UNIVERSITY OF CANBERRA

DISTINCTIVE BY DESIGN

MAPPING MENTAL HEALTH CARE

RESEARCH METHODOLOGY

NATSEM at the University of Canberra

FOR MENTAL HEALTH AUSTRALIA

BY: Yogi Vidyattama and Ingrid Hatfield
CONTACT: yogi.vidyattama@canberra.edu.au; policy@mhaustralia.org

This page is intentionally blank

CONTENTS

BACKGROUND

Mental Health Australia commissioned the National Centre for Social and Economic Research (NATSEM) at the University of Canberra (UC) to analyse existing datasets to shed light on barriers and inequities in access to mental health care in Australia. This document outlines NATSEM's approach, using data from the National Health Survey to estimate need for mental health services, and Medicare Benefits Schedule (MBS) data on service use to reflect access to mental health services.

Output from this analysis is presented in a series of online maps illustrating spatial variation at Statistical Area 3 (SA3) and Primary Health Network (PHN) level. A natural break classification based on Jenks algorithm was chosen to delineate five different categories depicted through colour coding from low to high rates of mental ill-health and service access. The algorithm is inherent in the ArcGIS mapping software to identify class boundaries where there are relatively big differences in the data values and hence, the groups of areas with similar values will be put together (ESRI, n.d.).

The project involved consultation with mental health and data experts regarding the meaning of various indicators including what is captured in different data sources. Through this process, it was decided the following measures would be used.

Measuring prevalence of mental ill-health and need for services

This research used measures of prevalence of mental ill-health as a proxy for need for mental health services. To estimate prevalence of mental ill-health, the research drew on data from the 2020-21 National Health Survey (NHS) undertaken by the Australian Bureau of Statistics (ABS) from August 2020 to June 2021 (ABS 2022). However, the sample size of Australian households in the 2020-21 NHS is not enough to allow reliable estimate at SA3 level or some areas at PHN level. Therefore, the estimate of prevalence of mental ill-health at this level is produced by NATSEM's Spatial Microsimulation (SpatialMSM) model based on the 2020-21 NHS combined with 2021 ABS Population and Housing Census and 2019-20 ABS Survey of Income and Housing (SIH). Detailed information about this modelling method used to calculate these estimates is provided below.

The NHS asks participants about their experience of psychological distress (measured by the Kessler Psychological Distress Scale K10), and long-term mental health conditions (defined in the survey as a diagnosable mental illness which has lasted or is expected to last 6 months or more).

This research used these measures from the NHS to estimate both the proportion of people experiencing high/very high psychological distress and long-term mental health conditions across each region in Australia.

Measuring access to mental health services

This research uses data on Medicare-subsidised mental health-specific services as an indicator of access to mental health care. These services are an important part of the mental health system, funded by the Australian Government and with high quality data. Medicare-subsidised services are only one component, however, of a much larger system of mental health services state and territory government funded and non-government services.

Medicare-subsidised mental health specific services are provided by general practitioners (GPs), psychiatrists, psychologists and other allied health professionals. The Australian Institute of Health and Welfare provides collated data on Medicare-subsidised mental health services use (AIHW 2022). To provide an indication of access to mental health services, this research mapped the proportion of the population across each region who accessed Medicare-subsidised mental health specific services in 2020-21.

This research also presents data on Primary Health Network (PHN) commissioned mental health services. The Australian Government funds PHNs to commission mental health services at a regional level to address local health service needs and gaps. PHN commissioned services provide essential alternate or complementary supports to Medicare-subsidised services, and deliver support to population groups who may not otherwise be able to access appropriate mental health care.

The Department of Health and Aged Care provided data from the Primary Mental Health Care Minimum Data Set on 'active clients' for mental health service use by PHN from 1 July 2021 to 30 June 2021. This data is presented according to the proportion of the population accessing PHN commissioned mental health services in 2020-21. This data is only presented at the PHN geographic level (not SA3).

Areas of Concern

Using the natural break classification, this research identified areas of concern by ranking areas into lower, medium and higher psychological distress (indicator of need) and lower, medium and higher MBS mental health service use (indicator of access). As previously discussed, the approach allows the researcher to group areas that have similar values in each of the indicators of need and access. Therefore, this grouping provided a clear differentiation of relative need and service access of a certain group of areas compared to other groups of areas. Given both groups of three are mutually exclusive, each category in the need indicator (lower, medium, higher) can have three possibilities in the access indicator (lower, medium, higher). Hence, there are nine combined categories.

