RPCSAs**) is proposed. The RPCSAs are customizable, meaning that it will be possible to modify the required properties and produce a different set of catchment areas. This work proposes an innovative method based on patient's preferences focus not only on spatially uneven distribution of providers but also emphasizes on the variations among population groups because of their different socioeconomic and demographic characteristics.

The problem of designing catchment areas will be cast in the framework of discrete optimization, as the minimization of an objective function that contains two terms. The first term captures the fact that users want geographies such that patients in one unit mostly attend services in the same unit. The second term captures the fact that units must be homogeneous along certain user specified dimensions, and variation of certain variables should mostly occur across, but not within, units. This integer programming problem is a hard combinatorial optimization problem (NP-COMPLETE) and we propose an approximate solution based on a combination of meta-heuristics and divide-and-conquer techniques.

Moreover, we expand the usual approach in health care performance comparison based on contrasting the average behaviour of each Health Care Service Area by adding the analysis of the variation, within and across these service areas, at different level of interest.

More precisely we identify clusters of PHN catchments called peer groups. This enables fairer comparisons of individuals PHN catchments and also provides a summary of the variation across Australia diverse metropolitan, regional and rural populations by presenting aggregate results for each peer groups. Reporting on variation across similar local areas aims to provide health care professionals with information about factors they could influence to improve the coordination of care in the community.

Ultimately, we will develop a ‘What If’ scenario analysis tool which enables healthcare planners to analyse local supply and demand of health workforce, in order to identify innovative policy options to increase access to health services and decrease healthcare costs.

Finally, the integration of the zoning scheme output with a visualization tool through a graphical user interface provides a framework suitable to analyse the patient pattern behaviour across space and time.

In order to highlight the main targets of this thesis a summary of the above research objectives has been put together:

RESEARCH STATEMENT

1. Formulate a mathematical model and develop a visualization tool that captures the diversity within and across geographical areas in order to define PHN and RPCSAs peer groups.
2. Provide a general purpose model that encodes the basic properties of a RPCSA and implement automated methods for the design of customizable RPCSAs.
3. A simulation model that allows for changes to key variables, such as travel criteria, service bundles, health professional delivering services, movement of doctors between areas to look at operationalising different policy options for improving the availability of services.

In all instances, the aim is to offer data and analytic tools to identify primary care clinician supply and needs in communities across Australia, with area that reflect patient's travel to primary care. RPCSAs data can help identify areas with low supply of primary care and Bulk Billing providers, and populations with relatively high health risk. This research also provides information about primary care utilization by the elderly.

Thus, the overall objective of this work is to develop generic approaches and procedures which can be used to produce a national database of primary care resources and utilization for small areas in Australia. [11]

GAPS IN THE CURRENT RESEARCH/KNOWLEDGE

Good access for health care for all populations, regardless of geography, remains a key goal of governments and societies internationally [12]. Access to health care service is often modelled using catchments to define regions where utilization of health care services occurs ([13]; [14]; [15]). Lots of efforts have been made to document potential accessibility, which is about the geographical linkage between people and essential services. At its core, the concept refers to the separation between services and population – how much distance, cost, time and effort are involved in reaching service facilities [16].

Access to health facility in a given location is influenced by many factors, including the availability of health services in the area (**supply**), the number of people living in that area (**demand**), the population's health status, the socio-economic and financial resources available to the population, people's knowledge about health and the health care system, and geographical impedance between population and health services [17].

Access to healthcare varies across space because of uneven distributions of healthcare providers and consumers (***spatial factors***), and also varies among population groups because of their different socioeconomic and demographic characteristics (***non-spatial factors***) [18]. Accordingly, *spatial* access emphasizes the importance of geographic barriers (distance or time) between consumer and provider, whereas *non-spatial* access stresses nongeographic barriers or facilitators such as social class, income, ethnicity, age, sex, etc. [19].

There were a few attempts to integrate the two categories variables, using various techniques such as the Principal Component Analysis [20], the Factor Analysis [18] and the Equity Model [21]. Furthermore, an important issue is that the competitions on either supply or demand side have not been considered in all these models. On the supply side, health facilities in the same or even different categories (clinics, doctor's office, etc.) may provide overlapping types of services. The facilities may compete for the same potential clients. On the demand side, different population subgroups can be classified by demographic, socioeconomic, or insurance status. These people may seek service opportunities at the same facilities of high demand. How to model the competition remains a research topic.

The framework of discrete optimization allows the implementation of spatial access in a reasonably simple process (i.e. a modular system avoiding interdependency) such that allow users to select more than a single indicator (types of services) and specify the properties (homogeneity of demographic and socioeconomic variables) that need to be satisfied.

Moreover, all the existing techniques don't take into account the effective spatial interaction between patients and providers. Thus, they do not address the determinants of service utilization patterns and hence have limited value for forecasting and planning.

After all, people are free to choose where they access Primary Health Care services and where they go depend primarily on their personal preferences.

Therefore, the main challenge of this research is to integrate spatial and non-spatial factors into one framework for assessing Healthcare Access and identifying shortage areas in a competitive system.

RESEARCH ARCHITECTURE

Although the main targets of this research are basically three:

- Define peer groups of different geographies (SA3/CRPCSA).
- Design customizable geographies (CRPCSA) to study the Geographic Variation of health indicators.
- Design a simulation model for planning and forecasting the Health Workforce distribution.

These components are embedded in an integrated structure of different stages and tools.

Figure 5

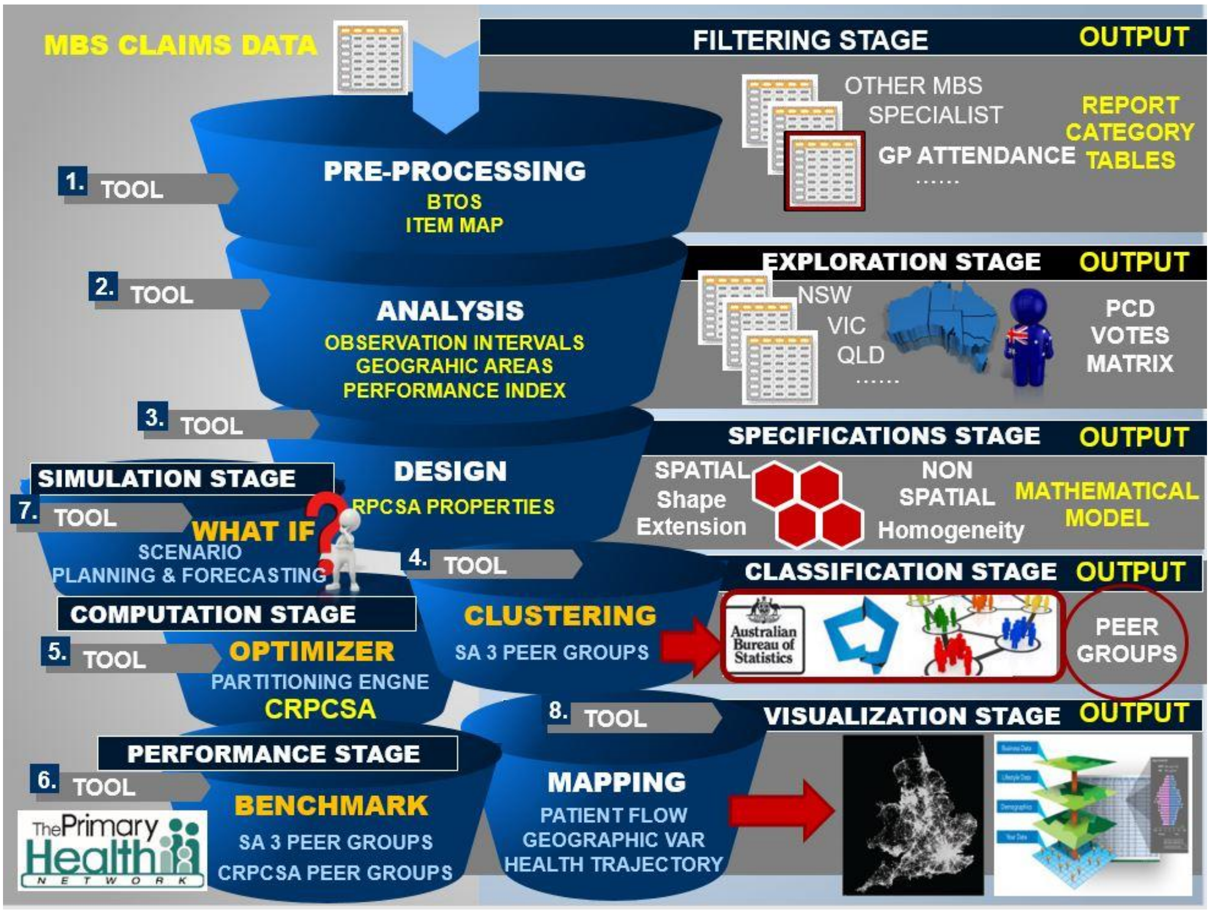


Figure 5 System Architecture for Customizable Rational Primary Care Service Areas

The Architecture components can be easily described in two main processes of distinct stages.

In the initial stage (Pre-processing) the data are filtered according to a specific report category (Indicator), such as:

- GP type use
- GP service utilization by residents of Residential Aged Care Facilities
- After hours GP service utilization
- Specialist service
- Access to services by type of services relative to need
- Vaccination rates for children
- Diagnostic and surgical procedures

Or any kind of health care activities and procedures based mainly on the criteria of High – Cost and High - Volume policy relevance and data availability. Then, the tables are sampled in one or more observation time frames for the selected administrative geographic areas. The final step in the Analysis process is to produce a votes matrix which summarizes people preferences to different locations, that contains all the information regarding the patient mobility within the study area.

The next stages are the core of the project:

In the Design Stage, the user determines the relevant RPCSAs' properties which include spatial units which are demographically, socially and economically homogeneous as well as physical properties such as the shape and the extension of the catchment area. The proposed model serves as a basis for the partitioning engine to produce Customizable Rational Primary Care Service Area (CRPCSA) and to compare the goodness in the benchmark module with other geographical units like SA3. Finally, the integration of the zoning scheme output with the Simulation tool provides a framework suitable for the analysis of the key variables to identify innovative policy options to increase access to health service and maximize the value of investments in health workforce.

All these tools are linked by several interfaces with a Graphical User Interface (GUI) useful to analyze the patient pattern behavior (Figure 6) and visualize the geographic variations in health care based on the patient's place of residence. The innovation is the graphical decomposition of its main variables in terms of the different dimensions of Access to Health Care Services (I personally call the 5 As) Figure 7 .

Figure 6 Mapping patient flow in R

In the next section of the document we further explore the multidimensional nature of access to healthcare and what has already been done in this field.

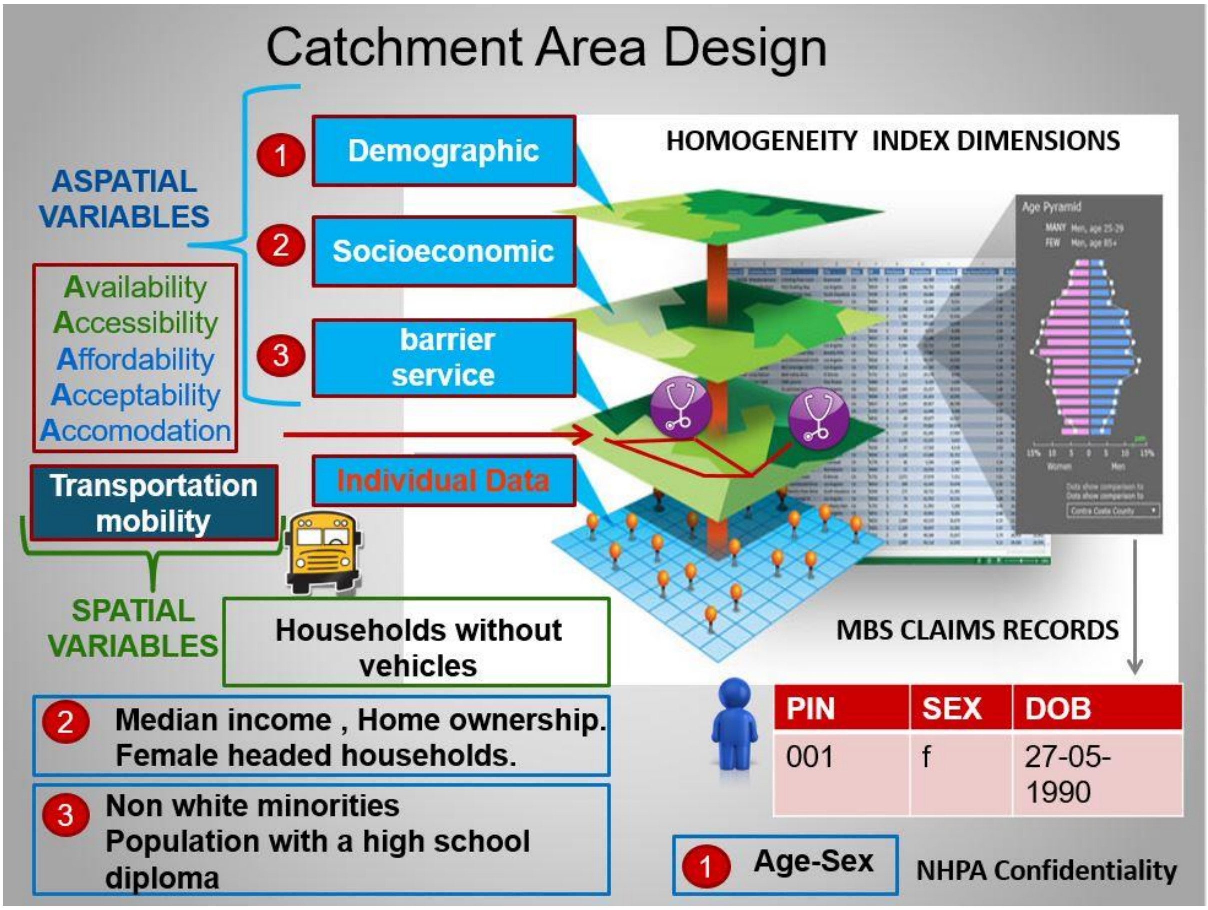


Figure 7 Access to Health Care - Homogeneity Index Dimensions

LITERATURE REVIEW

This review begins with a description of the issues that need to be considered when analysing ***potential*** and ***revealed accessibility*** to health care services – that is, patterns of health service utilization. Such patterns are the result of individual choices about when and where to use services, the geographical configuration of health care opportunities, and the mediating effects of medical referrals and regulations [22]. We then discuss how the **design of a catchment area** might be accomplished with a brief description of the common properties and objectives that must be satisfied. We conclude the review with a short description of the heuristic techniques used to solve the problem in this area of study and the limitations in the case of a RPCSAs.

ACCESS TO HEALTH SERVICES

The **service area** or **catchment area** for a health care provider is the geographic area that contains the bulk of population served. For a health care provider, the service area ties the client population to a geographic area, a neighbourhood or community or set of communities. Some health facilities have ***mandated service areas*** [22] in that they are required to serve the population living within a particular region, say, a county or set of Postal Codes. A typical example of mandated catchment areas are the public schools, as all children who live within a given area are required to attend a particular school. Such mandated areas are less common in the case of health services in Australia, where patients do not need to register with a single General Practitioner (GP) Table 1.

		Primary care physicians referral to access secondary care		
		Required	**Incentives**	**No requirement, no incentive**
Are patients required or encouraged to register with a primary care physician?	**Required**	Denmark, Finland, Ireland, Italy, Netherlands, Portugal, Slovenia, Spain		Czech Republic
	Incentives	Australia, New Zealand, Norway, Poland,	Belgium, France, Switzerland	
	No requirement, no incentive	Canada, Chile, United Kingdom	Mexico	Austria, Germany, Greece, Iceland, Israel, Japan, Korea

Table 2 OECD Health System Characteristics Survey, 2013

When choice of providers is not mandated, health care services have ***natural service areas*** that arises through individual decisions and medical referral patterns. The service areas for different health care providers typically overlap, reflecting the diversity of health care needs and choices among people living in the same areas. Patient origin information is crucial for identifying ‘natural’ service areas. Talen [23] has described a number of approaches to measuring potential accessibility and Gary Higgs provided an adapted framework to define health care catchment Table 3.

Approach	**Definition**	**Health example**
Container	The number of facilities contained with a given unit.	Number of GP surgeries in census ward
Coverage	The number of facilities within a given distance from a point of origin.	The number of hospitals 10 km from a population centroid.
Minimum distance	The distance between a point of origin and the nearest facility.	Distance between village centre and nearest pharmacy
Travel cost	The average distance between a point of origin and all facilities.	Average distance between centroid of census tract and all GP surgeries.
Gravity	An index in which the sum of all facilities (weighted by size or supply side characteristics) is divided by the frictional effect of distance.	All GP surgeries (weighted by list size) or those with, for example, specialised services or female GPs, divided by distance.

Table 3 Measurement of accessibility (with examples from health sector). Adapted from Talen. Gary Higgs

Container, coverage, minimum distance and travel cost are typically the spatial dimensions used by GIS (Geographic Information System) analyst to develop measures of access to health care services. All these methods rely on the accuracy of the geocoded information of the patient. The GIS analyst geocodes the addresses of patients who use the health care facility, and plots those address locations on a map Figure 8 [24]. To focus on the primary service area, we can plot the 80% or 90% of clients who live closest to the facility and identify the natural service area this way Figure 9.

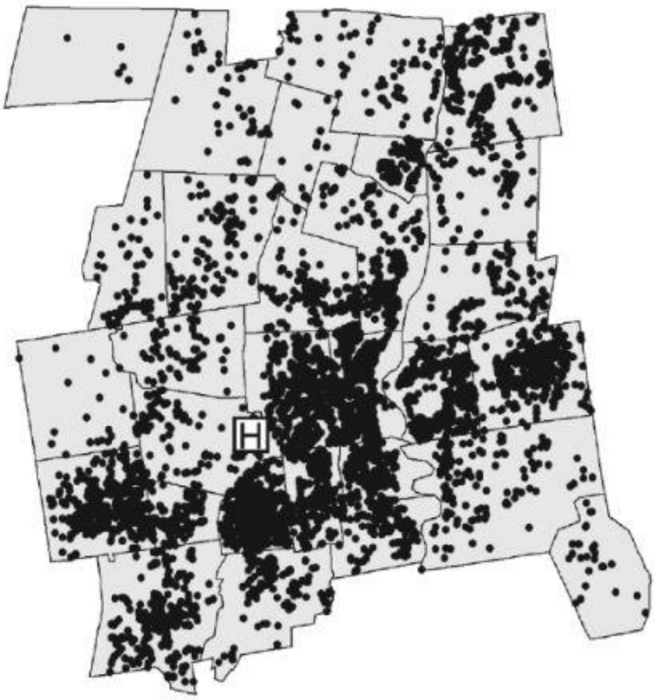


Figure 8 Hospital Service Area Based on Geocoded Patient Residential Locations for All patients. Source: GIS and Public Health 2nd ed Ellen K. Cromely. Guilford Press 2012

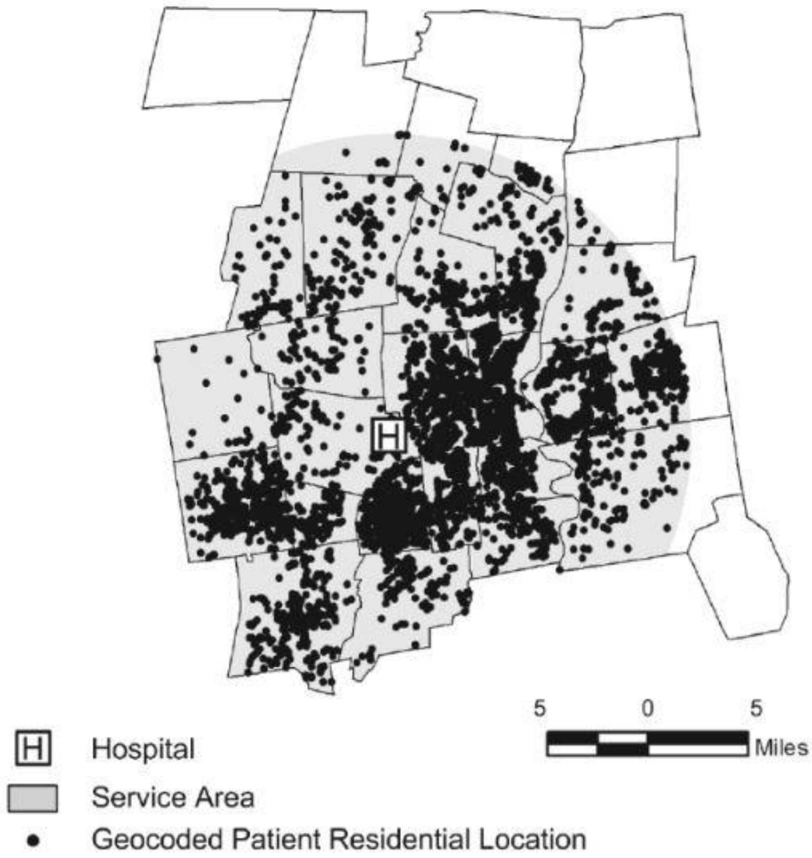


Figure 9 Hospital Service Area Based on Geocoded Patient Residential Locations Excluding Outliers. Source: GIS and Public Health 2nd ed Ellen K. Cromely. Guilford Press 2012

Nadine Schurman defined rural hospital catchments based on travel-time using road network data in a GIS environment [25]Figure 10.

Although maps of service areas are useful descriptive tools, they don't consider the actual interaction between the two sides of the demand and supply equation. The measure generally assumes that 'given a maximum range for the service being offered at a facility and assuming that every member of the population is a *potential user* of the service, the pattern of physically accessibility will depend only on the relative location of the population and the service facilities' [26]. As indicated previously, this could be represented as travel time, road or map distance.

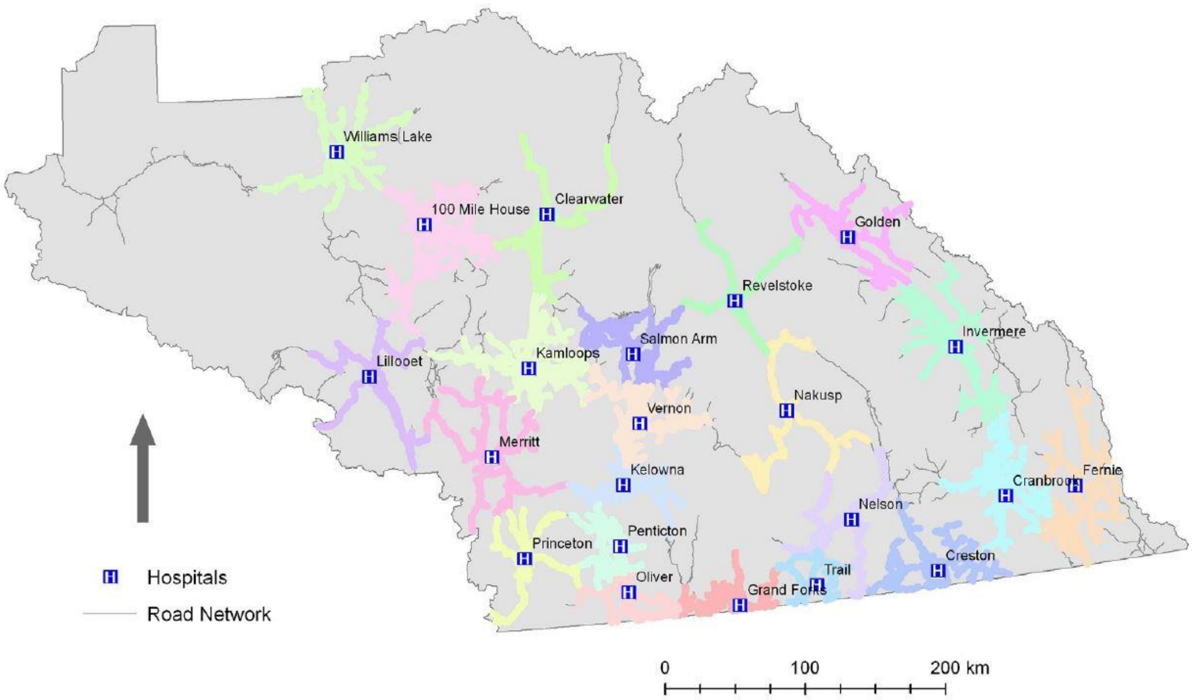


Figure 10 1-hour service areas for all hospitals. Data source: GIS Innovations Ltd

Khan [27] has reviewed the approaches taken to calculating potential access measures in a health context and provides a useful typology which acknowledges the dichotomy between potential and realised (revealed), spatial and aspatial measures. The typology may be conveniently show as a 2x2 matrix Figure 11.

Gugliardo [14] describes this matrix in terms of stages and dimensions. The two stages are ‘’potential’’ for care delivery, followed by ‘’realized’’ delivery of care. “Potential exists when a needy population coexists in space and time with a willing and able healthcare delivery system. Realized care, sometimes referred to as actualized care, follows when all barriers to provision are overcome”.

STAGES		
Potential	Realized	DIMENSIONS
Geographic Access Studies of distance and availability that do not consider utilization measures.	Geographic Access Utilization studies that consider spatial factors.	Spatial AVAILABILITY ACCESSABILITY
Social Access Studies of Affordability, culture and other non-spatial factor that do not consider utilization measures.	Social Access Utilization studies that consider affordability, culture and other non-spatial factors.	Aspatial AFFORDABILITY ACCEPTABILITY ACCOMODATION

Figure 11 Taxonomy of health care access. Mark F. Gugliardo

A number of barriers can impede progression from potential to realized access. Penchansky and Thomas [28] have grouped barriers into five dimensions:

- *Availability* defines the supply of services in relation to needs: Are the capacity and types of services adequate to meet health care needs?

- *Accessibility* describes geographical barriers including distance, transportation, travel time and cost. It highlights the geographical location of services in relation to population in need.
- *Accommodation* identifies the degree to which services are organized to meet clients' needs, including hours of operation, application procedures, and waiting times.
- *Affordability* identifies the degree to which services are organized in regards to people's ability to pay. Income levels and insurance coverage are critical aspects of affordability.
- *Acceptability* describes clients' views of health services and how service providers interact with clients. Services are acceptable if clients are well treated and satisfied, if providers and clients communicate openly, and if clients are confident about the quality of care delivered [22].

The last three dimensions are aspatial and reflect healthcare financial arrangement and cultural factor, whilst the first two are spatial in nature.

In a nutshell Availability refers to the number of local service points from which a client can choose and Accessibility is travel impedance (time, distance) between patient location and service points.

The *regional Availability* [29] is commonly expressed as the manpower ratio in terms of the number or size of health care manpower or facilities relative to the potential user population in a defined area (physician/population ratio, and its variants) [30]. However, the ratio approach is limited because assume there is no cross boundary flow of people accessing facilities in adjoining areas [31]. Another limitation of physician-to-population measures is that they assume all consumers have equal access to such facilities independent of how far away they live or work from the health care sites. At the outer boundaries of a catchment, for instance, the probability increases that a person will choose to use services in adjacent catchment. Thus they do not account the role of a ‘distance-decay’ effect on utilisation patterns. For these reasons several methods have been developed to incorporate simultaneously the Availability and Accessibility dimensions. Usually this two dimensions are considered together and are referred as ‘’Spatial Accessibility’’ (SA) [14].

These methods include the kernel density models [14] and gravity models [26] [32].

In kernel estimation, we begin by geocoding service providers to point locations. A circular window scans the map, and the density of the population served is computed within this window Figure 12.

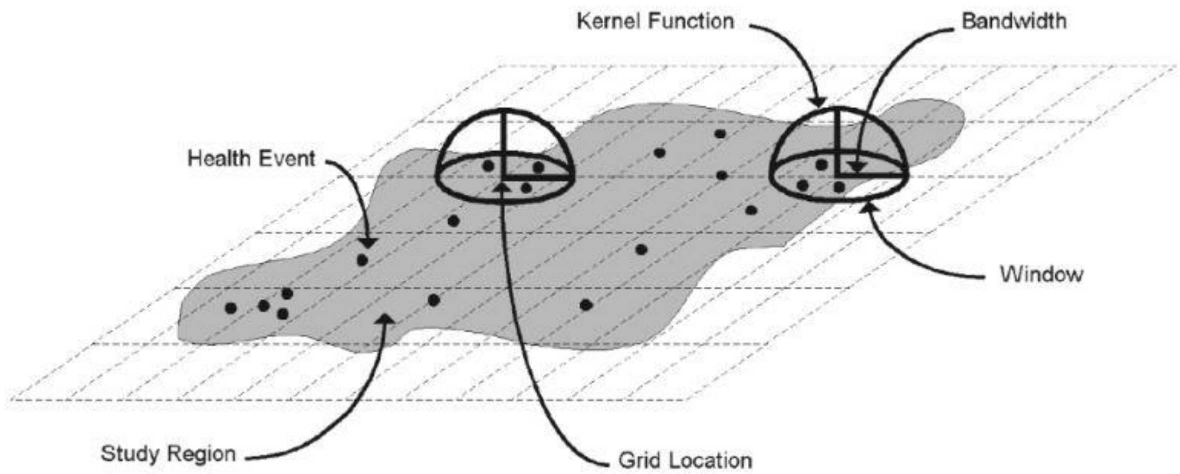


Figure 12 A schematic of the kernel estimation method - GIS and Public Health p 165

However, the kernel methods have important limitations. The mathematical form of the kernel function is somewhat arbitrary and perform poorly in rural areas where few providers exist [33]. On the other hand, the gravity models are conceptually more complete and more flexible to realize but less intuitive to interpret.

The gravity model is based on an analogy with Newton physics in which the interaction between places is directly related to their relative sizes or attractiveness and inversely proportional to the distance between them. Attractiveness depends on price, quality of services, accommodation, cultural appropriateness, and a host of service related factors [16]. Typically, the decrease in access with increasing distance or travel time is represented in terms of an impedance function. Impedance functions can have different functional forms including power, exponential, Gaussian, and log-logistic (s-shaped) [34] Figure 13. Researchers have used statistical methods to find an appropriate impedance function that best fits actual travel patterns. However, the estimation of the parameters has certain limitations. The effects of distance can vary over time and space, leading to errors in estimating potential accessibility. These methods were also used for predicting flows of people to health services [35].

