Identifying patterns of general practitioner service utilisation and their relationship with potentially preventable hospitalisations in people with diabetes: The utility of a cluster analysis approach

Author names and affiliations

1) Ninh Thi Ha

School of Public Health, Curtin University, Perth, Western Australia 6845, Australia

thininh.ha@student.curtin.edu.au

2) Mark Harris

School of Economics and Finance, Curtin University, Perth, Western Australia 6845, Australia

mark.harris@curtin.edu.au

3) David Preen

Centre for Health Services Research, School of Population and Global Health, The University of Western Australia, 35 Stirling Highway, CRAWLEY WA 6009, Australia

david.preen@uwa.edu.au

4) Suzanne Robinson

School of Public Health, Curtin University, Perth, Western Australia 6845, Australia

Suzanne.robinson@curtin.edu.au

5) Rachael Moorin

School of Public Health, Curtin University, Perth, Western Australia 6845, Australia and School of Population and Global Health, The University of Western Australia, 35 Stirling Highway, CRAWLEY WA 6009, Australia

r.moorin@curtin.edu.au

Corresponding author

Ninh Thi Ha

Email: thininh.ha@student.curtin.edu.au

School of Public health, Curtin University, Perth, Western Australia 6845, Australia

Abstract

Aims: We aimed to characterise use of general practitioners (GP) simultaneously across multiple attributes in people with diabetes and examine its impact on diabetes related potentially preventable hospitalisations (PPHs).

Methods: Five-years of panel data from 40,625 adults with diabetes were sourced from Western Australian administrative health records. Cluster analysis (CA) was used to group individuals with similar patterns of GP utilisation characterised by frequency and recency of services. The relationship between GP utilisation cluster and the risk of PPHs was examined using multivariable random-effects negative binomial regression.

Results: CA categorised GP utilisation into three clusters: moderate; high and very high usage, having distinct patient characteristics. After adjusting for potential confounders, the rate of PPHs was significantly lower across all GP usage clusters compared with those with no GP usage; IRR=0.67 (95%CI: 0.62-0.71) among the moderate, IRR=0.70 (95%CI 0.66-0.73) high and IRR=0.76 (95%CI 0.72-0.80) very high GP usage clusters.

Conclusions: Combination of temporal factors with measures of frequency of use of GP services revealed patterns of primary health care utilisation associated with different underlying patient characteristics. Incorporation of multiple attributes, that go beyond frequency-based approaches may better characterise the complex relationship between use of GP services and diabetes-related hospitalisation.

Keywords: Cluster analysis; primary health care; potentially preventable hospitalisation; diabetes; data linkage.

1. Introduction

Diabetes is an increasing public health issue causing a substantial burden on health care systems around the world [1]. In Europe, the number of people with diabetes was nearly 60 million in 2013, and is estimated to increase to 70 million by the early 2030s [2]. Similarly, in the United States the prevalence of diabetes was estimated at 29.1 million in the national report in 2014 [3]. In Australia, a country of approximately 24 million people, the prevalence of diabetes was about 1.2 million in 2014-15 [4] and is estimated to increase to 3.4 million by early 2030s [5].The condition costs the Australian Health system more than $AU6.5 billion each year [5]. Diabetes is considered an ambulatory care sensitive condition [5], and consequently enhancing primary health care to better manage diabetes has been a major approach in the health care system of Australia [5, 6].

The literature suggests that better primary health care delivery reduces the risk of hospitalisations for ambulatory care sensitive conditions in general [7-9]. With respect to diabetes, a recent systematic review indicated that regular primary care was associated with reduced risk of hospitalisation [10]. However, other aspects such as frequency of visits or access to primary health care show inconsistent results [10].

In Australia, primary care services, mainly provided by general practitioners (GP), are subsidised through a universal health insurance scheme, Medicare, on a fee-for-service basis [6]. Dedicated financial incentives have been provided under Medicare for GPs to provide comprehensive care for diabetes [6]. However, to our knowledge, limited research has evaluated patterns of utilisation of primary health care services for people with diabetes and their impact on health outcomes. Current studies are limited to examining the utilisation of primary health care based on single indicators such as frequency [6] or regularity of services used [11].

Since patterns of primary health care utilisation are likely to be complex, more advanced approaches that account for multiple factors are required to more accurately classify and discover meaningful patterns of primary health care utilisation by people with diabetes. K-mean cluster analysis, a data-driven approach, is capable of taking into account multiple dimensions simultaneously and is suitable for use with large datasets [12]. The technique can classify individuals with similar characteristics into homogeneous groups which can also maximise heterogeneity between groups

[12]. The technique has been applied to a variety of settings, for example, health behaviour [13]; health psychology [14]; health care cost analysis [12] and genetic classification [15].

Thus, our study aims to apply K-mean cluster analysis to identify GP utilisation patterns using multiple attributes of GP usage among people with diabetes. We will also examine the impact of identified GP utilisation patterns on the risk of potentially preventable hospitalisations (PPHs). Understanding patterns of GP utilisation and how they impact on health outcomes is useful for planning health care provision targeted to encouraging particular patterns in utilisation and enhancing the relationship between patients and their primary health care provider.

