Harmonising Incompatible Datasets to Enable GIS Use to Study Non-communicable Diseases in Tonga

Siu I. Fanga Jione[1] · Paul Norman[2]

Abstract

We investigated the use of geographic information systems (GIS) to study the prevalence of non-communicable diseases (NCDs) in association with area deprivation within Tongatapu, the largest island in Tonga. This is a case study to determine what is possible to achieve since various data issues influenced by institutional and organizational barriers exist, affecting data integration and GIS use. A data conversion framework was developed using geographic conversion tables (GCTs) to process existing data into a compatible format and create new geographies to analyse the prevalence of NCDs and area deprivation across different levels of geography. Area deprivation was higher in the rural district and was associated with the prevalence of NCDs. However, at lower levels of geography, the distribution and patterns of NCDs, and area deprivation were unclear. This was influenced by the methods of data collection, recording and dissemination. There is a need for a national action plan outlining the standard operating procedures for all stakeholders to adhere to, and thereby produce and disseminate comprehensive, reliable, and high-quality data. Otherwise, data will be collected for basic reporting but impractical for sophisticated analysis and research. We suggest an investigation into dasymetric mapping to disaggregate population data and develop automating processes for large national datasets.

Keywords Geographic Information Systems · Area deprivation · Non-communicable diseases · Geographic conversion tables · Tonga

✉ Siu I. Fanga Jione

1 Suva, Fiji

2 School of Geography, University of Leeds, Leeds, UK

Introduction

As the leading cause of premature deaths in Tonga (Tonga Department of Statistics et al., 2013a; WHO, 2018), non-communicable diseases (NCDs) continue to impose social and heavy financial burdens on families, the public health sector (WHO, 2018) and the economy. NCDs continue to reduce life expectancy, increasing the population of economically active people living with a disability. The burdens of NCDs alongside infectious disease create a "double burden" on a fragile health system (Anderson, 2013). In 2012, Tonga's national STEPwise survey on NCDs found 91% of respondents were overweight and 68% were obese (Tonga Ministry of Health, 2014). Additionally, the Demographic and Health Survey (DHS) highlighted that 98.7% of the 2,599 individuals surveyed in Tonga were at risk of NCDs (Tonga Department of Statistics, 2013a). In 2016, 83% of total deaths in Tonga were NCD related (WHO, 2018).

Responding to the growing concerns on the prevalence of NCDs and related deaths, Tonga introduced various legislation (Hufanga et al., 2012) and launched a national strategy to address NCDs in 2004 (Matoto et al., 2014). The establishment of the National Diabetes Centre followed focusing on the diagnosis and treatment of patients, mortality records and improving data management (Anderson, 2013). In 2007, the Tonga Health Promotion foundation was established to develop and implement health and wellness programs for women and youth (Tonga Health, 2016). Following the WHO guideline on the surveillance of NCD risk factors, Tonga conducted two nationwide STEPwise surveys in 2004 and 2012 (Tonga Ministry of Health, 2014). A national demographic and health survey (DHS) in 2012 provided key estimates of the overall state of health (Tonga Department of Statistics et al., 2013a). The 2016 *Hala Fononga* (Tonga Health, 2016) and the *2021 Tuiaki 'I he 'amanaki ki ha Tonga mo'ui lelei* (Tonga Ministry of Health and Tonga Health, 2021) National Strategies introduced a multistakeholder approach aligned to and strengthened existing global and regional frameworks (World Bank, 2014; WHO, 2014). Between 2004 and 2012, alcohol and tobacco consumption were reported to have decreased along with the percentage of individuals that were listed as overweight and obese. More economically active people were reported to be more involved in physical activities (Tonga Department of Statistics et al., 2013a), however, NCD related mortality remained high (WHO, 2018). It is unclear whether improvements were uniform or varied across social classes and areas, or whether the surveys were carried out with the same respondents over the same geographical areas for both surveys.

Fitzgibbon et al. (2012) criticised the design of health interventions, in that they are generalised and not tailored to specific populations with different religious background, values, ethnicity, gender, and in different geographic settings (Wang, 2020). The existing legislation, regulations (WHO, 2014; Tonga Health, 2016) and interventions (Kermode et al., 2015) continue to focus on the prevalence of NCD risk factors. There is also insufficient information on morbidity (illness) (Anderson, 2013; Tonga Ministry of Health, 2014), its distribution and associated risk factors across geographical areas and at different geographical

levels. Understanding the prevalence of NCDs by area and at varying geographical levels would be informative of the socio-economic issues at the smallest scale and the impact and effectiveness of existing interventions.

The current reporting of health data influence the way NCDs are perceived and addressed. Health data is often aggregated and reported into urban and rural areas (Wang, 2020), creating *problems of scale* (Openshaw, 1984). Only three villages make up the urban area (Fig. 1) of Tonga with a population that is less than 30% of the total population (Tonga Department of Statistics and SPC, 2011). The problems with the urban-rural dichotomy is that variability of health in relation to *urbanicity* is unclear, and the reliability and validity of measuring the effects of urbanization on NCD risk factors and interventions is dependent on the spatial scale used (Wang, 2020). Rather than focusing on urban and rural areas, there has been a lot of interest in the association of ill health and deprivation across different geographical areas (Ajebon & Norman, 2016; Seaman et al., 2019). Many high-income countries have used country specific census variables and administrative boundaries at a selected geographical level to develop indices to measure deprivation (Carstairs, 1995; Deas et al., 2003; Havard, 2008; Pampalon, 2009; Salmond & Crampton 2012). However, very few studies have explored the association of deprivation and NCDs in low- and middle- income countries like Tonga. Investigating the characteristics of where people live at different geographical levels and its association with the prevalence of NCDs may provide more insight to better understand the variation and magnitude of risk factors and the prevalence of NCDs.

Geographic Information Systems (GIS) is a powerful analytical tool that has been used in health research because of its ability to integrate multiple datasets for

Fig. 1 Urban and rural areas within the Tongatapu island division

analysis (Murad, 2018; Shafi et al., 2018). The increase in communicable and non-communicable diseases has led to growing interest in the use of GIS in the public health sector (Rodriguez et al., 2013; Kim et al., 2018; Murad, 2018; Shafi et al., 2018). However, despite the growing interest and progress in the use of GIS in public health (Wang, 2019), many institutional, organisational, and technical barriers exist (Göçmen & Ventura, 2010; Kim et al., 2018). Tonga for instance, has very few qualified GIS specialists and trained staff (H, 'Etika, personal communication, July 26, 2019) and are mostly based at the Ministry of Lands and Survey. The Ministry of Health and Statistics department mainly use geospatial technologies for data collection and simple reporting. Only a few government departments such as the GIS unit at the Ministry of Lands and Survey use GIS for advanced spatial analysis (Sifa, 2018). As a result, there are many datasets that exist but have been collected using different methods and have been processed, aggregated, and disseminated into different formats, geographies, and categories that make data integration difficult. This highlights that Tonga does not have an existing policy or framework for data agencies to adhere to for data quality, structure, access and use. Without any standard operating procedures to guide data collecting agencies on data access, management, processing, structure, and quality, the use of these datasets for research will be limited (Norman & Riva, 2012). Health data for instance, were not intended for spatial analysis (Fradelos et al., 2014) therefore, its current structure, quality and format influences its usability outside of the health sector. Other barriers to GIS use, are changes to administrative boundaries (Exeter et al., 2005) and area names (Marsh et al., 2000; Norman et al., 2003) which make data integration in a GIS and any longitudinal study difficult.

Several methods have been used to resolve these data issues, such as creating consistent geographies using different methods to suit the study purpose (Norman et al., 2003), disaggregating population data using dasymetric mapping (Eichhorn, 2020); creating consistent areas through time (CATT) enabling the comparison of census data over time and same areas (Exeter et al., 2005) and; creating geographic conversion tables (GCTs) to apportion, convert and link disparate datasets for areas with data inconsistencies (Wilson & Rees, 1999; Simpson & Yu, 2003), and areas that have gone through name and boundary changes (Wilson & Rees, 1999; Simpson, 2002) over time.

In this paper, we examine the existing use of GIS in Tonga and effectiveness (data integration, accessibility, and representation) in studying the distribution of NCDs. We identify the inconsistencies in the existing geographies and data structure that are relevant to the use of GIS in studying the distribution of NCDs on Tongatapu, the largest island in Tonga (Matoto et al., 2014; WHO, 2016; Sifa 2018; Catalan et al., 2020). The challenges are two-fold. Firstly, there is the problem of inconsistent GIS data structure and secondly, there are inconsistent methods of data collection, recording and disseminating of socio-economic data by different agencies. The paper explores these issues using the Tonga 2011 Census data and the 2012 Demographic and Health Survey (DHS) data. The rural district of Vainī and urban district of Kolofo'ou on Tongatapu make up the study area. Both districts are included in the DHS dataset, have inconsistencies in their GIS data and represent data for urban and rural areas. We present a data conversion framework using geographic

conversion tables (GCTs) and various data processing methods to convert, create new geographies and integrate georeferenced and non-georeferenced datasets. The geography of NCDs in association with area deprivation were analysed using the new geographies.