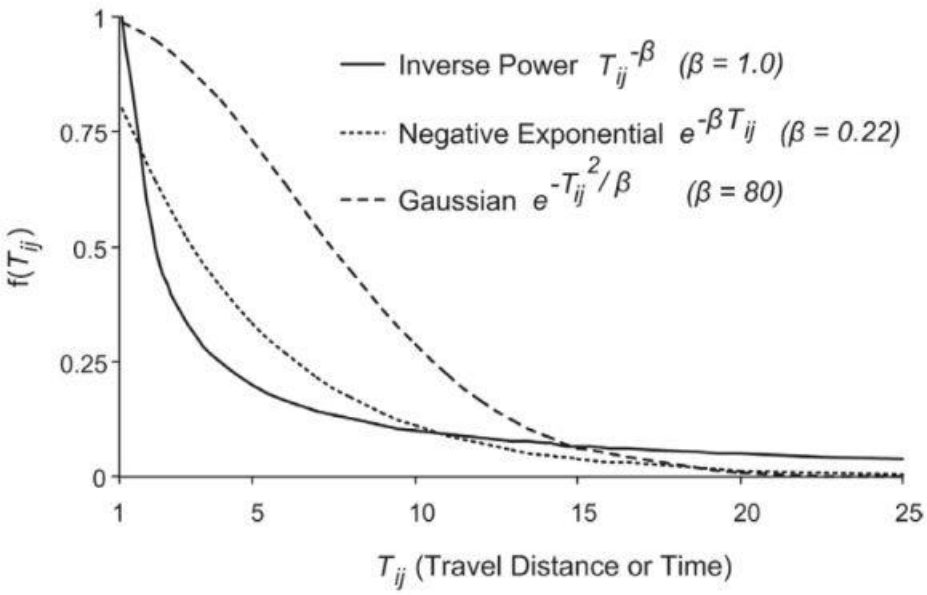


Figure 13 Different spatial impedance functions. Tij is a measure of travel time or distance. Source GIS and Public Health

In spite of its elegance, the gravity formulation is not intuitive to healthcare workforce policy makers, who prefer to think of spatial accessibility in terms of provider-to-population ratios or simple distance. Therefore, Luo and Wang [18] proposed an improvement of the gravity model, the two-step floating catchment area (2SFCA) to identify areas in Illinois that have a shortage of physicians. The **census tract** was chosen as the ***analysis unit*** for population, or **demand**. ZIP code area was chosen as *the analysis unit* for physicians, or supply. Population-weighted centroid of census tracts (based on block group - level population) were used instead of simple geographic centroids to represent population locations more accurately Figure 14 [32].

In their two-step process a provider-to-population ratio is first estimated for each provider location (ZIP code centroid). They constructed a threshold travel distance/time around each service provider j and computed the provider-to-population ratio Rj within that window (total number of providers assigned to the ZIP centroid is divided by the population living within that window). Second, for each population location i, they searched all provider locations within a threshold travel distance/time and sum up the Rj values for those providers. The resulting value was the accessibility score for the population at place i. Higher values indicated better spatial access to health care providers [18] Figure 15. The accessibility ratio obtained for each census tract is interpreted as the amount of the total services shared by one patient in the tract. Non-spatial factor can be combined with this technique in order to highlight inner city areas that have poor spatial measures of access on the FCA approach [18]. McGrail and Humphreys [36] used 2SFCA to evaluate the spatial accessibility of primary care physicians in Victoria, Australia.

However, these floating catchment area methods are still limited as they assume equal access within the catchment. In the absence of detailed information on individual address (or the socio-economic characteristics of people accessing such services), the demand is often estimated by assuming all people live at the centroid and so providing a crude estimate of accessibility [37].

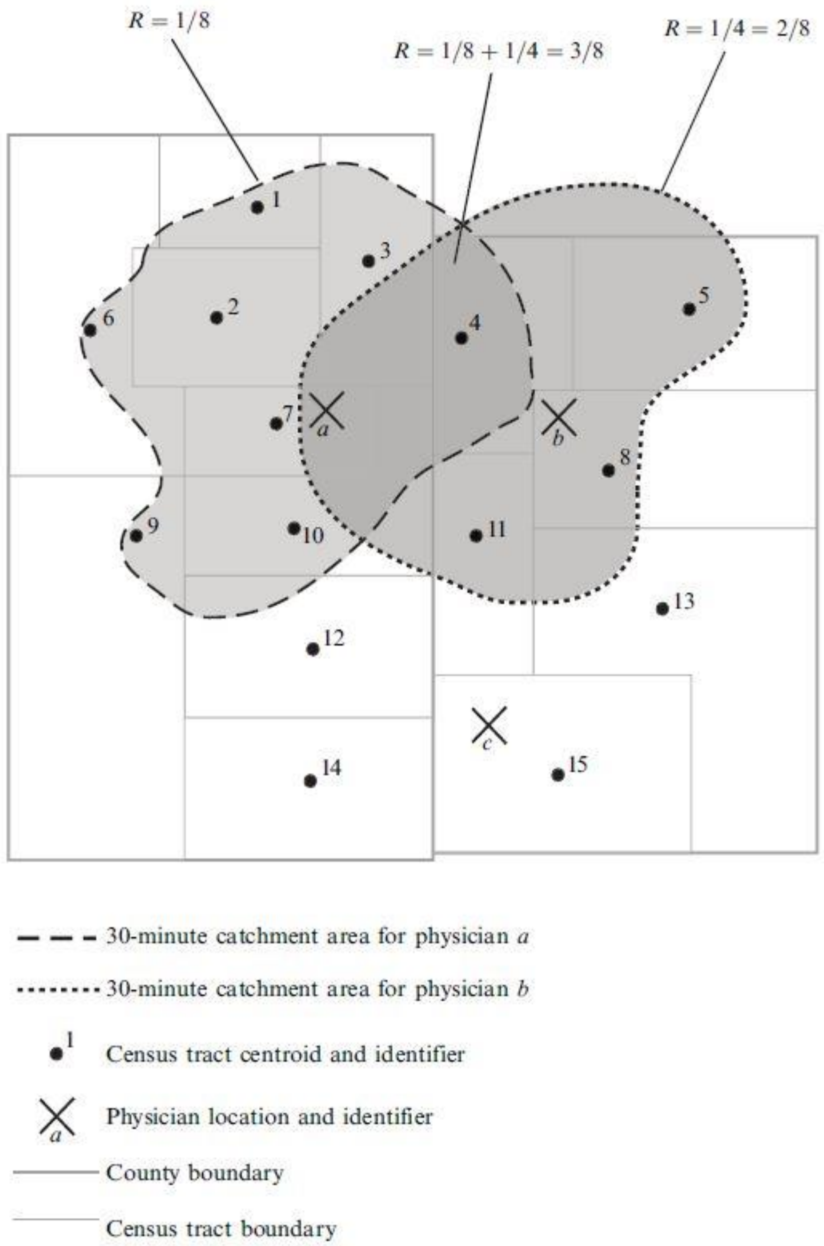


Figure 14 An early version of the floating catchment area (FCA) method. Luo W, Wang F (2003)

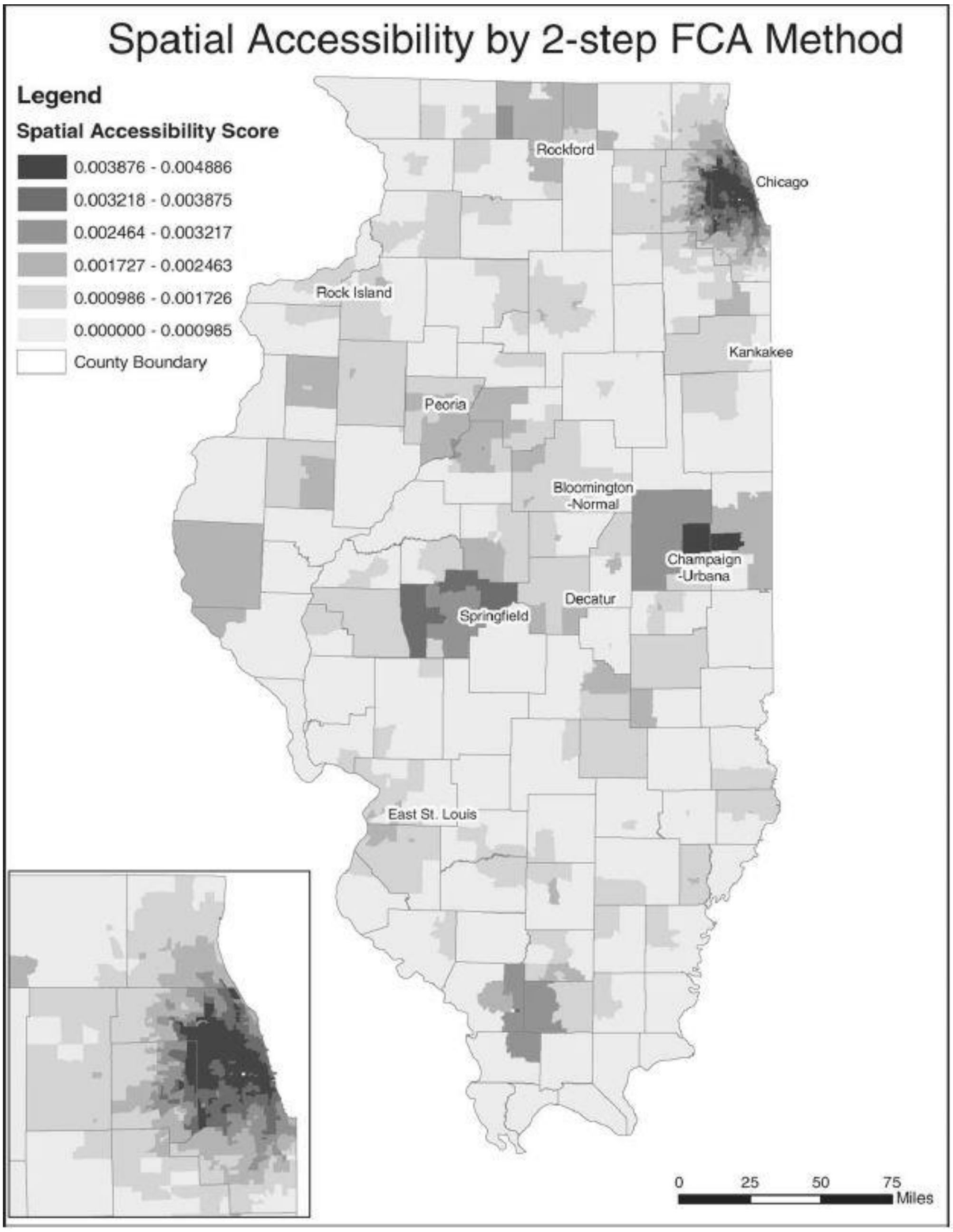


Figure 15 Spatial accessibility to primary care Illinois. Luo, Wang (2004)

Moreover, these measures are calculated from demand points based on where patients reside and not where they work for example. This may be particularly problematic for access to non-urgent services, such as General practitioner visit, where patients may attend such facilities during the working day.

Furthermore, many studies suggest that individuals in more remote settings accept lengthy travel as routine part of their lives [38]. In Australia, these remote areas are usually big with sparsely population and hence having no apparent primary care service.

Lastly, the selection of the analysis unit and the threshold travel distance is arbitrary usually based on pre-defined hypothesis (e.g. the average walking distance a person is prepared/able to travel to access a particular facility) or can be based on actual empirical evidence from those utilising the service. Often however, such data does not exist at a sufficient level of detail so this threshold tends to be calculated at specified increments (e.g. 10,15,20 min travel time).

In conclusion all the aforementioned techniques are only valid tool to explore the spatial accessibility (SA) of a given area and they don't provide any systematic procedure to design new geographies as it is not clear which criteria can be used to aggregate the specific unit, but most importantly, as indicated previously, they don't describe the effective spatial interaction between patients and providers. Therefore, studying the movements or flows between places can help to examine this issue.

In the next sections we explore how the catchment area design problem and the spatial interaction can be treated in geographical research.

CATCHMENT AREA DESIGN

In a geographical context the catchment area design is usually termed as Regionalization. Regionalization is a spatial process where the territorial units are grouped together following predetermined spatial rules and restrictions [39].

A first important distinction is made between basic spatial units (b.s.u) and zones [40].

- b.s.u: The smallest spatial unit for which data are available (e.g. Postal Code Areas, Statistical Local Areas, etc.).
- zone: A geographic area containing one or more b.s.u.

Thus a catchment area is nothing more than a zone with certain objectives and restrictions.

The most common restrictions are:

- **Contiguity**: A set of b.s.u who share the same boundary. It has to be possible to travel between two points in a zone without having to pass another zone [41].
- **Absence of holes**: This property prevents the formation of solutions with embedded areas [42].
- **Integrity**: This property prevents the use of a b.s.u in more than one zone, each unit has one and only one zone.
- **Capacity**: these constraints can be used to limit both the number of zones existing as the number of possible b.s.u within each zone [43].

The objectives frequently used are as follows:

- **Population size**: Ensuring that output zones exceed a particular population size threshold.

- **Homogeneity**: zones should have similar features, the distribution of the population with the respect to gender, age, ethnicity, should be common among the b.s.u within a zone [44].
- **Compactness**: The output zones must have a compact geometric shape close to a circle or square. In most applications, compact zones usually have geographically concentrated activity, therefore less travel, more service times. In other words, the term compactness expresses the desire for zones with minimal total travel times [45].
- **Balance**: This criterion expresses a relation of zones among each other and is motivated by the desire of an even distribution of people in the study area. However, sometimes strict upper or lower bounds for the size of zones are given. For example, on maximal travel times or minimal number of people living within the zone (population size) [46].

In vision of this qualitative description of the problem we can have a broad idea of the goal:

- **Catchment Areas Identification**: is the problem of grouping small geographical areas, called basic spatial units, into larger geographic clusters, called zones (catchment areas), in a way that the latter are acceptable according to relevant planning criteria.

Evidently, each criterion applied to zone design performs a constraint to the optimum solution with an additional increase of processing time and may also make the problem unfeasible.

Therefore, extensive use of criteria should be avoided if the study does not require such constraints, while through organisation of the zone design methodology is essential before aggregating process.

As Alvanides notes, "it is far more important to avoid the occasional straggling or spindly shaped output zone [that] it is to insist that all zones are approximately circular in shape, which is a highly unrealistic goal".

HEURISTIC APPROACH

A useful review of the extensive literature relating to techniques for constructing aggregations of geographical areas that are in some sense optimal is provided by Duque et al [47]. These methods include several disciplines, including computational geography, and several aspects of operation research including graph theory, combinatorial optimisation, and computational complexity [48].

The zone design problem is difficult to solve computationally. The most basic approach would be to enumerate all possible solutions that could be generated, and choose the one that best fits our optimal criteria. This is not feasible in practice, because the number of possible partition is usually effectively infinite, and each partition has to be constrained geographically. This means that a medium size problem of 53 b.s.u partitioned in 7 zones yields $1.22*10^{41}$ possible solutions [49]. Additionally, in terms of computational complexity the zone deign problem has been shown to be NP-Complete [50]. Because of that most methods in the literature employ a heuristic approach for finding an approximate solution in a reasonable time.

The heuristic approaches employed include:

- Clustering algorithms which merge only contiguous areas.
- Starting from seed areas and adding neighbours.
- Starting from a feasible initial solution and swapping zones in and out.
- Graph theoretic approaches.

These techniques of creating optimal geographies are usually called 'supervised regionalization methods' [47]. Over the last decade different software have been developed to aggregate a collection of input areas into a smaller number of spatially contiguous regions while optimising an objective function [51] , [52], [53]. Until recently a majority of this programs could not deal with flow or spatial interaction data [54], [55]. Moreover, their application in the health domain are relative lacking.

Therefore, the objective of this research is to fill these gaps and define a new generation of PCSA (RPCSA) that incorporates the common properties and objectives of a catchment area.

METHODOLOGY

The literature review leads us to make the approach to our problem in light of three key issues:

- Does the PCSA model can easily be generalized without jeopardizing their definition?
- If the first issue is resolved or if there is a way to get around it, are there efficient algorithms to solve this problem?
- Does the RPCSA model useful in practice and in decision situations?

In order to answer these questions, this research will:

- Formulate a model able to compare any two geographical areas in terms of socioeconomic, demographic and infrastructure variables.
- Identify peer group geographies for reporting the geographic variation of health indicators.
- Formulate a model to define a RPCSA.
- Identify and implement an optimization method to solve the problem.
- Analyze and evaluate the computational effort required and the quality of the results obtained with different scenarios.
- Compare the conclusions drawn by studying geographic variation using RPCSA and using other geographic boundaries, such as SA3.

In the next section I'll illustrate how to measure the diversity with Larger Areas in order to identify appropriate peer groups for PHN and enable fairer comparison within and across these service areas.

Then we introduce the problem of the Rational Primary Care Service Area design as a multi-objective partitioning problem and the Health Workforce Supply/Demand planning.

PEER GROUPING

Australia is a diverse country settle by a diverse population. To examine either the behaviour or characteristics across the entire population for findings regarding a subset of the people living in a specific area, it is crucial to have a complete count of the residents in that area.

‘’The population census is the most significant source of demographic, social and economic information about entire populations at national, regional and local scale. One of the most fundamental tenets of census taking is the total *confidentiality of the data* collected and the delinking of it from any way of identifying the individual from whom it was collected. Accordingly, census data are only made available in *aggregates* of *subgroups* or the population of geographical areas. In both academic and policy-related research it is important to use spatial units which are meaningful in relation to the particular issue under investigation. Traditionally, administrative units have been employed as the b.s.u for both research and planning. However, there has long been a questioning of the meaningfulness of administrative boundaries for many areas of social, economic and health planning’’ [56].

There are two ways in which geography of census data can be made more relevant to the needs of the user:

- Adopting a *census geography* which includes spatial units which are socially, economically and environmentally meaningful; and

- Adopting *b.s.u* as the smallest areas for which census data are available which are small enough to facilitate analysis being able to aggregate these building blocks into their own purpose-specific regions.

Australia has recently revamped its census geography based on the above considerations. A new Australian Standard Geographical Classification (ASGC) was developed by the Australian Bureau of Statistics (ABS) Figure 16. In addition to these regions the ABS also provides data for some non-standard areas, postcodes, suburbs, etc. The boundaries of these areas are based on meaningful social, economic and environmental relationships so that there is a degree of **homogeneity** in each of the units.

In the case of the SA3s the main criteria for the delimitation of their boundaries are [57]:

- POPULATION: SA3s are designed to have populations between **30.000** and **130.000**
- FUNCTIONAL: SA3s are often the functional areas of regional towns and cities with a population in excess of 20.000 or clusters of related suburbs around urban commercial and **transport hubs** within the major urban areas.
- IDENTIFYING REGIONS: the regional breakups have been designed to reflect regional identity. These are areas with both **geographic** and **socio-economic similarities**.

Basically, as stated by ABS, the aim of SA3s is to create a standard framework for the analysis of ABS data at the regional level through clustering groups of SA2s that ***have similar regional characteristics***.

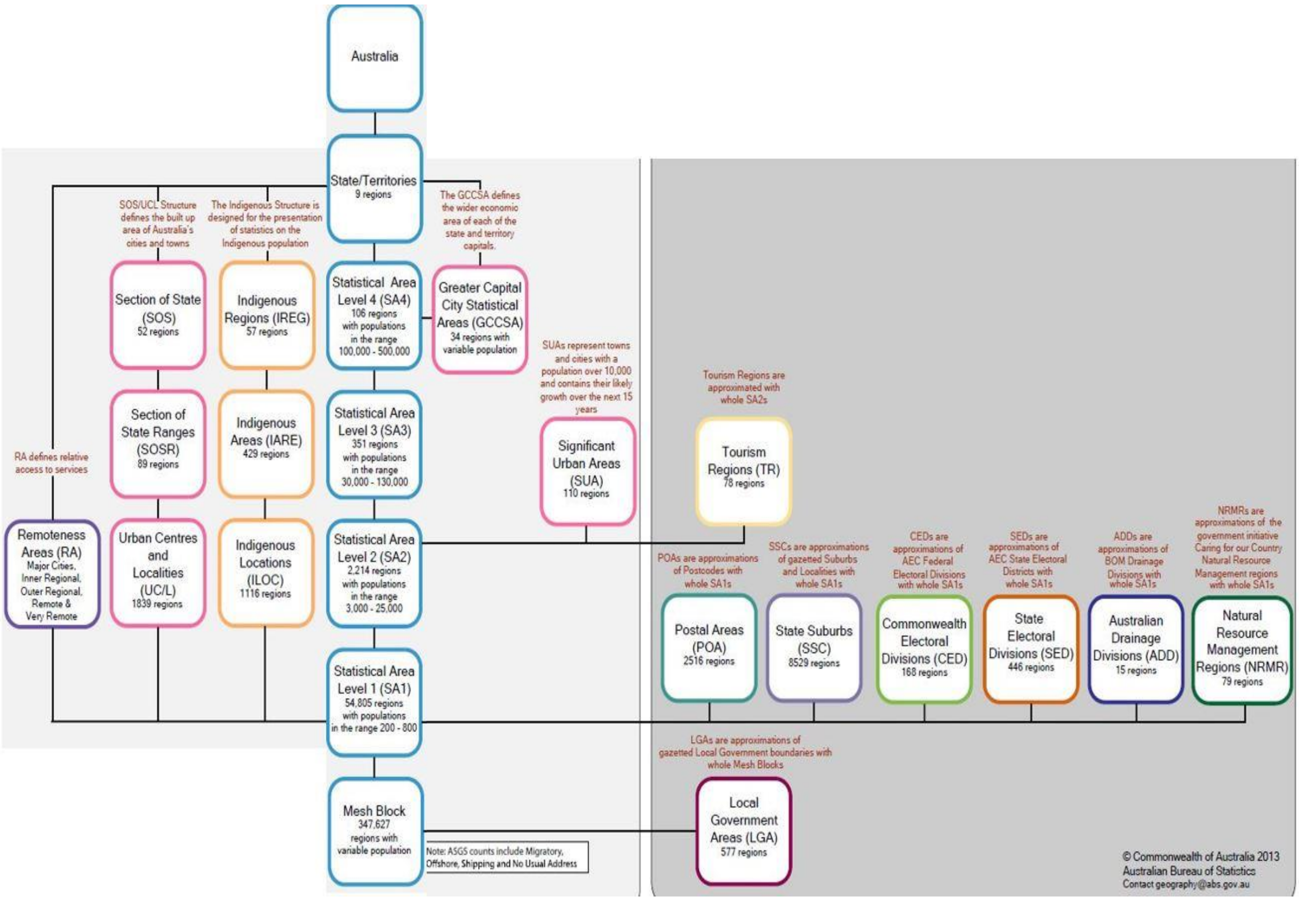


Figure 16 Australia Statistical Geography Standard (ASGS): structure and summary for Census

This leads us to the following research questions:

1. How can we define the concept of homogeneity?
2. Is the SA3s Geography effectively homogeneous?
3. How can we compare any two SA3s?

In order to answer these questions, it is necessary to:

1. Identify the relevant demographic, social-economic and geographic attributes needed to evaluate the degree of homogeneity in a specific area.
2. Then we move on to the definition of homogeneity and discuss the limitations of the different approaches employed in recent years. Therefore, we propose a novel

methodology to overcoming the drawbacks of these techniques and evaluate the degree of homogeneity for the SA3s Geography.

3. Finally, we formulate a metric to compute the dissimilarity between any two SA3.

HOMOGENEITY DIMENSIONS

The spatial distribution of population and settlements across a country and their interconnectivity and accessibility from urban to remote areas are important for delivering healthcare, distributing resources and economic development [58].

Human settlement is multidimensional. Coombes and Raybould [59] have identified three key dimensions to modern human settlement patterns which are all important for policy-makers to take into account when they are allocating resources or designing programmes. These are:

- *Settlement size*: ranging from metropolitan to hamlet;
- *Concentration*: ranging from dense to sparse; and
- *Accessibility*: ranging from central to remote.

In their view, these three dimensions need to be recognized and measured individually. However, in examining demographic, economic and social behaviour it may be that variation in people's level of accessibility to services such as education and health is more influential than whether or not they live in an urban area. As indicated in Figure 11, accessibility is a complex variable incorporating not only physical elements but also socio-economic differences in access to transport, cultural and other factors. While researchers are aware of the importance of both aspects in assessing health care, often the two types of factor are studied separately [60] [27]. Therefore, it is essential to develop a methodology that embodies the dual nature of access.

In this research, we propose a valuable framework [Figure 17] for reporting the PHNs Geographic Variation by SA3 level.

Basically, the model has three variables Table 4:

DIMENSION	VARIABLE	CATEGORY
GEOGRAPHIC	ARIA+	SPATIAL
DEMOGRAPHIC	AGE DISTRIBUTION	NON SPATIAL
SOCIO-ECONOMIC	SEIFA	

Table 4 PHN Framework SA3 variables

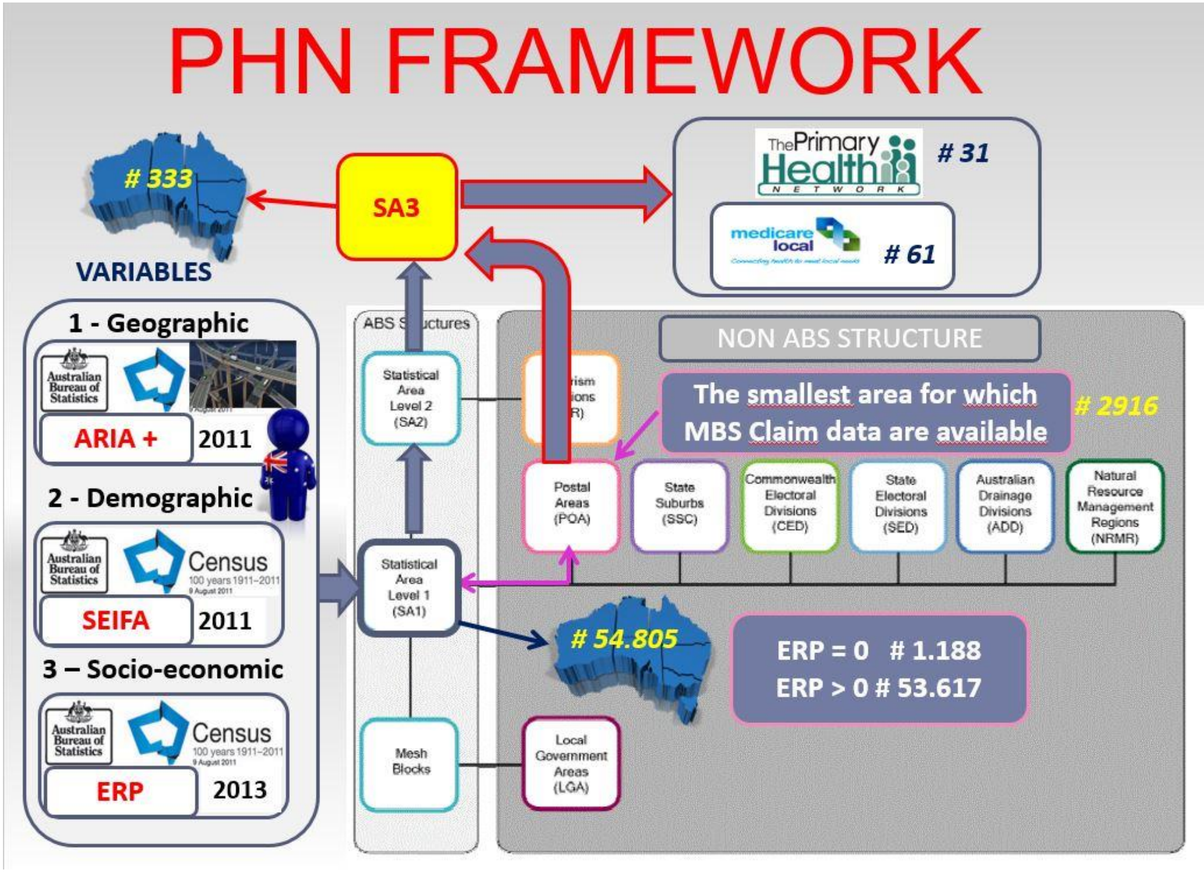


Figure 17 Primary Health Network Framework

GEOGRAPHIC

To measure the physical accessibility of a specific area, the ABS has included an accessibility classification as part of its geography.

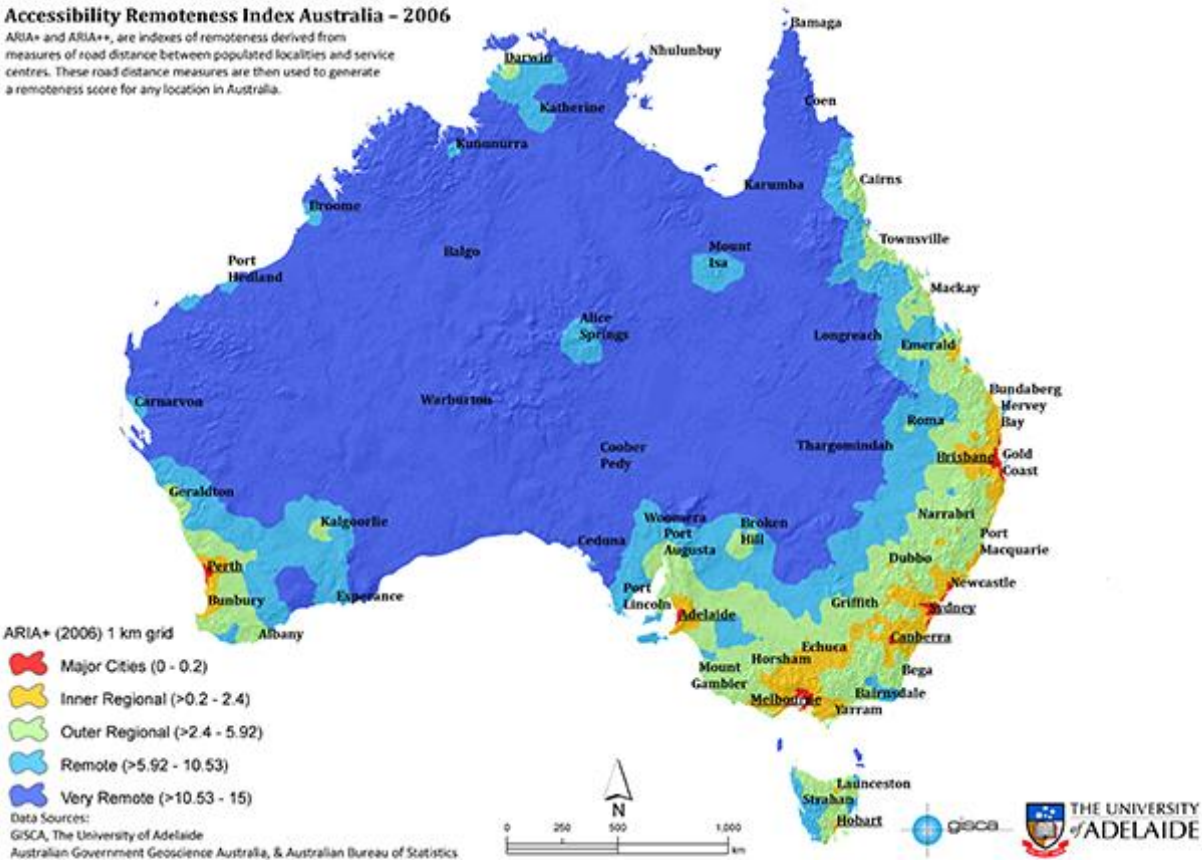


Figure 18 ARIA+. Accessibility Remoteness Index for Australia.