These nine categories were then regrouped to classify areas into five different levels of concern, as below:

Concern: lower need/higher service access
Moderate concern: medium need/higher service access; lower need/medium service access
High concern: higher need/higher service access; medium need/medium service access; lower need/lower service access
Very high concern: higher need/medium service access; medium need/lower service access
Severe concern: higher need/lower service access

This concern indicator provides a simple method to compare in relative terms areas' level of need for and access to mental health care, as indicated in the matrix below:

Figure 1: Classification of need and access categories to areas of concern

	Level of need (prevalence of high/very high psychological distress)		
Service Access (rate of use of Medicare-subsidised mental health services)	Lower	Medium	Higher
Lower	**High concern**	**Very high concern**	**Severe concern**
Medium	**Moderate concern**	**High concern**	**Very high concern**
Higher	**Concern**	**Moderate concern**	**High concern**

THE MODELLING METHOD USED FOR NEED ESTIMATION

The 2020-21 National Health Survey (NHS) collected information about health status and lifestyle factors from approximately 11,000 households around Australia. While data from the 2017-18 NHS is publicly available via the ABS, this research used the more recent 2020-21 data through the ABS DataLab. The data are at the individual unit level but contain location information at the SA4 geographical scale.

As stated earlier, this project intended to provide information at the SA3 and PHN levels. These area levels are considered as an appropriate level to communicate about health service provision. Although the data in ABS DataLab is at the individual level and contains the SA4 level geography, a direct estimation at this area level can be restricted due to the number of observations for each SA4. The spatial microsimulation is a small area estimation model that aims to overcome this issue. The model works by distributing the larger number of observations from a larger area to the smaller area, based on benchmarks constructed from reliable data. At NATSEM, the model is named SpatialMSM and it distributes the survey sample by allocating the different weights that represent the number of people that is likely to be represented by the observation in different small areas based on the information from Census data.

The SpatialMSM cannot be applied directly to the 2020-21 NHS since there is another restriction. The NHS data being in the DataLab environment means the data cannot be taken out to be linked to other databases, such as the ABS Census data which contain information at the smaller area level. Therefore, a methodology was adopted for this research that had been specifically developed for highly confidential data with relatively small sample size (Vidyattama et al, 2015). This methodology uses another survey that has similar demographic variables to the restricted survey that will be imputed with the characteristics of interest using regression and/or probability model. This other survey is the one that will then be distributed to the small area. This study utilises the ABS 2019-20 Survey of Income and Housing (SIH) as the other survey. Furthermore, rather than conducting the estimation process for SA3 and PHN separately, the project estimated the indicators at the smaller SA2 spatial unit – small areas which usually equate to a suburb in cities with an average of around 10,000 people. The ABS considers that this SA2 geography represents a community that interacts socially and economically. Given the geographical concordance between SA2 and SA3 and PHN, the indicators could then be aggregated to provide results at the SA3 and PHN area levels.

The first step of the estimation at SA2 level followed the reweighting process of Tanton et al (2011). This approach requires a Census (in this case the 2021 Census) for small area benchmarks and the unit record data from the ABS 2019-20 SIH. The reason for using this survey was the large number of observations and that it has been proven to be able to be reweighted to produce reasonable estimates. The benchmarked variables needed to be available on both the population Census and the survey, using the same definitions and the same categories. The benchmarks also needed to be related to the final variable that is required from the spatial microsimulation model. Although the current benchmark from

Census originated from a study about income and housing (e.g. poverty, rent and mortgage stress), it has been the most suitable to be used with the SIH, due to its completeness and yet able to produce a high convergence rate that will be discussed further in the validation section. This allows the study to capture most of the variables that are important for mental health care estimation such as income, household composition, age, sex and labour force status. The model used for this report uses nine benchmarks from the 2021 Census as indicated in Table 1.

