New Zealand Population Review, Vol 50, 42-70.

Accuracy Of Domicile Codes in New Zealand's Hospital Discharge Data and Implications for Urban-Rural Analyses

Te Tōtika o ngā Waehere Tauwāhi kei Ngā Raraunga Tuku i Ngā Hōhipera o Aotearoa me ngā Pānga ki Ngā Tātaritanga Tāone-Taiwhenua

GABRIELLE DAVIE,* JESSE WHITEHEAD,† RORY MILLER,‡ JUNE ATKINSON,§ SUE CRENGLE,** BRANDON DE GRAAF,†† ROSS LAWRENSON,‡‡ MICHELLE SMITH,§§ AND GARRY NIXON***

* Gabrielle Davie is a research associate professor in the Department of Preventive and Social Medicine at the University of Otago | Ōtākou Whakaihu Waka. Corresponding author: gabrielle.davie@otago.ac.nz

† Jesse Whitehead is a researcher at Te Ngira Institute for Population Research at the University of Waikato | Te Whare Wananga o Waikato.

‡ Rory Miller is a senior lecturer and the director of the Rural Postgraduate Programme of the Department of General Practice and Rural Health at the University of Otago | Ōtākou Whakaihu Waka.

§ June Atkinson is a senior analyst in the Department of Public Health at the University of Otago | Ōtākou Whakaihu Waka, Wellington.

** Sue Crengle is a professor of Māori health and a co-director of the Ngāi Tahu Māori Health Research Unit at the the University of Otago | Ōtākou Whakaihu Waka.

†† Brandon de Graaf is a research fellow in the Department of Preventive and Social Medicine at the University of Otago | Ōtākou Whakaihu Waka.

‡‡ Ross Lawrenson is a public health physician and professor in population health at the University of Waikato | Te Whare Wananga o Waikato.

§§ Michelle Smith is a research nurse in the Department of General Practice and Rural Health at the University of Otago | Ōtākou Whakaihu Waka.

*** Garry Nixon is a professor and associate dean in the Department of General Practice and Rural Health at the University of Otago | Ōtākou Whakaihu Waka.

Abstract

Inaccurate geospatial information can result in misleading conclusions. Manatū Hauora | Ministry of Health has recommended that, of their national collections, only the Mortality Collection (MORT) is suitable for rural-urban analysis of health outcomes. This paper analyses 48,644 deaths in hospital (2015–2018) and compares whether the domicile code recorded in the National Minimum Dataset (NMDS) of hospital discharges is consistent with that in MORT. While 16.5 per cent of rural residents had inconsistent domicile codes, this was higher for urban residents (21.6 per cent). Domicile inaccuracy resulted in incorrect rurality classification for 1.0 per cent and 13.6 per cent of the most urban and most rural residents, respectively.

Keywords: geospatial, accuracy, domicile, rural, hospitalisations

Whakarāpopotonga

Ka puta pea i ngā mōhiohio mokowā ā-nuku hē ngā whakataunga tuapeka. E ai ki te tūtohunga a te Manatū Hauora, kei te tika anake te Kohinga Matenga (Mortality Collection (MORT)) i waenga i ā rātou kohinga ā-motu, mō te tātaritanga taiwhenua-tāone o ngā putanga hauora. Ko tā tēnei pepa he tātari i ngā matenga 48,644 i ngā hōhipera (2015–2018) me te whakataurite mēnā he ōrite te waehere tauwāhi i tuhia i te Huinga Raraunga Mōkito ā-Motu (National Minimum Dataset (NMDS)) o ngā tukunga i ngā hōhipera ki tērā o MORT. Ahakoa kīhai i ōrite te ōrau 16.5 o ngā waehere tauwāhi o ngā kainoho tuawhenua, he nui ake kē tērā mō ngā kainoho tāone (21.6 ōrau). Nā te kore tōtika o te tauwāhi i puta he whakarōpūtanga taiwhenua hē mō te 1.0 ōrau, me te 13.6 ōrau o ngā kainoho tāone ake, mō ngā kainoho tuawhenua ake.

Ngā kupu matua: mokowā ā-nuku, tōtika, tauwāhi, taiwhenua, whakaurunga hōhipera

Geospatial analyses are an important tool for undertaking research and underpinning policy. The overwhelming increase in the availability of geographic data and evolving software and methodologies have put a spotlight on geospatial analyses (Edwards et al., 2014). Although there are considerable benefits gained from such analyses, the impact of spatial uncertainty and geocoding errors cannot be ignored. Poor quality collection of addresses and subsequent geocoding can result in misleading conclusions, particularly when there are differential inaccuracies in the collection and geocoding of addresses (Delmelle et al., 2022; Kinnee et al., 2020).

Administrative data provides a wealth of opportunities for research. An individual's encounters with a range of health providers and government agencies are increasingly being routinely collected and made available as data sets for research and policy purposes, widening the breadth and depth of questions that can be explored (International Journal of Population Data Science, 2023; Stats NZ, 2022). One limitation of administrative data is that address information is typically not geocoded at the time of the event, missing an opportunity for location clarification or correction. This, combined with the Modifiable Area Unit Problem (MAUP) – whereby aggregating individual level data to different geographic boundaries can produce vastly different results – is likely to result in geographic variation in the accuracy of address data and resulting geocoding quality due to differences in population density and address type by region (Openshaw, 1984).

Geocoded data are used for range of reasons including identifying a small geographic area that someone lives in or an area that an event occurred in (e.g., an injury), monitoring spatial variation in outcomes, and comparing outcomes between areas that are similar or different in some way. Thus, geocoded data can provide valuable information to inform both health service planning and delivery. For example, small geographic areas can be assigned measures relating to socio-economic disadvantage or assigned rurality using an urban-rural classification system. Small areas can also be aggregated geographically to compare administrative regions, such as the health regions of Te Whatu Ora | Health New Zealand or the catchments of health service providers. Spatial uncertainty and geocoding errors, including patterning of missingness, can thus result in exposure misclassification and/or inaccurate health effect estimates across a range of domains, resulting in misleading conclusions and poor public health outcomes (Jacquez, 2012; Kinnee et al., 2020).