2. Material and methods

2.1 Data sources

The Western Australian (WA) linked data used for this study comprised whole-of-population administrative health data linked at the individual level, for residents of WA aged 18 years or older who were registered at any time on the WA Electoral Roll [16]. The data included a complete set of WA Hospital Morbidity Data System (HMDS) records; Medicare Benefit Scheme (MBS) claim records; WA Electoral Roll (ER) records; and WA mortality records for each individual subsequent to their first ever WA Electoral Roll record. Details of each dataset have been described previously [17]. In brief, the datasets provide statutory information on all hospitalisations (HMDS), claims for medical services out-of-hospital including GP visits (MBS), dates individuals migrated in and out of WA or changed address while living in WA (Electoral Roll) and date/cause of death.

2.2 Study population

Annual panel data from 1998/1999 to 2003/2004 were constructed consisting of individuals with diabetes identified via HMDS or MBS data prior to the start of or in the baseline financial year (1998/99). Diabetes mellitus was determined using the International Classification of Disease (ICD), 9th edition-clinical modification (ICD-9-CM) codes in HMDS records and MBS claims indicative of the presence of diabetes as described elsewhere [17]. All individuals were observed annually from the baseline year to 30 June 2004, last year living in WA or death (whichever occurred first) for any change in GP utilisation, hospitalisations and clinical and demographic characteristics. GP utilisation and demographic and clinical characteristics were measured in the exposure year, and PPH outcomes measured in the following year. Only individuals who were alive and resident in WA for at least two consecutive years were included in the study. The couplet design (ie. comprising pairs of years, the exposure year followed by an outcome year) has been applied in recent publications [6, 17].

Ethical approval was provided by The University of Western Australia and Curtin University Human Research Ethics Committees who exempted the study from obtaining individual patient consent.

2.3 Study outcome and predictors

2.3.1. Diabetes related potentially preventable hospitalisations

The primary outcome measure was diabetes related potentially preventable hospitalisations (PPH) during the following-up year of each couplet.

Hospitalisations were deemed PPHs based on either their principal diagnosis being identified by the National Health Performance Framework [18] as a diabetes related PPH or identification by Davis et al [19] as associated with increased risk for people with diabetes. Principal diagnoses were captured using ICD-9-CM and Australian Modification ICD codes 10th revision (ICD-10-AM) codes included in the HMDS records (Appendix 1).

2.3.2. Variables for GP usage clustering

The goal of these cluster analyses was to identify patterns of GP service utilisation among people with diabetes. Candidate variables included in the cluster analyses were adapted from the customer relationship management framework proposed by Hughes (2005) [20] that capture both level of usage and strength of the relationship between patients acting as customers and GPs acting as primary care providers. Three main components suggested from the framework were Recency, Frequency and Monetary [20] which have been applied to healthcare data previously [21]. Since healthcare costs for Australia are covered by Medicare, with limited out of pocket payment from patients, the monetary component was not considered in our analyses. Greater recency and frequency are indicators of how well the relationship between patients with diabetes acting in the role of a customer and primary health care provider (GP) acting in the role of the service provider has been maintained [21].

In our study recency of GP usage consisted of three factors including: (i) the average time interval between access of health care service capturing the overall interaction between patients and GPs, (ii) the standard deviation from the average time interval capturing the extent of consistency in service utilisation, and (iii) the longest time interval between services capturing the extent that patients were out of coverage of primary care. Since the mean and standard deviation values may be driven by extreme values, two alternatives to the recency variable group were also considered in the cluster analyses including (A) mean time interval, mean absolute deviation from the mean and the longest time interval and (B) median time interval, median absolute deviation from median, and the longest time interval. The results of cluster analysis of the three groups of variables were compared in table 1. The time interval was determined between the date of a GP visit and the date of the previous health care service provided either from a GP or hospitalisation.

Frequency of GP usage was defined as the number of GP visits in a financial year. Those GP visits occurring within 14 days of the previous GP visit were counted as one GP usage to minimise over counting GP service utilisation, as those within 14 days of each other are likely to be associated with a single episode of care, for example where people may need to return to a GP to receive laboratory test results, rather than a subsequent discrete GP service as discussion with our GP experts.

All indicators were measured within financial years. However, a three-year look-back period was used, where necessary, to calculate the time interval between the first GP service in that year and the previous service. Three years was found to be the tie period that maximised capturing recency of GP utilisation for the cohort. Individuals having only one GP visit within a financial year were included in the cluster analysis if they had a previous health care service within the look-back period to enable the calculation of recency of GP usage.

2.3.3. Covariates

For this study, a number of individual characteristics were included to control for potential confounders in the relationship between GP usage cluster and PPHs. Demographic characteristics included were age group (18-44, 45-59, 60-74 and ≥75 years), gender, Indigenous status, quintile of the Census specific Socio-Economic Indexes for Areas (SEIFA) Index of Relative Socioeconomic

Disadvantage [22] and quintile of accessibility to services [23]. Diabetes complications were identified using ICD codes suggested by Young, Lin [24] and classified into four groups (0, 1, 2 and 3 or more complications) according to our previously published methods [17]. The number of comorbidities was summed from a list of comorbidities suggested by Holman et al. [25], excluding conditions classified as complications of diabetes. Regularity of GP visits was calculated as [1/(1+variance)] [9], where variance is a variance of the time interval between GP visits occurring within the financial year and classified into four quantiles. Number of specialist visits, and non-diabetes related hospitalisation were calculated within a financial year. Duration of diabetes was calculated in years.