Table 1: Benchmarks for the spatial modelling

	Benchmark	Description
1	NPRD_2*HIND_2	Number of Persons Usually Resident in Dwelling by Total Household Income (weekly)
2	TENLLD_2*HIND_2	Tenure and Landlord Type by Total Household Income (weekly)
3	HCFMD_2*HIND_2	Family Household Composition by Total Household Income (weekly)
4	RNTRD_2*HIND_2	Rent (weekly) by Total Household Income (weekly)
5	MRERD_2*HIND_2	Mortgage repayments by Total Household Income (weekly)
6	AGE_2*HIND_2	Age of person (15+) by Total Household Income (weekly)
7	HIED_2*HIND_2	Equivalised Total Household Income (weekly) by Total Household Income (weekly)
8	LFSP_2*AGE_2* SEX_2	Labour Force Status by Sex and Age of person (15+)
9	QALLP_2	Non School Qualification

In addition, in this report we:

- Used households from the Greater Capital City Statistical Area (GCCSA) to populate the SA2s in that GCCSA. This means for example, only households from Sydney were used to populate the SA2s in Sydney.

- Reduced the number of benchmarks if the model failed for an area. This is done according to the sequence in the table. The lower number of benchmarks means fewer constraints and a higher possibility of achieving an acceptable result. If the estimate is produced with less than 6 benchmarks, then the estimate is excluded from the overall database as unreliable.

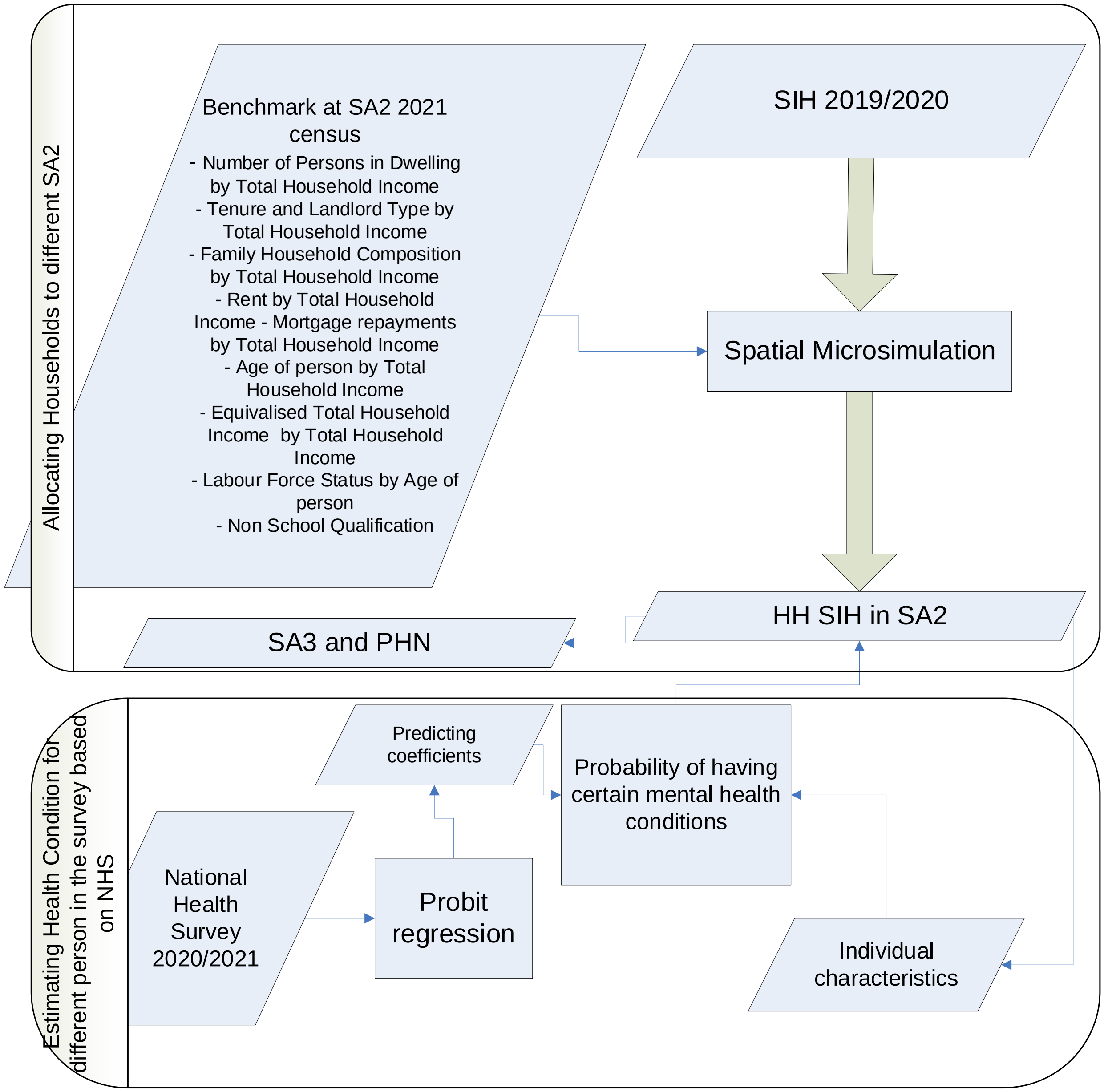
The technique then used a regression method to impute the specific conditions that were available from the NHS onto the synthetic database. The regression on variables of interest from the NHS produced the coefficients needed for the imputation of the variables onto the available unit record data. The regression used binomial independent variables of whether the individual is in the demographic groups mentioned above.