The Remoteness Index was initially developed at the University of Adelaide to quantify accessibility in non-metropolitan parts of Australia [61]. It is based on 11 338 population localities identified on the 1:250 000 topographic map series and uses road distances measured between each of these and the nearest of four different level of service centre. Figure 18 shows the distribution of remoteness categories and the classification is reported in [62] Figure 19.

Australia Remoteness Category	
Major Cities **High Accessible**	Locations with relative unrestricted accessibility to a wide range of goods and services and opportunities for social interaction.
Inner Regional **Accessible**	Locations with some restrictions to accessibility of some goods, services and opportunities for social interaction.
Outer Regional **Moderately Accessible**	Locations with significantly restricted accessibility of goods, services and opportunities for social interaction.
Remote	Locations with very restricted accessibility of goods, services and opportunities for social interaction.
Very Remote	Locational disadvantaged – very little accessibility of goods, services and opportunities for social interaction.

Figure 19 Australia: Remoteness Category

DEMOGRAPHIC

The age group distribution is based on the Estimated Residential Population Census of 2013 from ABS. The selection of the demographic variables affecting health needs is the same used by the ABS to cluster the Medicare Local [63] Table 5.

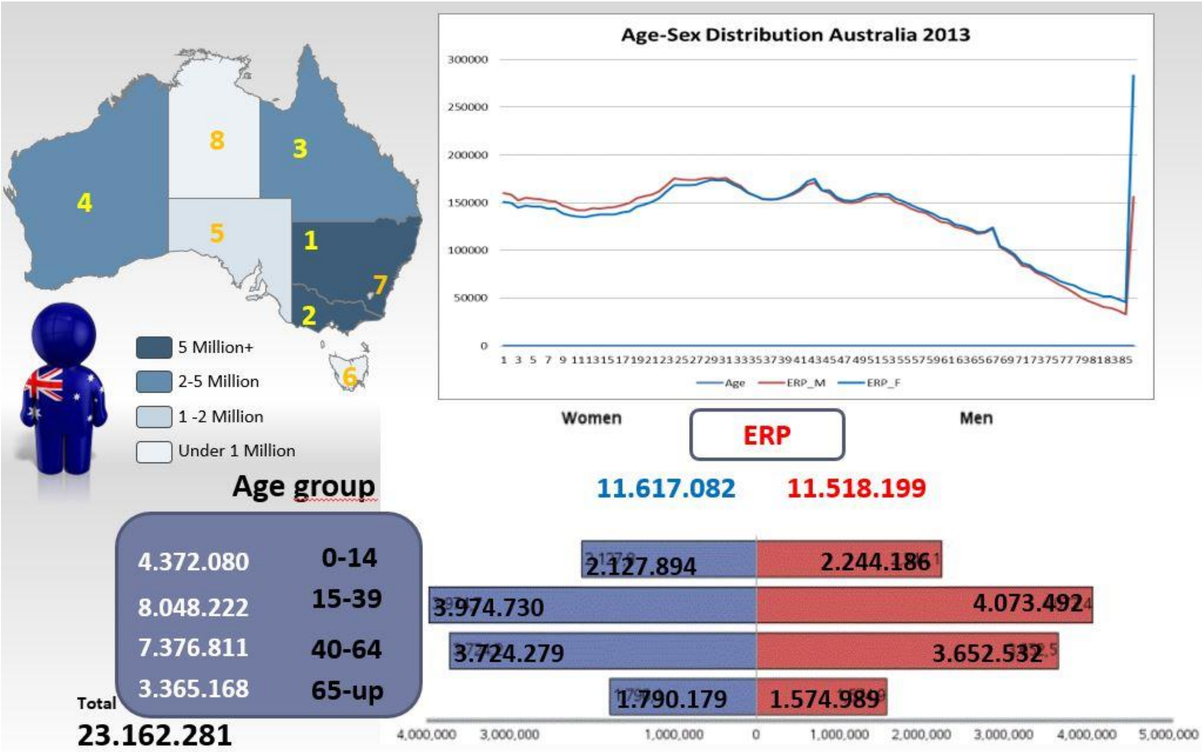


Table 5 ERP Australia - Age and Sex distribution 2013

SOCIO-ECONOMIC

For the socio-economic variables we propose the use of the Index of Relative Socio Economic Disadvantage (IRSD) [64]. The main reasons for selection of this approach are:

1. The same index has been used by the ABS to cluster the Medicare Local; and
2. It includes relevant variables to evaluate the economic and social conditions of people and households within an area.

Variable mnemonic	*Variable loading*	*Variable description*
INC_LOW	–0.90	% People with stated annual household equivalised income between $1 and $20,799 (approx. 1st and 2nd deciles)
CHILDJOBLESS	–0.85	% Families with children under 15 years of age who live with jobless parents
NONET	–0.81	% Occupied private dwellings with no internet connection
OCC_LABOUR	–0.75	% Employed people classified as 'labourers'
NOYR12ORHIGHER	–0.75	% People aged 15 years and over whose highest level of education is Year 11 or lower. Includes Certificate I and II
UNEMPLOYED	–0.74	% People (in the labour force) unemployed
LOWRENT	–0.73	% Occupied private dwellings paying rent less than $166 per week (excluding $0 per week)
ONEPARENT	–0.71	% One parent families with dependent offspring only
DISABILITYU70	–0.66	% People aged under 70 who have a long–term health condition or disability and need assistance with core activities
NOCAR	–0.56	% Occupied private dwellings with no cars
SEP_DIVORCED	–0.54	% People aged 15 and over who are separated or divorced
OVERCROWD	–0.52	% Occupied private dwellings requiring one or more extra bedrooms (based on Canadian National Occupancy Standard)
OCC_DRIVERS	–0.52	% Employed people classified as Machinery Operators and Drivers
OCC_SERVICE_L	–0.50	% Employed people classified as Low Skill Community and Personal Service Workers
NOEDU	–0.44	% People aged 15 years and over who have no educational attainment
ENGLISHPOOR	–0.34	% People who do not speak English well

Table 6 IRSD variables

For example, an area could have a low score if there are (among other things) many households with low income, many people with low qualifications or many people in low skill occupation.

This index is preferred in situation where the user wants to look at disadvantage and lack of disadvantage. An example would be where a user wants to ensure an allocation of funds goes to disadvantage areas. [65].

HOMOGENEITY INDEX

The multidimensional nature of Access poses different issues when we explore the diversity within larger geographical areas.

For larger geographies, such as SA3, a IRSD is created from the population weighted average of the SA1 scores within the larger area [64]. However, a single score for an area does not take into account the socio-economic diversity within that area. This index is an area based level measure, means it will mask some diversity at finer level of disaggregation. “It is incorrect to say that a person is a very disadvantage person if they live in a very disadvantage area. It is true that living in a very disadvantage area may disadvantage them to a certain extent, but it is possible that they are advantaged in many other respects such as having a good education and earning a high income, and are thus not typical of other residents in that area”. The diversity is important especially when comparing two larger areas.

Related to the issue above, it is usually desirable to use the smallest geographic unit possible when merging an index to another dataset such as ARIA+ and ERP. In the case of IRSD, the SA1 is the smallest unit available, thus when we conduct any sort of analysis is suggested [64] to consider using the distribution of SA1.

In view of the previous consideration, it is natural to ask the following question:

- Why do not we use the SA1 level to analyse the Geographic Variation?

The main reason is to prevent disclosure of information about individual or communities [66]. Identify disclosure occurs when someone learns something that they did not already know about another person, organisation or community through disseminated data [either from summarised data (e.g. tables of findings) or unit record data (microdata) released to researchers by agencies such as ABS, AIHW, DOH].

Maintaining an individual's confidentiality when using health data is a primary concern. One of the main concern is when characteristic about an individual can be inferred from statistical patterns or models (e.g. if regression model estimates are released, predicted values of an individual can be inferred). As a response to this issue, data can be rendered unidentifiable through aggregation. In this scenario, analysis is completed using SA1 geography, but results are aggregated results for reporting and map publication.

Considering that we must be aware of health-related data problems associated to individual confidentiality; it is fundamental to explore possibilities to assess the diversity within a geographic level higher than SA1 (e.g. SA3, Post code), by looking at its constituent SA1s.

Therefore, having a comprehensive view of the SA1s'distribution is of central importance to define the homogeneity within a larger area. This goal can be achieved by decomposing the different dimensions of access in multiple layers.

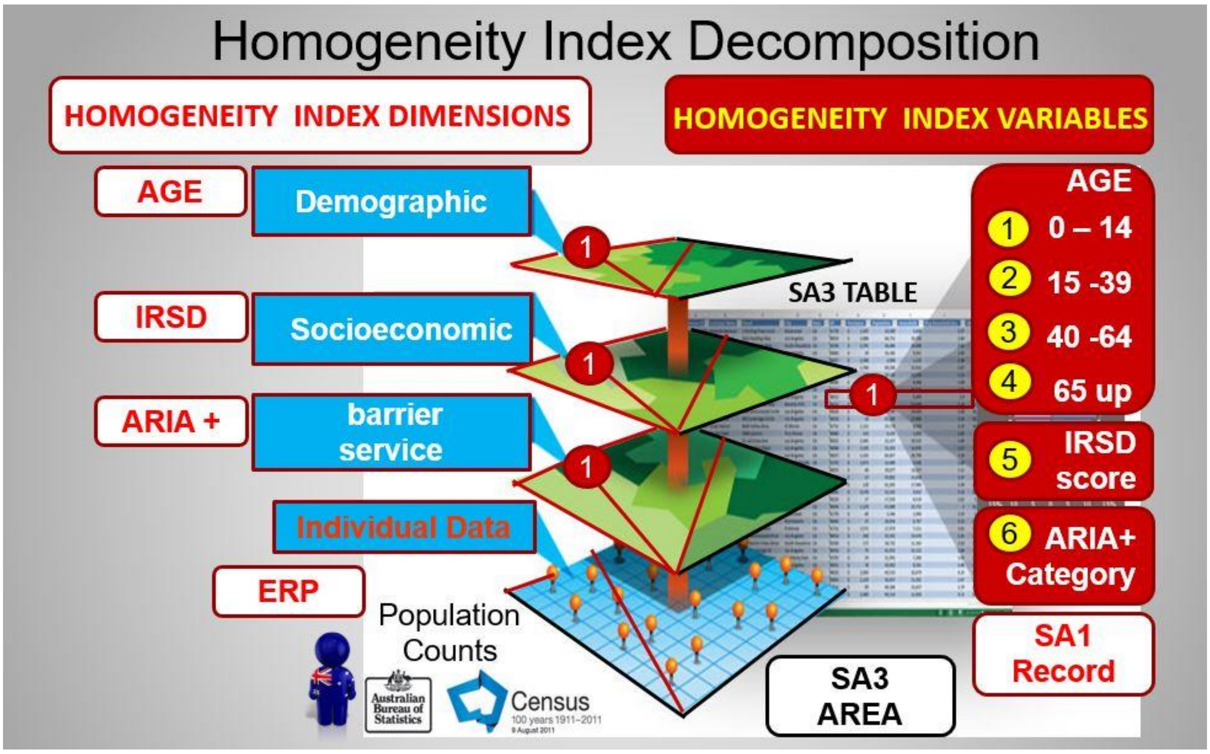


Figure 20 Homogeneity Index Decomposition

The Figure 20 illustrates the study area of a generic SA3 partitioned in three SA1s. Each SA1 is associated with a record of six values related to the dimensions (layers) of the Homogeneity Index.

However, it is important to remember that the IRSD scores are an ordinal measure, so care should be taken when comparing scores. For example, an area with a score of 500 is not twice as disadvantage as an area with a score of 1000, it just had more markers of relative disadvantage. For ease of interpretation, the ABS recommended using the index rankings and quantiles (e.g. percentiles, deciles) for analysis, rather than using the index score.

The Decile measure is defined below:

- Deciles: All areas ordered from lowest to highest score, the lowest 10% of areas are given a decile number of 1, the next lowest 10 % of areas are given a decile number of 2 and so on, up to the highest 10% of areas which are given a decile number of 10. This means that areas are divided up into ten equal sized groups, depending on their score.

Other commonly used quantiles include quintiles and quartiles; which quantile to use in analysis depends to the specific geography and variables.

Therefore, an appropriate selection of the best interval is crucial to identify a specific pattern in the distribution, or stated differently: The number of intervals into which we should divide the full range of values.

Too many intervals result in a raggedly shape distribution picture, which displays too much detail and makes it difficult to discern the essential pattern; while too few intervals summarize the data to a level that is too generalized, resulting in a loss of meaning full variations in the distribution's shape.

Therefore, the goal is to find the right balance between too many and too few intervals to display the best possible picture of the distribution's overall shape without losing sight of useful details.

In this context of the SA3 Peer grouping, we propose to use decile as a descriptive measure of a distribution. The justification of this choice can be explained looking at the IRSD score distribution.

Observing Figure 21 and Figure 22, it is clear that the first decile has a large spread of scores compared to the other deciles. This means that using quintiles or quartile, if there is specific interest in identifying disadvantaged SA1s, the likelihood to mask the characteristics of larger geographies is high.

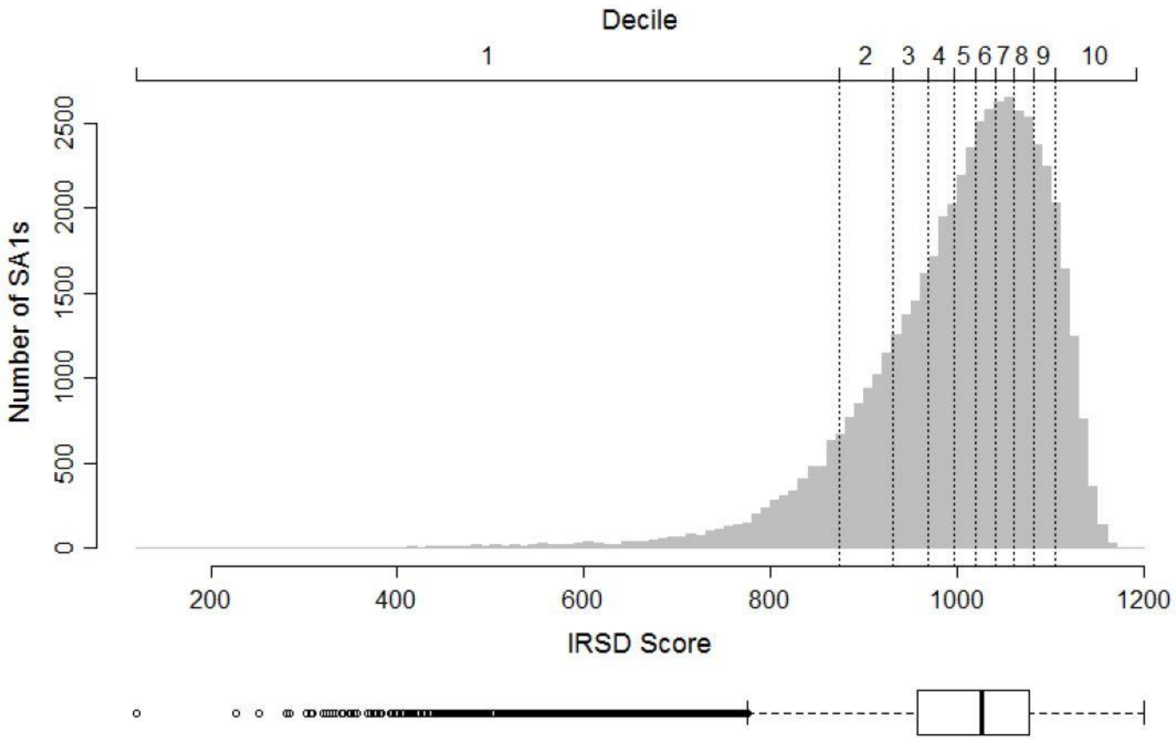


Figure 21 IRSD SA1 Distribution

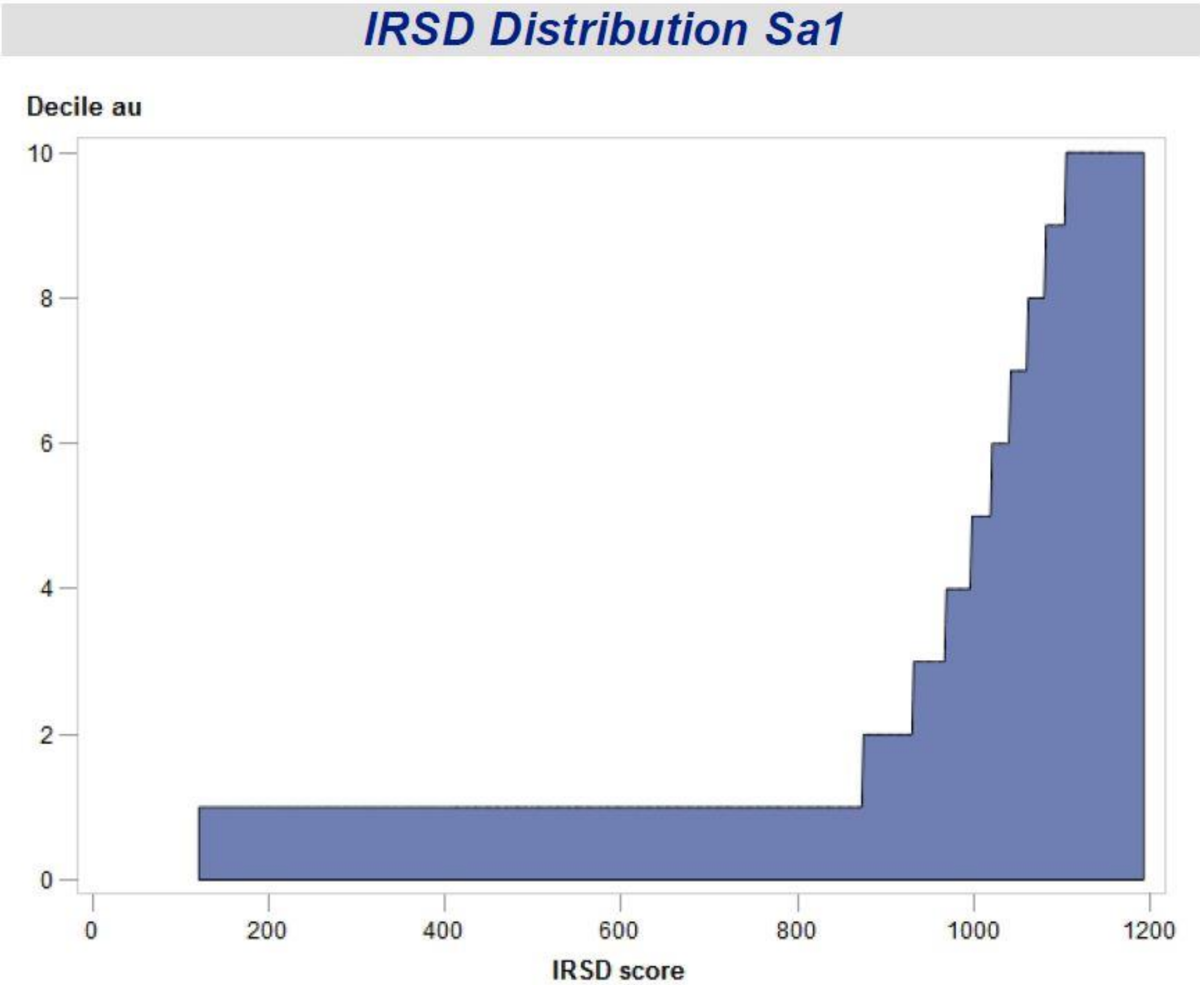


Figure 22 IRSD Decile Distribution spread

In view of the above explanation, it is quite plain that the total number of variables for a SA3 is one hundred Figure 23 (i.e. $10 \cdot 5 \cdot 4 = 200$), if we exclude the areas without IRSD score. For example, in the case of the IRSD dimension we have ten variables (11 with the excluded SA1). How can we plot the distribution of the IRSD based on the ERP? This can be accomplished by superimposing multiple thematic maps (population counts (ERP) and IRSD score).

The system collects the information about the area of interest as a collection of various thematic areas that can be linked together by the geography. This approach allows us to organise the complexity of the real world into a simple representation of spatial relationships.

Individual features such as the number of people living in a specific geography (SA1 ERP Figure 24) and the socio-economic characteristics (SA1 IRSD Figure 25), all these features from separate maps, represent different data from the same portion of the territory from which they come from (SA1). Consequently, we can combine this information and integrate them into a single layer. The resulting map [Figure 26] shows the IRSD SA1s decile distribution in Goulburn – Yass. The distribution reveals that the population living in this SA3 is not homogeneous.

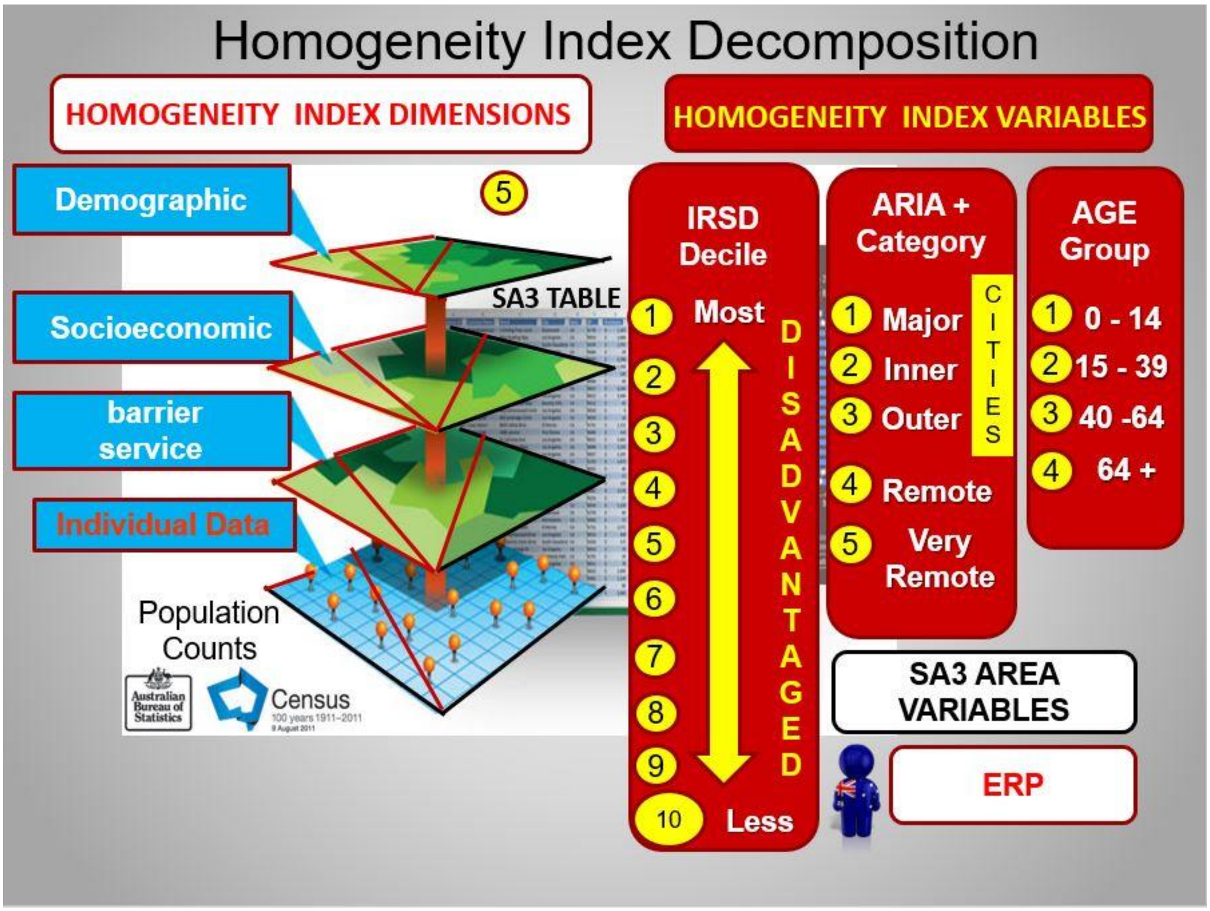


Figure 23 Homogeneity Decomposition SA3

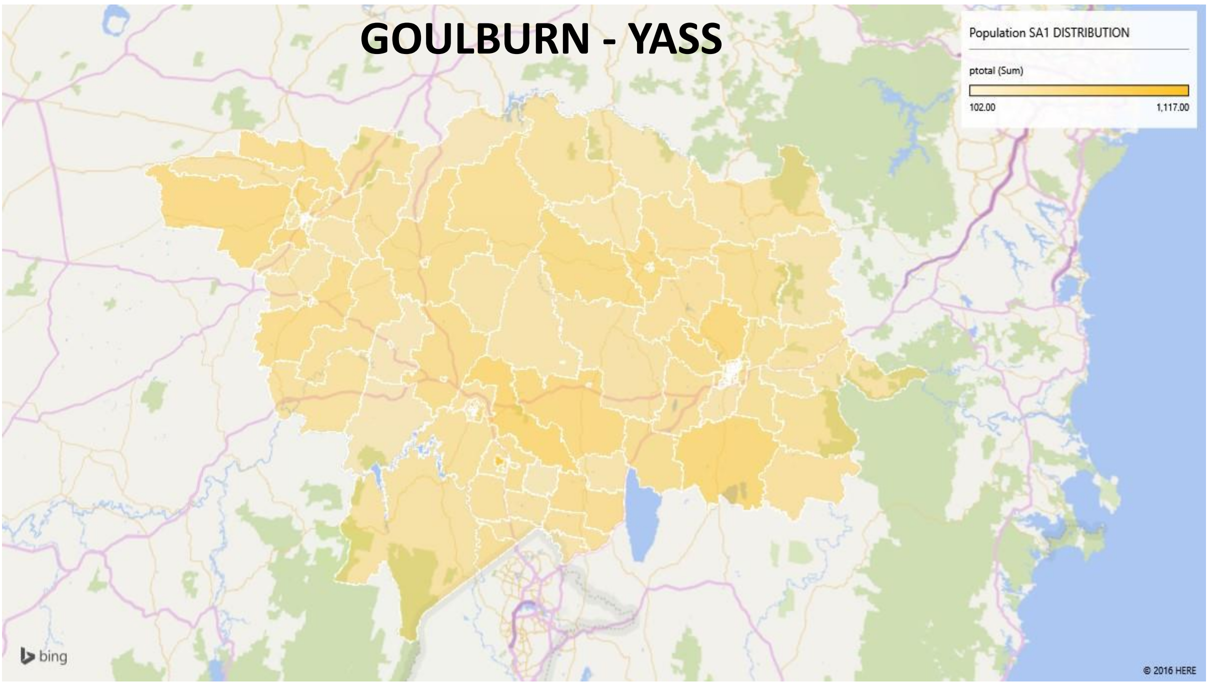


Figure 24 Goulburn-Yass SA1 ERP Distribution

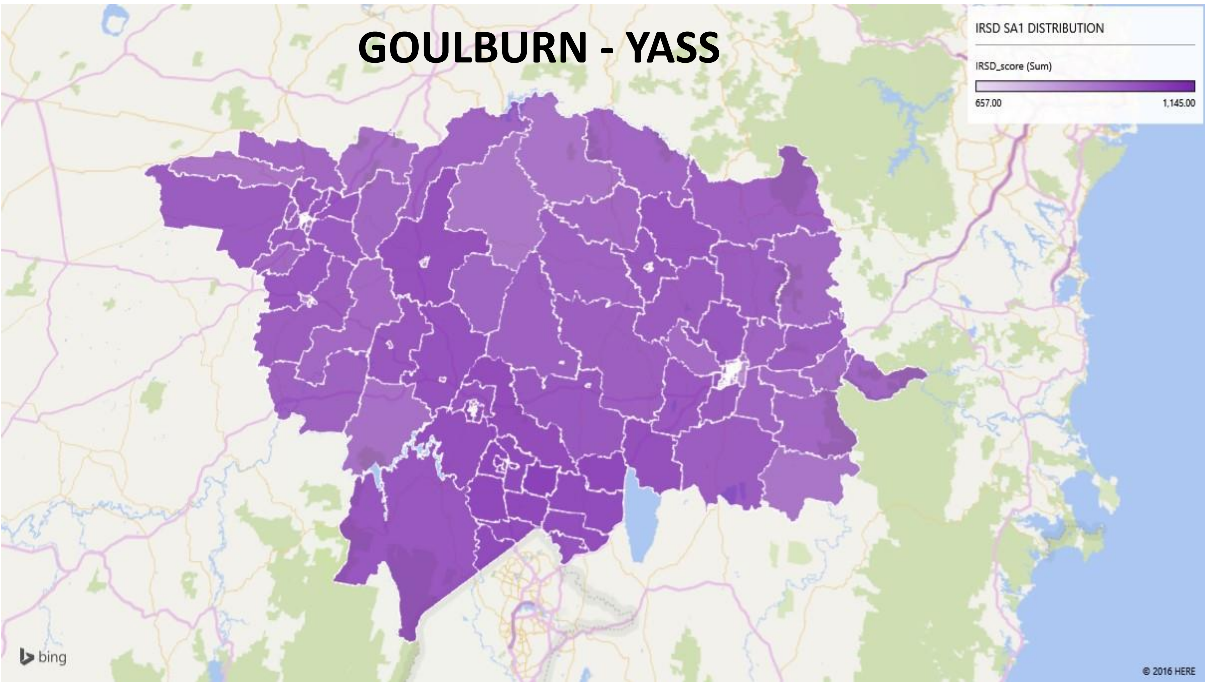


Figure 25 Goulburn - Yass IRSD population Distribution

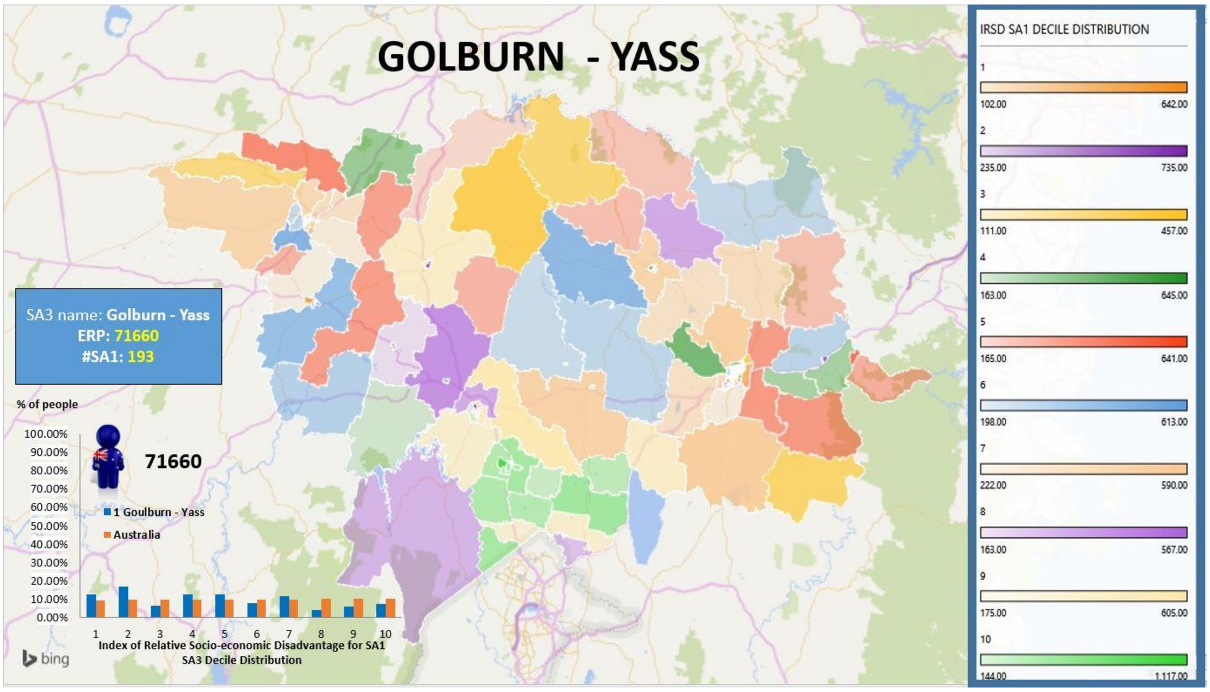


Figure 26 Golburn Yass ERP vs IRSD

Therefore, an important question is:

- **How can we define the Homogeneity of a distribution?**

Peter Radish and Wise [67] conducted an investigation on this issue and proposed a measure based on the SA1s concentration. The definition is based on the selection of a 'cut-off' for the distribution of SA1 level, such as the first decile or the first two deciles. Then, they calculate the proportion of SA1s within each larger area which falls within this cut-off.