2.4 Statistical analyses

Cluster analyses were conducted using different alternative combinations of recency and frequency of GP usage among those with at least one GP visit in a financial year. First, the values of the mean/median time interval, the standard deviation/absolute deviation of mean/median time intervals, longest time interval and frequency of GP visits were normalised by subtracting the minimum of each value and dividing that difference by the range of all values [12]. K-mean cluster analyses were then conducted on normalised values of recency and frequency of GP visits. The K-mean cluster approach was preferred as it is less susceptible to outliers in the data and is appropriate for use with large datasets [12]. The number of clusters was indicated using Calinski-Harabasz stopping rules for the options of 2 to 6 clusters, the large values of the Calinski-Harabasz pseudo-F index indicated distinct clustering [26]. Characteristics of final GP usage clusters were described using a box plot.

Both descriptive bivariate and multivariate analyses were performed. Descriptive analyses were used to summarise characteristics of participants among no GP usage and each GP usage cluster in the baseline year. The results were presented as the mean and standard deviation (SD) for continuous variables and percentage for categorical variables. Multivariate analyses were conducted using random-effects negative binomial regression model (NB) for panel data and zero-inflated negative binomial regression model (ZINB) with the inflated component contained in the intercept only. The Bayes Information Criterion (BIC) and Akaike Information Criterion (AIC) statistics were used to assess the fit of the model where NB with random effects was the preferred model compared to ZINB. We included Mundlak variables, defined as group-means of time-varying variables, to relax the assumption in the random-effects estimator that observed covariates were uncorrelated with the unobserved covariates [27, 28]. The group mean variables used were number of specialist visits and non-diabetes related hospitalisation. All analyses were conducted using STATA for Windows version 14.1.

3. Results

Clustering results

Table 1 presents summary results of cluster analyses with different groups of recency variables. The candidate group included mean time interval, mean absolute deviation from the mean, longest time to GP visit and frequency of GP visits; alternative A group included mean, standard deviation, the longest time interval and frequency; alternative B group included median, median absolute deviation from median, the longest time interval to GP visit and frequency of GP visits. Using the Calinski cluster stopping rule, all three groups identified three clusters. Compared with the candidate group, the other alternative groups had very high percentage of agreement in term of grouping subjects into a cluster with 99.3% in the alternative A group and 95.5% in the alternative B group. The candidate group also had highest Calinski F index value. Thus, the results of the candidate group were kept to present in this paper (Table 1). Figure 1 and Table 2 summarise the GP usage clusters from K-mean analyses. Three clusters were identified, including 1) moderate GP usage with mean time interval of approximately 10 months (296 days), standard deviation of about 4 months (115 days), the longest time interval of 14 months (404 days) and frequency of about 2 times a year; 2) high GP usage with mean time interval to a GP visits of 3 months (88 days), standard deviation of 1.5 months (48 days), the longest time interval of 5 months (147 days) and frequency of 3.7 times a year; and 3) very high usage with mean time interval of 1.5 months (40 days), deviation of 0.5 months (20 days), the longest time interval of 2 months (76 days) and frequency of visit approximately 7.8 times a year.

Characteristics of study population by GP usage cluster at the baseline year

Basic demographic and clinical characteristics of the study population are described in Table 3 by no GP usage and each GP usage cluster. The majority of the study population had high (n=17 077, 42.0%) and very high (n=15 858, 39.0%) GP usage, were aged 45 years or older (86.2%), and were more likely to be male (51%), non-indigenous (92.7%), moderate to least disadvantaged (51.6%), and living in areas with moderate to high accessibility to services (93.4%). Those with complications accounted for 43.3%% in the study population, higher in very high GP usage cluster (51.5%). The average number of comorbidities was 4.5 (SD3.6), the highest in those with very high GP usage cluster (mean 5.6; SD 3.5), followed by high GP usage cluster (mean 4.1, SD 3.5), no GP usage cluster (mean 3.5; SD 4.4) and moderate GP usage cluster (mean 3.2; SD 2.9). The average duration of diabetes was 6.4 (SD=4.3) years, similar duration across GP usage clusters and the no GP usage group. None and low regularity of GP visits were observed across GP usage clusters, except the very high GP usage cluster. High numbers of hospitalisations were observed among those with no GP usage (average of 3.4 admissions), followed by the very high GP usage cluster (0.8 admissions), high GP usage cluster (0.7 admissions) and moderate GP usage cluster (0.2 admissions).

Overall, the moderate GP usage cluster tended to be younger (25.1% aged 18-44 years, and 37.7% aged 45-60 years), male (62.6%), Indigenous (10.1%), live in less accessible areas (25.7%), compared with both the high and very GP usage cluster (Table 3). The moderate GP usage cluster was less likely to have complications (27.2%); had a lower number of comorbidities (3.2 (SD 2.9)); was less likely to have regular GP visits (20.5%) and had a lower number of hospitalisation (0.2; SD 0.8) compared with both high and very GP usage clusters The no GP usage group was quite comparable to other GP usage clusters in term of age, gender, complications and comorbidity distribution. However, the no GP usage group had a higher proportion of individuals who were indigenous (23.7%), in the highest disadvantage SEIFA quintiles (31.1%) and resided in very remote areas (20.1%).

