DEMENTIA BURDEN, GEOGRAPHICAL LOCATION and HOSPITALS IN AUSTRALIA

IBM PROFESSIONAL CERTIFICATE IN DATA SCIENCE

CAPSTONE PROJECT (WEEK 5)

(A health-related project)

Emmanuel Michael Mukinda Bukajumbe

September 2019

INTRODUCTION

Description of the problem and background. Alzheimer's disease (AD) is a neurodegenerative disease affecting the brain. It mainly occurs among the elderly that is, over 65 years of age (late onset AD) although it also occurs earlier (early onset AD).(1, 2) It is the second leading cause of morbidity and mortality in the Australia,(3) but has no cure. There are many factors associated with AD with ageing as the one most linked to the disease. Others factors include sex (more females are affected than males), cigarette smoking, alcoholism, chronic diseases such as diabetes mellitus, Huntington's disease and others, low education and low socioeconomic status, traumatic brain injury, anaemia, positive Apolipoprotein E-ε4 gene status, environmental pollutants.(1, 2, 4-7) In terms of the associated factors, there is a paucity of evidence concerning AD area (state) geographical differences in Australia. However, there is research evidence to suggest that geographical regions may differ in the progression of AD and its measurement.(8) A meta-analysis suggests that rural living (especially early-life rural living) as opposed to urban living, is associated with a high risk for AD.(9)

This project therefore primarily seeks to determine the link between AD and Australian states (Australian Capital Territory (ACT), New South Wales (NSW), Northern Territory (NT), Queensland (QLD), South Australia (SA), Tasmania (TAS), Victoria (VIC), and Western Australia (WA)) and their respective geographical locations. It secondarily briefly considers differences in terms of hospitals as well. The findings of the project may help the government of Australia and other governments to plan cities and/or housing in line with the geographical location of the different states and/or cities. It may also contribute to the planning of hospital infrastructure in Australia.

Description of the data and how they will be used to solve the problem. The data contain Australian states by latitude, longitude, and the prevalence of AD by state. The data also contain Australian cities by latitude and longitude and, the number of hospitals in the various states. The data on AD were obtained from Dementia Australia's electronic article on dementia in Australia and the prevalence estimates from 2019 to 2058(10) whereas the data on cities was obtained from LatLong.net, a website detailing the geographical coordinates of the cities(11).

Using various machine learning algorithms in Python (Python Software Foundation, Version 3.7), the data will be used to check the association between the number of hospitals (classified by hospital type) and the prevalence of dementia. They will be used to show whether there is an association between latitude and longitude of a particular state with the prevalence of dementia. The data will also be used to show the location of the major cities in the different states. Using the Foursquare Application Program Interface (API), the data will be used to show the hospitals found in the different key cities of states, with a comparison of the prevalence of AD in the cities and states.

METHODOLOGY

***Project design and selection*.** This is an ecological retrospective study involving all the 8 Australian states with the major cities in each state and, considering hospitals in Australia that is, three levels are studied – nationwide, state-wide and citywide levels. Selection of a state to be part of the study is purely due to its presence in Australia implying that all states had an equal chance of selection. Selection of a city for further scrutiny was arbitrary, but each state contributed one major city.

Relevant dates. The data were sourced from various websites(10, 11) in August and September 2019.

Variables. The predictor variables include number and types of hospitals, and the geographical coordinates (latitudes and longitudes) as obtained from the websites.(10, 11) The types of hospitals in the different cities was obtained using the Foursquare API. The outcome variable in the project is the burden of dementia. The primary focus of the project is on AD since, up to 80% of dementia cases are AD-related.(1) The dementia burden was obtained from prevalence statistics provided by Dementia Australia.(10) The burden per state was calculated as follows:

Dementia burden = (Number of cases in a state /Total number of cases in Australia) x 100%

Statistical analysis. The data were analysed using Python (Python Software Foundation, Version 3.7). Various modules and libraries were imported for the purpose of the analysis, including Pandas, Matplotlib, Scikit-Learn, Scipy, Folium, json, Nominatim, requests and Numpy. Quantitative variables (geographical coordinates, number of hospitals and burden

of dementia) and, qualitative variables (including hospital types, and states) were analysed as such. Data on dementia burden, cities and hospitals and types of hospitals were organized into Pandas dataframes, which were merged into a single Pandas dataframe.