One area of research where geospatial accuracy is particularly important is in rural-urban analyses of health outcomes. Until recently, most rural-urban analyses in New Zealand used generic rurality classifications developed by Stats NZ (Stats NZ, 2020). In 2022, a new rurality classification, the *Geographic Classification for Health* (GCH), was produced for New Zealand health research and policy purposes (Whitehead et al., 2022). The GCH is technically robust and aligns with a heuristic sense of what is rural in a health context. The 5-level GCH delineates three levels of rural (from least rural, R1, to most remote, R3) and two levels of urban (U1 representing major cities, and U2, provincial towns). The 2-level GCH combines R1–R3 and U1 and U2 into 'rural' and 'urban', respectively. Applying the GCH to large health data sets has identified rural-urban health outcome disparities that were previously masked by generic Stats NZ rurality classifications (Crengle et al., 2022; Nixon et al., 2023; Whitehead, Davie, de Graaf, Crengle, Lawrenson, et al., 2023). These recent findings align with international research from similar high-income countries with low population densities that found health outcomes in rural populations are generally poorer than those observed in urban populations (Australian Institute of Health and Welfare, 2023; Bremberg, 2020; Cross et al., 2021; Subedi et al., 2019). Since the development of the GCH, rural residents have been recognised as a priority population in Pae Ora legislation (Parliamentary Counsel Office, 2022) and a Rural Health Strategy has been developed by Manatū Hauora | Ministry of Health (MOH) (Ministry of Health, 2023) within which the GCH was used to define rural populations. As such, accurate geospatial information is required to accurately monitor the health outcomes of rural communities and design effective policy interventions.

New Zealand has a range of statistical geographies, from small to larger geographic units (Table 1, Figure 1). *Meshblocks* are Stats NZ's smallest geographic areas, each containing approximately 20–60 dwellings and 60–120 residents (Stats NZ, 2017). Prior to 2018, *census area units* (CAUs) were the next largest geographic unit. CAUs were phased out when Stats NZ's 2018 Statistical Standard for Geographic Areas (SSGA) introduced statistical areas 1 (SA1) and 2 (SA2). SA1s are aggregations of

Table 1: Small geographic units used in New Zealand of relevance to this study

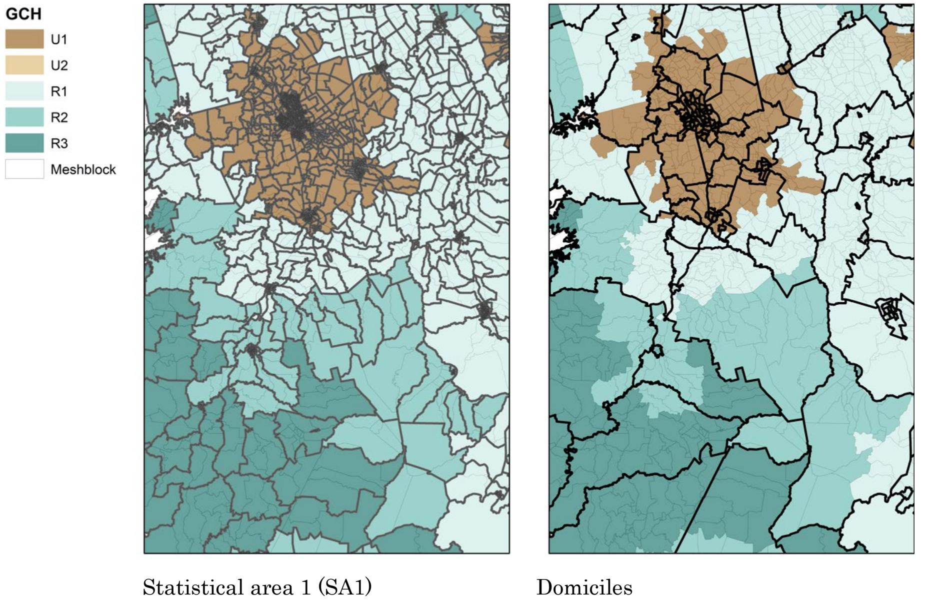
Geographic unit	Description	Number of units defined in 2018[1]	Approximate number of usual residents per unit[2]
Meshblock	Smallest geographic unit declared by Stats NZ. Vary in size from part of city block to large area of rural land.	53,589	60–120
Statistical area 1 (SA1)	New geography in SSGA2018. An SA1 contains at least one adjacent meshblock.	29,889	100–200
Statistical area 2 (SA2)	New geography in SSGA2018. An SA2 usually (but not always) contains aggregated adjacent meshblocks.	2,253	1,000–4,000
Domicile[3]	Used by Ministry of Health. 1:1 map to Stats NZ census area units (decommissioned in 2018). Built from meshblocks, not compatible with SA1s.	2,014	500–5,000

Notes: 1. Reference for meshblock, SA1 and SA2: *Statistical standard for geographic areas 2018* (Stats NZ, 2017).

2. Not all defined units contain residents; for example, 50,250 meshblocks contain > 0 residents.

3. Domicile codes based on 2013 census area units have been in use since July 2015 (Ministry of Health, 2021b).

Figure 1: Visualisation of the differences between SA1s and domiciles (both comprise multiple meshblocks) in urban and rural areas as defined by the Geographic Classification for Health (GCH)



meshblocks and are the smallest available geographic unit output. Use of SA1s rather than meshblocks minimises suppression of data when low frequencies are encountered. In 2018 there were 29,889 SA1s containing approximately 100–200 residents each, with a maximum population of 500 (Stats NZ, 2017).

Although defined at the SA1 level, like Stats NZ's Urban Accessibility Classification (Stats NZ, 2020), the GCH can be applied to meshblock-level data (with all meshblocks within an SA1 assigned the same value) and to SA2-level data (using population-weighted aggregation of SA1 values). While the GCH was developed at the SA1 level and analysts are encouraged to use this geographic unit wherever possible, it is recognised that this is not always possible and concordance files for the GCH at other geographic units are publicly available on https://blogs.otago.ac.nz/rural-urbannz/concordance-files/ and figshare (Whitehead, Davie, de Graaf, Crengle, Fearnley, et al., 2023).

The national collections, the responsibility of both the MOH and Te Whatu Ora | Health New Zealand, are the cornerstone for health research;

these administrative collections of health data span from registries such as the New Zealand Cancer Registry and the National Immunisation Register to the Pharmaceutical Collection (Te Whatu Ora | Health New Zealand, 2023b). Most collections contain a small area geography relating to an individual's residential address at the time of the event or activity. However, for some collections, like the Cancer Registry and the Immunisation Register, the geography 'at time of load' to the data warehouse is used. People in aged residential and long-term care, hospice and prison, for example, are exceptions for whom residential address is not used (Ministry of Health, 2021b).