Data

Two datasets were used to determine the association of the prevalence of NCDs and area deprivation. The 2011 census dataset (Tonga Department of Statistics, 2014) was accessed from the Department of Statistics website. The 2012 demographic and health survey (DHS) dataset (Tonga Department of Statistics et al., 2013b) is kept at the Secretariat of the Pacific Community (SPC) in Fiji and access can only be approved by the Tonga Statistics department or SPC. At the time of this study, the 2012 DHS was the only nationwide survey with data on NCDs that was collected and disseminated using geographies, thus supporting Anderson's (2013) argument of the lack of reliable and readily available data on morbidity. The survey provided an estimate of the general health of the Tongan population with limited information on NCDs. Individuals indicated whether they had sought medical treatment from a list of ailments. Data on individuals that selected any of the listed NCDs (respiratory illness, diabetes, hypertension and other NCDs) were extracted for this study. A total of 2,691 individuals make up the study population of which 80% were from Kolofo'ou and 20% from Vainī. To avoid large demographic differences between the two datasets, the 2011 census dataset was selected as it was collected a few months before the DHS data.

Existing Levels of Geography

Tonga has a dual land tenure system where the country is divided into estates that are either customary administered by the royal family and/or nobles, or freehold land administered by the government (Fig. 2). The existing geographies were created to follow the boundaries of these estates. The five levels of geographies excluding the estates from lowest to highest are census blocks, village, electoral constituency, district and island division. The lower levels nest within the higher levels. Figures 3 and 4 shows the existing geographies in the districts of Kolofo'ou and Vaini in the Tongatapu island division, of which this study is based on. Boundaries of the electoral constituency focuses on achieving an almost homogeneous population distribution, causing inconsistencies with some geographies overlapping two constituencies.

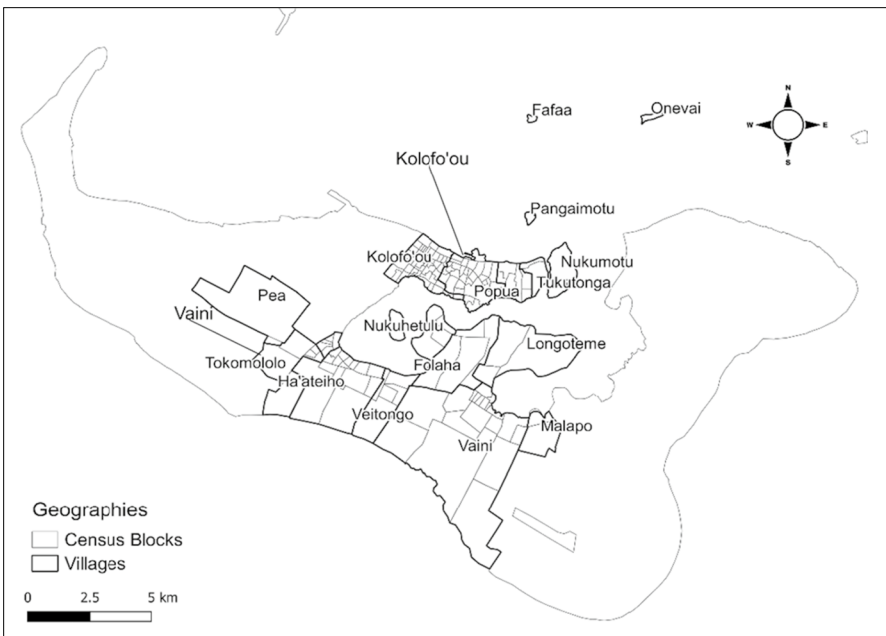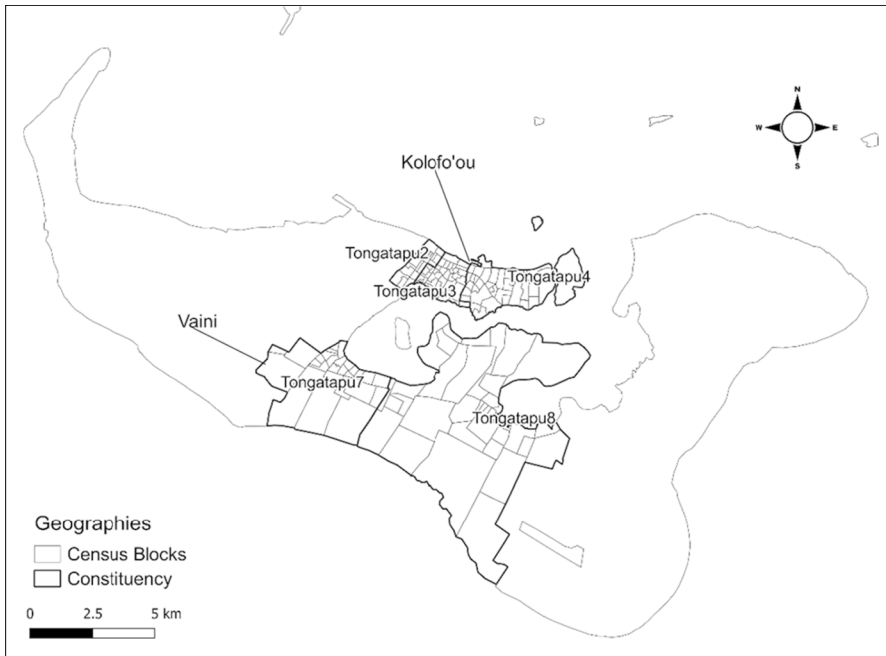**Fig. 2** Boundaries of census blocks and villages in the districts of Vainī and Kolofo'ou

Fig. 3 Boundaries of census blocks and constituencies in the districts of Vainī and Kolofo'ou

Fig. 4 Boundaries of the dual tenure system in the districts of Vainī and Kolofo'ou

Data Processing Methods and Results

This study found five types of data inconsistencies that limit the use of GIS for health research. A general framework (Table 1) for data conversion was developed to address these inconsistencies. This section describes these inconsistencies in detail along with the data processing methods used to address each one and the results.

Geographies Without Identification Codes

Integrating georeferenced and non-georeferenced data in a GIS requires a common attribute field within the datasets to enable a link between the two. The most common fields used are identification codes. While most existing levels of geographies have identification codes that link them to one another, the constituency, and estates – do not have any codes. Without identification codes, integrating data linked to these two geographies is challenging. Table 2 presents a description of the identification codes used for each level of geography illustrating how each relate to another by using the 7-digit code for census blocks. Census block code **1-3-08-090** is ordered from the highest level of geography to the lowest. The first digit represents the island division, then district within the division and then the third and fourth digits represent the village. The last three digits are codes for the census blocks within the village. There is no data structure

Table 1 A framework to link and convert inconsistent and incompatible datasets

Data Inconsistency	Enabler	Criterion/ variables	Method
Geographies without identification codes	Identification or reference codes	• ID Codes	• Assign codes to geographies, preferably alpha numeric codes for easy identification
Geographies without attributes	Best fit table	• None • Source geography • Target geography	• Use existing census blocks of settlements as building blocks • List in an old to new format, with source geography that best fits into the target geography
Settlements without geographical boundaries	Best fit table	• None • Source geography • Target geography	• Use existing census blocks of settlements as building blocks • List in an old to new format, with source geography that best fits into the target geography
Census blocks overlapping two constituencies	Weighted geographic conversion table	• Population or number of households • Area (Hectares)	• Identify source geography that will be converted into the target geography. • Identify weight criterion by using the percentage of population or number of households in the area overlap. • Use area (hectares) overlap in the absence of population and number of households. • Calculate the degree of hierarchy and degree of fit using Simpson's (2002) formula.
Inconsistent data entry methods	• Complete ID codes • Smallest geographical level	• ID Codes	• Record, process and disseminate using full ID codes • Ensure an almost geographically comprehensive dataset regardless of geographical level • Disseminate at smallest geography that can be aggregated into higher levels.

Table 2 A description of existing geographical codes

Level of geography	Code	Description
Country	TO	Country ID for Tonga
Division	1	Division ID for Tongatapu
Constituency	None	Currently no code
Estates	None	Currently no code
District	3	Code for Vainī district
Village	08	Code for Pea village
Census block	090	Census block in Pea

Table 3 New alpha numeric codes assigned to the constituencies

Constituency	New code
Tongatapu 2	CT2
Tongatapu 3	CT3
Tongatapu 4	CT4
Tongatapu 7	CT7
Tongatapu 8	CT8

that connects census blocks to the constituency or village to the estates that it nests into, other than geographical location. Due to the scope of this study, the estates boundary was not included in the data processing.

New Geographies

Alphanumeric codes were assigned to the constituencies covering Kolofo'ou and Vainī district (Table 3). The new code contains 3 new characters. The first letter C refers to Constituency and the second letter is the first letter of the division name, and the number indicates the constituency number. Assigning new codes enables data integration, linking lower levels of geography to the constituency levels and to other geographies. To illustrate the link between all levels of geography, a full code would now have an alphanumeric code identifying the division, constituency, district and village it nests in. For instance, 1-CT6-3-08-050.