Numerical/quantitative data were summarized using means, standard deviations and the maximum and minimum values. Data were also summarized using frequencies and/or percentages. To visualize the data, graphs (such as the bar graph) and plots (such as regression plots and box plots) and, maps were used. The correlation between any two numerical variables (reported as a correlation coefficient) that is, between dementia burden and any of, number of hospitals (for each type of hospital) or geographical coordinates was determined along with the p-values and β-statistic with the help of the Scipy library and, linregress.

A supervised machine learning algorithm from the Scikit-Learn module, linear regression, was used to determine the association of the number of hospitals (for each type of hospital) or, geographical coordinates with the burden of dementia since this, as the dependent variable in the project, is numerical. This was done in two steps; first in a simple linear regression (SLR) model including either number of hospitals or geographical coordinates as the independent variables and next in a multiple linear regression (MLR) model by including latitudes and longitudes to a SLR model with only number of hospitals as the independent variable. The results of the SLR and MLR models are reported using coefficients with their corresponding R-squared (R^2) scores to compare the models.

RESULTS

All the 8 Australian states were considered for the project. Table 1 shows the state study characteristics.

Table 1. Study characteristics for the different states in Australia.

State	AD burden (%)	No. of AD cases	No. of public Non-Psych hospitals	No. of public Psych hospitals	No. of private hospitals	Total No. of hospitals
ACT	1.327	5,932	3	0	0	3
NSW	33.381	149,250	214	8	205	427
NT	0.395	1,764	5	0	0	5
QLD	18.997	84,940	119	4	109	232
SA	8.398	37,551	75	2	56	133
TAS	2.521	11,270	22	1	0	23
VIC	25.671	114,779	148	3	169	320
WA	9.311	41,630	87	4	62	153
Total	**100.00**	**447,116**	**673**	**22**	**601**	**1,296**

Acronyms. Psych: psychiatric, non-psych: Non-psychiatric

There were 447,116 dementia cases in Australia at the time of data collection. NSW had the highest number of cases with 149,250 whereas NT had the least number with 1,764 cases. NSW also has the highest number of hospitals of any type (427) whereas ACT and NT have the lowest number of hospitals of any type with 3 and 5 hospitals respectively. The total number of hospitals in Australia at the time of data collection and analysis is 1,296. Figure 1 shows the dementia burden by Australian state.

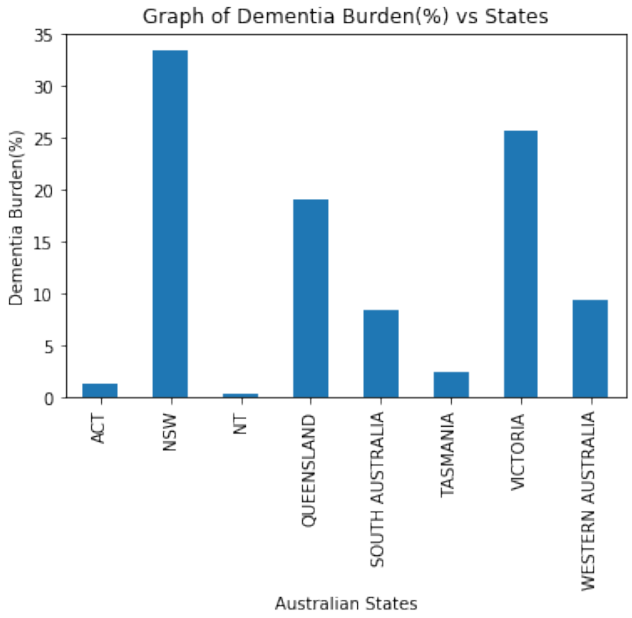

Figure 1. Graph of dementia burden by Australian States. ACT – Australian Capital Territory, NSW – New South Wales, NT – Northern Territory.