Another source of geospatial inaccuracy relates to the quality of the geocoding of 'usually resident' address in the MOH's national collections (Technical Reference Group – Ministry of Health and Health NZ, 2022). In New Zealand, health care users are assigned a National Health Index (NHI) number, with person-level information such as ethnicity, address, date of birth and, if applicable, death, recorded in the NHI database (Ministry of Health, 2009). One issue is that keeping addresses up to date in the NHI database relies on individuals being continuously asked about their address or them informing providers of an address change. Those who rarely engage with health care providers are less likely to have correct address information recorded in the NHI, thus creating biases. An additional issue is that vague or unrecognised addresses impact the accuracy of geocoding. As stated in the NHI Data Dictionary (Ministry of Health, 2009), the automatic assigning of domicile code using the provided address can "result in rural addresses being assigned to an urban domicile code where there is insufficient data to generate the correct code" (p. 18). As outlined in "User Guide: Geospatial data in the National Collections", vague addresses include rural delivery addresses (e.g., RD3, Palmerston), provision of local road name only, and addresses that relate to delivery points (e.g., dairies) and areas 'eligible' for, or currently under development for, housing (Technical Reference Group – Ministry of Health and Health NZ, 2022). In addition, some addresses are not recognised within the geocoding systems used, meaning that the address is unvalidated and thus cannot be assigned a geographic coordinate (Technical Reference Group – Ministry of Health and Health NZ, 2022). Those with no fixed address (e.g., people who are unhoused) also have poorly captured geocoded data, while data entry errors can also affect the quality of geocoded data.

The process of obtaining a geocode involves two steps: validation of the address and then geocoding the address to a defined geographic area (e.g., meshblock or domicile) (Technical Reference Group – Ministry of Health and Health NZ, 2022). Address validation web services used include Health-e-Address and eSpatial Address Management (eSAM). Data from Stats NZ, the MOH, NZ Post and Land Information New Zealand is used by eSAM to validate and standardise address and geospatial data (Te Whatu Ora | Health New Zealand, 2023a). Irrespective of the match quality, domicile is always populated in all national collections; for those with vague addresses, the nearest post office location may be used, unhoused people are normally assigned a domicile code relating to a larger health area (e.g., their local District Health Board), and overseas residents are coded as '9999'. In contrast, meshblock is left blank when addresses are 'too difficult' to assign to a meshblock and for overseas residents. The exception to this is MORT where addresses that cannot be geocoded are manually investigated and geocoded where possible. Thus, MORT is the only data set within the national collections that, according to the MOH, contains geocoded data considered to be of sufficient quality to permit meaningful analyses of rurality (Technical Reference Group – Ministry of Health and Health NZ, 2022). It is also important to note that addresses in MORT are sourced from the Births, Deaths and Marriages (BDM) Notification of Death for Registration (Ministry of Health, 2021a). Therefore, geospatial information in MORT (meshblock and domicile derived from meshblock) is more likely to be both up to date at the time of death, and to be the result of more accurate geocoding compared with the information in other national collections.

Since individuals' data can be linked across health data sets using encrypted NHI numbers, the 'gold standard' meshblock data within MORT can be compared with that in other health data sets such as the National Minimum Dataset (NMDS) of hospital discharges. Where events in these two data sets occurred at the same time, such as a death that occurred during a hospitalisation, the individual should have the same geographic information recorded in both data sets. Given that MORT contains higher-quality geospatial data due to address information being verified as up to date and a more rigorous process to investigate vague or 'difficult to geocode' addresses, then the investigation of deaths in hospital provides an opportunity to examine the quality of geocoded data in the NMDS. Since the smallest geographic unit collected in MORT and the NMDS (meshblock and domicile, respectively) differs, it is also possible to explore the impact that using domicile codes has on the accuracy of rural-urban analyses.

The objective of this study was, therefore, to assess the accuracy of geospatial information in the NMDS, arguably the most commonly used health data set in New Zealand, and to examine the impact of any inaccuracies on urban-rural analyses. Exploring whether inaccuracies and impacts differed by age and ethnicity was a secondary objective. To our knowledge, this is the first study to use a large national data set (MORT) to examine the accuracy of geospatial information held in different health data sets (NMDS) corresponding to the same major event (fatality). Given that rural addresses are, in general, less specific than those in urban settings and thus potentially more likely to be missing or incorrectly geocoded, it was hypothesised that residents considered rural in MORT would have higher rates of incorrectly recorded domiciles in the NMDS than urban residents.

Methods

A deidentified population-level data set containing linked information from MORT and the NMDS was extracted by the MOH in September 2020. The data set provided geographic information based on the usually resident address of those who had died in hospital between 1 July 2015 and 31 December 2018, and included the NMDS domicile, MORT meshblock and MORT domicile (derived from the MORT meshblock). NMDS discharge records within this period that were coded with a discharge type of 'DD' (died) or 'ED' (died while still in Emergency Department acute facility) were linked to MORT using the patients' NHI. Records that could not be matched to a death in MORT or that matched to a blank meshblock were excluded. GCH was obtained for all three geographic variables: MORT meshblock, MORT domicile and NMDS domicile using the GCH concordance tables (Whitehead, Davie, de Graaf, Crengle, Fearnley, et al., 2023).

Due to population growth, some meshblocks used in the 2013 Census split into two or more meshblocks in the 2018 Census. Where all such split meshblocks belonged to the same 5-level GCH category, the death was assigned that category. Deaths that had 2013 meshblocks recorded in MORT that split into two or more 2018 meshblocks with different 5-level GCH values were excluded. Meshblocks that were invalid or that related to areas that could not be assigned rurality (i.e., estuaries, lakes, islands) were also excluded.

Age and ethnicity were obtained from the NMDS discharge record. Age was provided in 5-year age groups to 85 years and over and was categorised as 0–24, 25–64, 65–74 and 75+ years. Ethnicity was categorised according to MOH protocols: if Māori was recorded in any of the three ethnicity fields in the NMDS, the individual was included as Māori; if not, they were considered non-Māori (Health Information Standards Organisation (HISO), 2017).

To assess the accuracy of address/geocoded information in the NMDS, the number and percentage of deaths in the NMDS with incorrect domicile codes (i.e., not matching the gold standard MORT code) were calculated overall and by both the binary and 5-level GCH. To explore the impact of aggregation error when using domicile for rural-urban analyses, comparisons between the GCH obtained using MORT meshblock and MORT domicile were also undertaken. To determine whether these differences are likely to have differential impacts, the above estimates were obtained by age and ethnicity. All percentages are presented with binomial exact 95% confidence intervals (CIs). Stata/SE version 17.0 was used for the analysis (StataCorp, 2021). Figure 1 was created in ArcMap10.8 (Esri, 2020), using the GCH and geospatial data on meshblock, SA1 and domicile (area unit) boundaries from the Stats NZ Geographic Data Service (Stats NZ, 2023).