Association between GP usage and the risk of hospitalisations

The preferred model was the panel negative binomial regression model based on information criterion (AIC and BIC). The results show that GP usage across all clusters had a protective effect against the risk of PPH in the following year after adjusting for all covariates. However, the greatest protective effect was observed for individuals in the moderate GP usage cluster (IRR=0.67 (95%CI: 0.62-0.71).

The average adjusted predictions indicate that on average 0.25 PPHs per year (95%CI: 0.24-0.27) can be expected for those in the moderate GP cluster; 0.26 per year (95%CI 0.259-0.27) for those in the high GP usage cluster and 0.29 per year (95%CI: 0.28-0.30) for those in the very high GP usage cluster, while those with no GP usage are estimated to have on average 0.38 hospitalisations per year (95%CI: 0.36-0.40) (Figure 2).

4. Discussion

This study aimed to reveal the latent pattern of GP contact using K-mean cluster analysis, a novel statistical technique, which overcomes many of the limitations associated with current studies by examining GP service use simultaneously across multiple attributes. Importantly we were able to include time intervals between service utilisations including average time interval, deviation of the time intervals and the longest time interval in assessing the patterns of GP service use which enhance the classification accuracy.

The rationale behind our exploration of incorporating multiple attributes to categorise GP use is our hypothesis that using frequency or regularity of GP contact alone may be too simplistic, since individuals that have the same number of visits or the same regularity in a year may have differences in the temporal distribution of visits. Shorter time intervals between services in combination with more regular provision may reflect “proactive care” and the strengthening of the relationship between patients and their GP. In turn, proactive care may allow the opportunity for continuous improvement in self-management skills and health literacy which may assist in the prevention and early treatment strategies in the primary care setting [6, 29]. The characterisation of GP utilisation based on multiple domains of GP use has not to our knowledge been previously reported and, we argue represents an advance on current single domain methods.

In our study, although the no GP usage group was comparable to other GP usage clusters in term of age and gender and disease severity, the group comprised higher proportion of disadvantage population (Indigenous status, highest disadvantage SEIFA and very remote). These findings highlight the existence of inequity in access of primary care for people with diabetes in particular sub-populations which have been previously reported in the literature

[30, 31] .The majority of individuals with diabetes were categorised in high or very high GP usage clusters. Those in high and very high GP usage clusters had high and very high recency and frequency of GP usage, respectively while those in the moderate GP usage cluster had both lower recency and frequency of contact. The clinical characteristics of each cluster differed significantly with those in the high or very high GP usage clusters more likely to have a higher number of complications and comorbidities compared with the moderate GP usage cluster. These results were in line with literature that showed higher health care service utilisation was observed among diabetes with multiple comorbidities and complications [32-34]. Thus, the multidimensional GP usage clusters identified in our study may be an indicator of patients' clinical characteristics which is driving their health care needs. This represents an improvement on other more simplistic measures such as frequency that do not correlate well with health outcomes [6, 10].

The literature does not show a consistent relationship between the level of primary health care and the risk of hospitalisation [7, 10]. While Comino et al. found that higher number of GP visits increased the risk of hospitalisation [6], other authors found an inverse relationship between the frequency of GP visits and hospitalisation [35]. Discordant results in the literature may be due to the complexity of the mechanism in the relationship between primary health care and hospitalisation, which may not be adequately captured by the number of GP visits [6]. Thus, use of a more complex measure of GP use, such as that developed in our study which incorporates several dimensions may be better suited to understand the risk of hospitalisation and help predict and contain the costs of healthcare for diabetes.

Our findings support the hypotheses that GP contact reduces the risk of hospitalisation. However, the effect was not linear for each additional level of GP usage, with the highest effect observed among those with moderate GP usage cluster. This may be explained by characteristics of GP usage cluster, those with moderate usage were likely to be younger, have fewer complications and comorbidities than those with high and very high GP usage. The results were also supported by the health demand model of Grossman where health is considered as a durable capital stock that depreciates with age and can be increased through investment in healthcare [36]. Thus, a finite lifetime increase in the depreciation rate of health may lead to an increase in demand for both preventive care and curative care [36, 37]. However, if primary health care can provide early treatment and prevention of illness, it would still be a substitute for hospital care in some instances [37].

Strengths and limitations of the study

The major strength of our study is that it was based on a large set of linked administrative data at the individual level that encompassed the whole-population and a comprehensive range of health care services. The linked whole-of-population data allowed us to assess changes in both exposure and outcomes at the individual level over the follow-up period. The panel data structure contained information on both within and between individual variations enabling us to control for the effect of unobserved covariates [38]. Our study also applied a novel advanced analytic approach, cluster analysis, and customer relationship management framework to reveal previously hidden patterns of primary health care utilisation. These approaches allowed us to examine primary health care utilisation across multiple attributes simultaneously, and thus characterise a measure of GP utilisation that may facilitate a better understanding of the influence of primary health care in reducing the risk of hospitalisations among people with diabetes.