For example, if the number of SA1s in the first decile within each larger area is less than the national value (0.1), indicates an under-representation of the SA1s in the most disadvantaged 10 % of the Australia-wide SA1 distribution. Similarly, a value greater than 0.1 represents an over-representation. This provides a natural benchmark for this measure in terms of defining what is a 'high' or 'low' concentration of SA1, but fails to address the issue in question.

Moreover, their definition is based on the SA1 concentration rather than the Estimated Residential Population and this can generate confusion, especially if we want to compare two SA3s in terms of their population. Even if we consider the population concentration we cannot conclude anything about the homogeneity of the entire area.

By looking at the histogram of Goulburn-Yass, we can definitely say that we have an over-representation of the population living in the first decile, but it does not tell you that in this area there are people with very different characteristics Figure 27.

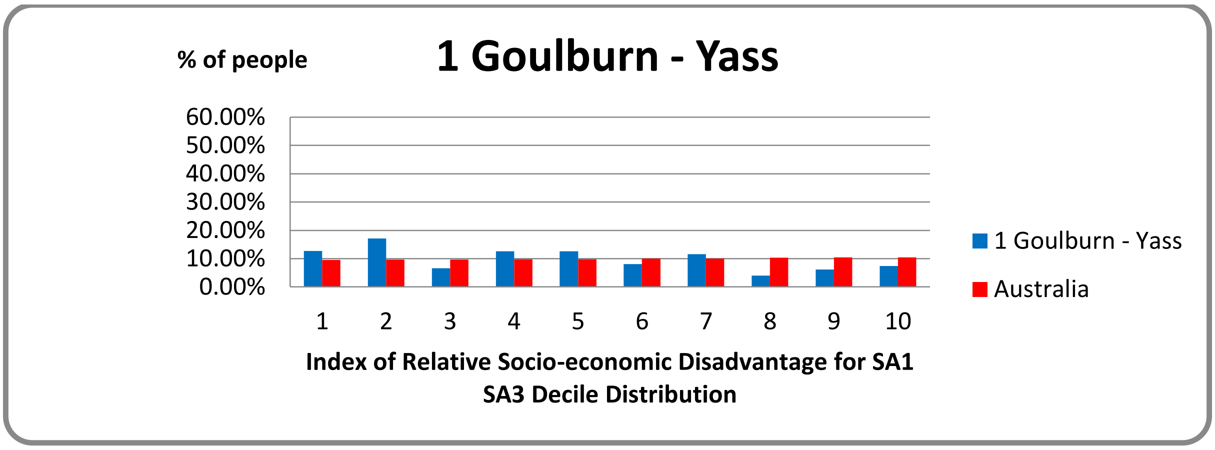


Figure 27 Goulburn - Yass IRSD Population Decile Distribution

An alternative measure is the population-weighted between area variance (PWVar) [68]. This is a standard variance, weighted to take account of differences in the population size of zones across different geographical configurations. PWVar is calculated over all the zones in a given geography and produces a single value. For a given homogeneity variable x, higher value of PWVar correspond to high between-area variances, and hence a lower within-area variance (and correspondingly high internal homogeneity). For more homogeneous areas, than, PWVar will be grater [69]. However, this method provides a global analysis of the partition and does not still solve the problem of evaluating the homogeneity of a distribution.

Another commonly used measure to numerically quantify the unequal distribution of social groups across areal units that comprise a city or country, is the Dissimilarity Index. However, this index is cumbersome to compute and interpret when the number of groups exceeds two. Thus, if one wants to generate an overall measure of segregation between ten social groups (e.g. decile), separate dissimilarity indices would have to be computed between all possible pairs of groups and averaged to get a single measure [70] .

A possible solution would be to apply the entropy Index [71], which can also be expanded to more variables simultaneously. However, we prove that in some cases this index is sensitive to only big changes in the scale and hence might be inappropriate in the case of clustering.

In general, we want an index easy to interpret and used with more than two variables, but above all able to capture the heterogeneity of the whole area of study. For example, in the case of Australia the index should returns the lowest value of homogeneity.

How can we achieve all these objectives simultaneously?

A possible solution to this problem is to classify the distributions' shape.

In the next section we provide a novel technique to classify and quantify the homogeneity of a discrete distribution.

SHAPE

The primary visual characteristic of a distribution is its shape, which shows us where the values are located throughout the spread. If we examine distributions closely, the number of possible shapes they can form is infinite.

Histograms do a good job of displaying the overall shape of a distribution while also making it easy to compare the magnitudes of individual intervals. What we cannot see, however, it's a measure of the distribution's centre, nor can we precisely see the distribution's spread.

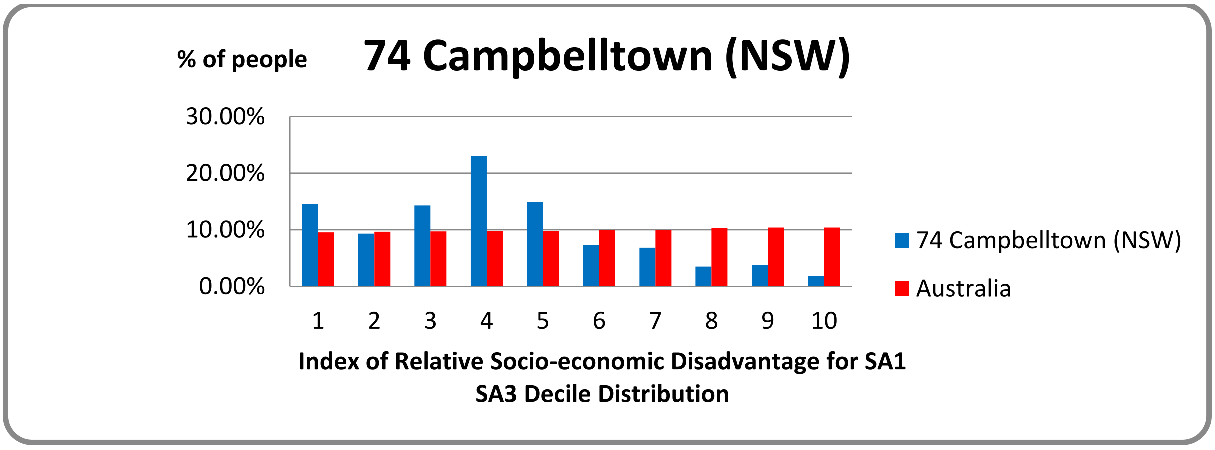


Figure 28 Campbelltown IRSD SA1 decile distribution

By looking out the histogram of Campbelltown Figure 28 , can you determine the median or the mean? Is Campbelltown a disadvantaged area or not? What are the Least or Most disadvantaged SA3s in Australia?

In order to answer to these questions, we propose a model that summarizes the complete set of values with two single numbers:

- Homogeneity Index (H): is a measure of the concentration of the samples. When the people are evenly distributed in all the bins (e.g. decile), we have minimum homogeneity. Conversely, when all the population is concentrated in a single bin, we have the maximum homogeneity.
- Concentration Location Index (CLI): In addition, an important information is where these values are noticeably higher than others. A possible solution is to classify the different distribution in terms of the IRSD deciles. For example, an index with a value equal to one means a very disadvantaged area. Conversely, a value equal to ten indicates a very wealthy area.

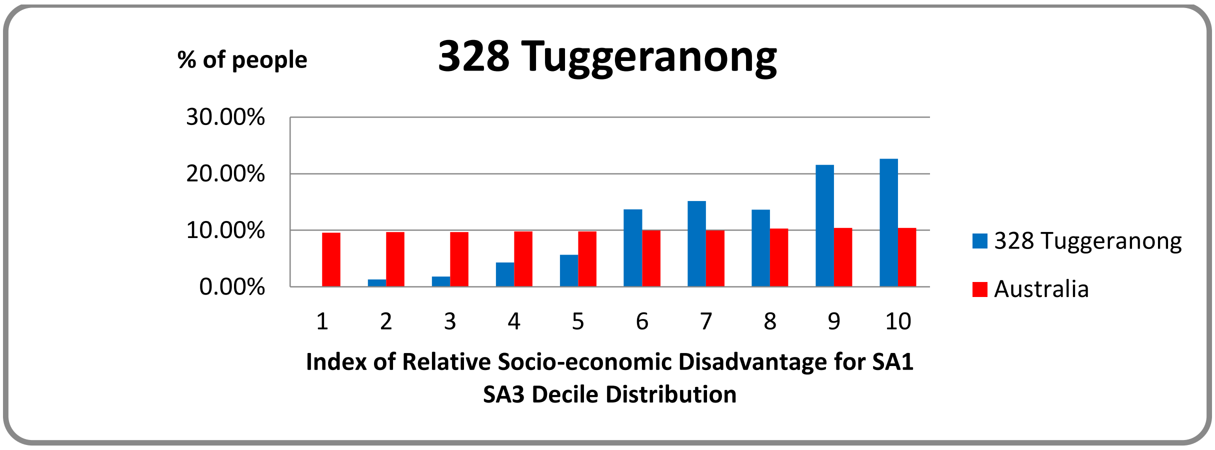


Figure 29 Tuggeranong IRSD ERP Decile Distribution

The easiest way to define the CLI is to identify the peak of a curve, but concentration don't always correspond to a predominant peak. For example, in the distribution of Tuggeranong Figure 29 there is a predominant peak on the right so the distribution is skewed to the left, but there are also high concentrations of values near the middle. So does the distribution belong to the last decile or not?

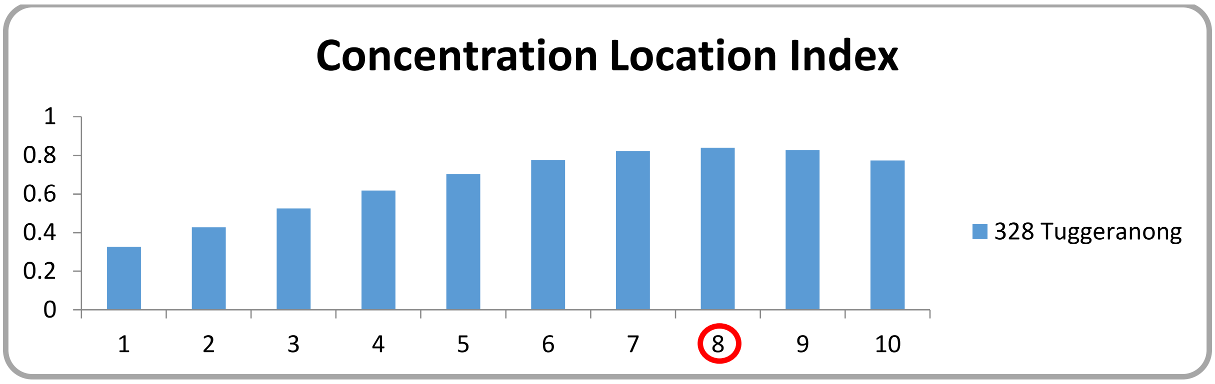


Figure 30 Tuggeranong IRSD Concentration Location Index

To classify the level of disadvantaged of a specific area we defined a function (the metric is based on the topological property of compactness) that ranks the single bins in terms of the distribution's concentration Figure 30. However, this is not enough as this metric can only give a score (1 to 10) to a distribution but it does not tell you if the population is evenly distributed or not. Thus, we introduced the Homogeneity function.

This function combines two terms to evaluate the degree of concentration and compactness of the samples in a distribution.

- Concentration Index: The first term quantifies the global concentration of the samples in a distribution. Basically, it is similar to the Gini Index with the substantial difference that can be visualized and applied to a multivariate distribution. It assumes a value equal to zero in the case of the uniform distribution and a value equal to one for a singleton (all the samples in a single decile).
- Compactness Index: This index quantifies the local compactness of the samples compared to a central value (Concentration Location Index). We have the maximum compactness in the case of a singleton, and the minimum compactness when the samples are evenly distributed or there are gaps in the values of the distribution.

The Concentration Index is a generalization of the Lorenz Curve. Max O. Lorenz [72] has introduced a special tool to visualize data. He represented the distribution of relative data (that is, data over their mean) by a curve, his celebrated Lorenz curve. We stress three important characteristics of this approach:

1. The Lorenz curve is visual: it indicates the degree of disparity “the more the curve is bent”.
2. Up to a scale parameter, the Lorenz curve fully describes the underlying distribution.
3. The pointwise ordering of Lorenz curves provides an ordering of disparity, the Lorenz dominance.

Generally, the Lorenz curve measures sort of scatter or dispersion with respect to a single variable. However, in many applications it is natural to consider several variables instead of a single one and to measure their dispersion in a multivariate setting.

A few attempts have been made in the literature to generalize the Lorenz Curve and Lorenz dominance for multidimensional data [73]. But besides a few rather special cases [74], none homogeneity measure have been proposed to extend the univariate Gini coefficient to a multivariate distribution. Furthermore, these definitions are difficult to interpret in a multivariate situation.

Our basic idea is to extend the usual view of the Lorenz Curve to a multivariate distribution and introduce a new definition of the Gini coefficient. Then, we discuss the limitation of this coefficient, and propose a remediation to overcome the related issue.

Therefore, the first question that comes in mind is:

What does a homogeneous/heterogeneous concentration curve look like?

The traditional approach is to generate a social gradient that runs from the poorest to the richest sections of the population, for example by dividing into deciles based on area level information, such as the IRSD, and then identifying the Lorenz Curve as being one defined by first ordering the wealth's (deciles) of the residents from the least populated to the most populated group (decile) and finally plotting the points. In addition, we plot the point (0,0) and linearly interpolate the 11 points to obtain the familiar bow shape curve.

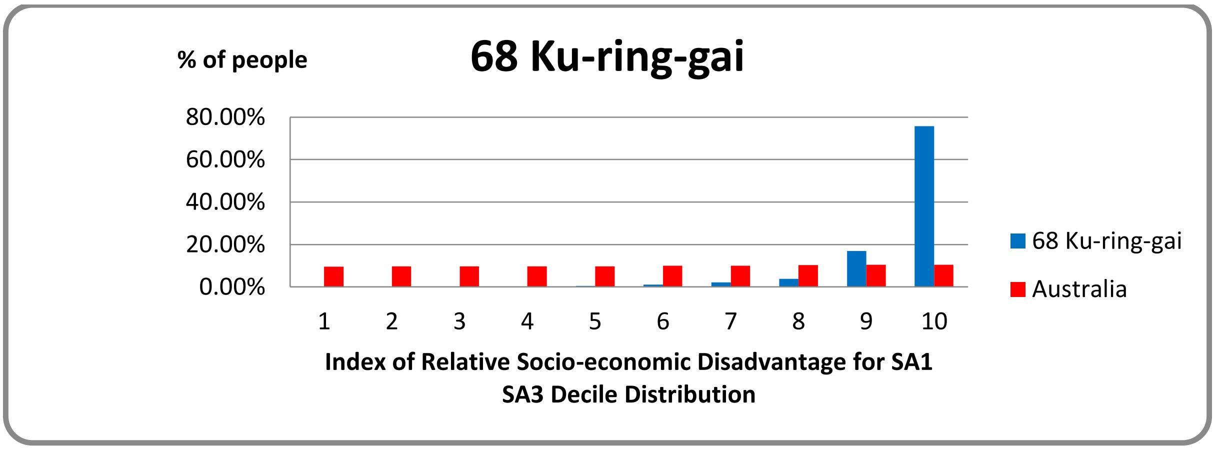


Figure 31 Ku-ring-gai Vs Australia IRSD Distribution

For example, if we compare the Australia's IRSD distribution to the Ku-ring-gai's one Figure 31, it is clear that the national Lorenz Curve is very close to the uniform distribution line (sometimes called 45-degree line) Figure 32 indicating that the concentration is evenly distributed in all the 'population groups', on the other hand if the bow it bent more for one distribution Figure 33, , than the concentration increases.

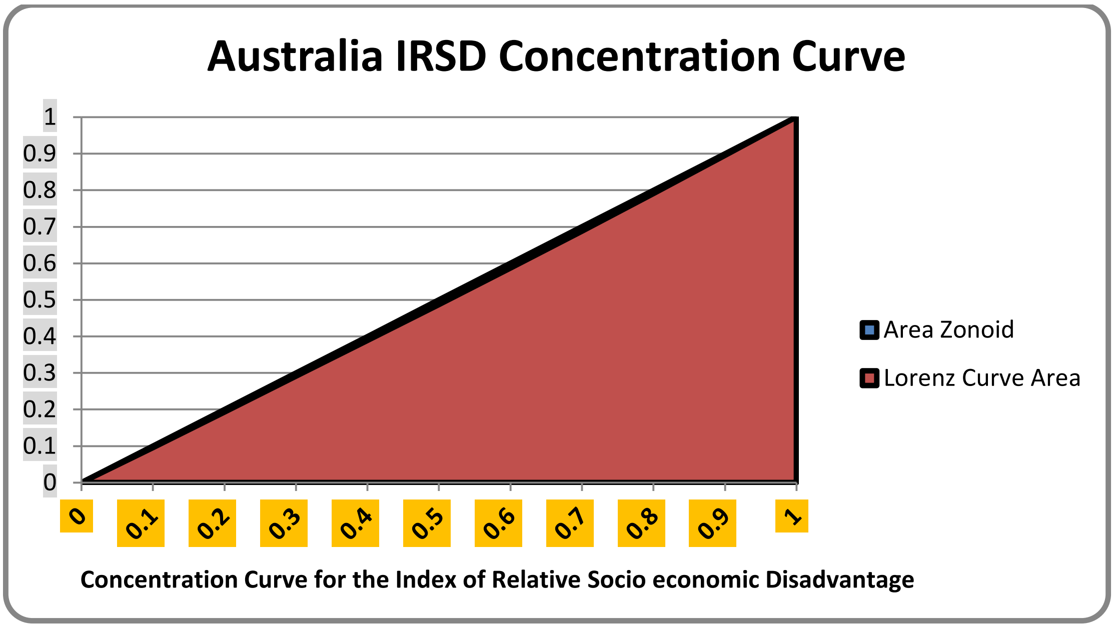


Figure 32 Australia IRSD Concentration Curve

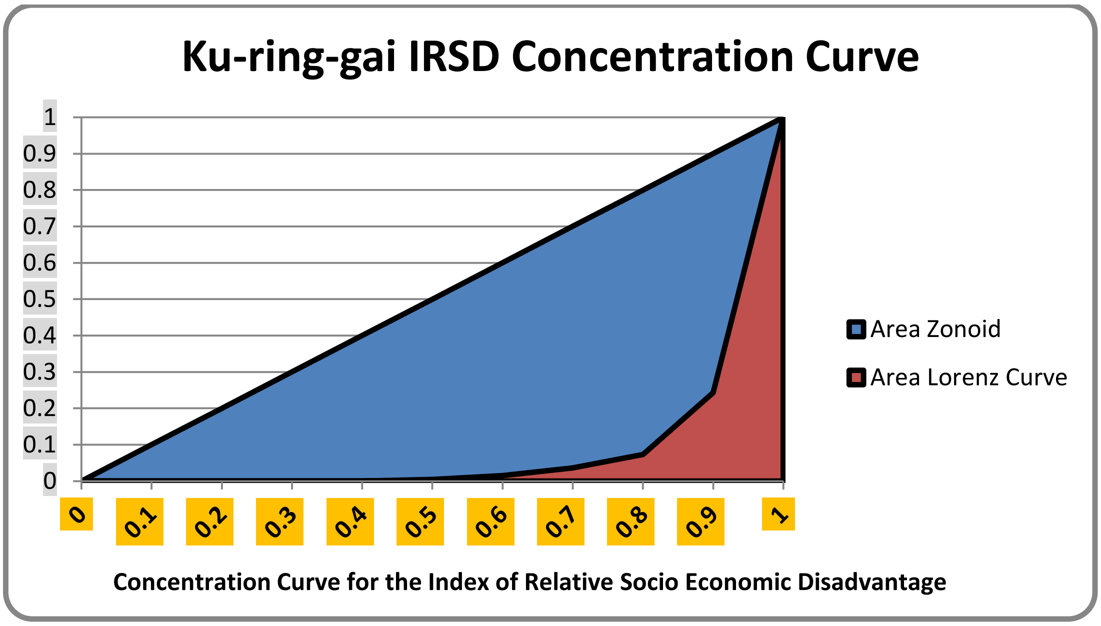


Figure 33 Ku-ring-gai IRSD Concentration Curve

Unfortunately, in the mathematical context no productive suggestions were made regarding how to provide a multivariate version of the graphical tool that Lorenz had provided in giving us his celebrated curve.

Therefore, we propose a graphical extension of Lorenz's curve in the bivariate case that can be generalized in any dimension.

The extension of the univariate Lorenz Curve to higher dimensions is not an obvious task. The three existing definitions were proposed by Taguchi [75] [74], Arnold [76], Koshevoy and Mosler [73], who introduced the concepts of Lorenz Zonoid (area between the Lorenz Curve of the distribution and the uniform distribution line) and Gini index Zonoid (the ratio between the Lorenz Zonoid and the area under the uniform distribution line).

In this work, using the definition of Lorenz and Gini Zonoid, a new formulation of the Lorenz Curve is given. This model can be easily interpreted as a convex linear combination of concentrated Lorenz Curves. Moreover, we propose an efficient way to compute the hyper-volume of Lorenz Zonoid using the integration definition of Henry Lebesgue.

The basic idea underlying this methodology is to consider the marginal distribution of the attributes separately and then plot the singles concentration curves.

In the case of Australia, if we want to plot the Lorenz Surface of the Aria+ and IRSD distribution Figure 34, we can consider the convex linear combination of these two curves Figure 35. The resulting Lorenz Zonoid volume is depicted in Figure 36.

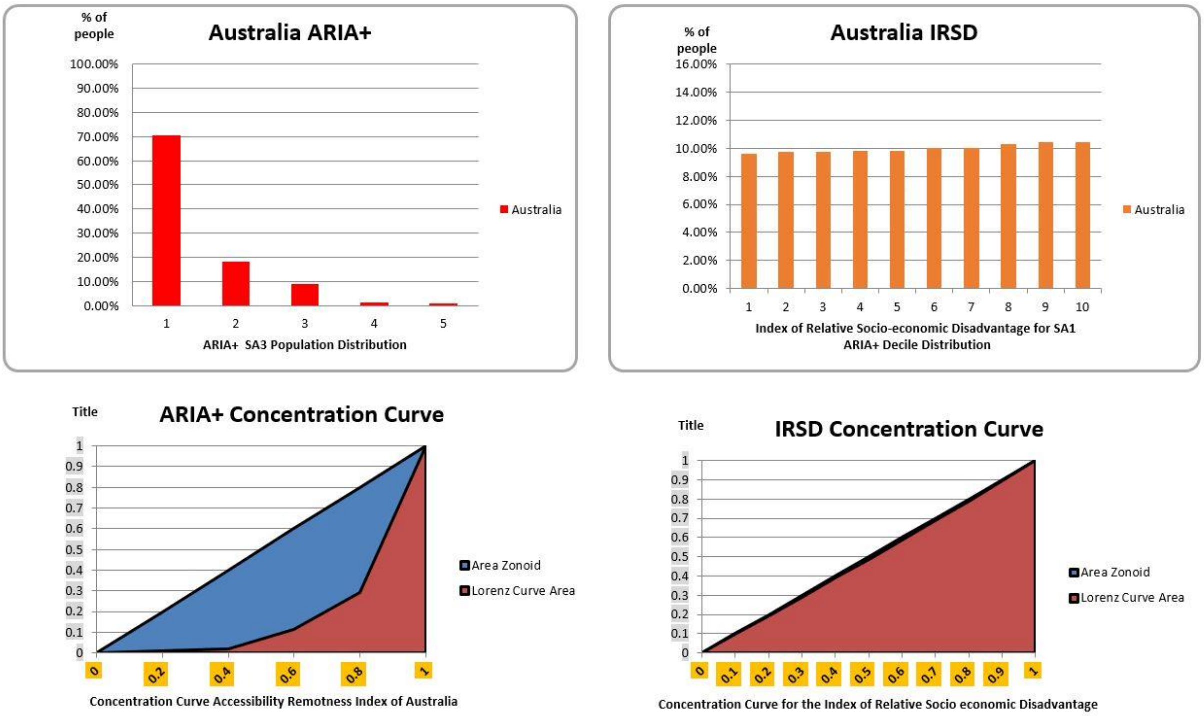


Figure 34 Australia ARIA+ & IRSD Concentration Curves

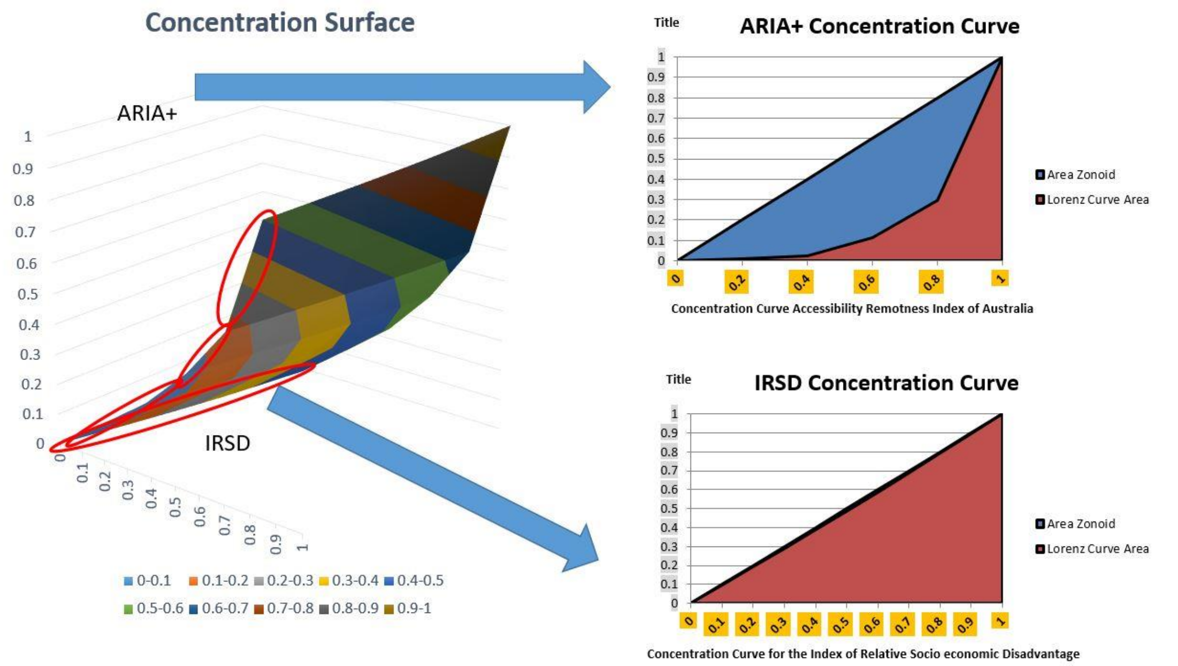


Figure 35 Australia ARIA+ & IRSD Concentration Surface

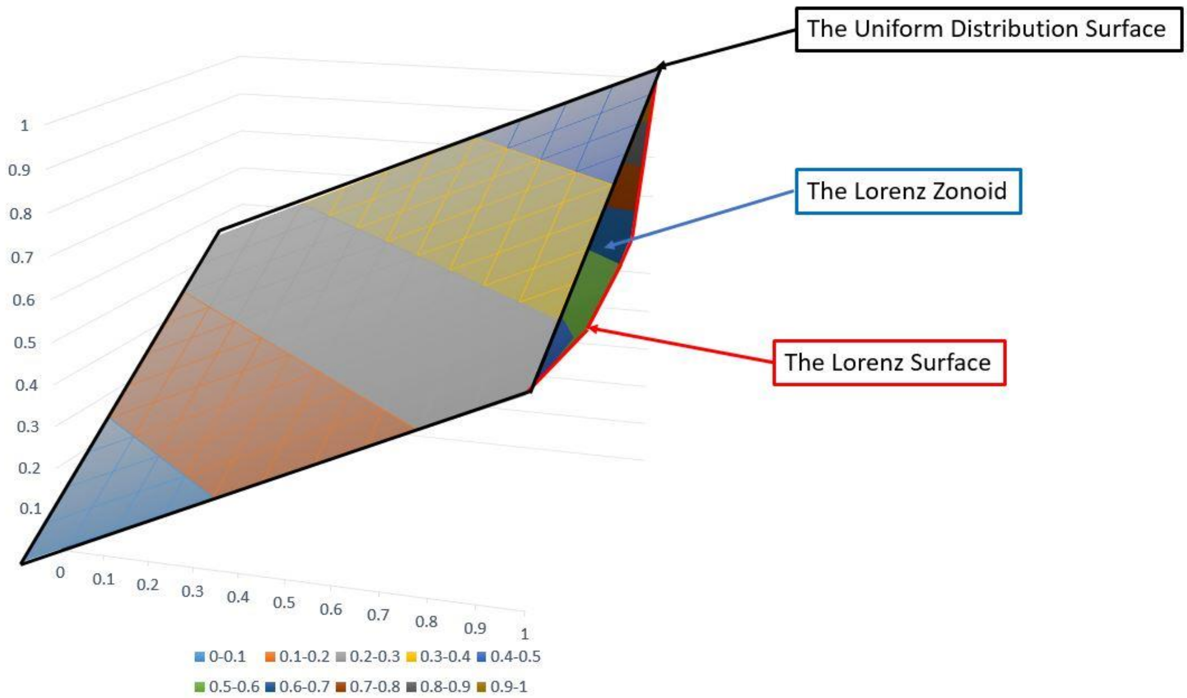


Figure 36 Australia Lorenz Zonoid ARIA+ & IRSD

However, looking only at the marginal distributions and the related concentration curves neglect possible dependencies in the value of attributes. In such a case, joint distributions have to be investigated and compared.