Results

There were 51,273 deaths in a New Zealand hospital over the period July 2015 to December 2018. Of these, no records were available in MORT for 1905 (3.7 per cent) of these in-hospital deaths. An additional 179 in-hospital deaths linked to MORT had no meshblock and/or domicile information. Of the remaining 49,189 in-hospital deaths, an additional 436 (0.9 per cent) were excluded as their 2013 MORT meshblock of usual residence split into more than one GCH category. In addition, 13 had meshblocks recorded that were not valid 2013 codes and 96 were not classifiable by rurality as the meshblocks represented ‘inland water’ or ‘island’. The analysis data set thus contained 48,644 in-hospital deaths (94.9 per cent of the original sample). The majority of those who died in hospital (30,995; 63.7 per cent) were aged 75 years or more; 18.8 per cent (8757) were aged 65–74 years, 16.1 per cent

(7841) aged 25–64 years, and 2.2 per cent (1051) were 0–24 years old. Of all the deaths in hospital, 10.5 per cent were Māori patients.

Overall, 20.5 per cent (95% CI, 20.1–20.9 per cent) of the 48,644 in-hospital deaths examined had an incorrect domicile code recorded in the NMDS (Table 2). A higher rate of inaccuracy in NMDS domiciles was observed for urban than for rural residents (21.6 per cent and 16.5 per cent, respectively). In the 5-level GCH classification, R2 residents had the lowest rate of NMDS domicile inaccuracy (12.6 per cent) while residents of the most urban (U1) areas had the highest rate of domicile inaccuracy (22.4 per cent).

Of the 48,644 in-hospital deaths, 2 had invalid domiciles recorded in the NMDS, 90 had a domicile indicating they were an overseas resident ('9999') and 44 had a DHB recorded as the domicile. Thus, 48,508 (99.7 per cent) remained.

Table 2: Domicile inaccuracy in the NMDS for those who died in hospital; overall, and by 2-level GCH and 5-level GCH

GCH from MORT meshblock	Deaths in hospital		Domicile same in NMDS and MORT?		% incorrect[1]	
	N	Col %	Yes	No	Est.	95% CI
Overall						
2-level GCH						
Urban	38,360	78.9	30,078	8,282	21.6	(21.2, 22.0)
Rural	10,284	21.1	8,582	1,702	16.5	(15.8, 17.3)
5-level GCH						
U1	27,030	55.6	20,986	6,044	22.4	(21.9, 22.9)
U2	11,330	23.3	9,092	2,238	19.8	(19.0, 20.5)
R1	7,513	15.4	6,177	1,336	17.8	(16.9, 18.7)
R2	2,342	4.8	2,047	295	12.6	(11.3, 14.0)
R3	429	0.9	358	71	16.6	(13.2, 20.4)

Note: 1. Assumes that the gold standard MORT is correct.

Using MORT domicile, 0.8 per cent (95% CI, 0.8–0.9 per cent) of urban residents were incorrectly classified as rural using the GCH from

NMDS domicile compared with the 5.0 per cent (95% CI, 4.6–5.5 per cent) of rural residents who were incorrectly classified as urban (Table 3). Using 2-level GCH, a similar number were classified as rural by MORT domicile (10,459) as by NMDS domicile (10,255), although this comparison hides the 848 deaths (322 urban and 526 rural) that were incorrectly classified when using GCH from NMDS domicile.

Comparison of the 2-level GCH obtained from MORT meshblock with that obtained from MORT domicile indicates that aggregation error affected rural residents more than urban residents (1.4 per cent incorrect compared with 0.9 per cent) (Table 3). Comparison of the 2-level GCH obtained from MORT meshblock with that obtained from NMDS domicile combines both errors (address and aggregation). The analysis indicates 6.0 per cent of rural residents were incorrectly classified as urban compared with 1.6 per cent of urban residents incorrectly classified as rural. Overall, 1229 of the 48,508 in-hospital deaths (2.5 per cent; 95% CI, 2.4–2.7 per cent) were coded to the wrong GCH category when the NMDS domicile code was used.

Considerable variation in NMDS address error was apparent by 5-level GCH strata; only 1 per cent of those classified as being U1 residents by their MORT domicile were not classified into the same GCH category using their NMDS domicile compared with 13.6 per cent of those classified as R3 residents by their MORT domicile (Table 4).

Since GCH is defined at the SA1 level and given domicile boundaries are incompatible with SA1 boundaries, obtaining GCH for domicile codes involves approximation, as well as aggregation. Thus, using the meshblock recorded in MORT to obtain a 5-level GCH is preferable and more likely to accurately represent the urban/rural status of individuals. Although the distribution by GCH MORT meshblock and GCH MORT domicile was almost identical (U1 to R3: MORT meshblock – 56 per cent, 24 per cent, 16 per cent, 5 per cent, 1 per cent – compared with MORT domicile – 56 per cent, 23 per cent, 15 per cent, 6 per cent, 1 per cent), cross-tabulations indicated the presence and extent of misclassification (Table 4). Minimal misclassification was observed between the GCH obtained from MORT domicile compared with MORT meshblock for U1 residents (0.1 per cent) whereas 12 per cent of R1 residents were misclassified (the vast majority of

Table 3: Inaccuracies in 2-level GCH obtained from domicile for those that died in hospital

2-level GCH	*N*	Col %	Urban	Rural	% Incorrect	
			Row %		Est.	95% CI
NMDS address error:[1] GCH from MORT domicile compared with NMDS domicile						
Urban	38,049	78.4	99.2	0.8	0.8	(0.8, 0.9)
Rural	1,459	21.6	5.0	95.0	5.0	(4.6, 5.5)
Aggregation error: GCH from MORT meshblock compared with MORT domicile						
Urban	38,262	78.9	99.1	0.9	0.9	(0.8, 1.0)
Rural	10,246	21.1	1.4	98.6	1.4	(1.2, 1.7)
Combined error: GCH from MORT meshblock compared with NMDS domicile						
Urban	38,262	78.9	98.4	1.6	1.6	(1.5, 1.7)
Rural	10,246	21.1	6.0	94.0	6.0	(5.5, 6.4)

Note: 1. Reasons include imprecise or out-dated addresses and/or issues with geocoding addresses.

these as R2). One in five R3 residents (20.7 per cent) had an incorrect GCH assigned when using MORT domicile with the majority of these misclassified as R2. Combining R2 and R3 reduces the overall level of inaccuracy slightly, from 3.1 per cent to 2.9 per cent, but more substantially reduces the inaccuracy in R2 and R3 (7.5 per cent and 20.5 per cent, respectively) to 4.5 per cent (R2/R3).