Our study has some limitations. Comorbidity was accessed by a simple count of conditions which may not well capture actual health care needs although the measure is frequently used in the literature [6, 34, 39]. The analyses were limited to Australian citizens in one Australian State, due to the reliance on the WA Electoral Roll, and those with a previous diagnosis of diabetes captured by our data. Thus, the result may not be fully generalizable to all individuals living with diabetes, since the Electoral Roll is known to under-represent some groups such as Indigenous Australians and those aged under 21 years of age [40]. However, the use of longitudinal Electoral Roll data provided the ability to accurately capture person-time at risk, due to capturing movement in and out of the state [40]. Limiting the study to a single Australian State is unlikely to have significantly influenced the findings, since Australia has a single public health system, Medicare. Similarly, our reliance on linked administrative health data to identify those diagnosed with diabetes limited the study to those who have previously accessed health services pathognomonic of diabetes and thus people living with diabetes who have never accessed diabetes-related health services are not represented. Individuals not included in our data are likely to be the lower severity patients who are less likely to need hospital care. These limitations are common and well-known in administrative datasets and, because of the features of the excluded patients, are likely to have limited effect on our examination of the pattern of primary care utilisation and the relationship between the patterns of utilisation on the risk of hospitalisation in previously diagnosed diabetes.

Through combining both temporal factors with measures of frequency of use of GP services our study revealed a latent pattern of primary health care utilisation. Incorporation of multiple attributes that go beyond a simplistic frequency-based approach may better characterise the complex relationship between use of GP services and diabetes-related hospitalisation. The study has demonstrated the ability of cluster analyses to provide a systematic formalised approach for exploring complex patterns of health service utilisation in large administrative datasets. Application the cluster analysis approach to other chronic conditions would be useful for accurate understanding patterns of service utilisation. Future studies should further examine temporal factors in the provision of primary health care and evaluate what combination of time between visits, regularity and frequency of access to primary care would best improve health outcome and contain costs.

Acknowledgements

The authors wish to thank the staff at the Western Australian Data Linkage Branch and Commonwealth Department of Health and Aging for providing the linked data for this project.

Funding: This work was supported by the National Health and Medical Research Council, project grant APP1078345.

References

[1] Van Loenen T, Faber MJ, Westert GP, Van den Berg MJ. The impact of primary care organization on avoidable hospital admissions for diabetes in 23 countries. Scandinavian Journal of Primary Health Care. 2016;34:5-12.
[2] Tamayo T, Rosenbauer J, Wild SH, Spijkerman AM, Baan C, Forouhi NG, et al. Diabetes in Europe: an update. Diabetes Res Clin Pract. 2014;103:206-17.
[3] Centers for Disease Control and Prevention. 2014 National Diabetes Statistics Report. Centers for Disease Control and Prevention; 2014.
[4] AIHW. How many Australians have diabetes? Canberra Australian Institute of Health and Welfare; 2017.
[5] Erny-Albrecht K, Bywood P, Oliver-Baxter J. The role of primary care in primary and secondary prevention of diabetes and related complications: PHCRIS Policy Issue Review. Adelaide Primary Health Care Research & Information Service; 2015.
[6] Comino EJ, Islam MDF, Tran DT, Jorm L, Flack J, Jalaludin B, et al. Association of processes of primary care and hospitalisation for people with diabetes: A record linkage study. Diabetes Research and Clinical Practice. 2015;108:296-305.
[7] Gibson OR, Segal L, McDermott RA. A systematic review of evidence on the association between hospitalisation for chronic disease related ambulatory care sensitive conditions and primary health care resourcing. BMC Health Serv Res. 2013;13:336.
[8] van Loenen T, van den Berg MJ, Westert GP, Faber MJ. Organizational aspects of primary care related to avoidable hospitalization: a systematic review. Family practice. 2014;31:502-16.
[9] Einarsdottir K, Preen DB, Emery JD, Kelman C, Holman CD. Regular primary care lowers hospitalisation risk and mortality in seniors with chronic respiratory diseases. J Gen Intern Med. 2010;25:766-73.
[10] Wolters RJ, Braspenning JCC, Wensing M. Impact of primary care on hospital admission rates for diabetes patients: A systematic review. Diabetes Research and Clinical Practice. 2017;129:182-96.
[11] Gibson D, Moorin R, Preen D, Emery J, Holman DA. Effects of the Medicare enhanced primary care program on primary care physician contact in the population of older Western Australians with chronic diseases. Australian Health Review. 2011;35:334-40.
[12] Liao M, Li Y, Kianifard F, Obi E, Arcona S. Cluster analysis and its application to healthcare claims data: a study of end-stage renal disease patients who initiated hemodialysis. BMC Nephrology. 2016;17:25.
[13] Conry MC, Morgan K, Curry P, McGee H, Harrington J, Ward M, et al. The clustering of health behaviours in Ireland and their relationship with mental health, self-rated health and quality of life. BMC Public Health. 2011;11:692.
[14] Clatworthy J, Buick D, Hankins M, Weinman J, Horne R. The use and reporting of cluster analysis in health psychology: a review. British journal of health psychology. 2005;10:329-58.
[15] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America. 1998;95:14863-8.
[16] Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. Aust N Z J Public Health. 1999;23:453-9.
[17] Ha NT, Harris M, Robinson S, Preen D, Moorin R. Stratification strategy for evaluating the influence of diabetes complication severity index on the risk of hospitalization: a record linkage data in Western Australia. J Diabetes Complications. 2017;31:1175-80.
[18] Australian Institute of Health and Welfare. A set of performance indicators across the health and aged care system. Canberra: Australian Institute of Health and Welfare; 2008.
[19] Davis WA, Knuiman MW, Hendrie D, Davis TM. Determinants of diabetes-attributable non-blood glucose-lowering medication costs in type 2 diabetes: the Fremantle Diabetes Study. Diabetes Care. 2005;28:329-36.
[20] Hughes AM. Strategic Database Marketing: McGraw-Hill Pub. Co.; 2005.