The burden of dementia. The mean burden of dementia is 12.50% (SD 12.25%). The maximum dementia burden is 33.38% in NSW and the minimum burden of dementia is 0.395% in NT. See table 2.

Hospitals. On average, there are more public non-psychiatric hospitals (84) than any other type of hospital. This number is followed by private hospitals (75) and, the lowest number of hospitals is the public psychiatric hospitals (3). Table 2 shows the summary descriptive statistics of type of hospitals and dementia burden in Australia.

Table 2. Descriptive statistics of hospital type and dementia burden in Australia

	Mean (%)	**Standard deviation or SD (%)**	**Median (%)**
Dementia burden	12.50	12.25	8.86
No. of public hospitals (NPsych)	84.13	74.601	81.00
No. of public hospitals (Psych)	2.75	2.66	2.50
No. of private hospitals	75.13	79.55	59.00
No. of dementia cases	55,899.50	54,751	3959.50

Acronyms – NPsych: Non-psychiatric, Psych – psychiatric

Figure 2 is a box plot of dementia burden showing its descriptive characteristics.

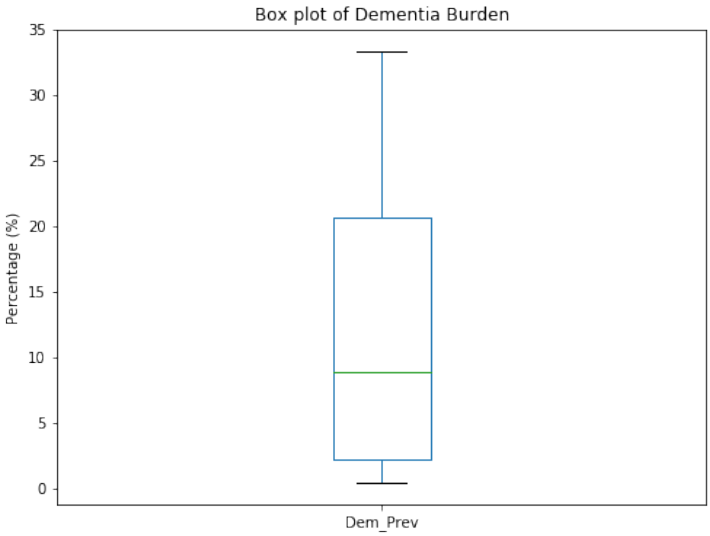

Figure 2. Box plot of dementia burden (Dem_Prev)

Correlation and regression statistics. Using linregress, the correlation coefficients, r's (for the association of number of hospitals and, of geographical coordinates, with dementia burden) and beta (β) coeffiecients SLR were obtained. For MLR assessing the association of number of hospitals with dementia burden including latitude and longitude, the Scikit-Learn module was employed. Table 3 shows the correlation and regression statistics.

The number of hospitals of any type correlates strongly and positively with the burden of dementia. See figure 3 (a to d). For each unit increase in the number of hospitals for all types of hospitals, there is an increase in burden of dementia; this is highest with public psychiatric hospitals. Each unit increase in number of public psychiatric hospitals is associated with a statistically significant 4.073 percentage increase in dementia burden. The increase in R^2-value is also significantly higher with public psychiatric hospitals than with the other types of hospitals. The inclusion of latitudes and longitudes in the SLR model with number of hospitals and dementia burden reduces the coefficient slightly. In addition, the R^2-value also increases slightly for the associations of number of public non-psychiatric hospitals and number of private hospitals, with dementia burden, but most for public psychiatric hospitals. However, there is a weak, statistically non-significant correlation between longitude and dementia burden, but no correlation between latitude and longitude. In line with this, linear regression shows that there is a non-significant association of longitude and dementia burden, but no association of latitude and dementia burden.