1 – Division
CT6 – Constituency
3 – District
08 – Village
050 – Census block

Fig. 5 Unpopulated census blocks in the Vainī district

Geographies Without Attributes

Census blocks were designed to cover populated areas of 50–60 households for census enumerators to cover during the census period. However, a few census blocks cover unpopulated areas and exist only to delimit areas geographically (Fig. 5). As a result, they do not exist in the census dataset but exist in the georeferenced dataset causing unequal number of records between the two datasets. The Vainī district has 61 records in the census data, and 68 in the georeferenced dataset. Seven census blocks exist but cannot be linked to any socio-economic data due to unequal number of records. This issue is only found in the rural district of Vainī and not in the urban district of Kolofo'ou.

Creating Equal Number of Records

In QGIS, the unpopulated census blocks were merged with adjacent populated census blocks using the *'eliminate selected polygon'* tool. To merge with the nearest area, three options were available: largest area, smallest area and largest common boundary. To ensure the merge did not extend to a census block in another village, only one village boundary was selected at a time. The ID codes of the populated census blocks were assigned to the output (merged census block) (Fig. 6) achieving equal number of records for the georeferenced dataset and census data.
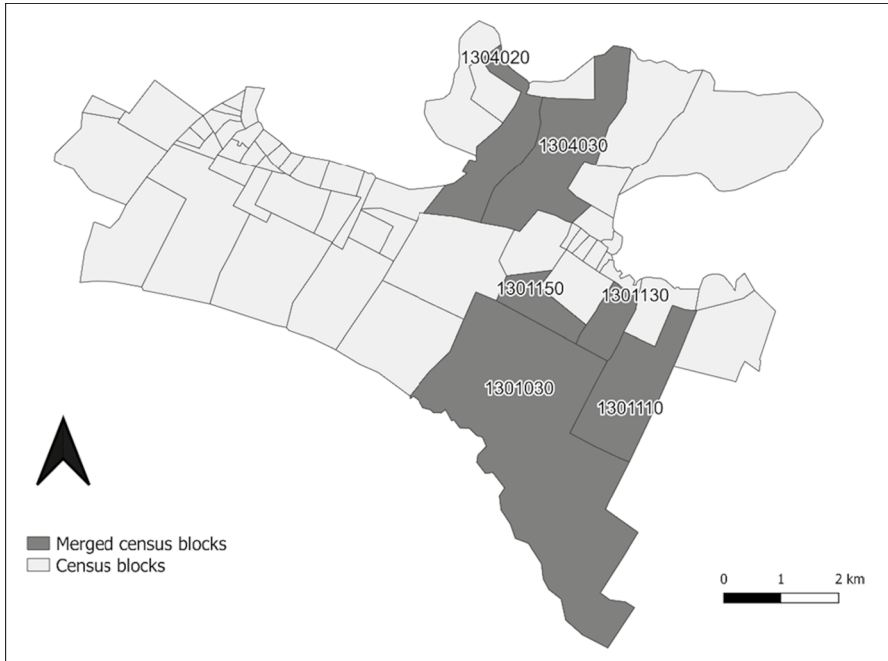
Fig. 6 Unpopulated census blocks merged with neighbouring populated census blocks

Table 4 Settlements without geographic boundaries in the two districts. 2011 Census data

District	Main village	Settlements	Population
Kolofo'ou	Kolofo'ou	'Amaile	585
		Pahu	589
		Fanga	945
		Fasi	1618
		Ngele'ia	1100
		Mataika	714
		Halaleva	**7**22
	Ma'ufanga	Houmakelikao	1488
		'Umusi & 'Anana	649
		Pili	540
		Halaleva	473
	Popua	Popua 'uta Fangaloto	508
Vainī	Pea Central	Lomaiviti	383
		Liahona	209
		Kahoua	303

Settlements Without Geographical Boundaries and Overlapping Two Villages

Settlements have emerged within villages and have expanded in size and population. They are often referred to as *semi-villages* and mistakenly classified as villages. In the Kolofo'ou district, seven settlements are in Kolofo'ou village, four in Ma'ufanga and one in Popua. In the Vainī district, three are in Pea village (Table 4). Socio-economic data for these settlements exist but are always aggregated into the villages and districts they nest within creating *problems of scale* (Openshaw, 1984). In the georeferenced dataset, settlements do not have administrative boundaries and exist as a collection of census blocks within a village. Without geographical boundaries, integrating georeferenced and socio-economic data for settlements is impossible because the attribute data cannot be linked to any settlement boundary or zone for spatial analysis.

Another issue associated with the lack of boundary is a settlement overlapping two villages and having more than one record in the census data. This is the case of Halaleva settlement (Table 4). In the georeferenced dataset, Halaleva is made up of six census blocks. Four census blocks are in Kolofo'ou and two are in Ma'ufanga village (Fig. 7). With this overlap, socio-economic data for Halaleva are split into two villages which can lead to misrepresenting their issues.
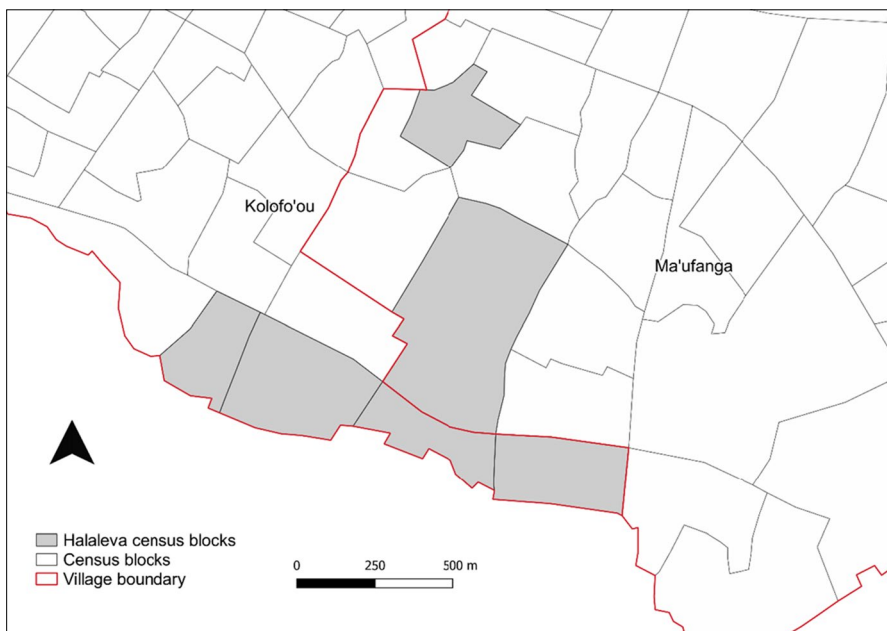
Fig. 7 Census blocks for Halaleva settlement overlapping the villages of Kolofo'ou and Ma'ufanga

Table 5 An extract of the best fit geography conversion table

Source Geography	Main village	Target Geography
1,101,200	Kolofo'ou	Ngele'ia
1,101,220	Kolofo'ou	
1,101,230	Kolofo'ou	
1,101,380	Kolofo'ou	
1,101,400	Kolofo'ou	
1,101,210	Kolofo'ou	Mataika
1,101,390	Kolofo'ou	
1,101,420	Kolofo'ou	
1,101,160	Kolofo'ou	Fasi
1,101,170	Kolofo'ou	
1,101,180	Kolofo'ou	
1,101,191	Kolofo'ou	

A Best Fit Conversion Table – Many to One Relationship

A best fit table (Wilson & Rees, 1999) was created to address the issues associated with settlements lacking a boundary and overlapping two villages. Best fit tables list the source and target geography in an old to new format (Wilson & Rees, 1999). The source geography is listed next to the target geography that it best fits into and will be converted to. Table 5 shows the source geography (census blocks), the main village in which data from the source geography was often aggregated into, and the target geography (settlements) in which the source geography best fits into. Despite Simpson's (2002) criticism on the possible errors of approximating population distribution caused by allocating socio-economic data to new geographies, the census blocks collectively make up each settlement and were easily merged without redistributing population. Therefore, the best fit conversion table was suitable.

Using the results of the best fit table, the census blocks that make up each settlement were amalgamated in QGIS to create individual villages with geographical boundaries. The census blocks that make up a single settlement were selected, and the inner boundaries were dissolved and then merged to create the boundary for each settlement. The boundaries of the old villages were then deleted from the main village layer, and the new settlement boundaries were merged with the original village layer. These processes were manually repeated for all the settlements until the final village layer contained all the newly created villages. The new geographies retained the names of the settlements. All the datasets were then updated with new ID codes (Table 6) for all the new geographies, ensuring they can be linked to existing geographies. In the Kolofo'ou district, three new geographies: Loto Kolofo'ou, Loto Ma'ufanga and Loto Popua maintained the main village codes as they are the original village centres. In the Vainī district, the boundary of Pea was dissolved to create new boundaries for the settlement of Lomaiviti, Liahona and Kahoua. Eleven new villages (Fig. 8 Map b) were created for the settlements in the Kolofo'ou district

Table 6 A list of the new geographies with new identification codes

District	Main villages	Village ID	New geography	New village ID
Kolofo'ou	Kolofo'ou	1101	Loto Kolofo'ou	1101
	Ma'ufanga	1102	Loto Ma'ufanga	1102
	Nukumotu	1103	Loto Popua	1105
	Popua	1105	Fanga	1111
	Tukutonga	1106	Fasi moe Afi	1112
	Pangaimotu	1107	'Amaile	1113
	Fafaa	1108	Pahu	1114
	'Onevai	1109	Mataika	1115
			Halaleva	1116
			Ngele'ia	1117
			Pili	1118
			Popua 'uta Fanga'uta	1119
			Houmakelikao	1120
			'Umusi/'Anana	1121
Vainī	Vainī	1301	Lomaiviti	1310
	Malapo	1302	Liahona	1311
	Longoteme	1303		
	Folaha	1304		
	Nukuhetulu	1305		
	Veitongo	1306		
	Ha'ateiho	1307		
	Pea	1308		
	Tokomololo	1309		

in addition to the three main villages: Loto Kolofo'ou, Loto Ma'ufanga, Loto Popua and Tukutonga.