Comparison of 5-level GCH obtained from MORT meshblock with that obtained from NMDS domicile indicates the combined impact of both NMDS address error and aggregation error (Table 4). Only 1.1 per cent of U1 residents were incorrectly classified by GCH using NMDS domicile compared with 29.8 per cent of R3 residents. Although only 1 per cent of those living in U1 had the incorrect GCH according to their NMDS domicile, this represents almost 300 deaths, a larger number than the 127 deaths in R3 residents that were incorrectly coded to other GCH categories. Sixteen

Table 4: Inaccuracies in 5-level GCH obtained from domicile for those that died in hospital

GCH from MORT domicile	*N*	Col %	GCH from NMDS domicile U1 row %	U2	R1	R2	R3	% Incorrect Est.	95% CI
NMDS address error:[1] GCH from MORT domicile compared with NMDS domicile									
U1	27,008	55.7	99.0	0.3	0.5	0.1	0.0	1.0	(0.9, 1.1)
U2	11,041	22.8	0.9	97.8	0.9	0.3	0.0	2.2	(1.9, 2.5)
R1	7,071	14.6	2.7	2.7	93.9	0.7	0.1	6.1	(5.6, 6.7)
R2	2,962	6.1	2.5	1.4	2.2	93.5	0.4	6.5	(5.6, 7.4)
R3	426	0.9	4.0	2.8	1.2	5.6	86.4	13.6	(10.5, 17.2)
Aggregation error: GCH from MORT meshblock compared with MORT domicile									
U1	26,967	55.6	99.9	0.0	0.1	0.0	0.0	0.1	(0.1, 0.2)
U2	11,295	23.3	0.0	97.1	2.9	0.0	0.0	2.9	(2.6, 3.3)
R1	7,490	15.4	0.9	0.9	88.0	10.1	0.1	12.0	(11.2, 12.7)
R2	2,330	4.8	0.0	0.2	3.7	93.0	3.5	7.5	(6.4, 8.6)
R3	426	0.9	0.0	0.2	7.5	12.9	79.3	20.7	(16.9, 24.8)

(table continued on the next page….)

GCH from MORT domicile	N	Col %	GCH from NMDS domicile U1 row %	U2	R1	R2	R3	% Incorrect Est.	95% CI
Combined error: GCH from MORT meshblock compared with NMDS domicile									
U1	26,967	125.2	98.9	0.3	0.6	0.1	0.0	1.1	(0.9, 1.2)
U2	11,295	52.4	1.0	95.4	3.3	0.3	0.0	4.6	(4.2, 5.0)
R1	7,490	34.8	3.6	2.9	83.6	9.8	0.1	16.4	(15.6, 17.3)
R2	2,330	10.8	2.6	1.5	5.1	87.6	3.3	12.4	(11.1, 13.9)
R3	426	2.0	4.2	2.6	8.0	15.0	70.2	29.8	(25.5, 34.4)

Note: 1. Reasons include imprecise or out-dated addresses and/or issues with geocoding addresses.

per cent (n = 1228) of R1 residents were incorrectly classified; of these, close to half (487; 39.7 per cent) were classified as urban (U1 or U2) with almost all the remainder (733, 59.7 per cent) classified as R2. Overall, 5.1 per cent (95% CI, 4.9–5.3 per cent) had the incorrect GCH category from the NMDS domicile when compared with the GCH obtained from the MORT meshblock.

For those under 75 years of age at the time of death, NMDS address error was less apparent between urban and rural residents (16.4 per cent and 14.2 per cent of domiciles incorrect, respectively) compared with those over 75 years where 24.5 per cent of urban residents had an incorrect domicile compared with 18.0 per cent of rural residents (Table 5). For Māori, the inaccuracy was 17.1 per cent for urban residents and 12.8 per cent for rural residents.

Table 5: Extent of domicile inaccuracy from NMDS address error for in-hospital deaths, by 2-level GCH, age and ethnicity

GCH from MORT meshblock	Domicile same in NMDS and MORT?		% Incorrect	
	Yes	No	Est.	95% CI
Age				
> 75 years				
Urban	11,611	2,279	16.4	(15.8, 17.0)
Rural	3,226	533	14.2	(13.1, 15.3)
75+ years				
Urban	18,467	6,003	24.5	(24.0, 25.1)
Rural	5,356	1,172	18.0	(17.0, 18.9)
Ethnicity				
Māori				
Urban	3,006	620	17.1	(15.9, 18.4)
Rural	1,290	189	12.8	(11.1, 14.6)
non-Māori				
Urban	27,072	7,662	22.1	(21.6, 22.5)
Rural	7,292	1,513	17.2	(16.4, 18.0)

Note: 1. Assumes that the gold standard MORT is correct.

Also, for those under 75 years of age, 1.0 per cent of urban residents were incorrectly classified as rural and 5.3 per cent of rural residents were incorrectly classified as urban (Table 6). The percentages for those over 75 years were slightly lower (0.8 per cent and 4.9 per cent, respectively). For Māori, 1.4 per cent of urban residents were incorrectly classified as rural and 4.8 per cent of rural residents were considered urban compared with 0.8 per cent and 5.1 per cent of non-Māori, respectively.

Table 6: Impact of NMDS address error on 2-level GCH obtained from domicile for in-hospital deaths, by age and ethnicity

GCH from MORT meshblock	*N*	Col %	Domicile same in NMDS and MORT? Urban	Rural	% Incorrect[1] Est.	95% CI
			Row %			
Age						
< 75 years	17,593		13,838	3,755	1.9	(1.7, 2.2)
Urban	13,775	78.3	99.0	1.0	1.0	(0.8, 1.2)
Rural	3,818	21.7	5.3	94.7	5.3	(4.6, 6.0)
75+ years	30,915		24,415	6,500	1.6	(1.5, 1.8)
Urban	24,274	78.5	99.2	0.8	0.8	(0.6, 0.9)
Rural	6,641	21.5	4.9	95.1	4.9	(4.4 5.4)
Ethnicity						
Māori	5,093		3,614		2.4	(2.0, 2.9)
Urban	3,592	70.5	98.6	620	1.4	(1.0, 1.8)
Rural	1,501	29.5	4.8	189	4.8	(3.8, 6.0)
non-Māori	43,415		34,639		1.7	(1.6, 1.8)
Urban	34,457	79.4	99.2	7,662	0.8	(0.7, 0.9)
Rural	8,958	20.6	5.1	1,513	5.1	(4.6, 5.5)

Note: 1. Assumes that the GCH from the MORT domicile is correct.

Using the 5-level GCH, NMDS address error resulted in notably higher domicile inaccuracy for those over 75 years of age compared with those younger across all levels of the GCH (Table S1).[1] For those in the most urban (U1) areas, 25.7 per cent of those over 75 years of age had incorrect NMDS domiciles compared with 16.5 per cent of those under 75 years of age. For Māori, those living in R1 and R2 areas had the lowest domicile inaccuracy (12.2 per cent and 11.7 per cent, respectively). For non-Māori, the comparable estimates were 18.5 per cent and 12.8 per cent, respectively.