Table 3. Correlation and regression statistics for the association of number of hospitals and, of geographical coordinates with dementia burden.

	β-coefficient (95% CI)	**r***	**Std error**	**P-value**	**R^2**
A) Hospital type					
Public NPsych					
a) SLR	0.162 (-0.706 – 1.030)	0.984	0.443	<0.001	0.968
b) MLR	0.158				0.988
Public Psych					
a) SLR	4.073 (3.993 – 4.153)	0.884	0.041	0.004	0.782
b) MLR	4.052				0.851
Private					
a) SLR	0.153 (-0.347 – 0.653)	0.995	0.255	<0.001	0.991
b) MLR	0.151				0.995
B) Geo coords					
a) Latitude	-0.082 (-1.332 – 1.168)	-0.053	0.638	0.902	0.003
b) Longitude	0.358 (-0.710 – 1.426)	0.259	0.545	0.535	0.067

Acronyms: r – correlation coefficient, * - obtained using linregress, which provides the β-coefficient along with the p-value: Std error – Standard error of the mean: R^2 – R-squared statistic: SLR – simple linear regression, MLR – multiple linear regression: Geo coords – geographical coordinates

N:B; The 95% confidence interval (CI) of the β-coefficient was estimated (calculated) using the Std error and z-score of 1.96 as follows: 95% CI = β-coefficient ± 1.96