Census Blocks Overlapping Two Constituencies

In 2010, 17 electoral constituencies were created for the general elections under the new electoral system (Electoral Boundaries Commission Act, 2010). The constituency boundaries were then used in the 2011 national census survey to aggregate and disseminate data for the 2014 general elections. With the creation of the new constituencies, we found that boundaries of four census blocks overlap into two constituencies shown in Fig. 9 map a. In the attribute table and census data, there is only one record of these census blocks nesting into one constituency as a one-to-one relationship. The same boundaries used in the 2014 general elections were used for the 2017 elections (Electoral Boundaries Regulations, 2017) and at the time this study was carried out, the boundaries had not been revised. We contend that the problem is a digitizing error and requires adjustment.
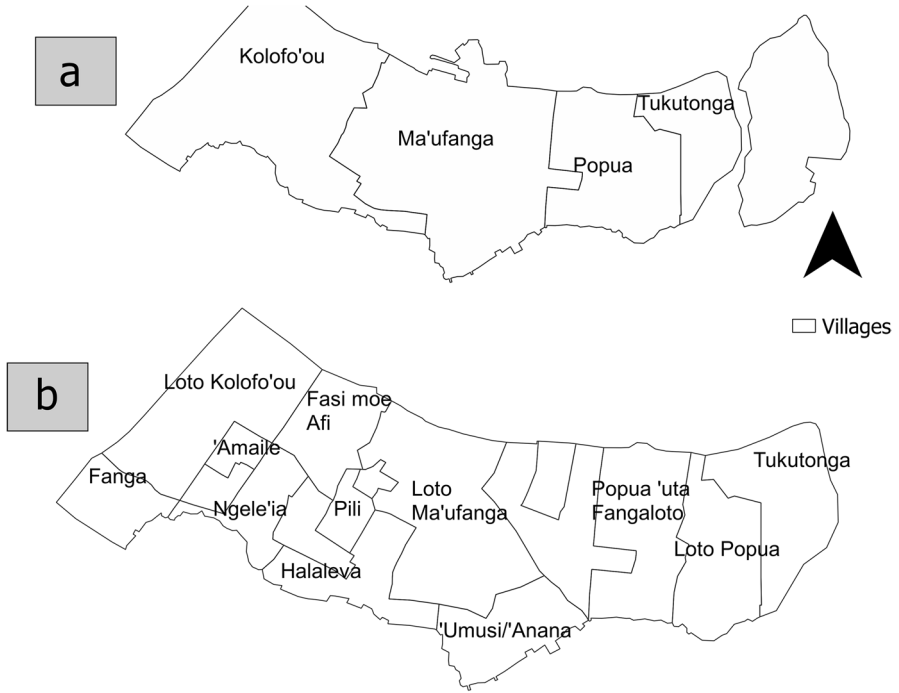
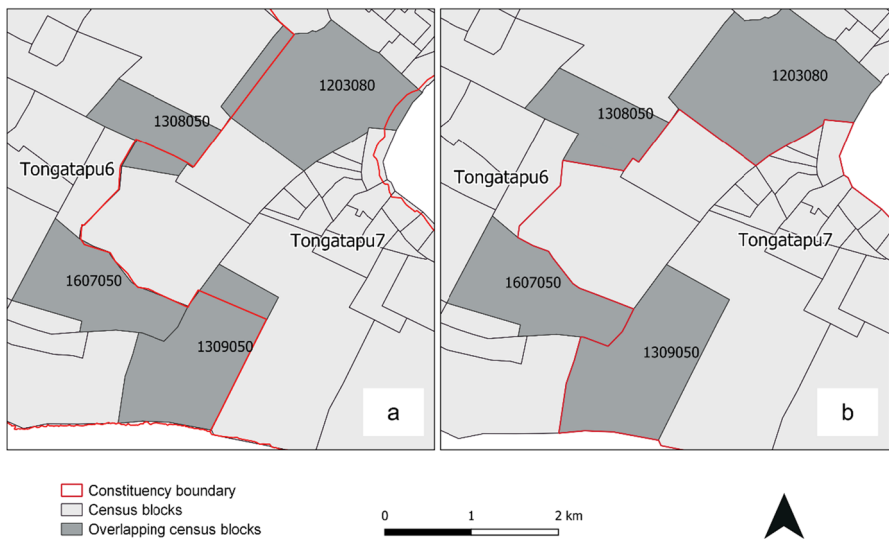**Fig. 8** Old and new village boundaries in the Kolofo'ou district

Fig. 9 Census blocks overlapping two constituencies and redefined boundaries using a weighted GCT

Table 7 Geographic conversion table using area as weight criterion

Source Geography	Source name	Target geography	Total Area (Ha)	Area overlap (Ha)	Weight
1,308,050	Pea-Lomaiviti 5	CT6	425.7	409.7	0.962415
1,308,050	Pea-Lomaiviti 5	CT7	425.7	16	0.037585
1,607,050	'Utulau-4	CT6	156.7	135	0.861519
1,607,050	'Utulau-4	CT7	156.7	21.7	0.138481
1,309,050	Tokomololo 5	CT6	220.4	32	0.145191
1,309,050	Tokomololo 5	CT7	220.4	188.4	0.854809
1,203,080	Tofoa 8	CT6	264.3	35.3	0.13356
1,203,080	Tofoa 8	CT7	264.3	229	0.86644

Weighted Table – One to Many Relationship

A weighted geographic conversion table was used to apportion the census blocks overlapping two constituency boundaries. The table consist of a source geography (census blocks), target geography (constituency) and a weight criterion (area in hectares) (Table 7). The weight criterion was used to calculate the overlap and determine where the boundary should be. Since the census and DHS data only have one record per census block and the overlap only exist in the geometry of the data – area was used as the weight criterion. In QGIS, the total area of the census blocks and area overlap were calculated in hectares. The weight was calculated by dividing the area overlap by the total area and it shows how much of the source geography overlaps the target geography. The source geography with a weight less than 0.4 was merged into the target geography with the higher weight. Using the result of the weighted GCT, the boundary of the census blocks was apportioned in QGIS to nest in the target geography (Fig. 9 map b).

Inconsistent Methods of Data Entry

Data collection for the DHS survey used census blocks and some village names, however, data entry only recorded the last two digits of the 7-digit census block code. With the missing identification codes of the higher levels of geography, integrating the DHS data with the georereferenced dataset becomes problematic for data users, particularly data processing for novice GIS users. This issue questions the reliability and usability of nationwide health data. Two records were recorded as block number 10 and another 110. Without the full code, one can easily be mistaken for the other or aggregated into another area if the village column is missing.

Data Cleaning

The full dataset was manually processed in Microsoft Excel so that each record contains the full code of each census block. The final georeferenced, census and DHS datasets used

in this study were all updated and consist of the new codes for the constituency, village and census blocks. Each geography can be linked to one another using the identification codes.

Constructing a Deprivation Index

Using available census variables, a deprivation index was developed for Tongatapu to determine possible association of deprivation with the prevalence of NCDs across different geographical areas. Low social class, unemployment, lack of car ownership and overcrowded households are commonly used variables to develop deprivation indices (Townsend, 1987; Carstairs, 1995; Seaman et al., 2019) in many western countries. However, they do not reflect socio-economic disadvantage in the Tongan context. From the census dataset, variables that reflect disadvantages in the Tongan context were not available in a compatible structure and format. Full access to the census microdata of the Tonga Population and Housing Census used by Catalan et al. (2020) would have provided variables that are more suited to measure standard of living. However, the variables were not available in time for this study. Four variables were selected; *% renting, % depend on neighbour's water source, % no strong income and % that cook with firewood* and all reflect socio-economic disadvantage to some extent. Two variables were available at the village level and two at the district level. The variables at village level were aggregated to district level to ensure the deprivation index consisted of variables in the same unit of analysis. Table 8 shows that the association between the variables ranged from a very weak negative to a positive correlation. Despite the variation and high uncertainty in correlation, the variables were selected for the purpose of this study.

In SPSS, the variables were standardized into z-scores and added (with equal weighting) to create the deprivation index, which is a continuous score ranging from below average (less deprived) to above average (more deprived). A negative z-score indicates a value that is below the average and a positive score is interpreted as above average. The deprivation scores were then categorized into five population weighted quintiles so that it can be analysed alongside the NCD data. The weight used was the total population of the districts. Moving along the deprivation scale, quintile 1 represents least deprived areas and quintile 5 represents most deprived areas. The variables were then assessed against the deprivation index and Table 8 shows a moderate correlation.