Moreover, although the Lorenz ordering is a way to quantify the degree of concentration in a distribution; it does not take into account the presence of gaps in the values (i.e. areas where there are few or no values relative to surrounding areas).

For example, by looking at the IRSD Decile distributions of Armadale Figure 37 and Cleveland Figure 38, they look completely different although they have roughly the same concentration curves Figure 39, Figure 40.

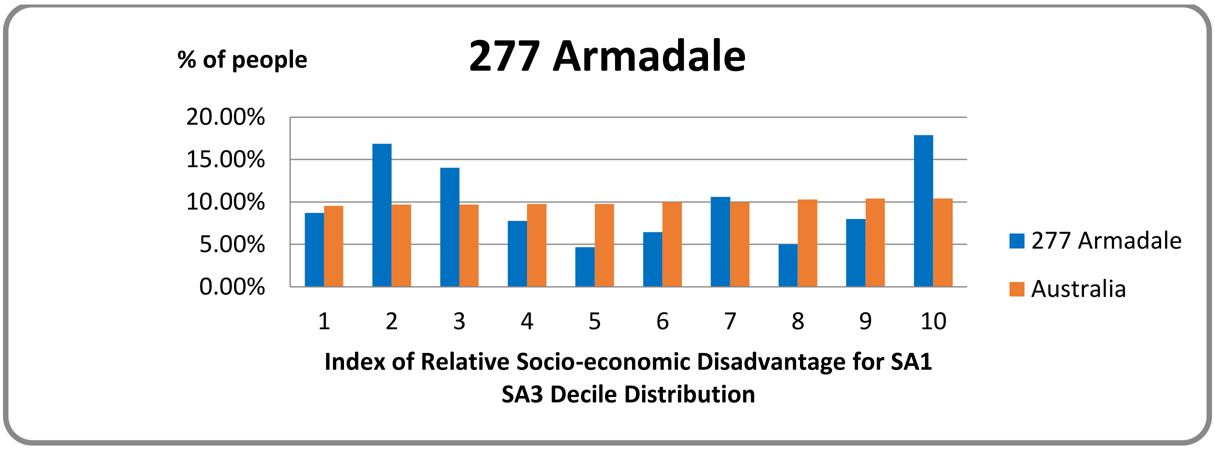


Figure 37 Armadale IRSD Decile Distribution

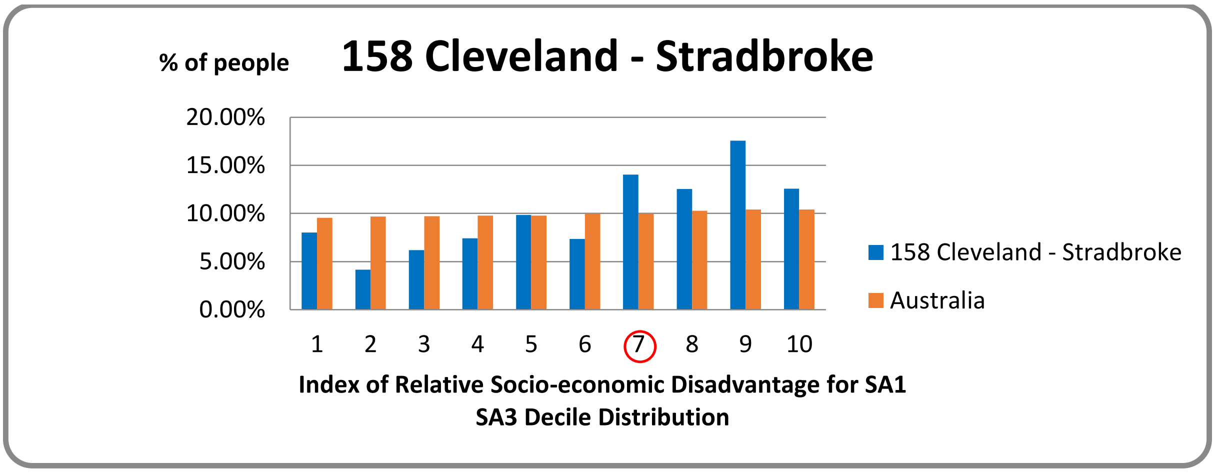


Figure 38 Cleveland IRSD Decile Distribution

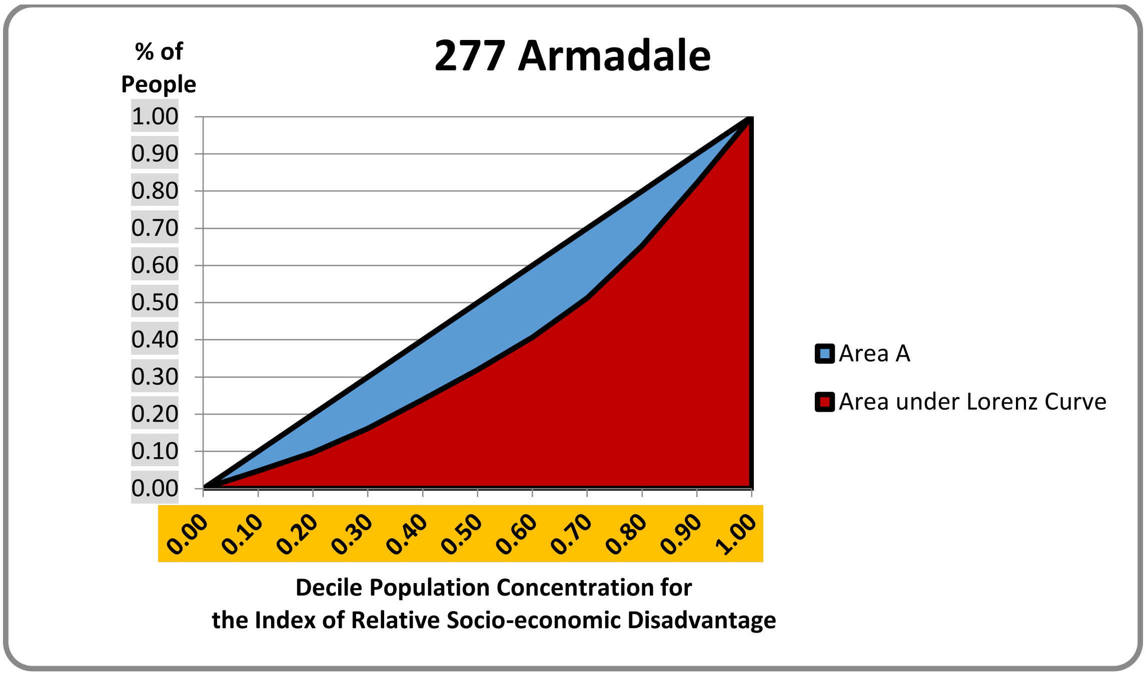


Figure 39 Armadale IRSD Concentration Curve

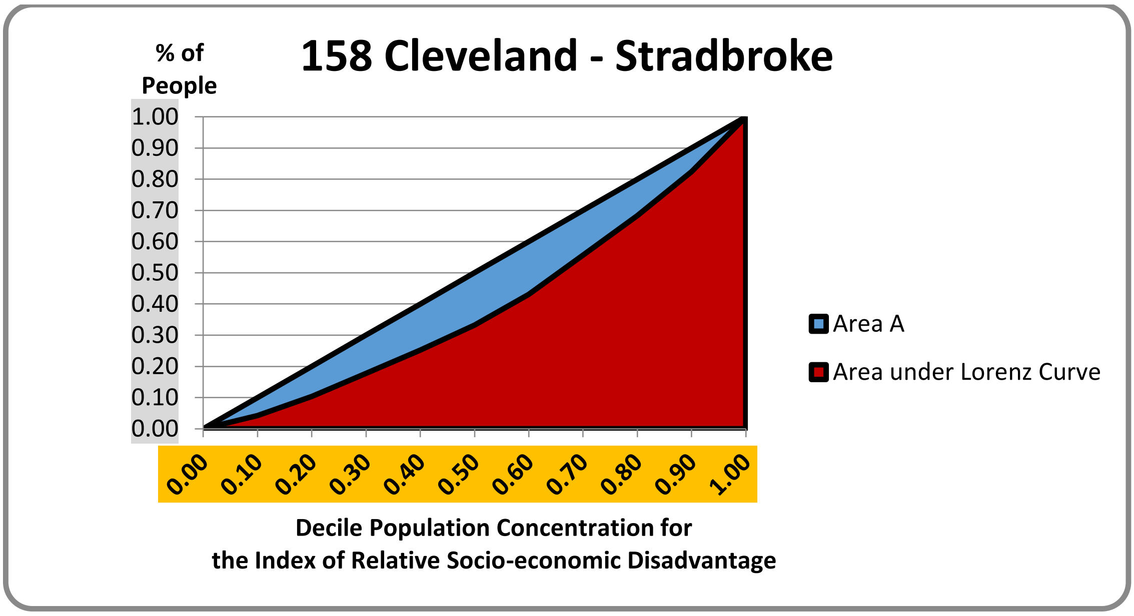


Figure 40 Cleveland IRSD Concentration Curve

How can we evaluate the compactness of a distribution around a central tendency measure such as the Concentration Location Index (CLI)?

A possible solution to this issue is to define an Index that quantifies the Local concentration of the samples (i.e. The Divergence Index). A high value of this index indicates that the samples are not tightly packed around the CLI, on the other hand a value equal to zero is representative of a singleton. This qualitative definition is based on the Jensen-Shannon divergence between the Distribution Spectra Energy of the geography and the Spectra Energy of the singleton.

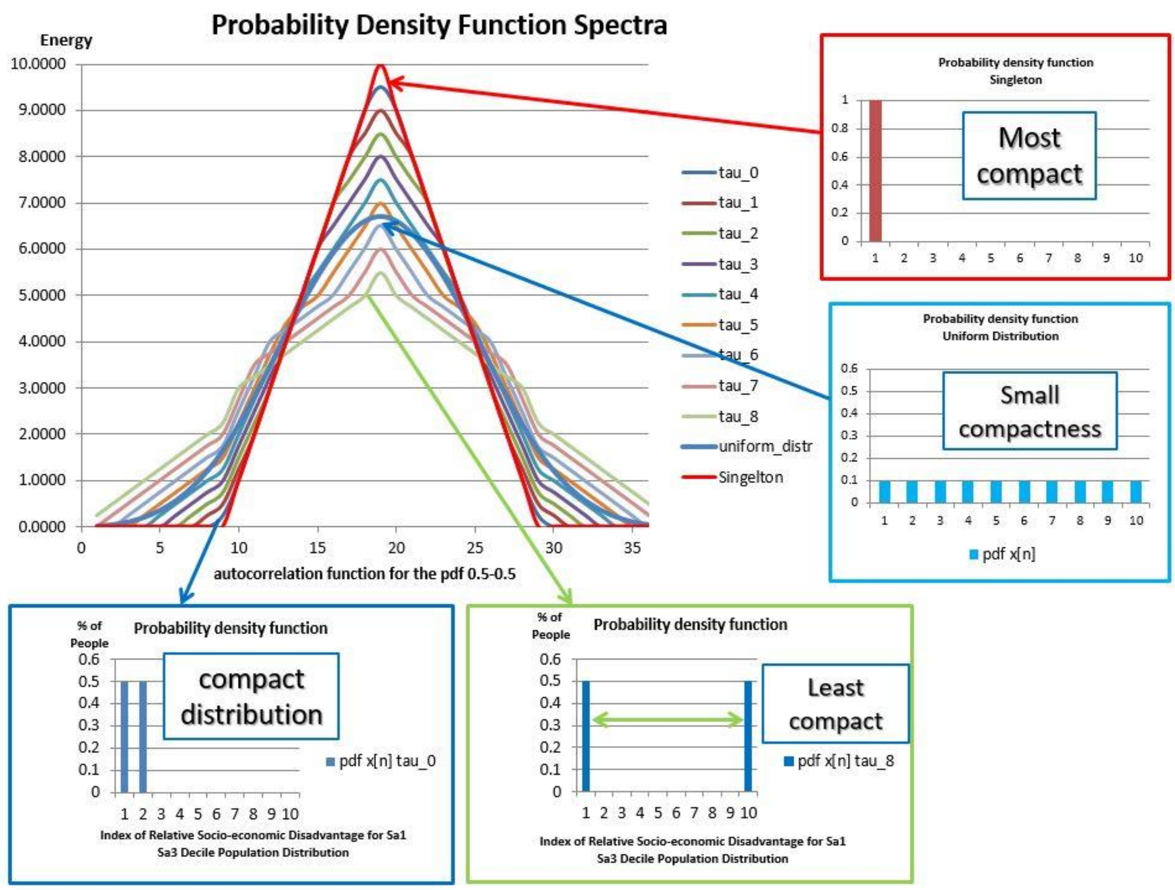


Figure 41 Spectra Energy

Figure 41 illustrates the energy spectral density of different distributions. When the energy is concentrated around a single bin (decile), the curve exhibits a high peak and a small bandwidth. Conversely, when the samples are spread out, the curve is flatter with a bigger bandwidth.

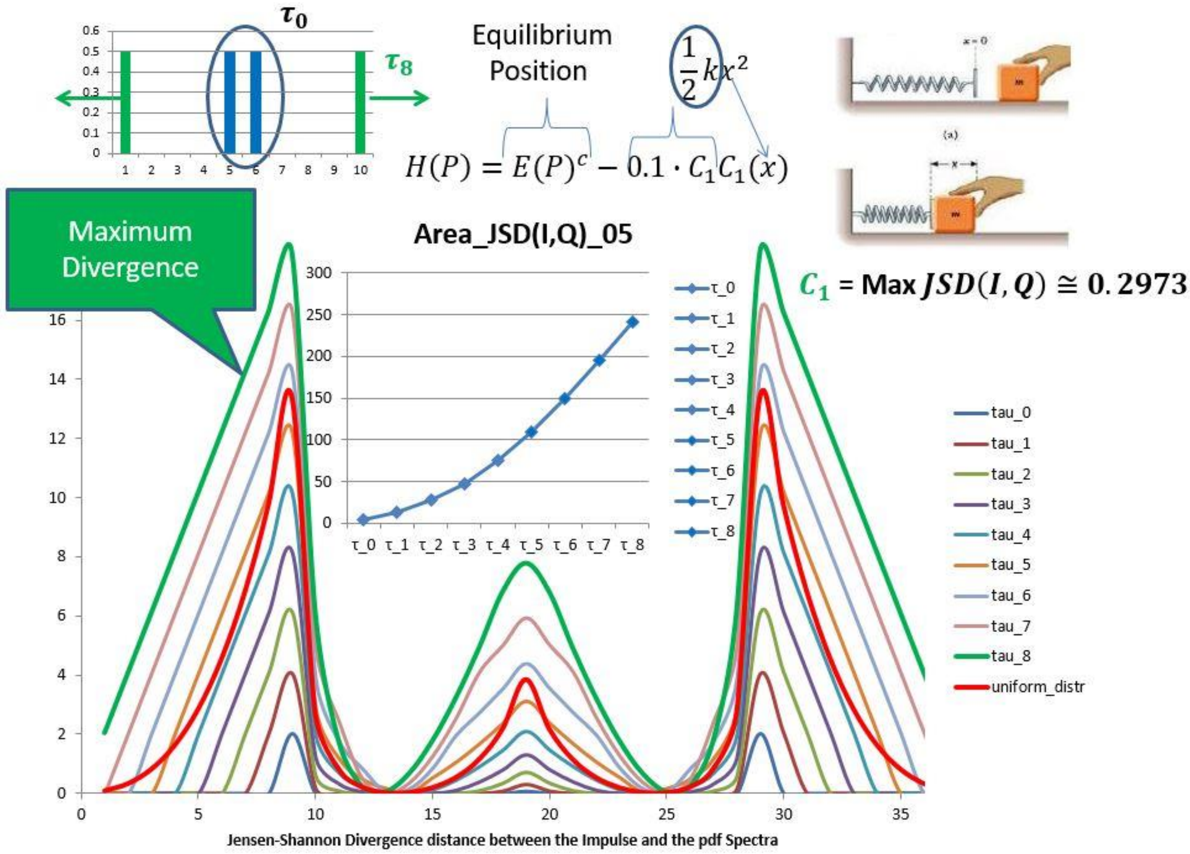


Figure 42 Spectra Energy Divergence

The configuration corresponding to the bimodal distribution is the least compact and, hence, the one with the highest Divergence Index Figure 42. This definition has been extended in the multivariate case (IRSD, AGE, ARIA+).

Figure 43 and Figure 44 depict the Divergence Index in the case of the Uniform and Bi-modal distribution for the variables ARIA+ and IRSD.

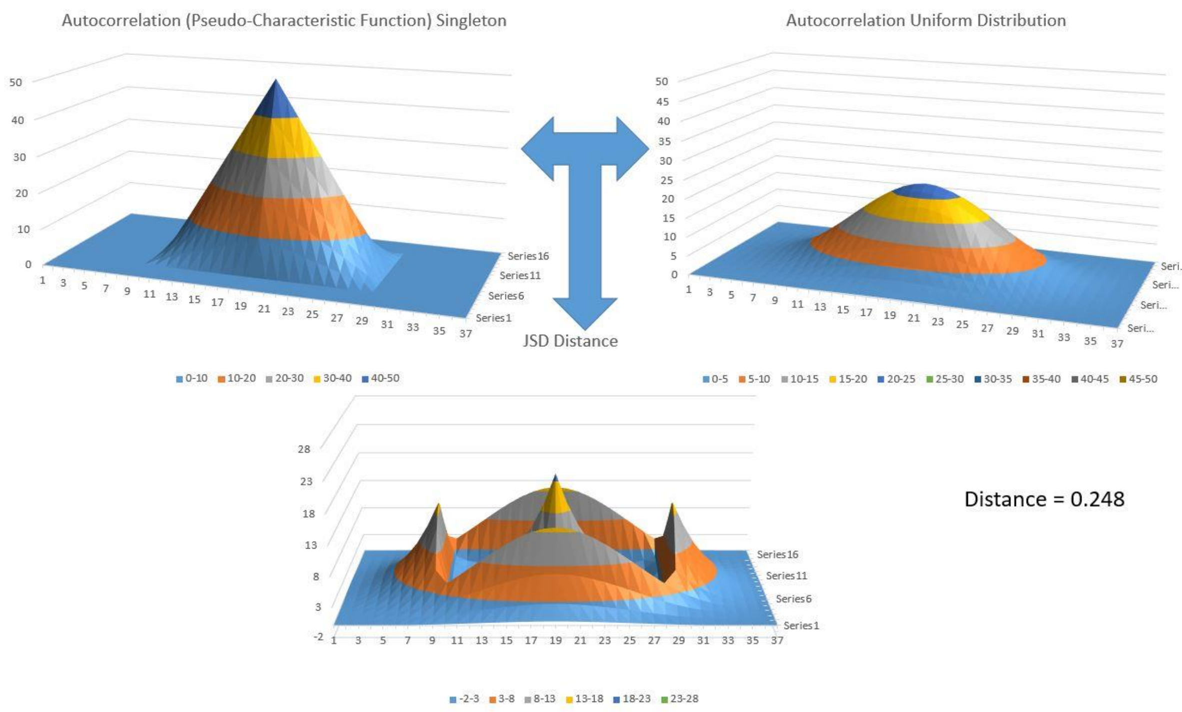


Figure 43 IRSD & ARIA+ Uniform Distribution Divergence

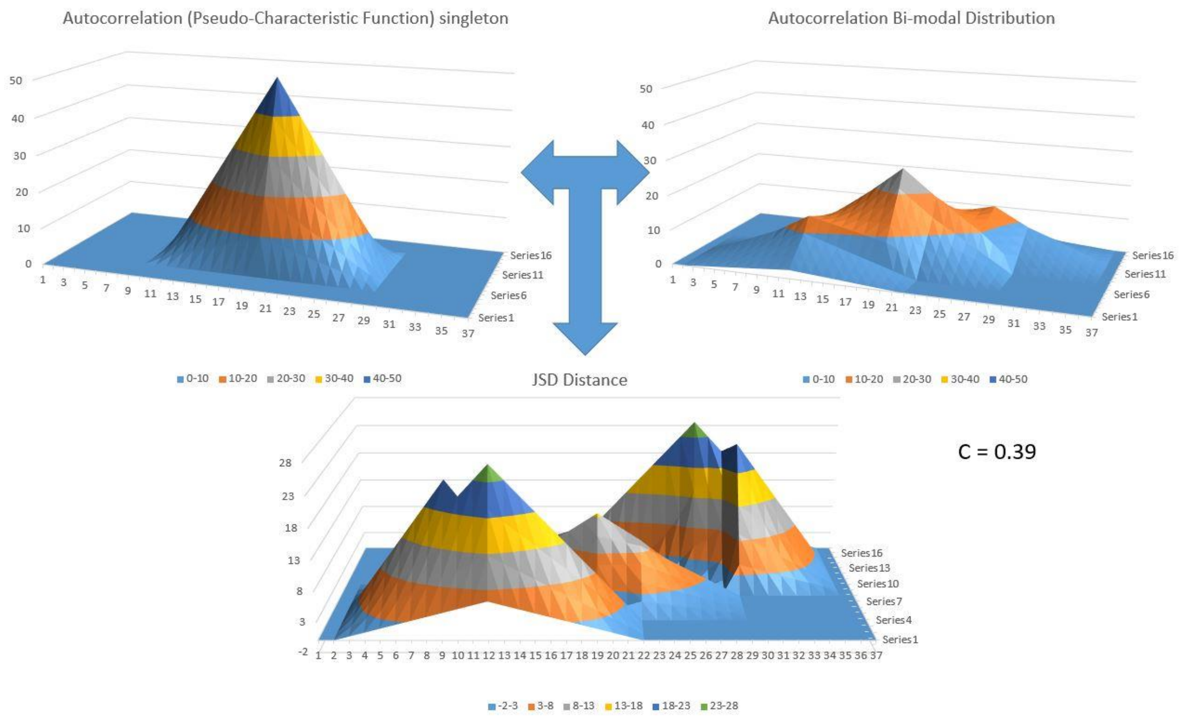


Figure 44 IRSD & ARIA+ Bi-modal Distribution Divergence

Figure 45 reports the classification of the SA3 in terms of the IRSD decile distribution. The pdf curve shows that a low value of Homogeneity ($H \leq 0.2$) usually corresponds to a Concentration Location Index closed to the central bin of the distribution ($4 \leq\ x0^* \leq 6$). Typical configuration satisfying this property are the bi-modal and uniform distributions. In such cases, the distribution "centroid" is not representative for the socio-economic classification of a specific area.

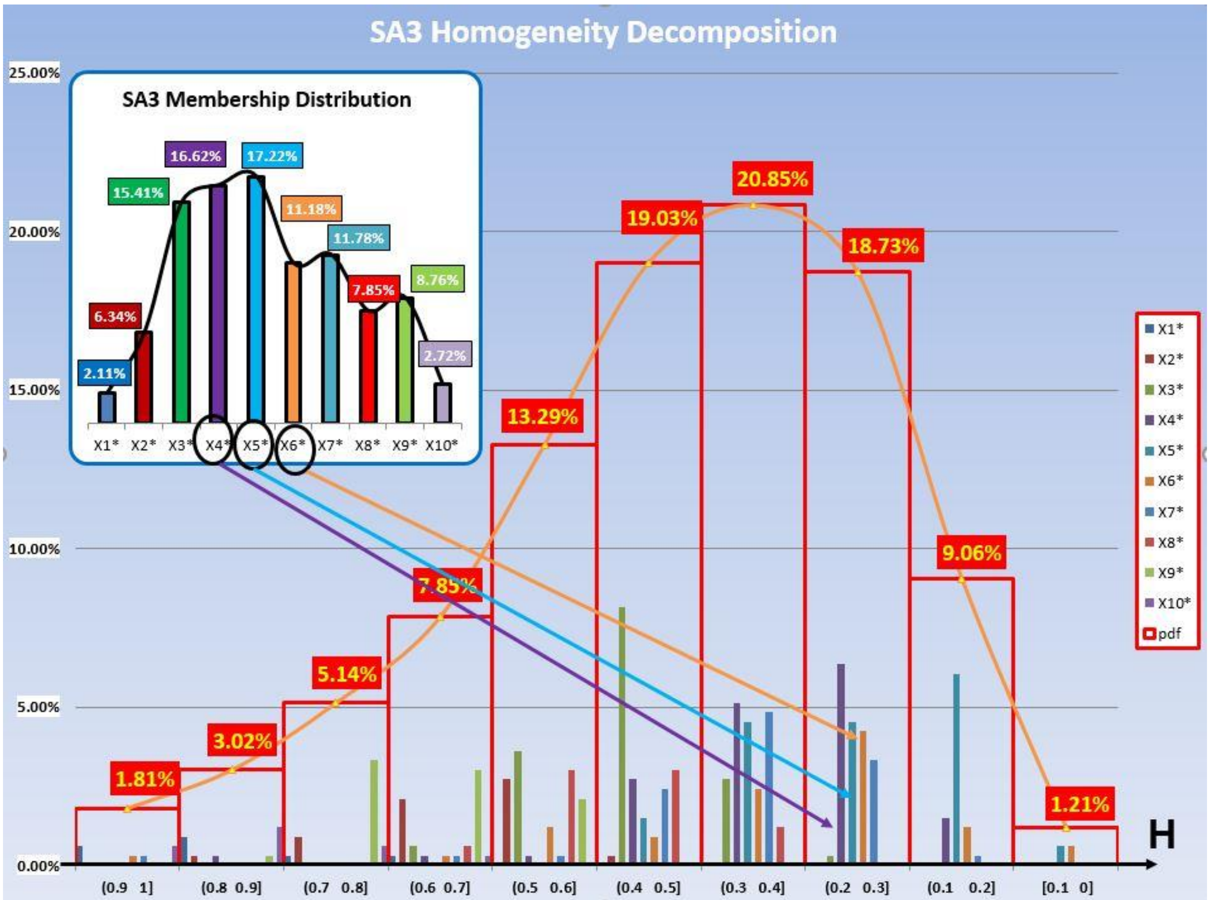


Figure 45 SA3 IRSD Membership Distribution

On the basis of this index, we evaluated the Homogeneity of the SA3 geography based on the IRSD decile distribution Figure 46, and we realize that nearly half of these geographic units

(49.85 %) cannot be easily classified in terms of the Concentration Location Index. Therefore, the SA3 catchment is not an appropriate unit of analysis for the IRSD.