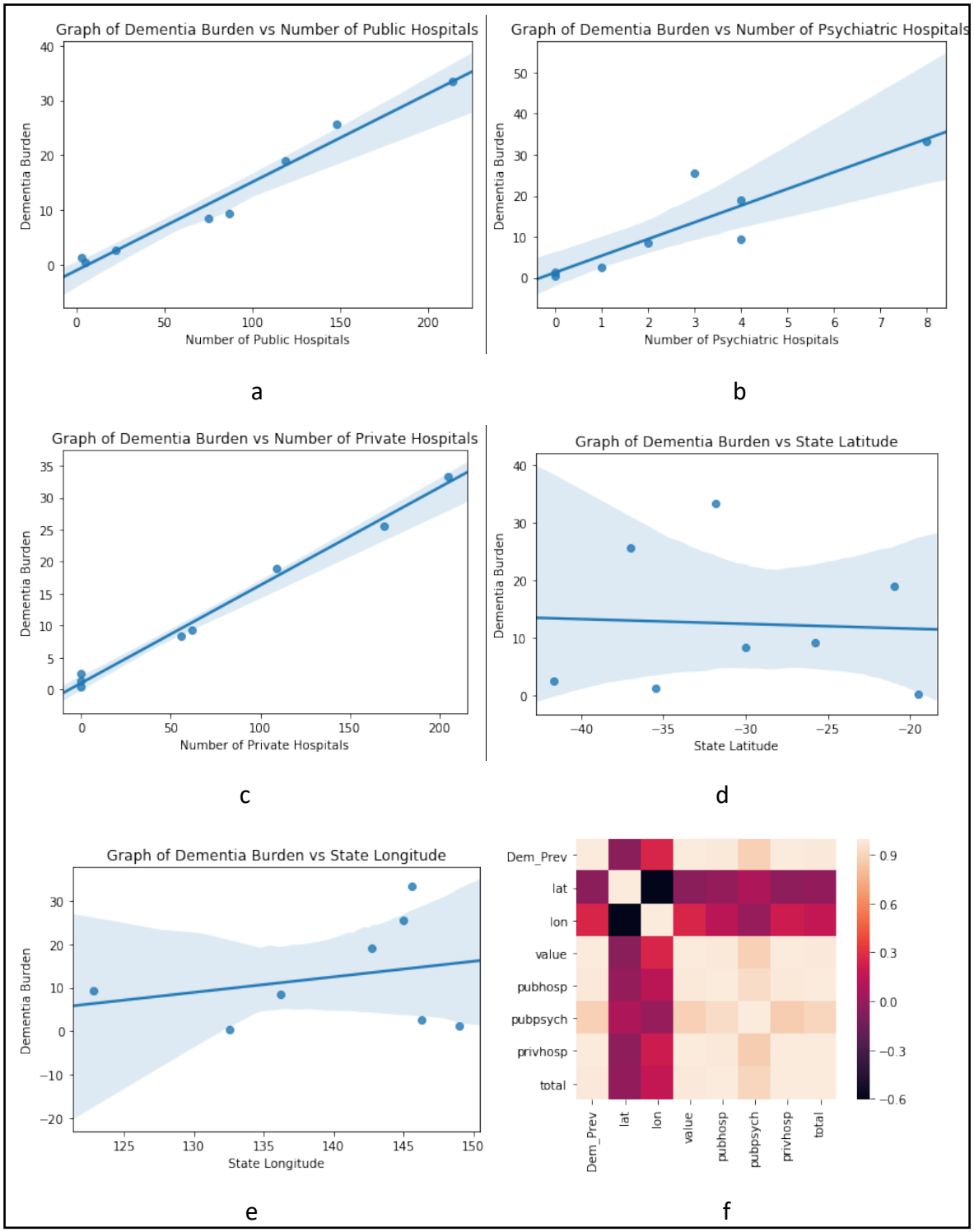

Figure 3. Regression plots of the association of number of hospitals and the burden of dementia (%). a – the association of number of public non-psychiatric hospitals with dementia burden, b – the association of the number of public psychiatric hospitals with dementia burden, c – the association of the number of private hospitals with dementia burden, d – the association of latitude with dementia burden, e – the association of longitude with dementia burden, f – heatmap for the correlation between any two numerical variables (Dem_Prev: Dementia burden, lat: latitude, lon: longitude, value: number of dementia cases, pubhosp: non-psychiatric public hospitals, psychhosp: public psychiatric hospitals, privhosp: private hospitals.)

***Australian cities and hospital differences*.** In general, most Australian cities are located on the coasts of various oceans and/or seas (figure 4) that is, the Indian Ocean, Southern Ocean, South Pacific Ocean, Arafura Sea, Timor Sea, Coral Sea and Tasman Sea. Because of the nature of this ecological study, involving Australian states and selected cities, clustering of cities carried out using density-based spatial clustering of applications with noise (DBSCAN), with a minimum of one sample, returns 28 clusters, which is the number of cities under study; the use of any minimum number of samples above 1 returns one cluster. Therefore, clustering may not significantly be informative in line with the study question. A few major cities of the Australian states (Canberra for ACT, Sydney for NSW, Darwin for NT, Brisbane for QLD, Adelaide for SA, Launceston for Tasmania, Melbourne for Victoria and Perth for Western Australia) were randomly selected from cities' data as per LatLong.net website(11) for a closer look at the hospitals in them.

Figure 4. Map of Australia showing the location of cities.

The Foursquare API was used to obtain json files that were in turn used to obtain hospital data. Only hospitals with the term 'hospital' as part of their name were considered in a radius of 5,000 metres of the each selected city. This is different for Darwin for which a radius of 15,000 metres was considered since no hospital was located with 5,000 metres. Table 4 shows the hospitals present in each city within the given radius. Given the nature of dementia, the table excludes children's hospitals and specific system-specialized hospitals such as eye hospitals and cardiology hospitals. Therefore, by name, the hospitals included in the table are not specific for any given specialty. Due to these reasons and the fact that the hospital search per city included those with the word 'hospital' in their name, the table is not exhaustive. This project assumes that the hospitals primarily handle all kinds of cases. The table therefore shows a wide choice of hospitals a patient with AD may be taken to for care.

Table 4. Hospitals within a 5,000-metre radius of the capital cities of Australian states.