Deprivation Index for Tonga

Deprivation was higher in populations that use firewood for cooking and those without a direct water source. It varied across the districts of Tongatapu. Urban districts were in the least deprived quintile and the rural districts ranged from average to the most deprived. Although these results are consistent with the findings in Norman (2010) and Datta and Singh (2016), it is not indicative of the level of deprivation because some of the variables had very weak correlation with one another. The urban district of Kolofo'ou appear to have low levels of deprivation but this is questionable and is subject to the 'ecological fallacy' problem. If one or two villages are in the *most deprived* category, data aggregation changes the scale of the problem at

Table 8 Relationship between the deprivation indicators and the deprivation index

Correlations (p values)	%Depend on neighbour's water source	%No strong income	%Energy for cooking with firewood	Deprivation Index
%Renting	-0.150	-0.538	-0.944**	-0.683
	(0.748)	(0.212)	(0.001)	(0.091)
%Depend on neighbour's water source	1.000	-0.462	0.195	0.630
		(0.297)	(0.675)	(0.130)
%No strong income		1.000	0.328	0.354
			(0.472)	(0.436)
%Energy for cooking with firewood			1.000	0.625
				(0.133)

** Correlation is significant at the 0.01 level (2-tailed) N=7

the district level. The results lack information on area deprivation at the lower levels of geographies that are often overlooked when analysing at higher levels of geography (Salmond & Crampton, 2012) and ignores the history of the areas (Jivraj et al., 2019).

Complementing existing literature (Carstairs, 1995; Deas et al., 2003), we found area deprivation to be associated with populations with low income. Catalán et al. (2020) identified poor populations in Tonga by combining deprivation and income to create a poverty index. Using five deprivation indicators: fridge ownership, adequate wall materials, landline phone access, flush toilet and having an indoor kitchen, poverty was lower in the urban areas and higher in the rural areas. However, using these variables to define an acceptable standard of living is framed from a privilege perspective disregarding the perceptions of those labelled poor and deprived (Fu et al., 2015).

Most deprivation indices use variables that are from a euro-centric viewpoint. Variables such as *% overcrowded households* and *% with no cars* do not reflect deprivation in Tonga. Overcrowded households are common in Tonga and indigenous populations (Fu et al., 2015). It was common with families that own large acres of land and farms, having the ability to provide for a large household. The use of *% with no cars* to measure deprivation is not applicable. in a few inhabited islets around Tongatapu that are a part of the Kolofo'ou district. Car ownership in these islands is irrelevant to their state of wellbeing (Shohaimi et al., 2004; Norman, 2010).

To create a deprivation index that is reflective of the Tongan context, it would be appropriate to add variables that reflect an individual's ability to; contribute and participate in social gatherings (Lee, 1998) and communicate in the local language (Cournane et al., 2015). Variables such as having ownership and access to plantations, livestock and traditional mats and having the necessary skills should be considered. Apart from household ownership, these are central to the *nofo 'a kainga* or the way Tongan people live and interact with each other (Morris, 2014). Crops and livestock are not only a food source, but they also play a fundamental role in Tongan festivals, ceremonies, and gift exchange. Ownership of plantations, livestock and mats reflects a household's ability to freely participate and contribute to social events without financial difficulty. Literacy of the Tongan language is also an important variable to consider. A person that cannot communicate within the community he or she lives in is subject to social deprivation (Cournane et al., 2015).

The variables selected in this study are not great indicators of deprivation both in the existing literature and in the context of Tonga. This was evident in the weak correlation between the variables. However, they were selected because they were linked to deprivation to some extent and available in one unit of analysis. More suitable variables were aggregated into different levels of geography and categories and could not be used for this study. For instance, marital status and literacy were aggregated into age group, employment was available at the division level. Household income and tenure were available at the district level. Variables on water and energy source, toilet facility and waste disposal were available at the village level. Therefore, variables at the village level were aggregated into the district level to align with the two variables at the district level.

There are obvious limitations because these variables does not reflect true area deprivation for Tonga. However, it highlights the existing data problems, including
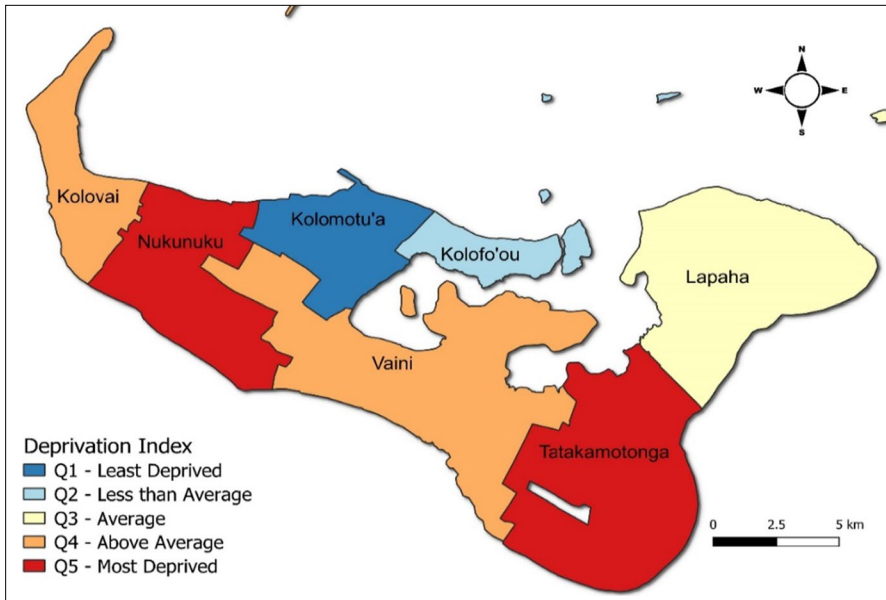
Fig. 10 Deprivation across the seven districts of Tongatapu

access to comprehensive national data in compatible geographies and format (Jivraj et al., 2019) and data access in general. If robust evidence is required to achieve the objectives of the Pacific Road map (World Bank, 2014), these issues must be addressed at a multi-stakeholder level for data agencies to produce quality data and provide easier access to data. This paper contributes to the current literature on area deprivation, by setting the groundwork for a Tonga specific deprivation index using geographic conversion tables to convert existing incompatible datasets into a usable format.

Deprivation and Non-communicable Diseases

We used the approach identified by Norman et al. (2003) and aggregated individual data into the geography best suited for the purpose of this study. To be on the same unit of analysis as the indicators of deprivation, data on NCDs were

Table 9 The relationship of the deprivation index and non-communicable diseases

Correlations (*p* values)	%Respiratory illness	%Diabetes	%Other
Deprivation score	0.404 (0.369)	0.716 (0.070)	0.451 (0.310)
%Respiratory illness	1.000	-0.275 (0.551)	0.651 (0.113)
%Diabetes		1.000	0.123 (0.793)

%Hypertension is omitted here
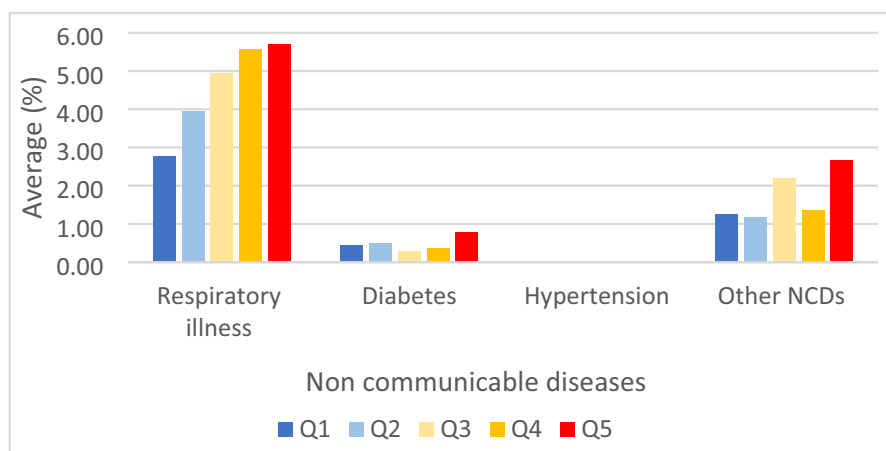
Fig. 11 Average rate of deprivation in association with NCD's in Tongatapu, 2011

aggregated from household data to census blocks and into the district level, imported into SPSS and merged with the census variables using the district codes. Figure 10 shows that the urban districts of Kolofo'ou and Kolomotu'a were the least deprived and deprivation varied in the rural areas. At the district level, it is unclear which villages are more deprived.

In determining the association between deprivation and NCDs, a correlation matrix (Table 9) revealed a strong correlation between deprivation and the prevalence of respiratory illness, diabetes and other NCDs. It is highly likely that they are more prevalent in areas that are more deprived. The spatial distribution of NCDs across deprived areas in Tongatapu was analysed by aggregating the NCD variables into the five deprivation quintiles in SPSS using mean statistics. Figure 11 shows the prevalence of NCDs increasing along the deprivation quintiles. Hypertension appears to be less common than the other NCDs as none of the respondents from the study areas selected hypertension. Complementing existing studies (Stafford & Marmot, 2003; Shohaimi et al., 2004; Di Cesare et al., 2013; Collins et al., 2018), NCDs were more prevalent in Q5 - the most deprived areas.