These included:

- Employed full time
- Employed part time
- Unemployed
- Not in labour force age 15-64 years
- Male
- Female
- Age 15-24 years
- Age 25-64 years
- In couple only household
- In couple with children household
- In single parent household
- In lone person household
- In household with equivalised income under $400/week
- In household with equivalised income between $400 to $1000/week
- In household with equivalised income between $1000 to $2000/week
- In household with equivalised income above $2000/week
- Different occupations
- each SA4

Given the variables of interest were binomial (two values – whether or not the person was experiencing high/very high psychological distress or long-term mental health condition/s), the model used was a probit regression model on the NHS 2020-21 database. The estimated coefficient for each independent variable listed above then allowed us to find the probability of the condition for each observation. We then applied the coefficients to the synthetic population estimated for the different SA2s. By using the SA2 synthetic population, we can utilise the individual fixed effect of each SA4 as one of the predictors in imputing all the necessary variables from the NHS. The flowchart of this process is shown in Figure 2.

Figure 2: The high/very high psychological distress and long-term mental health conditions at SA2 estimation process



VALIDATION

Validation of the modelling is essential. The validation of the small area estimates was carried out in three ways:

1. Looking at the proportion of areas for which we get convergence;
2. Comparing estimates from our spatial microsimulation model with estimates from the Census to identify how close our model predicts incomes from the Census at a small area level (unfortunately, the most important variable from the Census - prevalence of long term mental health condition/s, has a different measurement to the one in the NHS, and hence, cannot be used to validate the model estimates); and
3. a comparison of the aggregate number of the indicators that can be derived from the survey.

The first method of testing the reliability of our model is to look at the percentage of areas that provided estimates given a number of benchmarks. Reducing the number of benchmarks means that the model works (converges), but the estimates are not as good as when we have used more benchmarks. At some point, we decide that the estimate was not good enough to be published. Areas without reliable estimates are usually remote areas; or areas with very low populations (e.g. industrial areas or national parks). The proportion of areas that have converged in this model are shown in Table 2. It can be seen that nine benchmarks were mostly used to produce the estimates for small areas across Australia.

Table 2: Number of Benchmarks Used

	Number of Benchmarks used						
GCCSA	3 or 4	5	6	7	8	9	8 or more
1GSYD	0.0%	0.0%	0.3%	1.8%	3.1%	94.8%	97.8%
1RNSW	0.0%	0.3%	2.6%	2.7%	10.0%	84.5%	94.5%
2GMEL	0.0%	0.0%	1.5%	1.0%	1.4%	96.1%	97.6%
2RVIC	0.0%	0.2%	0.0%	2.7%	6.3%	90.7%	97.0%
3GBRI	0.0%	0.4%	2.1%	6.5%	8.3%	82.6%	90.9%
3RQLD	0.0%	0.6%	2.3%	5.3%	5.0%	86.7%	91.7%
4GADE	0.0%	0.0%	1.3%	2.1%	6.9%	89.7%	96.6%
4RSAU	0.0%	1.4%	0.4%	1.7%	22.6%	74.0%	96.5%
5GPER	0.0%	0.0%	0.6%	2.3%	2.0%	95.1%	97.1%
5RWAU	0.7%	1.3%	9.2%	5.0%	5.4%	78.4%	83.9%
6GHOB	0.0%	0.0%	0.0%	25.8%	14.2%	60.0%	74.2%
6RTAS	0.0%	0.0%	1.3%	10.1%	4.5%	84.1%	88.6%
7GDAR	0.0%	0.0%	7.8%	20.0%	6.5%	65.7%	72.2%
7RNTE	10.2%	34.4%	15.3%	14.8%	0.0%	25.2%	25.2%
8ACTE	0.1%	0.0%	5.5%	4.0%	4.7%	85.8%	90.5%
Australia	0.0%	0.3%	1.6%	3.2%	5.0%	89.8%	94.8%