Similarities were observed for those under and over 75 years of age in terms of patterns of inaccuracy between the 5-level GCH obtained from MORT domicile compared with NMDS domicile (Table S2).[2] Of those who died in hospital, a different distribution across the GCH was apparent by ethnicity with a lower percentage of Māori from U1 areas (43 per cent compared with 57 per cent of non-Māori) and a higher percentage from R2 areas (11 per cent compared with 6 per cent of non-Māori). For Māori, 3.5 per cent (95% CI, 3.0–4.1 per cent) of all deaths had an incorrect 5-level GCH assigned from the NMDS domicile; this was higher than that obtained for non-Māori (2.3 per cent; 95% CI, 2.2–2.5 per cent).

Discussion

This study has identified two important sources of inaccuracies to be aware of when using geospatial information within the NMDS: *address error,* where the domicile in the NMDS does not match the domicile recorded in MORT, and *aggregation error,* whereby the GCH category obtained from MORT domicile is different to the GCH category obtained from the gold standard MORT meshblock. When using GCH obtained from domicile in the NMDS, analyses will include the combined effect of both inaccuracies. Table 7 outlines the two sources of inaccuracies examined, their possible causes, impact on rural-urban analyses, and potential solutions.

Table 7: Summary of the issues, causes and potential solutions of address and geocoding errors in the NMDS and implications for rural-urban analyses

Issue	Description/Example	Possible cause(s)	What this study shows: Impact	Solution
NMDS address error	Occurs when an individual's domicile of residence in the NMDS is incorrect. Any health events/outcomes associated with this individual are thus incorrectly assigned to a domicile within which they did not reside.	Recorded address is not up to date (e.g., due to residential mobility), is imprecise (e.g., those who are unhoused), or is unable to be accurately geocoded (e.g., RD1). Data entry errors may also contribute.	Of those who died in hospital, one in five domiciles were incorrect in the NMDS. Although domiciles were more likely to be inaccurate for urban (22 per cent) than for rural (17 per cent) residents, less than 1 per cent of urban residents were misclassified as rural while 5 per cent of rural residents were misclassified as urban. Using the 5-level GCH, a relatively high proportion (14 per cent) of R3 residents were incorrectly misclassified as R2. Although domicile inaccuracy is substantive in the NMDS, the impact on rural-urban classification is less.	Recording accurate address information should be prioritised. This could include ensuring that address information is verified and updated at point-of-health-service utilisation and investing in accurate geocoding of rural addresses. Those concerned about rural/urban misclassification error from NMDS domicile should consider using a binary GCH or consider a 4-level GCH that combines the R2 and R3 categories, especially when numbers are small.

(table continued on the next page….)

Issue	Description/Example	Possible cause(s)	What this study shows: Impact	Solution
Aggregation error	A less accurate representation of an individual's actual context (such as rurality or area-level socioeconomic deprivation) is obtained when larger geographic units, rather than smaller geographic units, are used.	Geospatial data on usual residence within most health collections is only provided at the domicile level. Census area units that had a 1-1 concordance with domiciles were replaced in SSGA2018 with statistical areas. Both GCH and NZDep are derived from the census using SA1s, which are smaller than domiciles and do not align with the boundaries of domiciles (see Figure 1).	Overall, aggregation had a small impact on the assignment of GCH to deaths in hospital. Defining GCH using domicile rather than meshblock resulted in an incorrect 2-level GCH category for 1.4 per cent of rural residents and 0.9 per cent of urban residents. The disproportionate impact was more apparent using the 5-level GCH, with residents of R1, R2 and R3 having error rates of 12 per cent, 7 per cent and 21 per cent, respectively, compared with 0.1 per cent for U1 and 2.9 per cent for U2 residents.	Wherever possible, addresses (i.e., usually resident, location of incidence) should be geocoded to (x, y) coordinates and made available to researchers and analysts at the smallest available geographic unit. Geospatial data within health collections should be aligned to the SSGA and retired geographic units such as domicile should no longer be used.

Interestingly, our findings did not support the hypothesis that domicile inaccuracy was more common for rural addresses (Table 2). Instead, for people who died in hospital, this study found the domicile recorded in the NMDS was more likely to be incorrect for urban residents than for rural residents. The NHI Data Dictionary states that the automatic assigning of domicile code using the provided address can "result in rural addresses being assigned to an urban Domicile code where there is insufficient data to generate the correct code" (Ministry of Health, 2009). This may dissuade researchers and health policymakers from undertaking rural-urban analyses. However, our findings suggest that while there are issues with the accuracy of the geospatial information in the NMDS, this issue does not disproportionately affect rural addresses. That said, an additional important finding is that NMDS address error appears to have a differing impact on the rurality category assigned to urban and rural residents (Tables 3 and 4). Using the GCH category derived from MORT domicile as the gold standard, the GCH obtained from NMDS domicile was more likely to be accurate for residents of urban areas. This appears to echo evidence that geocoding rates (percentage of records geocoded) in the USA are lower in rural areas (Edwards et al., 2014).

Domicile inaccuracy is greatest in those aged over 75 years at the time of death in hospital and appears to occur to a lesser extent for Māori than for non-Māori, both in urban and rural areas (Table 5). The GCH obtained from NMDS domicile was incorrect for 2.4 per cent of Māori compared with 1.7 per cent of non-Māori (Table 6), which will be due to GCH being more likely to be incorrect for rural residents and a higher proportion of Māori living in rural areas.

Aggregation error (Tables 3 and 4) provides an example of the impact of averaging across geographic boundaries. As can be seen in Figure 1, a given domicile, by nature of its size and boundaries, can contain residents from more than one GCH category (defined at the SA1 level). This results in situations whereby the GCH category from the MORT domicile is different to that obtained from the gold standard MORT meshblock; this also disproportionately affects rural residents.

There are several possible reasons for the inaccuracies identified. NMDS address error is likely to result in addresses being inaccurately geocoded. While it has previously been assumed that this was more likely to be an issue in rural areas with RD addresses being more difficult to geocode, our findings point to greater inaccuracy for urban addresses (22 per cent compared with 17 per cent for rural addresses). The observed higher inaccuracy in urban areas could be a result of geographic factors. Urban domiciles are generally smaller than rural domiciles, meaning that in urban areas, geocoding inaccuracies are more likely to result in an individual being assigned to a different domicile. However, since an urban resident's domicile is more likely to be surrounded by domiciles that have the same GCH category, domicile inaccuracy has a greater impact on the GCH category for rural than for urban residents.

Aggregation error, whereby the GCH category of the MORT domicile differs from the GCH category of the MORT meshblock, is also the result of geographic factors, namely the Modifiable Area Unit Problem (Openshaw, 1984). Because the GCH was developed at the SA1 level, there are many instances where domiciles include SA1s that have different GCH categories (Figure 1). Therefore, an individual's GCH category derived from domicile in MORT may differ from that obtained from the GCH of the MORT meshblock. This is more likely to be an issue in more rural areas, where larger domiciles often cross GCH boundaries.