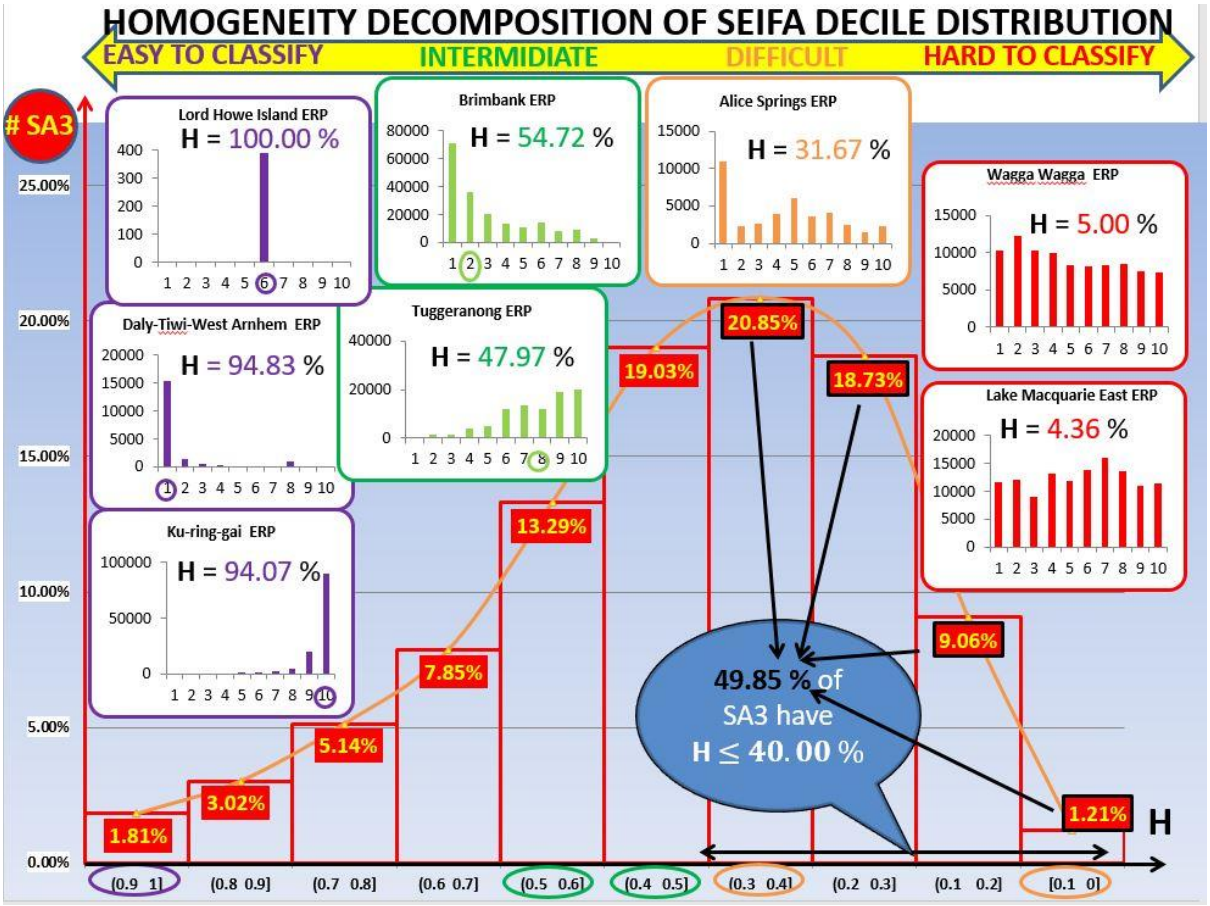


Figure 46 SA3 IRSD Homogeneity Distribution

Conversely, the Homogeneity of the SA3 geography based on the ARIA+ is reasonably good to classify the single areas Figure 47.

The next important stage is the formulation of a model able to compare any two geographic areas in terms of their attributes.

In the next section we propose a new metric to evaluate the degree of dissimilarity between any two SA3.

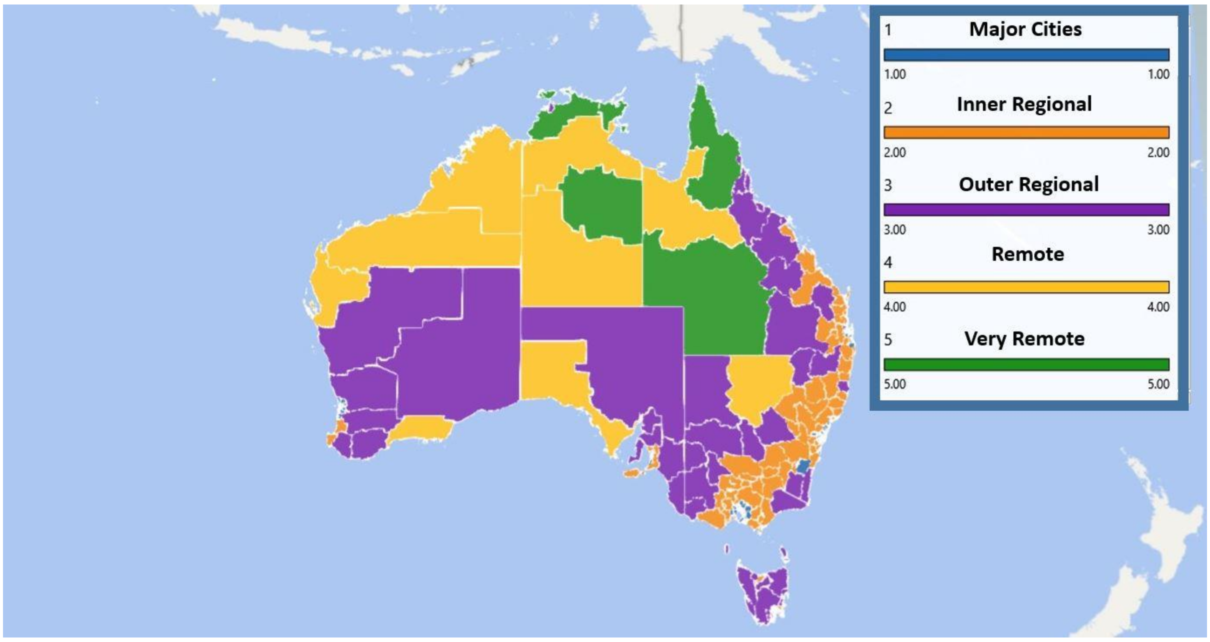


Figure 47 SA3 ARIA+ Classification

DISSIMILARITY DISTANCE

As discussed in the Research Case section, the main purpose of the clustering process is to identify groups of geographies, called peer groups catchment, that have similar geographic, demographic and socio-economic features. In the case of the SA3 geography, these features are the ARIA+, Age distribution and the IRSD respectively.

However, the clustering process is not a universal process as there are many groups of datasets and it can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Therefore, the fundamental issue in the SA3 clustering problem is the definition of a geographic cluster.

The solution to this problem must be sought in the selection of the algorithms input.

Basically, the algorithms can work over two kinds of input:

- Entity Matrix: The first is the matrix representing every entity and the values of its variables. In the case of the SA3, each row is an entity (SA3) and each column is a variable.
- Dissimilarity Matrix: The second is the matrix that express the similarity pair to pair between the SA3.

The selection of the right input often depends on the set of variables used in the model. However, looking only at the single variables neglect possible dependencies in the value of attributes. In such a case, a dissimilarity measure is preferred.

Therefore, we designed a distance which tells us the degree of dissimilarity between any two SA3.

This metric is composed of three terms:

- **Population size F_s:** The first term captures the property that the SA3s in the same cluster should have roughly the same number of residents; and
- **Population heterogeneity F_H**: the second term takes into account the socio-economic and geographical characteristics of the SA3.
- **Concentration Location F_C**: Finally, another important feature is where the concentration of the population is located in the distribution.

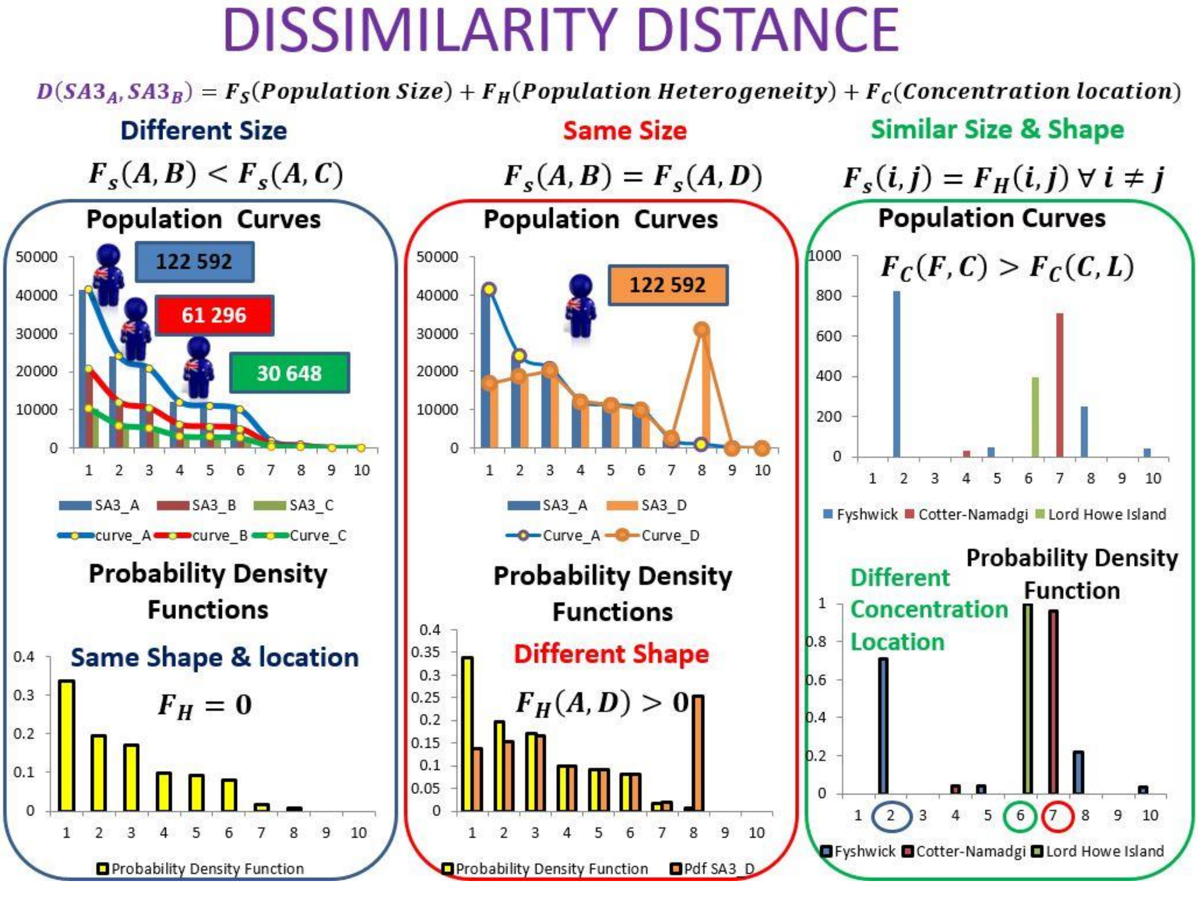


Figure 48 Dissimilarity Distance Composition

The graph on the left-hand side of Figure 48 shows a scaled version of a generic population curve in the IRSD decile distribution. The Red curve (B) is the Blue one (A) scaled by a factor of two and the Green one (C) by a factor of four. Therefore, they all have the same shape (probability distribution - $F_H(A,B) = F_H(B,C) = F_H(A,C) = 0$) but different size ($F_S(A,B) < F_S(A,C) < F_S(B,C)$).

Why do we care about the population size of a specific geography?

A critical issue with units of analysis is their population's size heterogeneity; the greater the heterogeneity the greater the likelihood of observing extra-variation attributable to the population's size rather than to variations in practice. Therefore, it is crucial to compare geographical areas with similar number of residents.

However, this is not enough as we may have geographies with same population size but different shapes. The graph in the middle of Figure 48 illustrates the comparison between two curves that have the same population size distance ($F_S(A,B) = F_S(A,D)$) but different probability distributions ($F_H(A,D) > 0$). It is clear that the Yellow curve is more disadvantaged compared to the Orange one.

Finally, we want to compare the samples concentration of two distributions in terms of the IRSD decile. The graph on the right-hand side of Figure 48 illustrates three SA3 distributions (Fyshwick, Cotter-Namadgi, Lord Howe Island) that are not overlapping with similar size and shape. Therefore, the distance in terms of population size and heterogeneity is exactly the same ($F_S(i,j) = F_H(i,j)\ \forall\ i \neq j$) but the residents of Fyshwick (F) are more disadvantaged compared to the population of Cotter-Namadgi (C) and Lord Howe Island (L) ($F_C(F,C) > F_C(C,L)$).

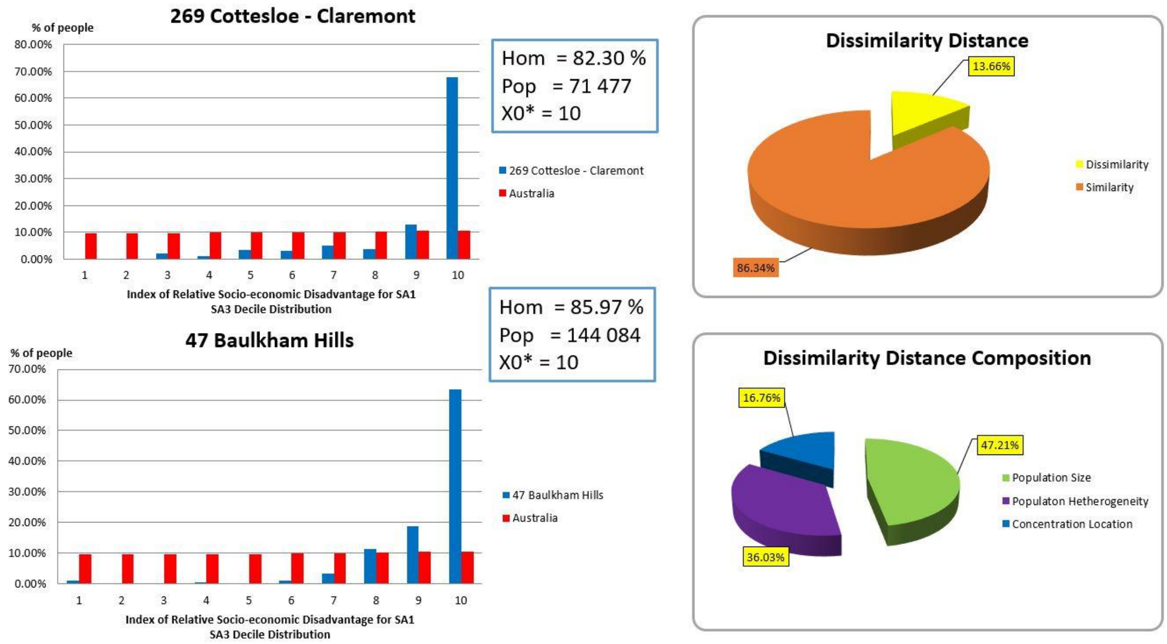


Figure 49 IRSD Dissimilarity Distance Decomposition

Figure 49 illustrates the comparison of two SA3 with the same Concentration Location Index (X0* = 10) and similar shape (82.30% ≅ 85.97%). The pie charts show the dissimilarity value (13.66 %) and the decomposition into its components (F_S= 47.21 %, F_H = 36.03 %, F_C= 16.76%). The most relevant contribution to the final value is the population size followed by the Heterogeneity and Concentration Location distance.

As indicated previously, comparing any two SA3 in terms of the single attributes may incur mistakes. For example, by looking out at the marginal distributions (IRSD Figure 50, ARIA+ Figure 51) of Eyre Peninsula and Outback we would say they are quite similar (IRSD similarity = 78.49 %, ARIA+ similarity = 82.80 %).

However, if we examine the joint distributions closely Figure 52, we draw different conclusions Figure 53 (similarity IRSD & ARIA+ = 66.84%).

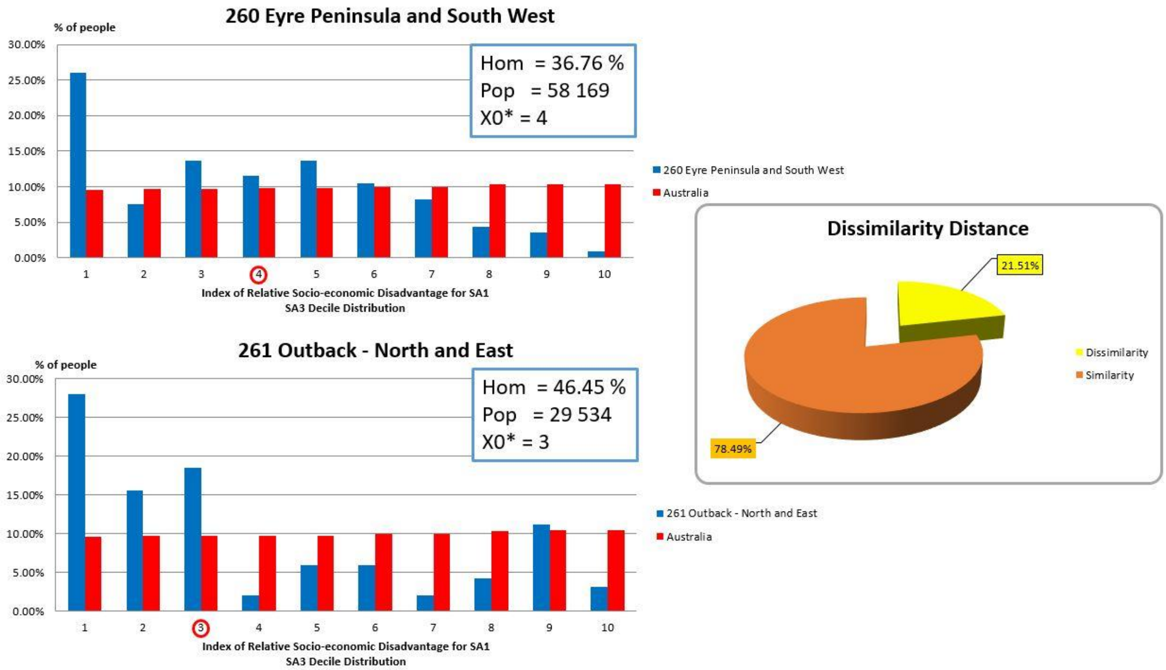


Figure 50 IRSD Dissimilarity

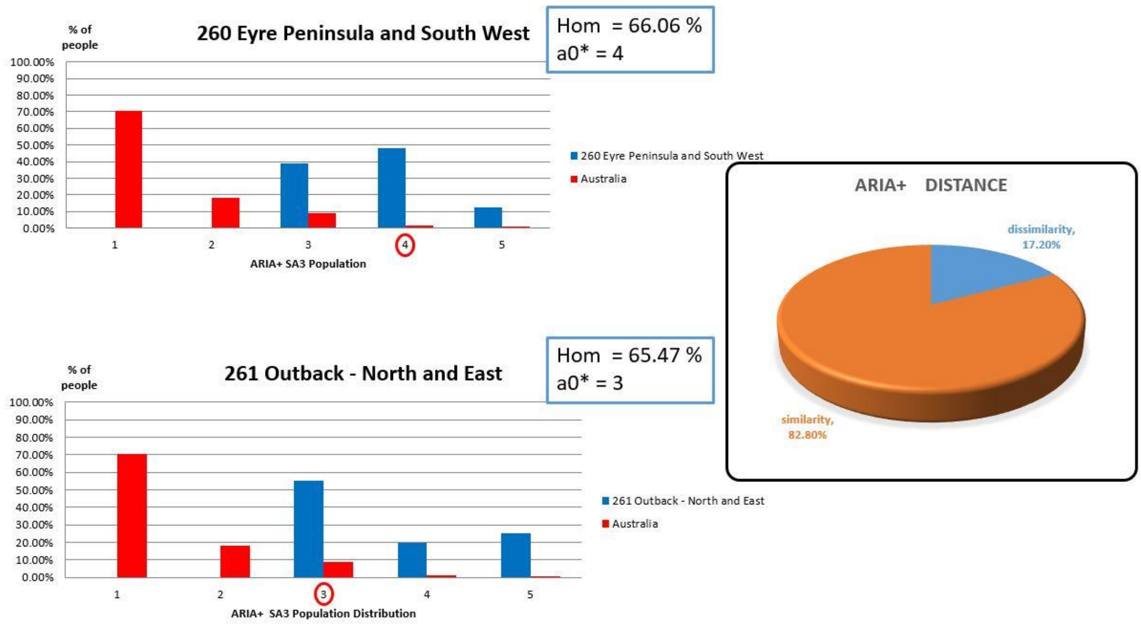


Figure 51 ARIA+ Dissimilarity

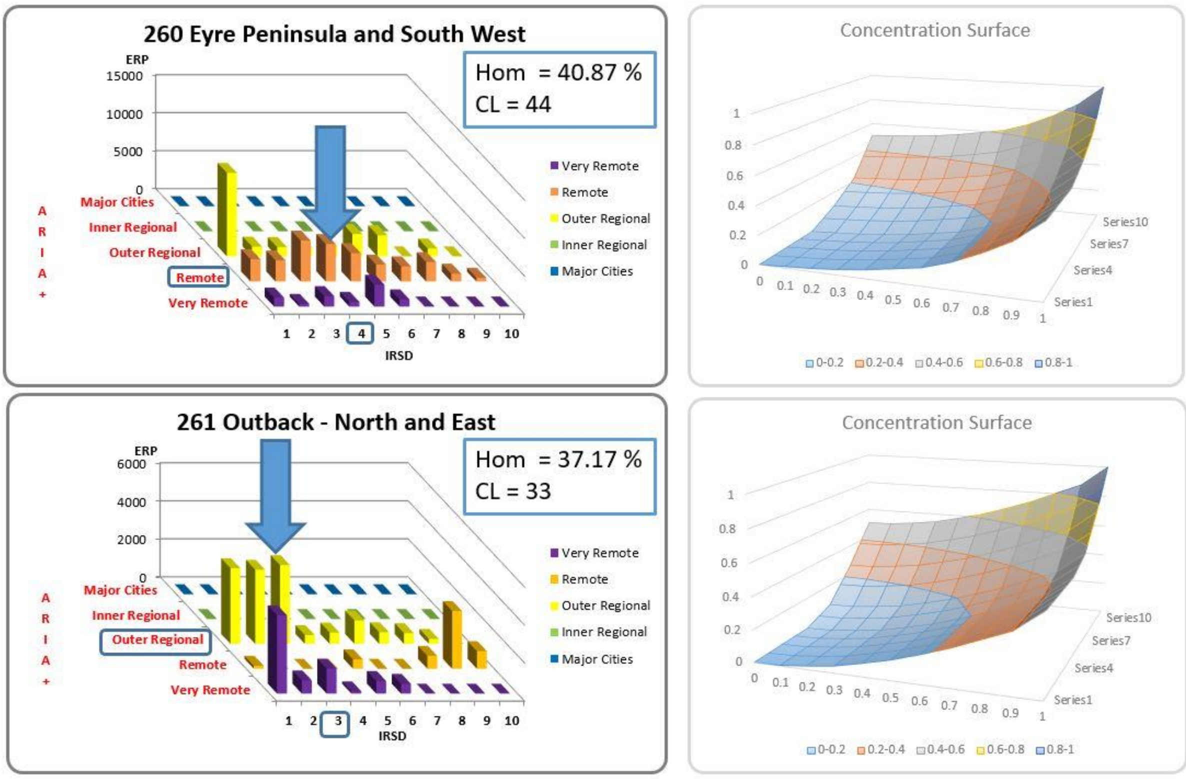


Figure 52 ARIA+ Vs IRSD Joint Distributions

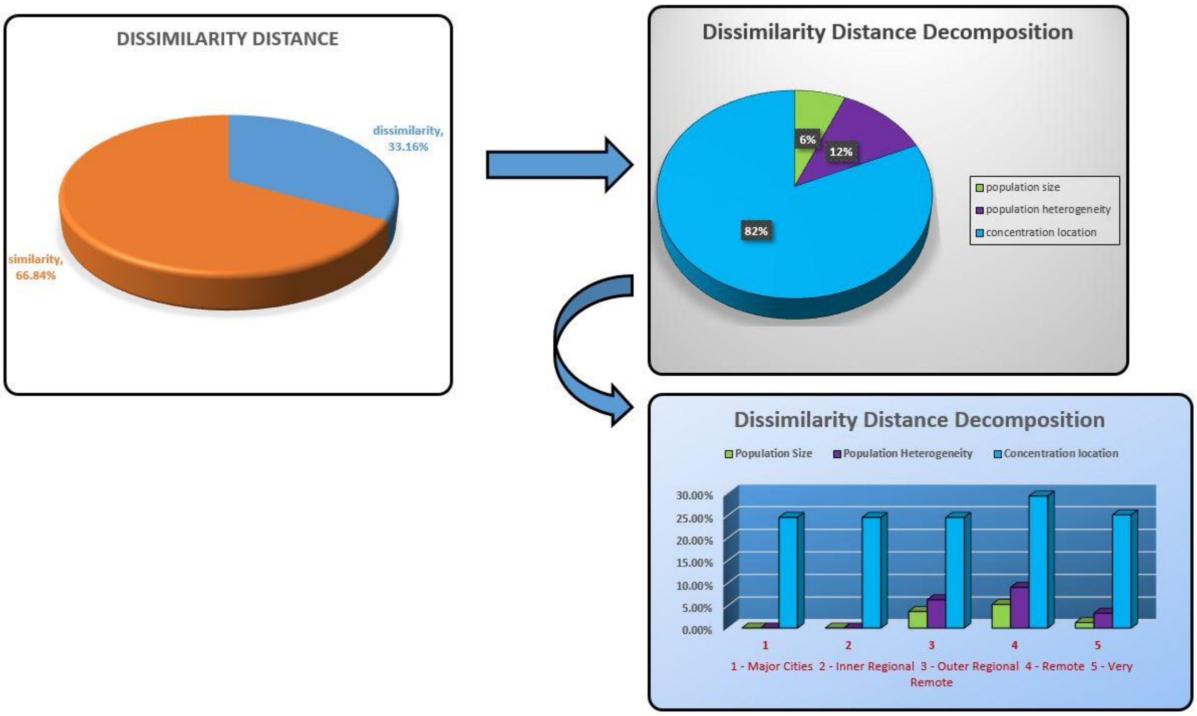


Figure 53 Joint Distribution Dissimilarity Decomposition

RATIONAL PRIMARY CARE SERVICE AREA DESIGN

The goal of our research is to examine the catchment area design problem using a more general approach suitable to define Rational Primary Care Service Area (RPCSA). As a literature review suggests, there is a wide range of objectives that, in general, are in conflict. However, the majority of the publications focused only on potential accessibility whilst neglecting the nature of population flow between places. Technically, these flows between places represent spatial interactions between patients (demand) and health providers (supply), which can be relevant to detect patterns of health service utilisation. Our emphasis is to devise a unified approach to handle such interactions. Through their specification, they enable us to examine the factors such as the cost of travel between origin and destinations or the attractiveness of destinations ([77] [78]) acting to influence flows within systems. These interactions can be better explained by using the definition of the Localization Index.

In the next sections we introduce the notion of Location Index and the RPCSA problem design.

LOCALIZATION INDEX

In considering the patients' preferences about where to use services and the geographical configuration of health care facilities, we can draw an analogy of the problem to the plant location problem.

The plant location problem is a particular instance of location-allocation problems that involve two basic sets of elements; there is a set of demand points spatially distributed throughout a study region and a set of facilities to serve them. The demand points could be population, school children and elderly; while the corresponding facilities would be hospitals or health centres and facilities for the elderly. The general problem seeks the maximum population which can be served within a set of stated service areas (catchment areas) given a limited number of facilities. For example, given the spatial distribution of the population and the time period over which the interactions are measured, deciding which people go to which health provider so that the patients' flow (patients' preferences) is maximised.

Basically, there are two types of decisions that must be made:

- Location decisions determine where to establish the facilities (catchment area) that will offer service to the residents within the catchment area; whereas
- Allocation decisions dictate how to satisfy the residents' demand from the established facilities (patients' flow within the catchment).

In considering the catchment area design framework, we can imagine that the demand points are the b.s.u (e.g. Postal Area POA) and the service area is a zone (catchment area) with a fixed number of health providers (e.g. GPs), so that a study region (e.g. Australia) is covered by a set of zones. These elements can be best represented by the simple concept of a directed graph whose nodes are the POAs and whose edges represent the strength of the patients' flow from one Postal Area (POA) to GPs in different postal areas.

A set of catchment areas corresponds to a partition of the graph in connected sub-graph. The optimal set of catchment areas is the partition that maximise the patients' flow.

A trivial solution to the problem would be to consider the whole study area as a single catchment area. However, in aggregating adjacent basic spatial units to form a larger zone, the original individual attribute values of the basic units are replaced by a single value. The uniqueness of each b.s.u and the variation for the whole area is often lost, resulting in a very heterogeneous catchment area.

In the following stylized graphic Figure 54, we have four POA labelled 2001 through 2004. The POAs contain different groups of Medicare beneficiaries (e.g. Age groups) and providers who provided some Medicare service within a study period. Ideally, the population inside a POA obtains all of its primary care from clinicians within the POA; in such a case the optimal partition has many catchment areas as the total number of POAs (4 zones).

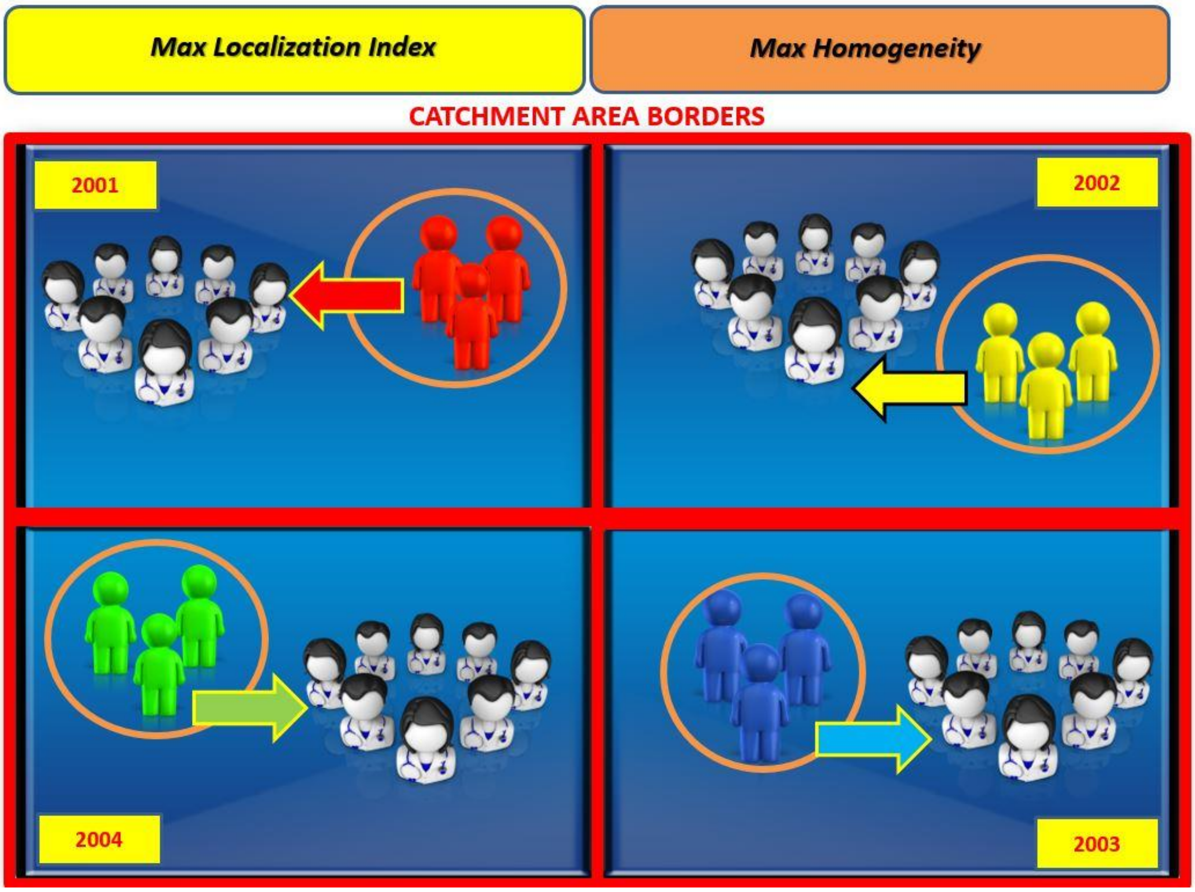


Figure 54 Catchment area design optimal solution

However, it is likely that most patients seek care in other areas and the overall demand (represented by the arcs) within the catchment is *uncover* Figure 55. Conversely, if we select the whole area Figure 56 we maximize the number of people served or *covered* within the catchment (max Localization Index) but the residential segregation is very heterogeneous (min Homogeneity Index).