Note: G means Greater (Capital Cities Areas); R means the Remainder (of the State/Territory)

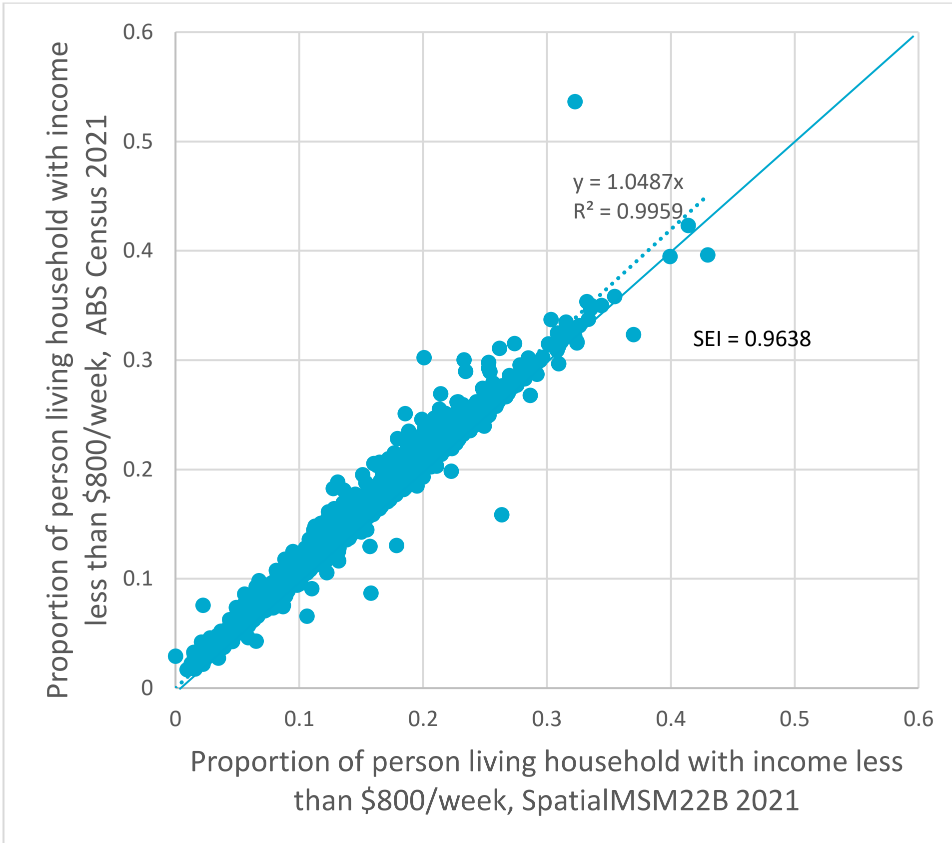
Based on this result, we decided to use the estimate produced using 6, 7, 8 or 9 benchmarks. Areas where results could not be derived using less than 6 benchmarks were removed from the analysis.

Another method to validate estimates at the small area level is to compare small area estimates with small area data that come from reliable sources. The measure for the validation is the standard error around identity (SEI) (Edwards and Tanton 2012). As mentioned above, the ideal variable to validate this study is the Census long term mental health condition. However, the Census data is based on the question " Has the person been told by a doctor or nurse that they have any of these long-term health conditions”.[1] This means the Census variable will be affected by access to a doctor or nurse to assess the person, while the variable that is being estimated from the NHS is based on self-report. The two variables therefore are likely to have a different distribution. Thus, to validate the small area estimates, we have calculated the proportion of people living in a household with income less than $800/week as well as household with equivalised income less than $300 a week from both the Census and from the model (at the time of the study the model was known asSpatialMSM22B). The two income thresholds were chosen based on the closest half median income measured available directly from the Census.