[21] Lee EW. Data Mining Application in Customer Relationship Management for Hospital Inpatients. Healthcare Informatics Research. 2012;18:178-85.
[22] Australian Bureau of Statistics. Census of Population and Housing: Socio-Economic Indexes for Areas Canberra: Australian Bureau of Statistics; 2001,2006&2011.
[23] Australian Bureau of Statistics. Census of Population and Housing: Socio-Economic Indexes for Areas. Canberra: Australian Bureau of Statistics; 1981-2006.
[24] Young BA, Lin E, Von Korff M, Simon G, Ciechanowski P, Ludman EJ, et al. Diabetes Complications Severity Index and Risk of Mortality, Hospitalization, and Healthcare Utilization. The American journal of managed care. 2008;14:15-23.
[25] Holman CD, Preen DB, Baynham NJ, Finn JC, Semmens JB. A multipurpose comorbidity scoring system performed better than the Charlson index. J Clin Epidemiol. 2005;58:1006-14.
[26] Caliński T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics. 1974;3:1-27.
[27] Chamberlain G. Multivariate regression models for panel data. Journal of Econometrics. 1982;18:5-46.
[28] Mundlak Y. On the Pooling of Time Series and Cross Section Data. Econometrica. 1978;46:69-85.
[29] Nuti LA, Lawley M, Turkcan A, Tian Z, Zhang L, Chang K, et al. No-shows to primary care appointments: subsequent acute care utilization among diabetic patients. BMC Health Serv Res. 2012;12:304.
[30] Bywood P, Katterl R, Lunnay B. Disparities in primary health care utilisation: Who are the disadvantaged groups? How are they disadvantaged? What interventions work? . PHCRIS Policy Issue Review. Adelaide: Primary Health Care Research & Information Service; 2011.
[31] Davy C, Harfield S, McArthur A, Munn Z, Brown A. Access to primary health care services for Indigenous peoples: A framework synthesis. International Journal for Equity in Health. 2016;15:163.
[32] Struijs JN, Baan CA, Schellevis FG, Westert GP, van den Bos GA. Comorbidity in patients with diabetes mellitus: impact on medical health care utilization. BMC Health Services Research. 2006;6:84.
[33] van Oostrom SH, Picavet HSJ, de Bruin SR, Stirbu I, Korevaar JC, Schellevis FG, et al. Multimorbidity of chronic diseases and health care utilization in general practice. BMC Family Practice. 2014;15:61-.
[34] Fisher K, Griffith L, Gruneir A, Panjwani D, Gandhi S, Sheng L, et al. Comorbidity and its relationship with health service use and cost in community-living older adults with diabetes: A population-based study in Ontario, Canada. Diabetes Research and Clinical Practice. 2016;122:113-23.
[35] Booth GL, Hux JE. Relationship between avoidable hospitalizations for diabetes mellitus and income level. Archives of internal medicine. 2003;163:101-6.
[36] Grossman M. On the Concept of Health Capital and the Demand for Health. Journal of Political Economy. 1972;80:223-55.
[37] Tian W-H, Chen C-S, Liu T-C. The demand for preventive care services and its relationship with inpatient services. Health policy (Amsterdam, Netherlands). 2010;94:164-74.
[38] Hsiao C. Panel data analysis—advantages and challenges. TEST. 2007;16:1-22.
[39] Calderón-Larrañaga A, Abad-Díez JM, Gimeno-Feliu LA, Marta-Moreno J, González-Rubio F, Clerencia-Sierra M, et al. Global health care use by patients with type-2 diabetes: Does the type of comorbidity matter? European Journal of Internal Medicine. 2015;26:203-10.
[40] Youens D, Preed D, Harris M, Moorin R. The importance of historical address information in longitudinal studies using administrative health data. International Journal of Epidemiology. 2017;In press