Urban-rural analyses are essential to provide evidence that will support the objectives of the Pae Ora legislation and the Rural Health Strategy for healthier rural communities. The finding that the impact of inaccurate NMDS domiciles on assigned rurality is greater for rural populations is concerning as this means that unknown inaccuracies are being disproportionately introduced. While the scale of these inaccuracies is relatively small (between 4 per cent and 7 per cent of rural residents misclassified as urban in this study), they need to be addressed. The findings of this study also have important implications beyond our direct focus on rural health research and policy. While incorrect NMDS domicile can still produce a GCH classification that is largely consistent with that derived from MORT meshblock, domicile inaccuracy is also likely to have an unknown impact on the accuracy of other area-based information such as the New Zealand Index of Socioeconomic Deprivation (NZDep) or other geographically based exposures/risk factors (Salmond & Crampton, 2012). For NZDep at least, due to high levels of spatial heterogeneity (Salmond & Crampton, 2002), it can reasonably be assumed that the impact will be substantial. This could have a wide-reaching impact as area-level socio-economic deprivation is a commonly used variable within health research.

There are also wider implications for the accuracy of information relating to usual residential address held within Stats NZ's Integrated Data Infrastructure (IDI). The IDI links administrative data on individuals from a range of different sources across education, income, benefits, migration, justice and health, and includes an 'address notification' table in which addresses from high-quality sources, including the NHI database, are prioritised (Stats NZ, 2022). Interestingly, a comparison of IDI meshblocks relating to usual residence with gold standard census meshblocks, rated both the NHI and PHO data sets as having the highest levels of agreement (over 70 per cent) compared with data from other sources including the Ministry of Education and Ministry of Social Development (Gibb & Das, 2015). Given the substantive lack of agreement, even for the health data sets, inaccuracies in collecting and coding addresses are likely to be having a much broader impact upon research and policymaking across a wide range of areas. One identified strategy to improve the quality of location information in the IDI was to improve the quality of address information collected by source agencies (Gibb & Das, 2015).

The two types of error discussed above have different solutions. Address error (domicile inaccuracy) is likely to be partly caused by incorrect addresses that have not been updated. Ensuring that addresses are regularly updated in national health collections should be a priority. In addition, if it is assumed that the underlying residential addresses in NMDS and MORT were the same in our data set, then there is a clear discrepancy in the effectiveness of address geocoding between these two national collections. One solution is to improve the collection of NMDS addresses, ideally by recording geographical latitude and longitude coordinates, which have been previously described as the "superior basic code" (Delmelle et al., 2022; Rushton et al., 2006), to accurately reflect where residents actually live. Better awareness is needed of the statement in the NHI Data Dictionary that: "Care should be taken to record accurate and useful residential addresses, since Domicile codes will be automatically assigned using this information" (Ministry of Health, 2009, p. 18).

Aggregation error requires a different solution. Our findings highlight how the spatial units that are used to aggregate individuals' address information can introduce inaccuracies. When New Zealand's SSGA was adopted in 2018 (Stats NZ, 2017), census area units were replaced by statistical areas and thus the 1-to-1 match between CAUs and domiciles was broken. Typically census data is now available at the SA2 level, with Stats NZ no longer making population data freely accessible at the domicile/CAU level. We strongly recommend that the health sector move to coding using the SSGA. This will ensure that the spatial units for numerators and denominators align, allowing researchers and policymakers to more effectively and efficiently integrate health, population and contextual information into their analyses. Until such time, we suggest that researchers and policymakers using the NMDS consider using a combined GCH R2/R3 category in their analyses to reduce the impact of misclassification error.

To our knowledge this study is the first of its kind to compare geospatial data for the same major health event, fatality, across two national health data sets that are commonly used in health research and to inform policy. Since the latest MORT data available at the time of this study was deaths registered in 2018, generalisation of the patterns observed to current practices should be made with some caution. That said, little has changed regarding collection of addresses and geocoding in the last few years. It should be noted that comparisons have been made at the NMDS domicile level, meaning that any within NMDS domicile inaccuracies will have been masked. This means that the level of discordance identified is likely an underestimate – as is our estimate of the impact of address error on GCH classification. Based on MOH recommendations, this analysis has assumed that the MORT meshblock/domicile is more accurate than the NMDS domicile due to the manual processing of missing and vague addresses. However, this assumption needs validating as the addresses in MORT may not have been correctly recorded on the Notification of Death for Registration.

Future research should expand on the work presented here to consider the impact of changes in address, meshblock, SA1 and domicile over time, and by extending the scope of analysis to include comparison between more than two national collections. A similar analysis comparing PHO enrolment meshblock at date of death with MORT meshblock would be informative. In addition, exploring the impact of address error on other area-based contextual information – such as NZDep – will be important for understanding the implications of geospatial inaccuracies on both health research and health resource distribution. Further research should also seek to understand reasons for the finding that people aged 75 years or older who died in hospital were substantially more likely to have a domicile incorrectly recorded in the NMDS than those who were younger.

Conclusion

This study provides valuable insight into the accuracy of NMDS domiciles and the impact of this on rurality classification and subsequent analyses. Considerable improvement is clearly required to improve the collection of patient addresses as the current level of domicile inaccuracy within New Zealand's hospital discharge data will be having an impact on research, planning and policy. Prioritisation needs to be given to investment in a geospatial strategy that supports the collection of address using geographical coordinates and ensures that government departments use meaningful spatial units when aggregating health data.

Notes

1 'Table S1: Comparison between domicile recorded in NMDS and MORT for those who died in hospital by 5-level GCH, age and ethnicity' is in the supplementary notes, which are available from the corresponding author on request.

2 'Table S2: Comparison between 5-level GCH obtained from domiciles in NMDS and MORT for those that died in hospital by age and ethnicity' is in the supplementary notes, which are available from the corresponding author on request.

Acknowledgements

We wish to thank Chris Lewis, Senior Information Analyst, Health New Zealand for the data extraction and his invaluable advice. This study was funded by a research grant from the Health Research Council of New Zealand (HRC 19/488).