Figure 3 indicates that we achieved a very close estimate (0.9959 R-squared and 0.9638 SEI). In Figure 3, the vertical axis is the estimate from the Census; and the horizontal axis is the estimate from our model for each SA2. If the Census and our model gave exactly the same result for all areas, we would see all points on the 45 degree line (shown as a solid line in Figure 3). The SEI is the variability of the estimates around this 45 degree line (the line of identity). Achieving a good result using Household equivalised income is more difficult for this model since it is only being used partially as benchmark number 7. Nevertheless, the SEI shows an acceptable result of 0.70. The R-squared is the correlation between the Census and model estimates, and is much higher at 0.98.

[1] https://www.abs.gov.au/statistics/health/health-conditions-and-risks/health-census/2021

Figure 3: Validation of proportion of persons living with equivalised income less than $300/week (Spatial MSM and Census data)



The last validation of the estimates compared the estimated indicators at the aggregate level to the estimates from the survey (see Table 3). These estimates for larger areas from the survey have enough sample size on the survey to be released from the DataLab. The indicator that was able to be released for this study was the long term mental health condition. The comparison in Table 3 results show that the estimates generated by this study's methodology were reasonable.

As shown in Table 3, the estimate for Brisbane can be considered an under-estimate. The estimate for Sydney may also be a slight under-estimate and Rest of Western Australia a slight overestimate, but not to the extent of the difference for Brisbane. However, the difference may also be due to the different demographic and socio-economic composition found in 2021 Census compared with the benchmark used in the model, which was based on the estimated resident population (ERP) at December 2020. The ERP is based on adjusted 2016 Census counts, as is the NHS.

Table 3: Validation using reliable aggregate results (SpatialMSM and Survey data)

GCCSA	Prevalence of Long term mental health condition		Accuracy
	From survey	From Model	(Survey / Model)
Greater Sydney	0.192	0.152	1.265
Rest of NSW	0.297	0.278	1.069
Greater Melbourne	0.232	0.253	0.917
Rest of Victoria	0.298	0.318	0.938
Greater Brisbane	0.286	0.204	1.403
Rest of Queensland	0.254	0.218	1.166
Greater Adelaide	0.264	0.250	1.054
Rest of South Australia	0.289	0.257	1.126
Greater Perth	0.268	0.242	1.104
Rest of Western Australia	0.274	0.339	0.808
Greater Hobart	0.274	0.243	1.125
Rest of Tasmania	0.273	0.266	1.029
Northern Territory	0.204	0.212	0.962
Australian Capital Territory	0.285	0.269	1.059

Similar validation calculations were also conducted within the ABS DataLab for the indicator of high/very high psychological distress. At the time of reporting, this data was not able to be released outside of the DataLab. However, the validation did show a similar variation between area results from the survey and estimates from the model.

These results also show that all the estimates provided in this report are modelled, and that the modelling process introduces errors. While all efforts have been made by NATSEM to get reasonable estimates, including validation of the estimates, as shown in this section, no estimate should be treated as perfect. All estimates suffer from model error, and survey error from the original ABS survey data. Other methods may produce different estimates, due to different assumptions and methods. The method we use is deterministic, meaning the estimates can be reproduced using the same method, data, benchmarks and assumptions we have used – there is no probabilistic (random) element in our model.

The authors are happy to be contacted to further discuss the methods used in the modelling.

REFERENCES

ABS (2022), National Health Survey: First Results methodology, available at https://www.abs.gov.au/methodologies/national-health-survey-first-results-methodology/2020-21#about-this-survey

AIHW (2022). Mental Health: Medicare-subsidised services, available at https://www.aihw.gov.au/mental-health/topic-areas/medicare-subsisded-services

Edwards, K. L., & Tanton, R. (2013). Validation of spatial microsimulation models. *Spatial microsimulation: A reference guide for users*, 249-258. https://link.springer.com/chapter/10.1007/978-94-007-4623-7_15

ESRI (n.d.) Data Classification method, ArcGIS Pro 3.0, available at https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm

Tanton, R., Vidyattama, Y., Nepal, B., & McNamara, J. (2011). Small area estimation using a reweighting algorithm. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(4), 931-951. https://doi.org/10.1111/j.1467-985X.2011.00690.x

Vidyattama, Y., Tanton, R., & Biddle, N. (2015). Estimating small-area Indigenous cultural participation from synthetic survey data. *Environment and Planning A*, *47*(5), 1211-1228. https://doi.org/10.1177/0308518X15592314