State	Selected City	Hospital name	State burden of dementia (%)
Australian Capital Territory	Canberra	Calvary Hospital	1.327
New South Wales	Sydney	Sydney Hospital	33.381
		East Sydney Private Hospital	
		St.Vincent's Hospital	
		St.Luke's Hospital	
		St.Vincent's Private Hospital	
		Royal Prince Alfred Hospital	
		Mater Hospital	
		Wolper Jewish Hospital	
		Rozelle Hospital	
		Mosman Private Hospital	
		Greenwich Hospital	
Northern Territory	Darwin*	Royal Darwin Hospital	0.395
		Darwin Private Hospital	
Queensland	Brisbane	Queensland Private Hospital	18.997
		St.Andrew's Hospital	
		Mater Adult Hospital	
		St.Vincent's Hospital	
		Mater Private Hospital	
		Mater Mother's Hospital	
		The Wesley Hospital	
		Royal Brisbane & Women's Hospital	
		Princess Alexandrian Hospital	
South Australia	Adelaide	Adelaide Memorial Hospital	8.398
		New Royal Adelaide Hospital	
		Women's and Children's Hospital	
		Calvary Wakefield Hospital	
		Calvary North Adelaide Hospital	
		Parkwynd Private Hospital	
		Royal Adelaide Hospital	
		St.Andrew's Hospital	
		Calvary Rehabilitation Hospital	
		Glenside Campus Hospital	
		Burnside War Memorial Hospital	
Tasmania	Launceston	Calvary Hospital	2.521
		St.Vincent's Hospital	
		Launceston General Hospital	
Victoria	Melbourne	The Alfred Hospital	25.671
		The Royal Melbourne Hospital	
		St.Vincent's Hospital	
		Epworth Freemasons Hospital	
		Royal Women's Hospital	
		Mercy/St.Vincent's Private Hospital	
		The Avenue Hospital	
		Melbourne Private Hospital	
Western Australia	Perth	Royal Perth Hospital	9.311
		Mount Hospital	
		Princess Margaret Hospital	
		Hollywood Private Hospital	
		St.John of God Mt Lawley Hospital	
		King Edward Memorial Hospital	
		Sir Charles Gairdner Hospital	
		South Perth Hospital	

*There is no hospital within 5,000 meters of Darwin. Darwin hospitals shown in the table are found within a 15,000-metre radius.

Discussion

Results summary. The project sought to determine the association of geographical coordinates and, of hospital numbers with dementia in the 8 Australian states. NSW has the highest dementia burden whereas NT has the lowest. There is a strong positive correlation between the number of hospitals and dementia burden. Each unit increase in the number of hospitals is associated with an increase in the dementia burden. This is highest with the public psychiatric hospitals. Geographical coordinates seem to affect the association of hospital numbers with AD burden. However, latitude is not associated with dementia burden, but longitude is associated with dementia, but the association is not statistically significant. Most Australian cities are located on the coasts of oceans and/or seas.

Possible explanation and/or inference. The burden of AD is highest in states with the highest population, but lowest in states with the lowest population. For instance, NSW, which has the highest dementia burden, has a population of 7.89 million people. Victoria with the second highest dementia burden has a population of 6.27 million people. ACT and NT, which have the lowest dementia burden, have populations of 406,692 and 245,786 respectively.(12) When major cities and other urban are considered, Sydney (in NSW) and Melbourne (in VIC) have the highest populations (4,391,674 and 3,999,982 respectively) and Kingaroy (in QLD) and Mudgee (in NSW) have the lowest populations (9,808 and 10,483 respectively).(13) The major Australian cities are located along the coastline (Figure 3). One could ask whether the coastline environment increases the risk for dementia or not. This may be associated with associated increased trade and/or economic activities in these cities. The association between hospital number and dementia burden may therefore be confounded by the level of these activities. Notably though, an inclusion of latitude and longitude in the linear regression model with a significant change in R^2-value for public psychiatric hospitals suggests a role of geography in AD. The lack of association of latitude with dementia burden and a weak association of longitude with dementia burden may suggest that this is a spurious association. Longitude is weakly correlated with dementia burden in this study; latitude is not. This is in contrast to a study comparing the dementia mortality in Italy, Chile and Newzealand, which suggests that high latitudes in Italy are