References

Australian Institute of Health and Welfare. (2023). *Rural and remote health*. https://www.aihw.gov.au/reports/rural-remote-australians/rural-and-remote-health#Heath%20status

Bremberg, S. (2020). Rural-urban mortality inequalities in four Nordic welfare states. *Scandavian Journal of Public Health*, *48*(8), 791–793. https://doi.org/10.1177/1403494820921684

Crengle, S., Davie, G., Whitehead, J., de Graaf, B., Lawrenson, R., & Nixon, G. (2022). Mortality outcomes and inequities experienced by rural Māori in Aotearoa New Zealand. *The Lancet Regional Health – Western Pacific*, *28*, Article 100570. https://doi.org/10.1016/j.lanwpc.2022.100570

Cross, S. H., Califf, R. M., & Warraich, H. J. (2021). Rural-urban disparity in mortality in the US From 1999 to 2019. *JAMA*, *325*(22), 2312–2314. https://doi.org/10.1001/jama.2021.5334

Delmelle, E. M., Desjardins, M. R., Jung, P., Owusu, C., Lan, Y., Hohl, A., & Dony, C. (2022). Uncertainty in geospatial health: Challenges and opportunities ahead. *Annals of Epidemiology*, *65*, 15–30. https://doi.org/https://doi.org/10.1016/j.annepidem.2021.10.002

Edwards, S. E., Strauss, B., & Miranda, M. L. (2014). Geocoding large population-level administrative datasets at highly resolved spatial scales. *Trans GIS*, *18*(4), 586–603. https://doi.org/10.1111/tgis.12052

Esri. (2020). *ArcMap 10.8*. https://www.esri.com/en-us/arcgis/products/arcgis-desktop/resources

Gibb, S. J., & Das, S. (2015). *Quality of geographic information in the Integrated Data Infrastructure*. https://www.stats.govt.nz/research/quality-of-geographic-information-in-the-integrated-data-infrastructure/

Health Information Standards Organisation (HISO). (2017). *HISO 10001:2017 Ethnicity Data Protocols*. Ministry of Health. https://www.health.govt.nz/publication/hiso-100012017-ethnicity-data-protocols

International Journal of Population Data Science. (2023). *Bringing together and analysing data from multiple sources*. https://ijpds.org/catalog/category/bringing-together-analysing-data

Jacquez, G. M. (2012). A research agenda: Does geocoding positional error matter in health GIS studies? *Spatial and Spatio-temporal Epidemiology*, *3*(1), 7–16. https://doi.org/https://doi.org/10.1016/j.sste.2012.02.002

Kinnee, E. J., Tripathy, S., Schinasi, L., Shmool, J. L. C., Sheffield, P. E., Holguin, F., & Clougherty, J. E. (2020). Geocoding error, spatial uncertainty, and implications for exposure assessment and environmental epidemiology. *International Journal of Environmental Research and Public Health*, *17*(16), Article 5845. https://doi.org/10.3390/ijerph17165845

Ministry of Health. (2009). *National Health Index Data Dictionary* (version 5.3.) https://www.health.govt.nz/publication/national-health-index-data-dictionary

———. (2021a). *Mortality Collection Data Dictionary* (v1.80). https://www.health.govt.nz/publication/mortality-collection-data-dictionary

———. (2021b). *National Minimum Dataset (Hospital Events) Data Dictionary* (v7.9.3). https://www.health.govt.nz/publication/national-minimum-dataset-hospital-events-data-dictionary

———. (2023). *Rural Health Strategy*. https://www.health.govt.nz/publication/rural-health-strategy

Nixon, G., Davie, G., Whitehead, J., Miller, R., de Graaf, B., Lawrenson, R., Smith, M., Wakerman, J., Humphreys, J., & Crengle, S. (2023). Comparison of urban and rural mortality rates across the lifespan in Aotearoa/New Zealand: A population-level study. *Journal of Epidemiology and Community Health*, *77*(9), 571–577. https://doi.org/10.1136/jech-2023-220337

Openshaw, S. (1984). The modifiable areal unit problem. In *Concepts and Techniques in Modern Geography, Issue 38.* Geobooks.

Parliamentary Counsel Office. (2022). Pae Ora (Healthy Futures) Bill. Section 40B Rural Health Strategy. https://legislation.govt.nz/bill/government/2021/0085/latest/whole.html

Rushton, G., Armstrong, M. P., Gittler, J., Greene, B. R., Pavlik, C. E., West, M. M., & Zimmerman, D. L. (2006). Geocoding in cancer research: A review. *American Journal of Preventive Medicine*, *30*(2, Supplement), S16–S24. https://doi.org/https://doi.org/10.1016/j.amepre.2005.09.011

Salmond, C., & Crampton, P. (2002). Heterogeneity of deprivation within very small areas. *Journal of Epidemiology and Community Health*, *56*(9), 669. https://doi.org/10.1136/jech.56.9.669

———. (2012). Development of New Zealand's deprivation index (NZDep) and its uptake as a national policy tool. *Canadian Journal of Public Health*, *103*(8 Suppl 2), S7–S11. https://pubmed.ncbi.nlm.nih.gov/23618071/

StataCorp. (2021). *Stata Statistical Software: Release 17.*

Stats NZ. (2017). *Statistical standard for geographic areas 2018*. https://www.stats.govt.nz/methods/statistical-standard-for-geographic-areas-2018

———. (2020). *Urban accessibility – methodology and classification.* https://www.stats.govt.nz/methods/urban-accessibility-methodology-and-classification

———. (2022). *Integrated Data Infrastructure.* https://www.stats.govt.nz/integrated-data/

———. (2023). *Geographic data service*. https://datafinder.stats.govt.nz/

Subedi, R., Greenberg, T. L., & Roshanafshar, S. (2019). Does geography matter in mortality? An analysis of potentially avoidable mortality by remoteness index in Canada. *Health Reports*, *30*(5), 3–15. https://doi.org/10.25318/82-003-x201900500001-eng

Te Whatu Ora | Health New Zealand. (2023a). *Addressing and geocoding*. https://www.tewhatuora.govt.nz/our-health-system/digital-health/health-identity/addressing-and-geocoding-overview/

———. (2023b, 12 July). *Collections*. https://www.tewhatuora.govt.nz/our-health-system/data-and-statistics/nz-health-statistics/national-collections-and-surveys/collections/

Technical Reference Group – Ministry of Health and Health NZ. (2022). *User guide: Geospatial data in the National Collections* (Version 1.0).

Whitehead, J., Davie, G., de Graaf, B., Crengle, S., Fearnley, D., Smith, M., Lawrenson, R., & Nixon, G. (2022). Defining rural in Aotearoa New Zealand: A novel geographic classification for health purposes. *New Zealand Medical Journal*, *135*(1559), 24–40. https://nzmj.org.nz/journal/vol-135-no-1559/defining-rural-in-aotearoa-new-zealand-a-novel-geographic-classification-for-health-purposes

———. (2023). *Geographic Classification for Health – Concordance Files* [Dataset]. figshare. https://doi.org/10.6084/m9.figshare.22728851.v1

Whitehead, J., Davie, G., de Graaf, B., Crengle, S., Lawrenson, R., Miller, R., & Nixon, G. (2023). Unmasking hidden disparities: A comparative observational study examining the impact of different rurality classifications for health research in Aotearoa New Zealand. *BMJ Open*, *13*(4), Article e067927. https://doi.org/10.1136/bmjopen-2022-067927