Results: Figures

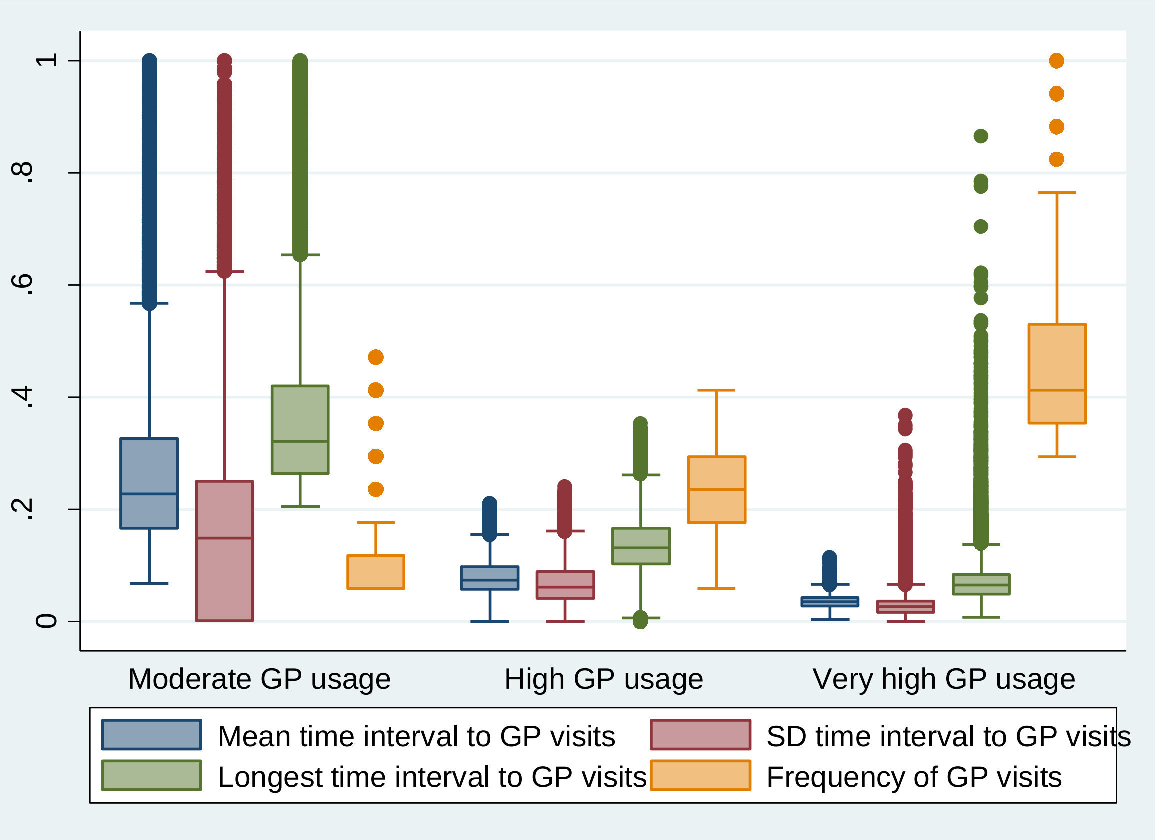

Figure 1. GP usage by clusters

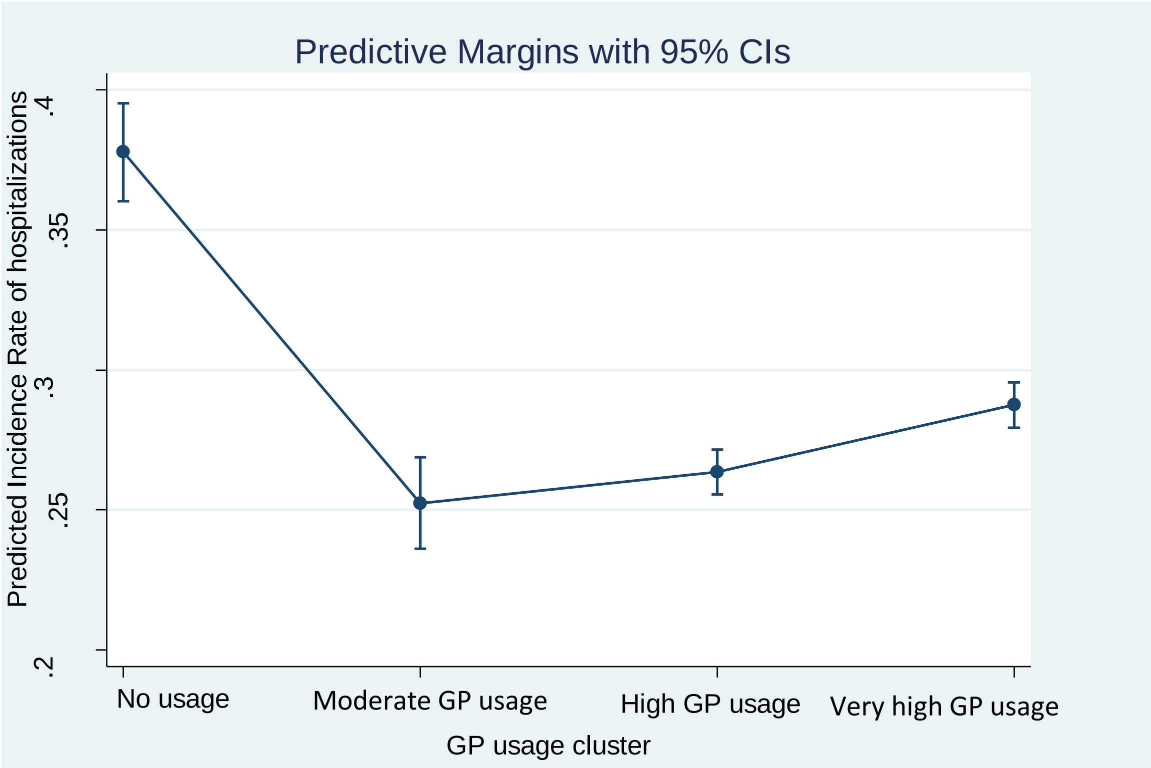

Figure 2. Predictive margins the incident rate of diabetes related PPH

Results: tables

Table 1 Cluster analysis outputs with different groups of recency variables

	Group of Indicators used in K-mean cluster		
	Candidate group	**Alternative A group**	**Alternative B group**
Mean	✓	✓	
Median			✓
Mean absolute deviation from the mean		✓	
Median absolute deviation from median			✓
standard deviation	✓		
The longest time to GP visit	✓	✓	✓
Frequency of GP visits	✓	✓	✓
Cluster stopping (Cali´nski rule)	133805	132616	129095
Number of clusters	3	3	3
% of agreement vs. group 1 (Kappa values)	-	99.3%	95.5%

Table 2 GP usage clusters summary

Clusters	**Mean** (days)	**SD** (days)	**The longest** (days)	**Frequency of GP visits**
Moderate usage				
Min	75	0	225	1
Mean	296.8412	115.0688	404.0527	1.919529
Max	1093	744.5834	1095	8
High usage				
Min	1	0	1	1
Mean	88.19658	48.81665	147.0608	3.716618
Max	230	178.1975	387	7
Very high usage				
Min	5.2	0	9	5
Mean	39.71341	20.81995	76.12468	7.819856
Max	124.75	273.0432	947	17

Table 3 Characteristics of study population by GP usage cluster

Characteristics	**No GP usage** (N, (%))	**Moderate GP usage** (N, (%))	**High GP usage** (N, (%))	**Very high GP usage** (N, (%))
N (%)	4 198 (10.3)	3 492 (8.6)	17 077 (42.0)	15 858 (39.0)
Age group (years)				
18-44	781 (18.6)	877 (25.1)	2,668 (15.6)	1178 (7.4)
45-59	1059 (25.2)	1316 (37.7)	5,649 (33.1)	3543 (22.3)
60-74	1183 (28.2)	1,016 (29.1)	6,655 (38.9)	7465 (47.1)
≥75	1175 (28.0)	283 (8.1)	2,105 (12.3)	3672 (23.2)
Gender				
Female	1679 (40.0)	1307 (37.4)	7,912(46.3)	9002 (56.8)
Male	2519 (60.0)	2185 (62.6)	9,165 (53.7)	6856 (43.2)