In view of the previous example, we can state the problem:

- **RPCSA identification**: is the problem of aggregating small geographical areas (e.g. POA), into larger geographic clusters (catchment areas) in a way that the latter are homogeneous and the overall demand is maximized.

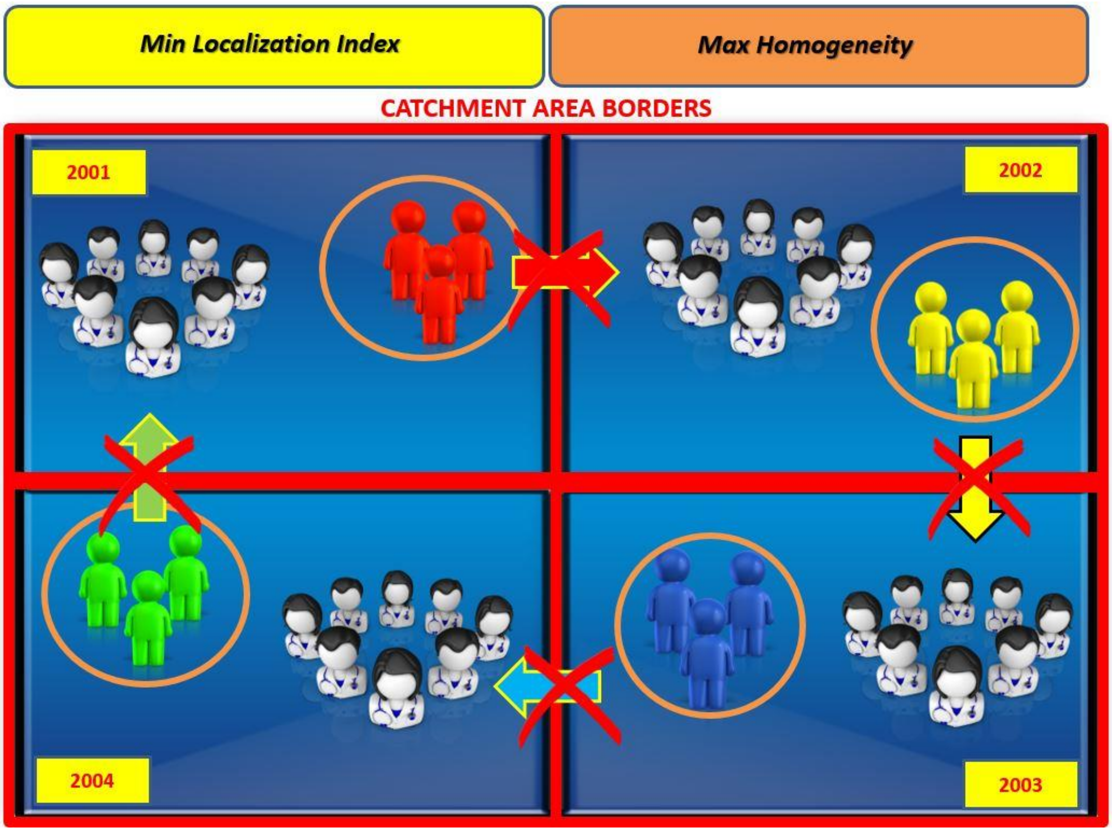


Figure 55 Catchment Area design max homogeneity

By looking at the problem statement, the optimization of two competing objectives enables us to address a fundamental question:

- How do we decide which is the best solution, when we want to maximize two criteria that are clearly in conflict?

In order to answer to this question, we formulate the problem of the RPCSA design as a multi objective set partitioning problem (MOSP).

In the next section we introduce the general setting for the RPCSA design and the related issues.

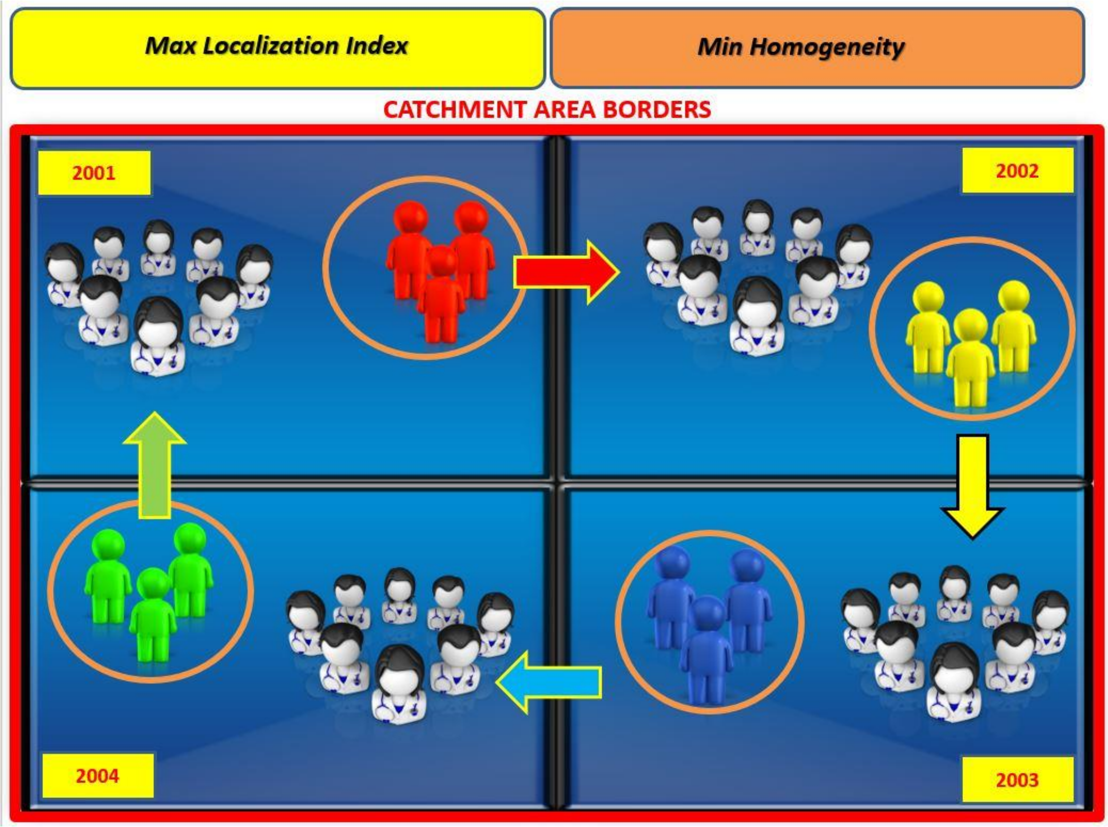


Figure 56 Catchment Area design max Localization Index

MULTI – OBJECTIVE SET PARTITIONING

As indicated earlier, in the general setting of the RPCSA problem, there is no single optimal solution that simultaneously optimizes the Localization and Homogeneity Indices. In these cases, the decision makers are looking for the "most preferred" solution, in contrast to the optimal solution. These situations occur fairly often in Multi-Objective Mathematical Programming Problems (MOMP), where the concept of optimality is replaced with that of Pareto optimality or efficiency. The Pareto optimal (or efficient, non-dominated, non-inferior) solutions are the solutions that cannot be improved in one objective function without deteriorating their performance in at least one of the rest [79].Therefore, the rational decision maker is looking for the most preferred solution among the Pareto optimal solutions of the MOMP.

In this work, the two competing objectives are engaged, proposing a new multi-objective framework in which a mathematical technique (rational generating functions [80] [81]) is used to produce the set of Pareto optimal solutions for smaller instances of the problem, in order to identify the most preferred solution among these alternatives.

The framework involves the application of a model to a study area which has been partitioned into zones. The definition of these zonal boundaries involves the selection of the scale of the study (e.g. the number of POAs within a RPCSA) and the aggregation of the data to match the choice of scale (the number of RPCSAs in the study area). In nearly all cases, there are an incredibly large number of alternative scales and aggregations which could be used. These considerations lead to the definition of two crucial problems:

- **The Scale Problem** arises because of uncertainty about the number of zones needed for a particular study.
- **The Aggregation Problem** arises because of uncertainty about how the data is to be aggregated to form a given number of zones.

The effects of these problems are known to researchers as *the scale effect* and *the zoning effect* [82].

- **The scale effect** is the variation in results that can be obtained when data for one set of b.s.u are progressively aggregated into fewer and larger units.
- **The zoning effect** is the variation in numerical results arising from the grouping of b.s.u into larger units to form a given number of zones.

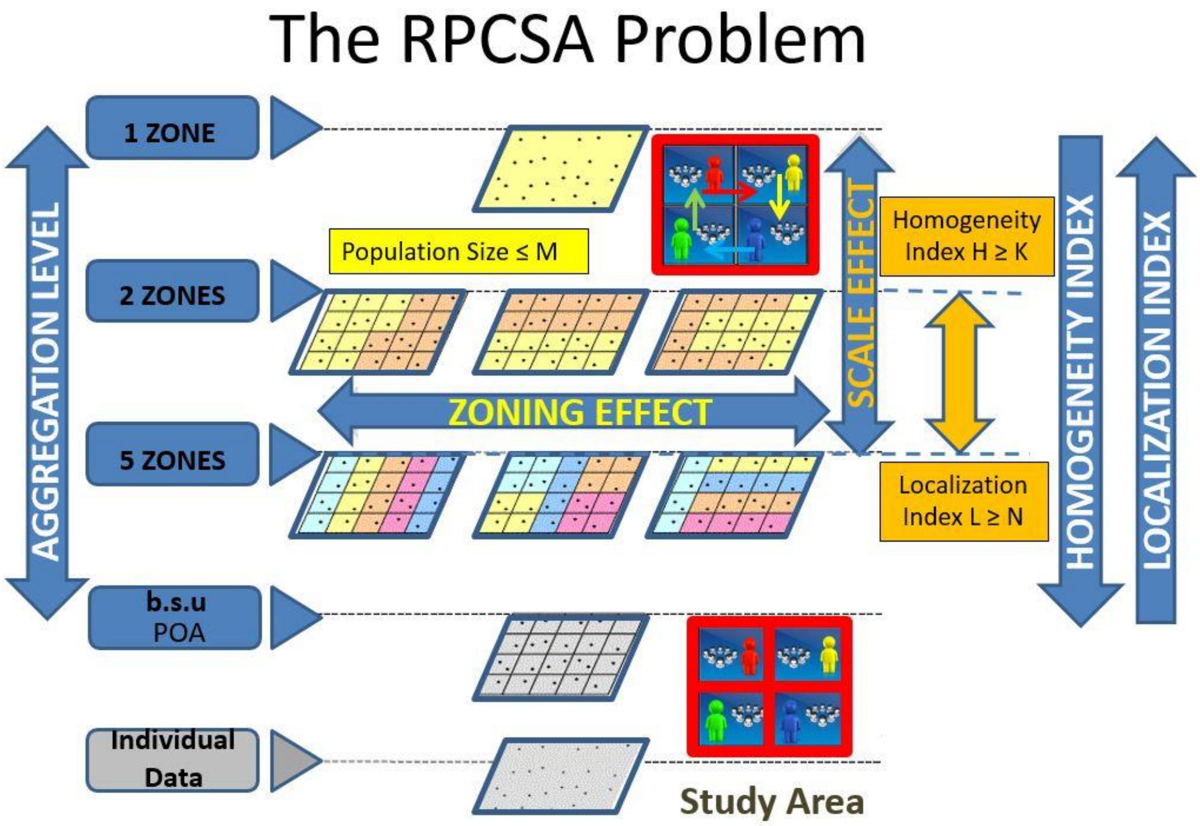


Figure 57 The RPCSA Problem

Figure 57 presents how the zoning and scale effects affect the Homogeneity and Localization indices. In this example, the study area consists of 20 POAS and the aggregation occurs for 5,2 and 1 output zone. At this point, it is worth noting that the POAs' health-related dataset is a summarized form of the individual level characteristics in order to protect individual confidentiality. By aggregating the b.s.u into 5,2 or 1 out zone (scale effect), both objectives go in opposite directions: the most aggregate level is the one that maximises the Localization index but it is also the least Homogeneous. Furthermore, as indicated in the Peer grouping section, a critical issue with units of analysis is their population's size heterogeneity; the greater the heterogeneity the greater the likelihood of observing extra-variation attributable to the population's size rather than to variation in practice. Therefore, the variation within the 5 output zones is smaller than the variation within the 2 output zones (zone effect). Thus, in the Design Process we should disaggregate a very heterogeneous population into very homogeneous sub-populations (maximising the Homogeneity Index) with a similar population size, while maximizing the overall demand covered (maximising the Localization Index). This can be done only as a compromise: for example, by imposing different thresholds for the aforementioned objectives to control the scale effect and fixing a limit to the population size to control the zoning effect.

Therefore, the problem is to define a general *Performance Index*, in order to provide a "measure of partition performance" in terms of the **objectives**, **constraints** and a **predefined target values**, so that by optimizing this index an optimum partition will be obtained Figure 58.

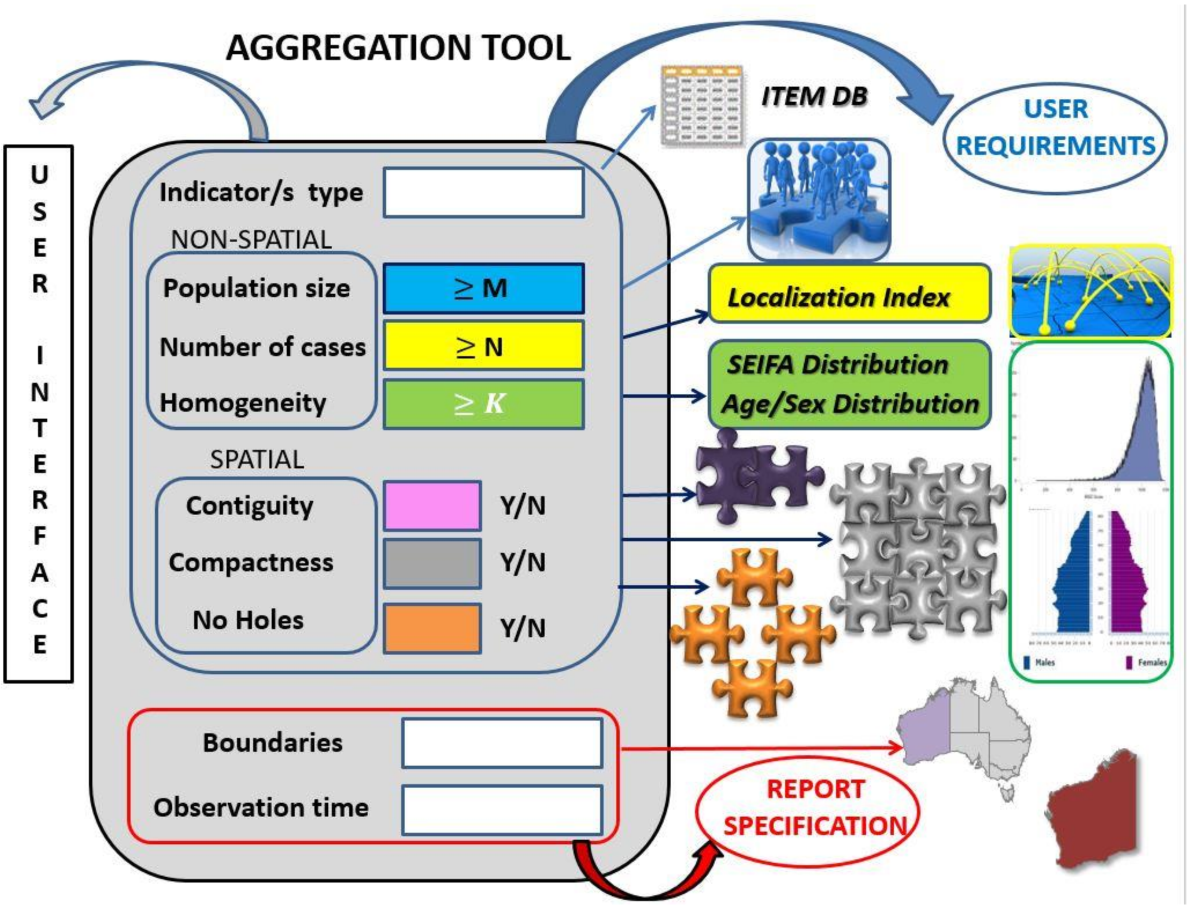


Figure 58 RPCSA Design Tool

ECOLOGICAL STUDIES

As mentioned in the research case section, an important outcome of this work is to provide a valuable framework to study the geographic variation of health indicators. The analytical arsenal used to study and report variation is based on measuring differences in the amount of services provided within a defined geographic unit of analysis (catchment area). Analytically, variation can be measured in terms of the differences in age and sex standardized rates of the indicator under study. However, when the geographic variation in the utilization of health care is related to the socioeconomic status of area populations several issues should be considered.

The most common approach is to calculate summary measures of inequality, such as concentration indices, that make use of the full distribution of both health care and income, and illustrate results using concentration curves with income as a ranking criterion. This approach uses individual-based economic data.

However, not all countries have income data available at the individual level. In this research we overcome this limitation by studying access to health care using area-level information. Studies that use area-based data are called *ecological studies* and should always be considered as exploratory in nature. Case example of how to construct proxy measures of individual socio-economic status can be found [83].

As a matter of fact, following this orientation we should care in making assumptions about individuals based on data that have been summarized for areas.

For example, we might examine the relationship between low income and the utilization rate of a specific indicator (e.g. GP bulk billing) using data for health regions across Australia. We might find a positive association suggesting that areas with high percentages of people with low income also have a high number of visits. It may be that the people in the Health records are not the same people with low income in a given area, and there is no way of determining this without individual-level data.

Whenever individual data are summarized for areas, the statistic of interest depends on the area boundaries used, and if different boundaries are used, even for the same individual-level data, different statistic can result. This is commonly referred to as the Modifiable Areal Unit Problem (MAUP). There are no solutions for MAUP but a general rule we could follow to minimize the MAUP effects in analyses using area-based data is to adopt the smallest possible areas for which the data are available. The larger the areas, the smoother the data will be, so using the smallest possible areas will help to minimize the effects of zoning somewhat. Therefore, ecological studies can be very useful and informative, especially for developing hypothesis, but to properly assess any associations found in an ecological study, there must be a follow up study using individual-level data.

In this research we propose a new approach that combine individual-level data (demographic variables in the health records – Age/Sex) with area-based Census data (IRSD SA1 distribution). The methodology basically involves the specification for the number of cases and the population included in the area while maximising the homogeneity distribution of the Census data.

HEALTH WORKFORCE SUPPLY/DEMAND PLANNING

Workforce issues affect the health care systems of many countries [84]. Since the late 1990s, federal agencies have been involved in an effort to improve the designation of areas and populations that are underserved [85]. To ensure a suitable, affordable health workforce for the future, innovation and reform are essential. Without nationally – coordinated reform, Australia is likely to experience (HWA [86]):

- Workforce shortages.
- Maldistribution of the medical workforce, resulting in less accessible services for Australians in rural, remote and outer metropolitan regions.
- An inefficient training system characterised by bottlenecks and insufficient capacity.
- A continued reliance on poorly-coordinated skilled migration to meet essential workforce requirements.

Multiple geographical classifications have been used for health workforce programs:

— District of Workforce Shortage (DWS) [87]

— Modified Monash Model (MMM) [88]

— ASGC-RA/ ASGS-RA [62]

— RRMA [89]

— Access Relative to Need ARN [90]

Geographic information systems (GIS) have been used to demonstrate methods for improving the designation of health professional shortage areas and medically underserved areas, especially with respect to primary care physician services [91] [18]. However, the evaluation of these approaches in the literature review concludes that these methods do not effectively address the supply and demand sides in assessing spatial and non-spatial factors for healthcare access. Nevertheless, there is a continuing emphasis on calculating physician/population ratios for administrative units, although there is no generally accepted ratio to identify high need areas. This occurs because the physician/population ratio is scale dependent and the greatest variability in the ratio occurs at the local scale. The larger the radius of the catchment, the fewer areas are identified as shortage areas. As the radius increases, more physicians and population are located within the catchment area, and this ratio converges to the ratio for the study area as a whole.

To overcome this limitation, we propose a new methodology for designating Health Professional Shortage Areas that incorporates the concept of *natural catchments*. The proposed method describes a two-step process. First we must identify the RPCSA based on the planning criteria, and then the data on providers and populations within the areas are collected and evaluated.

More precisely, we will conduct a supply and demand study, focusing on the GP workforce initially, to analyse:

- How far patients currently travelling to access Primary Health Services (PHC)?
  - Where are PHC services currently being accessed?
  - Where are patients located?
- What distances should patients be reasonably expected to travel to access certain PHC services?
  - What redistribution of the workforce or service location would need to occur to implement these *natural catchments*?
  - How would migration or training flows affect redistribution?
  - Would alternative service models be more appropriate to meet demand in some catchment areas?

Furthermore, this research involves adding greater information about doctor and practice characteristics, as well as longitudinal data which would allow the following analyses:

- We want to see if we can merge the medical data into practice information and also be able to look at referral patterns, i.e. GP specialist clusters that would provide information on access to secondary care.
- We are interested in doctors and practice characteristics and how these impact on distribution and serviced offered. E.g. whether trained in Australia or overseas, doctor age and sex, whether subject to limited registration, etc.
- We would like to be able to analyse the GP data over time and look specifically at movements between city and rural and remote areas particularly for overseas-trained doctors before and after they cease limited registration.
- We would like to be able to analyse training, i.e. where training is currently occurring, and where it could potentially occur outside metro areas.

The final stage of this process is to take this information and develop a simulation model that allows for changes to key variables, such as travel criteria, service bundles, health professional delivering services, movement of doctors between areas to look at operationalising different policy options for improving the availability of services.

DATA SOURCES AND SECURE ACCESS

PEER GROUPING

The ARIA+ and SEIFA index data are publicly available from the Australian Bureau of Statistics [92] [93]. On the other hand, the smallest geographic level for which the Estimated Residential Population (ERP) is the SA2. However, the SA1 ERP data set was generously contributed by William Watson, Head of the NHPA Research & Analysis Unit.

RATIONAL PRIMARY CARE SERVICE AREA

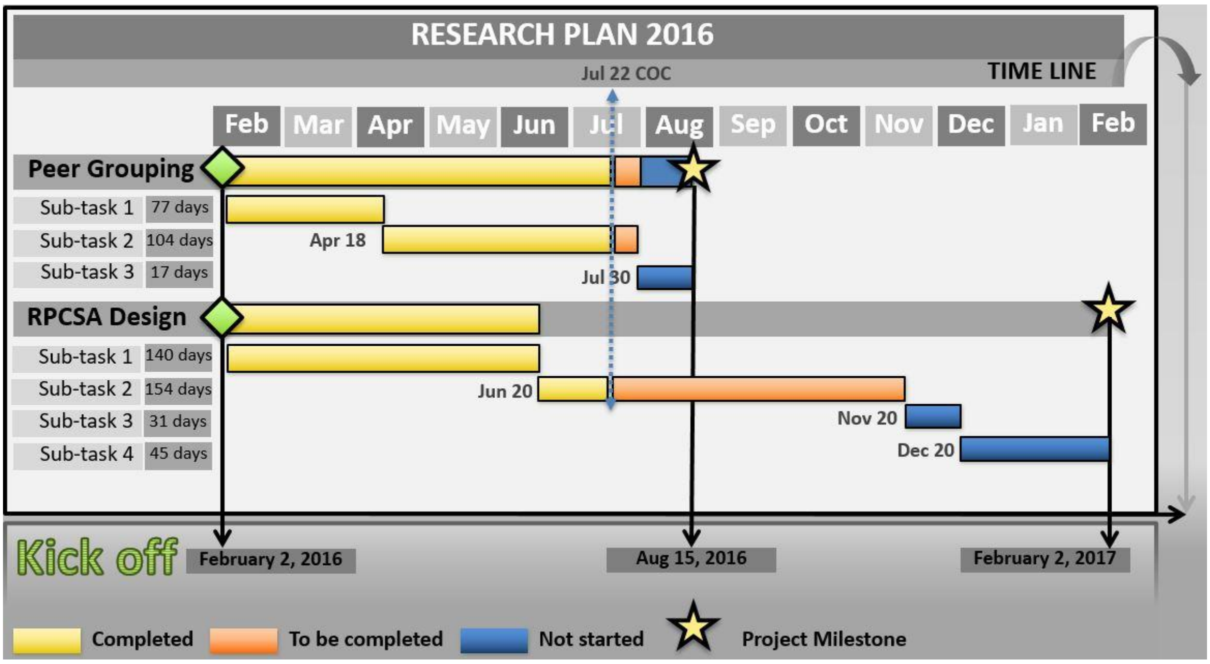
We will use individual level Medical Benefit Schedule data made available to us by NHPA, and accessed at NHPA offices at 1 Oxford Street in Sydney. MBS data allow to identify, for each person, the number of GP services accessed as well as the POA, age and gender of the patient and the POA of the GPs. Hence MBS data will be used to construct the patient flows given as input to the algorithm for the construction of the RPCSAs. We will also take advantage of the fact that several years of data are available to validate the approach and select a stable method.

HEALTH WORKFORCE SUPPLY/DEMAND PLANNING

The Health Analytics and Health Workforce Branch of the Department of Health will provide the geocoded and practice information of the Medicare providers for the Scenario analysis tool.

RESEARCH PLAN

The features of the research plan include:



Peer Grouping

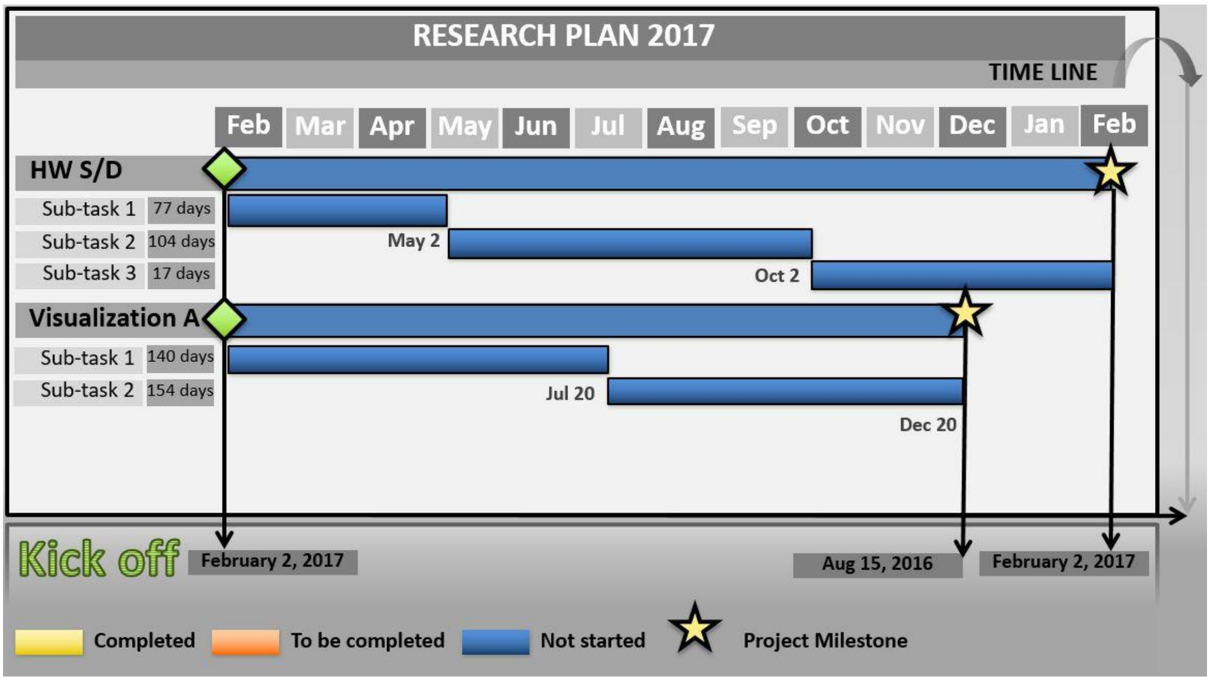
1. Draft of proposed methodology for peer grouping geographic areas.
2. Apply peer grouping methodology to existing geographic areas reported by NHPA such as ABS SA3 for Primary Health Networks.
3. Produce a technical report on peer group methodology.

RPCSA Design

1. Draft of proposed methodology for RPCSA design.
2. Apply RPCSA design methodology to MBS data and produce a first set of RPCSAs.

3. Apply peer grouping methodology to the set of peer grouped RPCSAs.

4. Perform a series of comparisons between the conclusions drawn by studying geographic variation using RPCSAs and using other geographic boundaries, such as SA3.



Health Workforce Supply/Demand planning

1. Draft of proposed methodology for Health Workforce Supply/demand planning.

2. Apply the proposed methodology to design a simulation model and produce a set of RPCSAs based on the planning criteria.