Distribution of NCDs Across Different Geographical Levels

At the census block level, the prevalence of NCDs is unclear (Fig. 12). The data were aggregated into village and district level using the new geographies which showed problems of scale. The results does not show the true prevalence of NCDs and is due to the data structure as a result of method of data collection for reporting health.
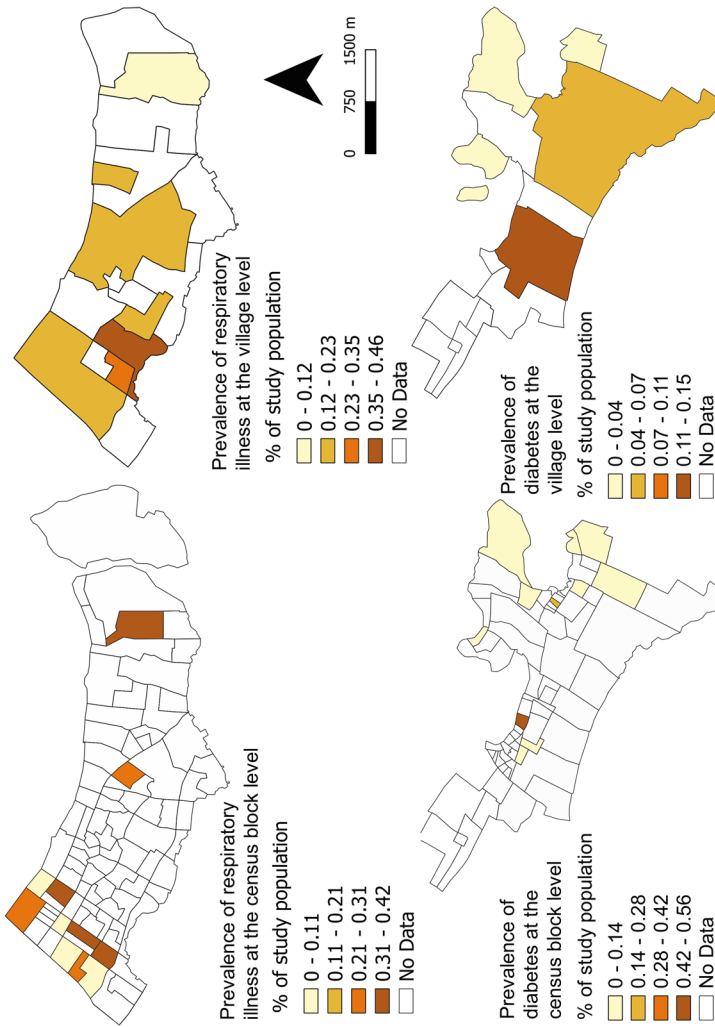
Fig. 12 Problems of scale displayed as data on NCDs is aggregated from census block to village level

Prevalence of respiratory illness at the village level

% of study population
- 0 - 0.12
- 0.12 - 0.23
- 0.23 - 0.35
- 0.35 - 0.46
- No Data

Prevalence of diabetes at the village level

% of study population
- 0 - 0.04
- 0.04 - 0.07
- 0.07 - 0.11
- 0.11 - 0.15
- No Data

Prevalence of respiratory illness at the census block level

% of study population
- 0 - 0.11
- 0.11 - 0.21
- 0.21 - 0.31
- 0.31 - 0.42
- No Data

Prevalence of diabetes at the census block level

% of study population
- 0 - 0.14
- 0.14 - 0.28
- 0.28 - 0.42
- 0.42 - 0.56
- No Data

0 750 1500 m

Discussion

The use of GIS in Tonga has been limited to a few government departments and utility companies. Using GIS to support health research has not been widely recognised. While investigating the use of GIS to study NCDs, several data issues were found to limit data integration and the use of GIS. Complementing Göçmen and Ventura (2010), we found that data structure and characteristics of available georeferenced and non-georeferenced data influenced by organizational and institutional factors were the main barriers.

The first data issue presented is the creation of the electoral constituencies without identification codes. This is problematic when integrating multiple datasets where a spatial join relies on identification codes. Additionally, when boundaries of constituencies change (Exeter et al., 2005) but their names remain the same becomes problematic when comparing statistics over time and space (Norman et al., 2003). A slight shift in a boundary can have a significant impact on the redistribution of population between the old and new boundaries (Wilson & Rees, 1999). Without knowledge of boundary changes, can lead to inaccurate interpretations on demographic changes. Ideally, the lower levels of geography should be able to be linked and can be aggregated into the old and new constituencies using ID codes (Wilson & Rees, 1999).

Using different data types from different data collecting agencies have always been difficult for data users (Göçmen & Ventura, 2010). This is because data collection agencies collect, process and disseminate data in a format that suit the purpose of GIS in their organisations rather than for the broad use of data by different end users. The issue of unequal number of records for census blocks in two different datasets, is subject to data handling errors and misinterpretation of data coverage by novice data users. Merging the census blocks that cover unpopulated areas with adjacent populated areas was essential to achieve equal number of records. Although this raises the problem of presenting socio-economic data over unpopulated areas which may be misleading, the choice of the unit of analysis and precise data labelling becomes critical when presenting data, so that the *problems of scale* are avoided (Openshaw, 1984).

Complementing the findings by Norman et al. (2003) and Norman and Riva (2012), this study found similar boundary problems. The existence and expansion of settlements without geographical boundaries within larger villages, present *problems of scale*. Although data for these settlements exist, settlements have not been recognised as an official level of geography in Tonga nor have they been officially named as a village, even though the population of many urban settlements are higher than some rural villages. As a result, socio-economic data for these settlements must be aggregated into a higher level of geography before it can be used in a GIS. However, different aggregation methods change the scale of data (Openshaw, 1984) affecting prioritisation of issues and areas at risk, and resource allocation for these settlements. The best-fit table (Wilson & Rees, 1999) provided a solution to create new boundaries for the settlements and existing villages that the settlements nest in, without altering population distribution in the

area. The boundaries of the census blocks that make up each settlement were used to create the boundary for the new geography so the population in the original settlements remain the same for the new geography.

Data processing was carried out manually and was time consuming and could be resolved by using automated systems AZTool software designed to create census output areas for the UK (Martin, 2003; Mokhele et al., 2016). The tool enables integrating data from two different geographies by creating zones for a specific purpose that is not subject to the MAUP problem of aggregating data from one geography to another to produce a new geography (Openshaw, 1984; Mokhele et al., 2016) used the AZTool software, to create optimized census areas for data collection and dissemination in South Africa. The results created output areas that were socially homogeneous and were consistent for both rural and urban areas using two criterion – type of housing and tenure to determine homogeneity in each output zone. However, suggesting the AZTool to create output areas in Tonga has its challenges. Firstly, Tonga does not have an institutional data framework to provide guidance to existing organisations. Secondly, the output may not conform to Tonga's dual tenure system and may suggest land division (Latu & Dacey, 2014).

The fourth problem identified in this study were census blocks overlapping two constituencies. This is a problem when studying spatial patterns of socio-economic variables at the census block level across different constituencies. Overlapping geographies cause misrepresentation and misinterpretation of data if not resolved (Martin et al., 2002). For instance, a census block that overlaps two constituencies can change the information on population distribution or other variables studied at the census block level in the two constituencies. Without the census block label, an overlapping census block could easily be misinterpreted as two separate census blocks in two different constituencies –when in fact it is only one.

The lack of awareness on the impact of these digitizing errors on data use has not been a priority since the introduction and slow development of GIS in Tonga. Without a framework to guide data management, quality control and use agreed upon by stakeholders, existing data problems and the production of inconsistent and low-quality datasets will continue (Barndt, 1998). Using the weighted GCT to resolve this issue is a limited approach (Norman et al., 2003) because the required variables – population and number of households per census block were unavailable. Although Exeter et al. (2005) created consistent areas through time (CATTs) using population and number of households to merge and dissolve boundaries, we use area as a weight criterion because the overlap exists only in the geometry of the GIS data caused by digitizing errors. Therefore, using area as the criteria to determine the boundary was suitable in this case. The weighted GCT created consistent records in the GIS and census data. In QGIS, each overlapping census block was assigned to a single constituency according to the results of the weighted GCT. Despite Simpson's (2002) criticism that using area as a criterion may cause errors in approximating population, using area as the criteria, demonstrated how to resolve issues of overlapping geography in the absence of the required and suitable variables.

The final problem found in this study, were the problems of data collection and inconsistent methods of data recording. Firstly, health data was collected using only a few census blocks from some villages. The DHS data was aggregated and reported

into urban and rural areas (Tonga Department of Statistics et al., 2013a), generalizing the status of health in Tonga. Based on the health data used and existing geographies, there are no clear patterns or distribution of NCDs at different geographical levels. When the data was aggregated from census block level to the village, district and urban areas, it created the *problem of scale* and *ecological fallacy problem* (Openshaw, 1984). With the issues on boundary changes discussed earlier, an expansion of urban areas in the future may challenge the spatio-temporal analysis of health data. Cataife (2014) found that using the smallest area statistics provided a clearer picture of the prevalence of obesity between affluent and less affluent areas. Therefore, identifying the geography of NCDs in lower levels of geographies can highlight populations that are deprived of certain interventions, useful for a more targeted approach.