associated with a risk for dementia. However, in the this study, Newzealanders, particularly women in the North Islands had a greater risk for dementia than those in the South Islands.(14) This could explain why longitude and not latitude, albeit, not statistically significantly, is associated with dementia burden in Australia given that Newzealand is a Australia's immediate neighbour. Environmental conditions such as air pollution(14, 15) as suggested by some studied could be a key factor in this case especially with the fact that the role of vitamin D in protection against dementia is controversial. Whereas some studies support its role,(14, 16) others show that its deficiency is not associated with dementia as is the case with an 18-year prospective study(17). The more industrialized states are the ones with the highest dementia burden. This may imply that many people grow up in the industrialized cities, work in them and retire in them that is; they grow old in them.

In addition the association of public psychiatric hospitals with dementia burden may not be surprising since dementia is considered a mental illness as shown in the American Psychiatric Association published Diagnostic and Statistical Manual of Mental Disorders fifth edition criteria (DSM-V criteria) in which it is considered a neurocognitive disorder.(4, 18) However, psychiatric hospitals are few compared to other hospitals. Therefore, the study may not be conclusive regarding this finding.

Practical components deduced from the data. Three practical components are deduced and, presented from the project.

a) *The research component*. Given the knowledge on the relationship between the number of hospitals and dementia burden, more research is required to study this phenomenon further. In addition, further research on the effect of geographical location on this association or on dementia burden per se is required. In line with this, the centres or hospitals where research on AD burden in line with geographical location can be carried out are shown in table 4.
b) *The business component*. Given the relationship between the number of hospitals and dementia burden, if one were interested in setting up a private clinic that is highly specialized in handling dementia cases, this project may provide a clue on where the hospital may be set up. According to this study, the highest dementia burden is in the most densely populated cities and states.

c) *Public health policy component*. The project shows differences in hospital numbers according to states and differences in dementia burden. Interestingly, where the dementia burden is high, the number of hospitals is also high. From a policy standpoint, the questions concerning this observation could rotate around adequacy of hospitals in terms of handling cases that may contribute to dementia or AD in the future or, dementia itself. Building other hospitals away from the densely populated areas to deal with the dementia burden may help reduce the work burden of the hospitals currently handling the dementia or AD cases. One reason for the high number of hospitals being associated with dementia burden could be a high number of referrals from other medical centres in addition to people growing up, working, and retiring in the highly industrialized and densely populated cities. In addition, the number of psychiatric hospitals is small, compared to the number of other hospitals, but is it too small? Should more public psychiatric hospitals be set up? Should the human resource capacity handling dementia be increased?

Limitations of the project. There may be other ecological factors that can affect the dementia burden or the association of hospitals and/or hospital numbers with dementia, apart from geographical coordinates that were not considered in the project. Their inclusion in the analysis may provide more information and more conclusive inferences.

Strengths of the project. Data were readily available from websites and/or webpages. In addition, the use of the Foursquare API provided quick and current and/or updated hospital data.

CONCLUSION

In this project, the number of hospitals is strongly positively correlated with dementia burden in Australia. This may be associated with population differences and may be due to the government's response to population health needs. Each unit increase in number of hospitals is associated with an increase in dementia burden. The data also suggest a possible role of geographical coordinates of states or cities in the AD burden, but this could be spurious given the fact that the study finds no correlation between latitude with dementia burden and only a weak correlation between longitude and dementia burden. Further

research involving other ecological factors and in-hospital data potentially associated with dementia is, recommended.