3. Produce a technical report, focusing on

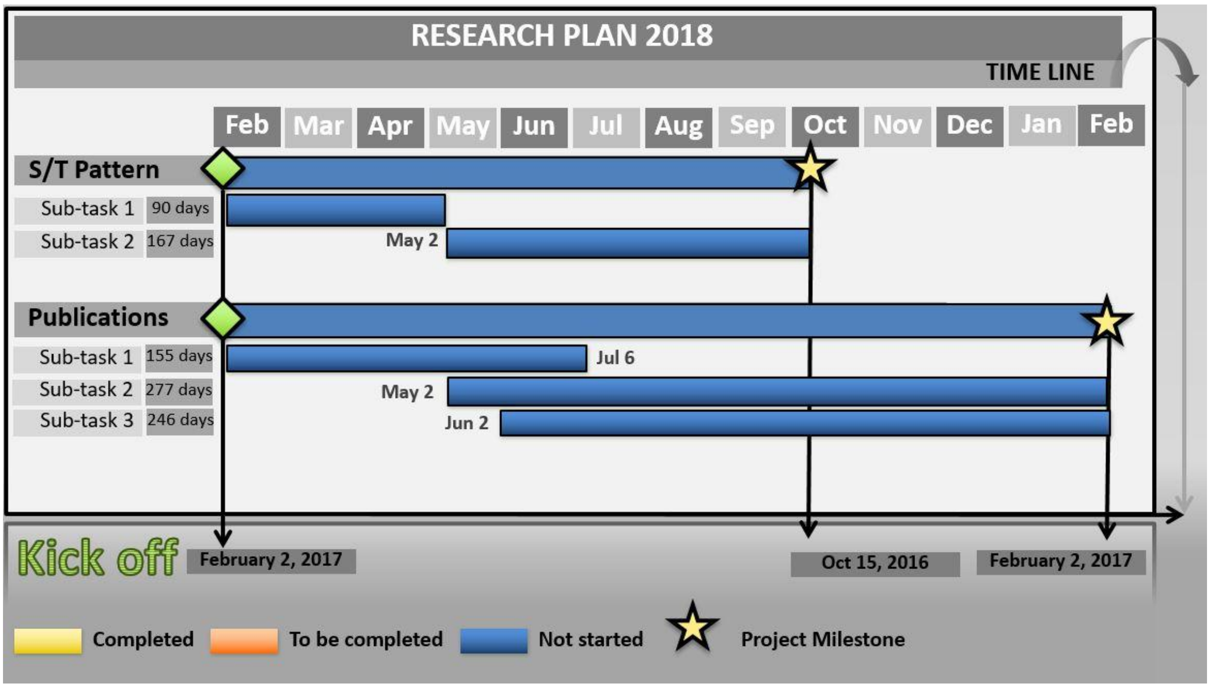
   — The current distribution of the medical workforce and patient access in Australia.

— The findings of scenario analysis conducted using the tool model.

— Options for strategic policy innovation that could support improved distribution of, and access to, medical services.

VISUALIZATION & ANALYSIS

1. Determine a methodology to analyse and visualize the patients flows (i.e. where people access what type of care).
2. Apply the proposed methodology to a set of RPCSAs to visualize flows based on the discovered regions and related attributes.



SPATIAL AND TEMPORAL PATTERN ANALYSIS

1. Determine a methodology to analyse both geographic and temporal patterns of health care service use.
2. Apply the proposed methodology to MBS dataset in order to study patients' trajectories through the entire health care system (when, how often and what type of care people seek).

PUBLICATIONS

1. Peer Grouping
2. RPCSA Design
3. Health Workforce Supply/Demand Planning

FUTURE WORK

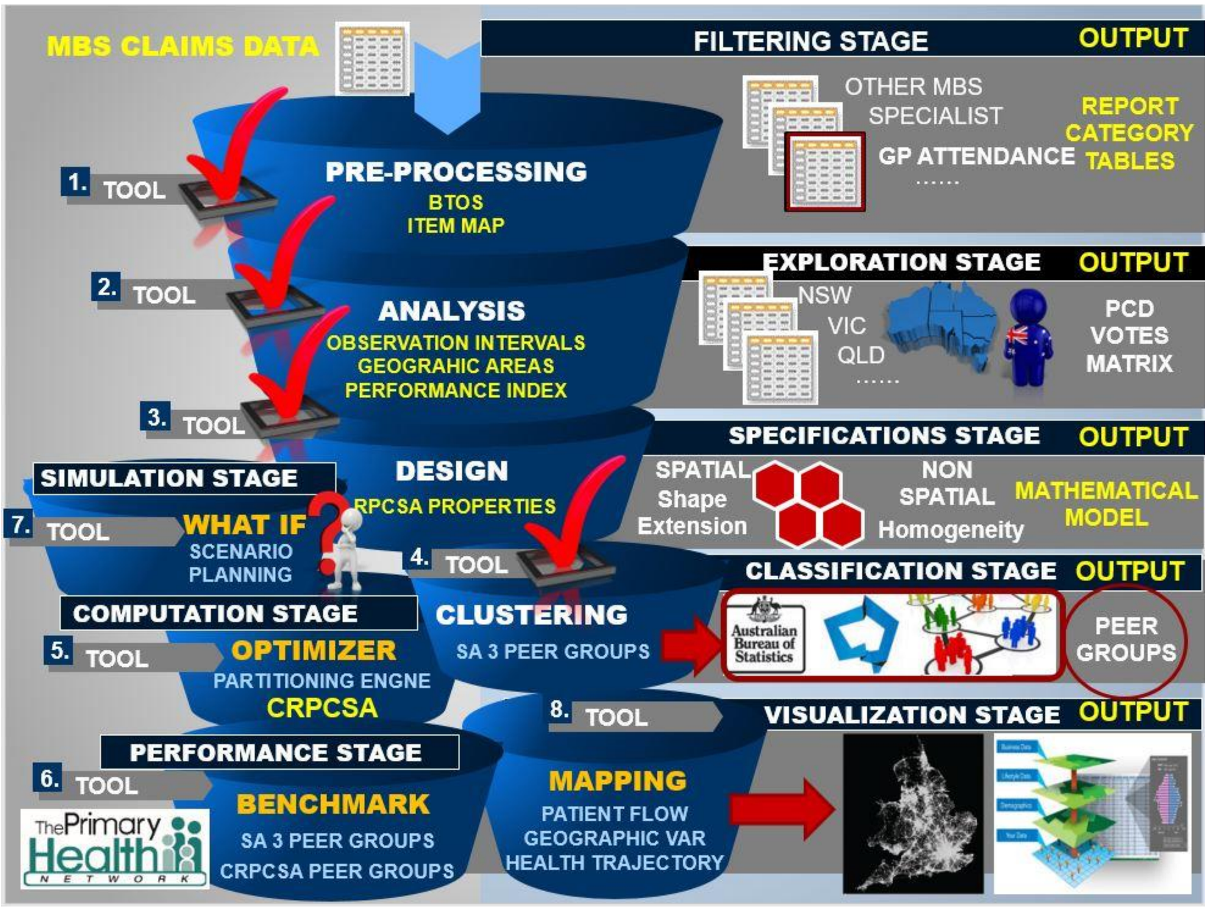


Figure 59 Research plan progress

The first three stages of the project have been accomplished and we are currently working on the final phase of the SA3 peer grouping (Classification stage).

We have built the patient flows (GP Attendance) needed by the graph partitioning algorithm (Filtering and Exploration Stages) to produce the first set of RPCSAs. A mathematical formulation of the problem has been given, although the user requirements should be clarified and documented to generate the corresponding specifications. For example, the selection of the target value for the Localization and Homogeneity indices and a range for the population size.

Regarding the Peer grouping, we proposed and implemented a method for comparing any two geographical areas in terms of Demographic, Socio-economic and Infrastructure variables. Furthermore, we developed an interface for displaying the aforementioned variables and the diversity within and across geographical areas.

Therefore, the next stages of this work are:

- Finalizing the Peer Grouping Classification.
- Computational Stage (Second Semester).
- Performance Stage (Second Semester).
- Simulation Stage (Third Semester).
- Visualization Stage (Fourth & Fifth Semester).

References

[1] Fisher. Es.Wenneberg DE, Stukel TA.Gottlied DJ. Lucas FL, Pinder EL, “The implications of regional variation in Medicare Spending: part 1: The Content, Quality, and Accessibility of Care,” *Annals of Internal Medicine,* vol. 138, pp. 273-287, 2003.

[2] Skinner JS, Wennberg JE, Fisher ES, “The Efficency of Economic Research, Working Papes,” National Bureau of Economic Research, Cambridge, 2001.

[3] Fisher. Es.Wenneberg DE, Stukel TA.Gottlied DJ. Lucas FL, Pinder EL, “The implications of regional variations in Medicare Spending:part : health outcomes and satisfaction with care,” *Annals of Internal Medicine,* vol. 138, no. 4, pp. 288-298, 2003.

[4] “Serious Science,” 4 July 2015. [Online]. Available: http://serious-science.org/geographic-variations-in-medical-practice-4035.

[5] ACSQHC and AIHW, “Exploring Healthcare Variation in Australia: Analyses from an OECD stusy,” Australian Commission on Safety and Quality in Health Care and Australian Institute of Health and Wealfare, 2014.

[6] O. -. t. O. f. E. C. -. o. a. Development, “Geographic Variations in health care use in 13 countries: A synthesis of findings,” in *Geographic Variations in Health Care What Do we know and What Can Be Done to Improve Health System Performance?* , OECD, 2014, p. 31.

[7] NHPA, “National Health Performance Authority 2013: Health Communities: Australians' experiences with primary health care in 2011 - 11,” 2013.

[8] NHPA, “Potentially preventable hospitalisations in 2013 - 2014,” National Health Performance Authority, 2015.

[9] S.Mzumdar, X.Feng, P.Konings, I.McRae nad F.Girosi, “A brief report on Primary Care Service Area catchment geographies in New South Wales, Australia,” *International journal of health geographies,* p. 13.1:38, 2014.

[10] e. a. Goodman DC, “Primary care service areas: anew tool for the evaluation of primary care services.,” *Health Serv Res,* vol. 38(1 Pt 1), pp. 287-309, 2003 Feb.

[11] “The Dartmouth Insitute - PCSA,” [Online]. Available: http://tdi.dartmouth.edu/research/evaluating/health-system-focus/primary-care-service-area.

[12] Dussoult. G. & Franceschini. M.C, “Not enough there, too many here: understanding geographical imbalances in the distribution of the health workforce.,” *Human Resources for Health,* no. 4(12), 2006.

[13] Luo, W. , & Whippo. T,L, “Variable catchment sizes for the two-step floating catchment area (2SFCA) method.,” *Health & Place,* no. 18(4), pp. 789-795, 2012.

[14] M. F. Gugliardo, “Spatial accessibility of primary care: concepts, methods and challenges.,” *International Journal of Health,* no. 3(3), 2004.

[15] McGrail, M.R, “Spatial accessibility of primary health care utilising the two step floating catchment area method: an assessment of recent improvements.,” *International Journal of Health,* no. 11(50).

[16] Cromley, E.K.& S.L. McLaffenty, “Analyzing Acces to Helath Services,” in *Gis and Public Health*, Guilford Press, 2011, p. 310.

[17] Aday, L.A & R.Andersen, “A framework for the study of access to medical care.,” *Health Services research,* no. 9, pp. 208-220, 1974.

[18] Fahui Wang, W. Luo, “Assessing Spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas.,” *Health & Place,* no. 11, pp. 131-146, 2005.

[19] Joseph, A.E., Phillips, D.R., “Accessibility and Utilization-Geographical Perspective on Helath Care Delivery.,” *Happer & Row Publishers,* 1984.

[20] Mc Grail and Humphreys, “The index of rural access: an innovative integrated approach for measuring primary care access.,” *BMC Health Services Research,* no. 9:124, 2009.

[21] Wang F, Tang Q., “Towards equal accessibility to services: A quadratic programming approach,” in *18th Conference on GeoInformatics*, Beijing China, 2010.

[22] Cromley, E.K.& S.L. McLaffenty, “Gis and Public Health,” p. 328, 2011.

[23] E. Talen, “Neighbourhoods as service providers: A methodology for evaluating pedestrian access,” *Environment and Planning B:Planning and Design,* vol. 30, pp. 181-200, 2003.

[24] Parker, E., & Campbell, J., “Measuring access to primary medical care: Some examples of the use of geographical information systems.,” *Health & Place,* vol. 4, no. 2, pp. 183-193, 1998.

[25] Schuurman, Nadine; Fiedler, Robert S; Grzybowski, Stefan CW; Grund, Darrin, “Defining rational hospital catchments for non-urban areas based on travel-time,” *International Journal of Health Geographics ,* vol. 43, no. 5, 2006.

[26] Joseph, A.E. and Bantock, P.R., “Measuring potential physical accessibility to general practioners in rural areas: A method and case study,” *Social Science and Medicine,* vol. 16, pp. 85-90, 1982.

[27] A. Khan, “An integrated approach to measuring potential spatial access to health care services,” *Socio-economic Planning Sciences,* vol. 26, no. 4, pp. 275-287, 1992.

[28] Penchansky, R., & Thomas, J.W., “The concept of access: Definition and relationships to consumer satisfaction,” *Medical care,* vol. 19, no. 2, pp. 127-140, 1981.

[29] A.E Joseph and D.R Phillips, “Acessibility and utilization: Geographical Perspectives on Health Care Delivery,” *Harpen & Row,* 1984.

[30] Knapp,K.K nad Hardwick,K., “The availability and distribution of dentists in rural ZIP codes and primary care shortage areas (PC- HPSA) ZIP codes: Comparison with primary care providers,” *Journal of Public Health Dentistry,* vol. 60, no. 1, pp. 43-48, 2000.

[31] Makuc, D.M.,Hoglund,B.,Ingrm,D.D.,Kleinman,J.C,and Feldman,J.J, “The use of Health Service Areas for measuring provider availability,” *The Journal of Rural Health,* vol. 7, no. 4, pp. 347-356, 1991.

[32] Luo,W.& F.Wang, “Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region.,” *Environment and Planning B,* vol. 30, pp. 865 - 884, 2003.

[33] Yang,D.H,R. George & R.Mullhen, “Comparing GIS-Based Methods of Measuring Spatial Accessibility to Health Services.,” *Journal of Medical Systems,* Vols. 23-32, pp. 23-32, 2006.

[34] Thill, J.,& Kim,M, “Trip-making, induced travel demand and accessibility,” *Journal of Geographical Systems,* vol. 7, no. 2, pp. 229-248, 2005.

[35] Martin, D., & Williams, H, “Market area analysis and accessibility to primary healthcare centres,” *Environment and Planning A,* vol. 24, no. 7, pp. 1009-1019, 1992.

[36] McGrail, M.R.& J.S.Humphreys, “Measuring spatial accessibility to primary care in rural areas: improving the effectiveness of the two-step floating catchment area method,” *Applied Geography,* vol. 29, pp. 533-541, 2009.

[37] Hewko, ., Smover-Tomic, K.E, and Hodgson, M.J., “Measuring neighbourhood spatial accessibility: Does aggregation error matter,” *Environmental and Planning A,* vol. 34, pp. 118-1206, 2002.

[38] Kwan,M.,& Weber,J., “Individual accessibility revisited:implications for geographical analysis in the twenty-first century,” *Geographical Analysis,* vol. 35, no. 4, pp. 341-353, 2003.

[39] Gigg, D., “The Logic of Regional Systems,” *Annals of the Association of American Geographers,* vol. 55, pp. 465-491, 1965.

[40] Openshaw, S., “A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling,” *Transitions of the Institute of British Geographers,* vol. 2, pp. 459-472, 1977.

[41] Ricca, F., Scozzari, A., & Simeone, B., “Weighted Voronoi region algorithms for political districting.,” *Mathematical and Computer Modelling,* vol. 48, no. 9, pp. 1468-1477, 2008.

[42] Datta, D., & Figueira, J., “Graph partitioning by multi-objective real-valued mataheuristics: A cmparative study.,” *Applied Soft Computing,* vol. 11, no. 5, pp. 3976-3987, 2011.

[43] Ji, X., & Mitchell, J., “Branch-and-price-and-cut on the clique partitioning problem with minimum clique size requirement.,” *Discrete Optimization,* vol. 4, no. 1, pp. 87-102, 2007.

[44] Salazar-Auguilar, M., Rios-Mercado, R., & Gonzalez-Velarde, J., “A biojective programming model for designing compact and balance territories in commercial districting,” *Transportation Research Part C,* vol. 19, no. 5, pp. 885-895, 2011.

[45] Menthora, A., Johnson, E., & Nemhauser, G., “An otimization based heuristic for political districting,” *Management Science,* vol. 44, no. 8, pp. 1100-1114, 1998.

[46] Silva de Assis L, Morelato Franca P, Luiz Usberti F, “A redistricting problem applied to meter reading in power distribution network.,” *Comput Oper Res,* vol. 42, pp. 65-75, 2014.

[47] Duque, J.C, Ramos,R, Surinach,J., “supervised regionalization methods: a survey,” *International Regional Science Review,* vol. 30, no. 3, pp. 195-220, 2007.

[48] Raffensberg, J.F, “Choosing collection areas for the NZ census, unplished research report, University of Canterbury, New Zeland”.

[49] Fernando Bacao,Victor Lobo, Marco Painho, “Applying genetic algorithm to zone design,” *Soft Comp,* vol. 9, pp. 341 - 348, 2005.

[50] Altman M, “The Computational Complexity of Automated Redistricting: Is Automation the Answer?,” *Rutgers Comput and Technol Law J,* vol. 23, pp. 81-142, 1997.

[51] Guo D., “Regionalization with dynamically constrained agglomerative clustering and partitioning,” *International Journal of Geographical Information Science,* vol. 22, no. 7, pp. 801-823, 2008.

[52] Cocking S, Martin D, “Zone design for environment and health studiesusing pre-aggregated data,” *Social science & medicine,* vol. 60, no. 12, pp. 2729-42, 2005.

[53] Thomas Talbot, “Aggregation without aggravation: determining spatial contiguity and joining geographic areas using hashing,” in *Proceeding of the SAS forum 2010*, 2010.

[54] Guo D, Jin H., “iRedistrict: Geovisual analytics for redistricting optimization,” *Journal of Visual Languages & Computing ,* vol. 22, no. 4, pp. 279-89, 2011.

[55] Guo D, Zhu X., “Origin-destination flow data smoothing and mapping,” *Visualization and Computer Graphics, IEEE Transaction on.,* vol. 20, no. 12, pp. 2043-52, 2014.

[56] Graeme Hugo, “Census Geography,” in *Spatially Integrate Science*, Handbook of Research Methods and Applications in Spatially Integrated Social Science, 2014, pp. 110-111.

[57] ABS, “Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas,” ABS Australian Bureau of Statistics, Canberra, 2011.

[58] Catherine Linard, Marius Gilbert,Abdislan M.Noor, Andrew J.Tatem, “Population Distribution, Settlment Patterns and Accessibility across Africa in 2010,” *Guy J-P Schumann,* 2012.

[59] Coombes, M. and Raybould, S., “Public policy and population distribution: Developing appropriate indicators of settlement patterns,” *Environmental and Planning C: Government and Policy,* vol. 19, no. 2, pp. 223-248, 2001.

[60] Carr-Hill, R., Smith,P.,Martin,S.,Hardman,G., “Allocating resources to health authorities:development of method for small area analysis and the use of inpatient services.,” *British Medical Journal,* vol. 309, pp. 1046-1409, 1994.

[61] Bamford, E., Dunne, L. and Hugo, “Accessibility/Remotness Index of Australia (ARIA),” Department of Health and Aged Care, Occasional Papers Series No.6, Canberra:Commonwealth Department of Health and Aged Care, 1999.

[62] ABS, “Australian Statistical Geography Standard (ASGS). Volume 5 Remotness Structure,” ABS Australian Bureau of Statistics, Canberra, 2011.

[63] National Health Performance Authority 2013, “Health Communities: Australian's experiences with primary health care 2010-11,” Commonwealth of Australia 2013, Canberra, 2013.

[64] Brian Pink, “Socio-economic Indexes for Areas (SEIFA) 2011,” ABS Australian Bureau of Statistics, Canberra, 2013.

[65] ABS, “Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2011,” [Online]. Available: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/2033.0.55.001main+features100052011.

[66] Centre for Epidemiology and Evidence , “Health Stats NSW: Privacy issues and the reportin of small numbers,” NSW Ministry of Health, Sydney, 2015.

[67] Peter Radish and Phillip Wise, “Socio-economic Indexes For Areas: Robustness, Diversity Within Larger Areas and the New Geography Standard,” ABS, Canberra, 2012.

[68] Martin,D.J, “Optimising Census geography: the separation of collection and output geographies,” *Internation Journal of Geographical Information Science,* vol. 12, pp. 673-685, 1998.

[69] Martin Ralphs and Lyndon Ang, “Optimised geographies for data reporting: zone design tools fro Census output geographies,” Statistics New Zealand Working Paper No 09-01, Wellington:Statistics New Zeland, 2009.

[70] White,Micheal J, “The Measurement of Spatial Segragation,” *American Journal of Sociology,* vol. 88, pp. 1008-19, 1983.

[71] H. Theil, “Statistical Decomposition Analysis,” Amsterdam:North Holland.

[72] M. Lorenz, “Methods of measuring the concentration of wealth,” *Publication of the American Statistical Association,* vol. 9, pp. 209-219, 1905.

[73] Koshevoy,G.,Mosler,K, “The Lorenz Zonoid of a Multivariate Distribution,” *Journal of the American Statistical Association,* vol. 91, p. 434, Jun 1996.

[74] Wolfgang Eichorn, Manfre Krtscha, “Multivariate Lorenz Majorization and Hetherogeenity Measures: An Axiomatic Approach,” *Analyse & Kritik,* vol. 29, pp. 206-207, 2007.

[75] T. Taguchi, “On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two-dimensional case-I,” *Annals of the Institute of Statistical Mathematics,* vol. 24, pp. 355-382, 1972.

[76] B. Arnold, “Pareto Distributions,” *International co-operative Publishing House,* 1983.

[77] Fotheringham,A.S, “Some theoretical aspects of destination choice and their relevance to rpoduction-constrained gravity models,” *Environmental and Planning A,* vol. 15, no. 8, pp. 1121-1132, 1983b.

[78] Singleton,A.,Wilson,A.,O'Brien,O, “Geodemographics and spatial interaction: an integrated model for higher education,” *Journal of Geographical Systems,* pp. 1-19, 2010.

[79] E. M, “Efficency and Non dominance,” in *Multicriteria optimization*, Heidelberg, Springer, 2005, pp. 23 -24.

[80] W. K. Barvinok A, “Short rational generating functions for lattice point problems,” *J Am Math Soc,* vol. 16, pp. 957-979, 2003.

[81] P. J. Blanco V, “A new complexity result on multiobjective linear integer programming using short rational generating functions,” *Optim lett,* vol. 6, pp. 537-543, 2012.

[82] S. Openshaw, “Ecological fallacies and the analysis of area census data,” *Environmental and Planning A,* vol. 16, pp. 17-31, 1984.

[83] Mah-jabeen Soobader, Felicia B. LeClere, Wilbur Hadden and B. Maury, “Using Aggregate Geographic Data to Proxy Individual Socioeconomic Status: Does size matter?,” *American Journal of Public Helath,* vol. 91, pp. 632-636, 2001.

[84] W. H. Organization, “World health report 2006,” World Health Organization, Geneva,Swizerland, 2011.

[85] T. J. H. G. M. R. R. L. R. T. D. e. a. Ricketts, “Designating places and populations as medically underserved: A proposal for a new approach,” *Journal of Health Care for the Poor and Underserved,* vol. 18, no. 3, pp. 567-589, 2007.

[86] H. W. A. 2012, “Health Workforce 2025 - Doctors,Nurses and Midwives - Volume 1,” Health Workforce Australia, Adelaide, 2012.

[87] A. G. D. o. Health, “District of Workforce Shortage (DWS),” Australian Government Department of Health, [Online]. Available: http://www.doctorconnect.gov.au/internet/otd/publishing.nsf/content/dwsfactsheet.

[88] A. G. D. o. Health, “Modified Monash Model,” Australia Government Department of Health. [Online].

[89] AIHW, “RRMA classification,” Australian Institute of Health and Welfare. [Online].

[90] A. I. o. H. a. W. 2014, “Access to primary health care relative to need for Indigeneous Australians,” AIHW, Canberra, 2014.

[91] Juarez, P.D., Robinson, P.L., & Matthews - Juarez P., “100% access, zero health disparities, and GIS: An improved methodology for designating health Professions shortage areas,” *Journal of Health and Social Policy,* vol. 16, pp. 155-167, 2002.

[92] A. B. o. Statistics, “1270.0.55.005 - Australian Statistical Geography Standard (ASGS): Volume 5 - Remoteness Structure, July 2011,” [Online]. Available: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.005July%202011?OpenDocument.

[93] A. B. o. Statisitcs, “2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2011,” [Online]. Available: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012011?OpenDocument.

[94] Aday, L., & Anderson, R., “Equity of access to medical care: A conceptual and empirical overview,” *Medical Cae,* vol. 19, pp. 4-27, 1981.

[95] M. F. Gugliardo, “Spatial accessibility of primary care: concepts, methods and challenges,” *International Journal of Health Geographic,* vol. 2, 2004.

[96] Penchansky R, Thomas JW, “The concept of Access,” *Med care,* vol. 19, no. 2, pp. 127-140, 1981.

[97] Luo W, Wang F , “Measures of spatial accessibility to health care in a GIS environment: Sysnthesis and a casre study in Chicago Region,” *Environment and Planning B,* vol. 30, no. 6, pp. 865-884, 2003.

[98] McGrail, M.R.& J.S.Humphreys, “Measuring spatial accessibility to prim,ary health care services: Utilising dynamic catchment sizes.,” *Applied Geography,* vol. 54, pp. 182-188, 2014.

[99] Hugo, G..J., Griffith, D., Rees, P., Smailes, P., Badcock, B. and Stimson, R., “Rethinking the ASGC:Some conceptual and practical issues,” *Monograph Series 3, National Key Centre for Social Application of GIS University of Adelaide,* 1997.

[100] G. Corrado, “Measurement of Inequality of Incomes,” *The Economic Journal,* pp. 124-126, 1921.

[101] T. Taguchi, “On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two-dimensional case-II,” *Annals of the Insitute of Statistical Mathematics ,* vol. 24, pp. 599-619, 1972.

[102] E. M, “Efficency and Nondominance,” in *Multicriteria optimization*, Heildeberg, Springer, 2005, pp. 23-24.

List of Figures

Figure 1 Population size of geographic units in participating OECD countries, 2011 or latest year. Source: National data submitted for the OECD project on Medical Practice Variations. 6
Figure 2 Medicare Local catchment 7
Figure 3 Average non-hospital MBS expenditure for very high and frequent GP attendance, by Medicare Local catchment, 2012 - 2013. NHPA analysis of Department of Human Services, Medicare Benefits statistics 2012 - 2013 Data extracted November 2014 8
Figure 4 Primary Health Network - Australian Department of Health 9
Figure 5 System Architecture for Customizable Rational Primary Care Service Areas 15
Figure 6 Mapping patient flow in R 17
Figure 7 Access to Health Care - Homogeneity Index Dimensions 18
Figure 8 Hospital Service Area Based on Geocoded Patient Residential Locations for All patients. Source: GIS and Public Health 2nd ed Ellen K. Cromely. Guilford Press 2012 22
Figure 9 Hospital Service Area Based on Geocoded Patient Residential Locations Excluding Outliers. Source: GIS and Public Health 2nd ed Ellen K. Cromely. Guilford Press 2012 23
Figure 10 1-hour service areas for all hospitals. Data source: GIS Innovations Ltd 24
Figure 11 Taxonomy of health care access. Mark F. Gugliardo 25
Figure 12 A schematic of the kernel estimation method - GIS and Public Health p 165 27
Figure 13 Different spatial impedance functions. Tij is a measure of travel time or distance. Source GIS and Public Health 28
Figure 14 An early version of the floating catchment area (FCA) method. Luo W, Wang F (2003) ... 30
Figure 15 Spatial accessibility to primary care Illinois. Luo, Wang (2004) 31
Figure 16 Australia Statistical Geography Standard (ASGS): structure and summary for Census 40
Figure 17 Primary Health Network Framework 42
Figure 18 ARIA+. Accessibility Remoteness Index for Australia. 43
Figure 19 Australia: Remoteness Category 44
Figure 20 Homogeneity Index Decomposition 48
Figure 21 IRSD SA1 Distribution 50
Figure 22 IRSD Decile Distribution spread 51
Figure 23 Homogeneity Decomposition SA3 52
Figure 24 Goulburn-Yass SA1 ERP Distribution 53
Figure 25 Goulburn - Yass IRSD population Distribution 53
Figure 26 Golburn Yass ERP vs IRSD 54
Figure 27 Goulburn - Yass IRSD Population Decile Distribution 55
Figure 28 Campbelltown IRSD SA1 decile distribution 57
Figure 29 Tuggeranong IRSD ERP Decile Distribution 58
Figure 30 Tuggeranong IRSD Concentration Location Index 59
Figure 31 Ku-ring-gai Vs Australia IRSD Distribution 61
Figure 32 Australia IRSD Concentration Curve 62
Figure 33 Ku-ring-gai IRSD Concentration Curve 62
Figure 34 Australia ARIA+ & IRSD Concentration Curves 64
Figure 35 Australia ARIA+ & IRSD Concentration Surface 64
Figure 36 Australia Lorenz Zonoid ARIA+ & IRSD 65

Figure 37 Armadale IRSD Decile Distribution........66
Figure 38 Cleveland IRSD Decile Distribution........66
Figure 39 Armadale IRSD Concentration Curve........67
Figure 40 Cleveland IRSD Concentration Curve........67
Figure 41 Spectra Energy........68
Figure 42 Spectra Energy Divergence........69
Figure 43 IRSD & ARIA+ Uniform Distribution Divergence........70
Figure 44 IRSD & ARIA+ Bi-modal Distribution Divergence........70
Figure 45 SA3 IRSD Membership Distribution........71
Figure 46 SA3 IRSD Homogeneity Distribution........72
Figure 47 SA3 ARIA+ Classification........73
Figure 48 Dissimilarity Distance Composition........75
Figure 49 IRSD Dissimilarity Distance Decomposition........77
Figure 50 IRSD Dissimilarity........78
Figure 51 ARIA+ Dissimilarity........78
Figure 52 ARIA+ Vs IRSD Joint Distributions........79
Figure 53 Joint Distribution Dissimilarity Decomposition........79
Figure 54 Catchment area design optimal solution........82
Figure 55 Catchment Area design max homogeneity........83
Figure 56 Catchment Area design max Localization Index........84
Figure 57 The RPCSA Problem........86
Figure 58 RPCSA Design Tool........87
Figure 59 Research plan progress........98

List of Tables

Table 1 OECD (2014), Geographic Variations in Health Care What Do we know and What Can Be Done to Improve Health System Performance? OECD Health Policy Studies, OECD Publishing........5
Table 2 OECD Health System Characteristics Survey, 2013........20
Table 3 Measurement of accessibility (with examples from health sector). Adapted from Talen. Gary Higgs........21
Table 4 PHN Framework SA3 variables........42
Table 5 ERP Australia - Age and Sex distribution 2013........45
Table 6 IRSD variables........46

END