Secondly, the DHS data on NCDs were recorded using names of district, village and the last two digits of the census block codes rather than the whole code. This not only limits the ability of the data to be integrated with other datasets at the census block level, but it also causes confusion for data users because many census block codes have the same last two digits. Accessing data in the incorrect format with incorrect geocoding information affects the quality and use of data (Barndt, 1998).

The inconsistent method of collecting and disseminating health data, aggregating census data and inconsistent GIS data did not produce clear gradients of deprivation and NCDs in the two districts. However, it highlighted existence of low-quality health data and lack of data on morbidity that Anderson (2013) pointed out.

Institutional Framework

Geographic Information System can either be central to the day-to-day operation of an organisation or used for basic reporting and data visualisation (Campbell, 1999). The role of GIS in an organisation determines the quality of data, usability and transferability (Somers, 1998). Recognising the importance of data quality and the challenges associated with the current data structure due to barriers highlighted by Göçmen and Ventura (2010), we provide and suggest a general institutional framework (Fig. 13). It aims to provide an enabling environment for data use and integration. Of particular importance, is the institutional body made up of representatives from data collecting agencies and data users. Their main role is to deliberate and agree on the procedures for data collection, processing, and dissemination that agencies must adhere to. To guide data management, the framework consists of four main areas: methods of data processing, achieving quality data, data structure and available resources to produce consistent quality datasets.

Conclusions

This study identified several data inconsistencies limiting the use of GIS for advance spatial analysis and research related to the prevalence of NCDs and its association with area deprivation. These limitations are also the main barriers
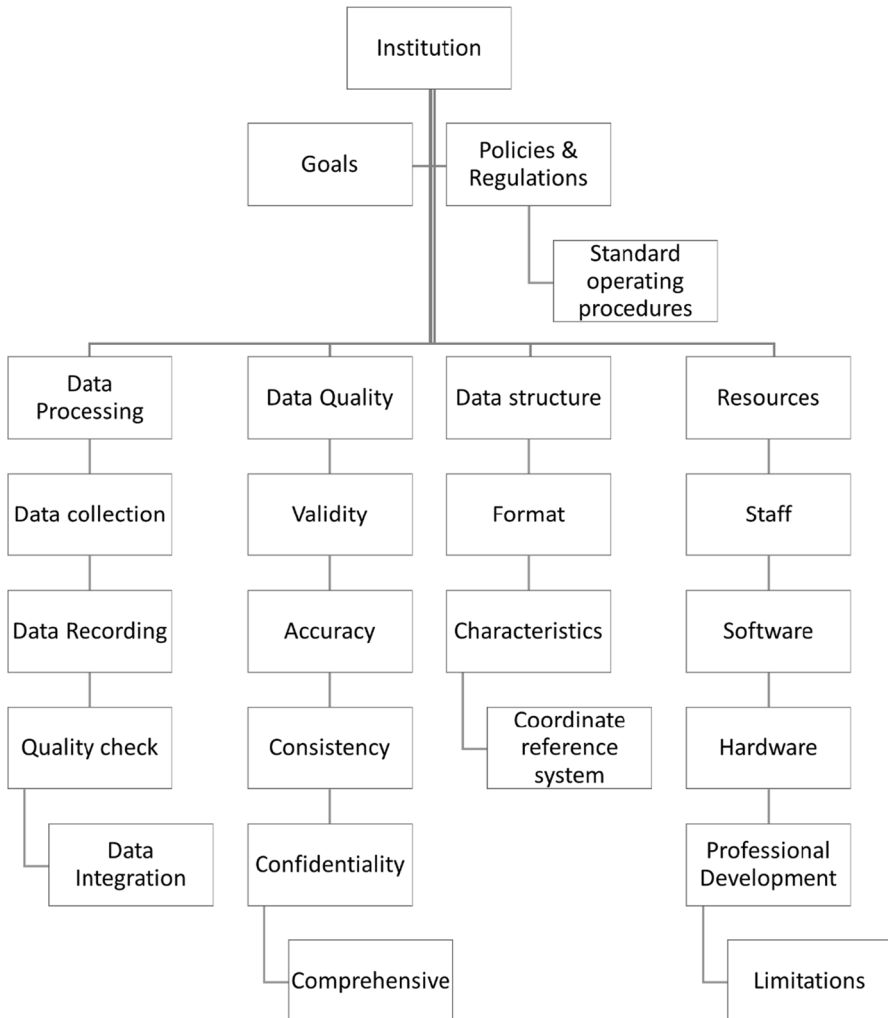
Fig. 13 An institutional framework for multi-stakeholder standard operating procedures

for the slow diffusion of GIS into various sectors and the lack of GIS awareness and use in general. Various data collecting agencies collect and process data into the format that is suitable for their data use and reporting. Therefore, the disseminated data are aggregated and grouped into different levels of geographies or incompatible formats. Collecting and aggregating data from census blocks to urban and rural areas and inferring the results are homogeneous, emphasised the *problems of scale* and *ecological fallacy* with the existing methods. Geographic conversion tables and selected data processing methods were used to create compatible datasets for data integration and analysis.

The existing data inconsistencies limits the freedom to use the data for more sophisticated use other than basic reporting. The relevant census variables required to develop a deprivation index were not available in the desired unit of analysis, limiting the findings of this study. Despite this, this study contributes to the current literature on deprivation in that it created the first deprivation index for Tonga – a Pacific Island country. Supporting existing literature, deprivation was found to be higher in rural areas and NCDs was more prevalent in deprived areas. However, because of the quality of data, the results were unable to show the distribution of NCDs at the smallest level of geography.

The methodology employed can be reproduced in other areas of health research with similar data types. However, we recommend the development of automation processes for data processing and conversion of large national datasets, exploring the dasymetric mapping technique to disaggregate population data using land use data, and using longitudinal data (Jivraj et al., 2019). Additionally, an in-depth study on deprivation in Tonga, creating a deprivation index at low geographical levels using variables that are relevant to Tonga is an important area of research.

Declarations

References

Ajebon, M., & Norman, P. (2016). Beyond the census: a spatial analysis of health and deprivation in England. *GeoJournal, 81*(3), 395–410. https://doi.org/10.1007/s10708-015-9624-8

Anderson, I. (2013). *T*he economic costs of noncommunicable diseases in the Pacific Islands. A rapid stocktake of the situation in Samoa, Tonga and Vanuatu. *Health and Nutrition and Population (HNP) Discussion Paper* (pp. 1–75). The World Bank.

Barndt, M. (1998). Public participation GIS – Barriers to implementation. *Cartography and Geographic Information Systems*, *25*(2), 105 – 112. https://doi.org/10.1559/152304098782594607

Campbell, H. (1999). Institutional consequences of the use of GIS. In P. Longley, M. Goodchild, D. Maguire, & D. Rhind (Eds.), *Geographical information systems* (pp. 621–631). Wiley.

Carstairs, V. (1995). Deprivation indices their interpretation and use in relation to health. *Journal of Epidemiology and Community Health, 49*(Suppl 2), S3–S8.

Cataife, G. (2014). Small area estimation of obesity prevalence and dietary patterns. A model applied in Rio de Janeiro. *Health & Place, 26,* 47–52. https://doi.org/10.1016/j.healthplace.2013.12.004

Catalán, H., Fifita, V., & Fainga'anuku, W. (2020). Small-area multidimensional poverty estimates for Tonga 2016 drawn from a hybrid Hierarchical Bayesian estimator. *Applied Spatial Analysis and Policy, 13,* 305–328. https://doi.org/10.1007/s12061-019-09304-8

Collins, P., Elia, M., Kurukulaaratchy, R., & Stratton, R. (2018). The influence of deprivation on malnutrition risk in outpatients with chronic obstructive pulmonary disease (COPD). *Clinical Nutrition, 37,* 144–148.

Cournane, S., Byrne, D., Conway, R., O'Riordan, D., Coveney, S., & Silke, B. (2015). Social deprivation and hospital admission rates, length of stay and readmissions in emergency medical admissions. *European Journal of Internal Medicine, 26,* 766–771.

Datta, S., & Singh, K. (2016). Analysis of child deprivation in India: Focus on health and educational perspectives. *Economic Analysis and Policy, 50,* 120–130.

Deas, I., Robson, B., Wong, C., & Bradford, M. (2003). Measuring neighbourhood deprivation: A critique of the Index of Multiple Deprivation. *Environment and Planning Government C: Government and Policy, 21,* 883–903.

Di Cesare, M., Khang, Y., Asaria, P., Blakely, T., Cowan, M., Farzadfar, F., & Ezzati, M. (2013). Inequalities in non-communicable diseases and effective responses. *Lancet, 381,* 585–597.