References

1. Flemming KD, Jones Jr, L.K., editor. Chapter 30. Mild Cognitive Impairment and Alzheimer's Disease by Khan, Q.U.A., Taner, D. in; Mayo Clinic Neurology Board Review. Clinical Neurology for Initial Certification and MOC, page 275-284

New York, USA: Oxford University Press; 2015.
2. Halter JB, Ouslander, J.G., Studenski, S., High, K.P., Asthana, S., Supiano, M.A., Ritchie, C., editor. Chapter 66. Dementia Including Alzheimer Disease by Carlsson C,M., Gleason C.E., Puglielli L., Asthana S in; Hazzard's Geriatric Medicine and Gerontology: pages 991-1013. 7th ed. USA: McGraw Hill Education; 2017.
3. Australia D. Dementia statistics Australia2019 [Available from: https://www.dementia.org.au/statistics.
4. Daroff R, Ankovich, J., Mazziota, J.C., Pomeroy, S.L., editor. Chapter 95 Alzheimer's Disease and other Dementias by Peterson R., Graff-Radford J in; Bradley's Neurology in Clinical Practice; pages 1380-1399. 7th ed: Elsevier; 2016.
5. de la Torre JC. Alzheimer's disease is a vasocognopathy: a new term to describe its nature. Neurol Res. 2004;26(5):517-24.
6. Hong CH, Falvey C, Harris TB, Simonsick EM, Satterfield S, Ferrucci L, et al. Anemia and risk of dementia in older adults: findings from the Health ABC study. Neurology. 2013;81(6):528-33.
7. Kivipelto M, Mangialasche F, Solomon A, Group FS. Pointing the FINGER at multimodal dementia prevention - Authors' reply. Lancet. 2015;386(10004):1627.
8. Henley DB, Dowsett SA, Chen Y-F, Liu-Seifert H, Grill JD, Doody RS, et al. Alzheimer's disease progression by geographical region in a clinical trial setting. Alzheimer's Research & Therapy. 2015;7(1):43.
9. Russ TC, Batty GD, Hearnshaw GF, Fenton C, Starr JM. Geographical variation in dementia: systematic review with meta-analysis. International journal of epidemiology. 2012;41(4):1012-32.
10. Australia D. Dementia in Australia; Prevalence Estimates 2019 - 20582019 28/08/2019:[https://www.dementia.org.au/files/documents/2019-58-Dementia-prevalence-S-T.pdf pp.].
11. LatLong.net. Cities in Australia: LatLong.net; 2019 [Available from: https://www.latlong.net/category/cities-14-15.html.
12. Australia P. State Population Australia: Population Australia; 2019 [Available from: http://www.population.net.au/states/.
13. Wikipedia. List of cities in Australia by population 2019 [updated 17/08/2019. Available from: https://en.wikipedia.org/wiki/List_of_cities_in_Australia_by_population.
14. Russ TC, Murianni L, Icaza G, Slachevsky A, Starr JM. Geographical Variation in Dementia Mortality in Italy, New Zealand, and Chile: The Impact of Latitude, Vitamin D, and Air Pollution. Dementia and Geriatric Cognitive Disorders. 2016;42(1-2):31-41.
15. Russ TC, Gatz M, Pedersen NL, Hannah J, Wyper G, Batty GD, et al. Geographical variation in dementia: examining the role of environmental factors in Sweden and Scotland. Epidemiology. 2015;26(2):263-70.
16. Sommer I, Griebler U, Kien C, Auer S, Klerings I, Hammer R, et al. Vitamin D deficiency as a risk factor for dementia: a systematic review and meta-analysis. BMC Geriatrics. 2017;17(1):16.
17. Olsson E, Byberg L, Karlström B, Cederholm T, Melhus H, Sjögren P, et al. Vitamin D is not associated with incident dementia or cognitive impairment: an 18-y follow-up study in community-living old men. The American Journal of Clinical Nutrition. 2017;105(4):936-43.
18. Sachdev PS, Mohan A, Taylor L, Jeste DV. DSM-5 and mental disorders in older individuals: an overview. Harvard review of psychiatry. 2015;23(5):320-8.