Eichhorn, S. (2020). Disaggregating population data and evaluating the accuracy of modeled high-resolution population distribution – The case study of Germany. *Sustainability, 12,* 3976. https://doi.org/10.3390/su12103976

Electoral Boundaries (2017). *Regulations*. Attorney General's Office.

Electoral Boundary. (2010). *Commission Act*. Attorney General's Office.

Exeter, D., Boyle, P., Feng, Z., & Flowerdew, R. (2005). The creation of consistent areas through time (CATTs) in Scotland 1981–2001. *Population Trends, 119,* 24–36.

Fitzgibbon, M., Tussing-Humphreys, L., Porter, J., Martin, I., Odoms-Young, A. & Sharp, L. (2012). Weight loss and African-American women: A systematic review of the behavioural weight loss intervention literature. *Obesity reviews, 13,*193–213.

Fradelos, E., Papathanasiou, I., Mitsi, D., Tsaras, K., Kleisiaris, C. & Kourkouta, L. (2014). Health based geographic information systems (GIS) and their applications. *Acta Informatica Medica, 22*(6), 402–405.

Fu, M., Exeter, D., & Anderson, A. (2015). The politics of relative deprivation: A transdisciplinary social justice perspective. *Social Science & Medicine, 133,* 223–232.

Göçmen, Z., & Ventura, S. (2010). Barriers to GIS use in planning. *Journal of the American Planning Association, 76*(2), 172–183. https://doi.org/10.1080/01944360903585060

Havard, S., Deguen, S., Bodin, J., Louis, K., Laurent, O., & Bard, D. (2008). A small-area index of socio-economic deprivation to capture health inequalities in France. *Social Science & Medicine, 67*(12), 2007–2016.

Hufanga, S., Carter, K., Rao, C., Lopez, A., & Taylor, R. (2012). Mortality trends in Tonga: as assessment based on a synthesis of local data. *Population Health Metrics*, *10*(14), [no pagination].

Jivraj, S., Murray, E., Norman, P., & Nicholas, O. (2019). The impact of life course exposures to neighbourhood deprivation on health and well-being: a review of the long-term neighbourhood effects literature. *The European Journal of Public Health, 30*(5), 922–928.

Kermode, M., Armstrong, G., Gloede, S., Filiai, S., 'Ofanoa, R., Hufanga, S., & Tukia, O. (2015). *Public health evidence related to obesity in Tonga*. A literature review. Unpublished.

Kim, D., Zhang, Y., & Lee, C. (2018). Understanding needs and barriers to using geospatial tools for public health policymaking in China. *Geospatial Health, 13*(594), 79–85.

Latu, L., & Dacey, S. (2014). *LIS and customary land tenure: The Tongan Approach.* 19th Annual Conference of the National Advisory Committee on Computing Qualifications (NACCQ 2006), 10 July, Wellington.

Lee, H. (1998). Creating their own culture: diasporic Tongans. *The Contemporary Pacific, 10*(1), 1–30.

Marsh, A., Gordon, D., Heslop, P., & Pantazis, C. (2000). Housing deprivation and health: A longitudinal analysis. *Housing Studies, 15*(3), 411–428.

Martin, D., Dorling, D., & Mitchell, R. (2002). Linking censuses through time: problems and solutions. *Area, 34*(1), 82–91.

Martin, D. (2003). Extending the Automated zoning procedure to reconcile incompatible zoning systems. *International Journal of Geographical Information Science*, *17*(2), [no pagination].

Matoto, V., Viney, K., Roseveare, C., Colaguiri, R., & Marais, B. (2014). Burden and spectrum of disease in people with diabetes in Tonga. *Public Health Action, 4*(2), S44–S49.

Mokhele, T., Mutanga, O., & Ahmed, F. (2016). Development of census output areas with AZTool in South Africa. *South African Journal of Science, 112*(7/8), 1–7.

Morris, N. (2014). Probability, uncertainty and the Tongan way. In P. Liljedahl, S. Oesterle, C. Nicol, & C. Allan (Eds) *Proceedings of the joint meeting of PME 38 and PME-NA 36*, *[no date]* (pp. 241–248). PME.

Murad, A. (2018). Using GIS for determining variations in health access in Jeddah city, Saudi Arabia. *International Journal of Geo-Information, 7*(254), [no pagination].

Norman, P. (2010). Identifying change over time in small area socio-economic deprivation. *Applied Spatial Analysis and Policy, 3*(2–3), 107–138.

Norman, P., Rees, P., & Boyle, P. (2003). Achieving data compatibility over space and time: creating consistent geographical zones. *International Journal of Population Geography, 9,* 365–386.

Norman, P., & Riva, M. (2012). Population health across space and time: the geographical harmonisation of the Office for National Statistics Longitudinal Study for England and Wales. *Population Space and Place, 18,* 483–502.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and planning A, 16,* 17–31.

Pampalon, R., Hamel, D., Gamache, P., & Raymond, G. (2009). A deprivation index for health planning in Canada. *Chronic Diseases in Canada, 29*(4), 178–191.

Rodriguez, R., Hotchkiss, J. & O'Hare, A. 2013. Geographic information systems and chronic kidney disease: racial disparities, rural residence and forecasting. *Journal of Nephrology, 26*(1), 3–15.

Salmond, C., & Crampton, P. (2012). Development of New Zealand's Deprivation Index (NZDep) and its uptake as a national policy tool. *Canadian Journal of Public Health, 103*(2), S7–S11.

Seaman, R., Riffe, T., Leyland, A., Popham, F., & van Raalte, A. (2019). The increasing lifespan variation gradient by area-level deprivation, a decomposition analysis of Scotland 1981–2011. *Social Science & Medicine, 230,* 147–157. https://doi.org/10.1016/j.socscimed.2019.04.008

Shafi, Q., Qadir, J., & Jalal, F. (2018). Using GIS to develop health information system (HIS) for Srinagar city, Jammu and Kashmir. *International Journal of Advanced Remote Sensing and GIS, 7*(1), 2589–2602.

Shohaimi, S., Welch, A., Bingham, S., Luben, R., Day, N., Wareham, N., & Khaw, K. (2004). Area deprivation predicts lung function independently of education and social class. *European Respiratory Journal, 24,* 157–161.

Sifa, L. (2018). *GIS application for post tropical cyclone Gita*. Pacific Island GIS & RS User Conference, 26–30 November, Suva.

Simpson, L. (2002). Geography conversion tables: A framework for conversion of data between geographical units. *International Journal of Population Geography, 8,* 69–82. https://doi.org/10.1002/ijpg.235

Simpson, L., & Yu, A. (2003). Public access to conversion of data between geographies, with multiple look up tables derived from a postal directory. *Computers Environment and Urban systems, 27,* 283–307.

Somers, R. (1998). Developing GIS Management Strategies for an organization. *Journal of Housing Research, 9*(1), 157–178.

Stafford, M., & Marmot, M. (2003). Neighbourhood deprivation and health: does it affect us all equally? *International Journal of Epidemiology, 32,* 357–366.

Tonga Department of Statistics and Secretariat of the Pacific Community (SPC). (2011). *Tonga 2011 Census of Population and Housing, Volume 2: Analytical Report*. Secretariat of the Pacific Community.

Tonga Department of Statistics, Tonga Ministry of Health, SPC and UNFPA. (2013a). *Tonga Demographic and Health Survey, 2012*. Secretariat of the Pacific Community.

Tonga Department of Statistics, Tonga Ministry of Health, SPC and UNFPA. (2013b). *Tonga Demographic and Health Survey data, 2012*. Secretariat of the Pacific Community.

Tonga Department of Statistics. (2014). *Census of Population and Housing data*. Tonga Department of Statistics.

Tonga Health Promotion Foundation (Tonga Health). (2016). *Hala Fononga. Tonga National Strategy to prevent and control non communicable disease (2010–2015)*. Ministry of Health.

Tonga Ministry of Health (TMoH). (2014). *Kingdom of Tonga NCD Risk Factors STEPS report*. WHO.

Tonga Ministry of Health (TMoH) and Tonga Health. (2021). Tonga strategic roadmap for emergency and disaster risk management 2021-2023. Tonga: Tonga Ministry of Meterology, Energy, Information, Disaster Management, Environment, Climate Change and Communciations.

Townsend, P. (1987). Deprivation. *Journal of Social Policy, 16,* 25–46.

Wang, F. (2020). Why public health needs GIS: a methodological overview. *Annals of GIS, 26*(1), 1–12.

Wang, F. (2019). Why public health needs GIS: a methodological overview. *Annals of GIS, 26*(1), 1–12.

Wilson, T., & Rees, P. (1999). Linking 1991 population statistics to the 1998 local government geography of Great Britain. *Population Trends, 97,* 37–45.

World Bank (2014). *Pacific NCD Roadmap*. Available online: https://ncdalliance.org/resources/pacific-ncd-roadmap Accessed 6 Aug 2019.

World Health Organization (WHO). (2014). *Western Pacific Regional Action Plan for the prevention and control of Noncommunicable diseases (2014–2020)*. World Health Organization.

World Health Organization Regional Office for The Western Pacific (WHO). (2016). *Health information systems in the Pacific at a glance 2016*. World Health Organization.

World Health Organization (WHO). (2018). *Noncommunicable diseases country profiles 2018*. World Health Organization.