Indigenous status				
No	3084 (76.3)	2911 (89.8)	15,197 (93.8)	14978 (96.5)
Yes	961 (23.7)	329 (10.1)	1,003 (6.2)	549 (3.5)
SEIFA				
Highest Disadvantage	1285 (31.4)	631 (18.4)	23,240 (19.2)	3435 (21.8)
High disadvantaged	1037 (25.3)	918 (26.7)	4,797 (28.4)	4558 (28.9)
Moderate disadvantage	573 (14.0)	593 (17.3)	2,381 (14.1)	2185(13.8)
Less disadvantage	544 (13.5)	561 (16.3)	2,754 (16.3)	2416(15.3)
Least disadvantage	645 (15.7)	728 (21.2)	3,691 (21.8)	3158 (20.0)
Accessibility				
Very remote	825 (20.1)	251 (7.3)	611 (3.6)	79 (1.2)
Remote	172 (4.0)	90 (2.6)	355 (2.1)	184 (1.1)
Moderate	268 (6.5)	265 (7.7)	946 (5.6)	659 (4.2)
Accessible	210 (5.1)	273 (7.9)	1,027 (6.1)	695 (4.4)
Highly accessible	2619 (63.9)	2,552 (74.3)	13,926 (82.6)	14036 (89.1)
Complication severity level				
No complication	1957 (46.6)	2,543 (72.8)	10,845 (63.5)	7694 (48.5)
1 complication	746 (17.8)	385 (11.0)	2,372 (13.9)	2638 (16.6)
2 complications	577 (13.7)	322 (9.2)	1,804 (10.5)	2266 (14.3)
3+ complications	918 (21.9)	242 (6.9)	2,056 (12.0)	3260 (20.6)
Number of comorbidity				
Mean (SD)	3.5 (4.4)	3.2 (2.9)	4.1 (3.4)	5.6 (3.5)
Duration of diabetes (years)				
Mean (SD);	6.7 (4.4)	6.3 (4.2)	6.1 (4.2)	6.5 (4.4)
Regularity quantiles				
No regularity	4,198 (100.0)	2,776(79.5)	3,315 (19.4)	0
Quantile 1		716 (20.5)	6,684 (39.1)	287 (1.8)
Quantile 2			4,719 (27.6)	2,972 (18.7)
Quantile 3			1,497 (8.8)	5,917 (37.3)
Quantile 4			862 (5.0)	6,682 (42.1)
Diabetes related PPH				
Mean (SD)	2.5 (17.5)	0.07 (0.38)	0.25 (2.6)	0.25 (1.02)

Table 4 Association of GP usage pattern and potentially preventable hospitalisation with and without adjustment for other covariates (results from random effects negative binomial regression)

	Multivariate NB		Adjusted multivariate NB		ZINB	
	IRR	(95%CI)	IRR	(95%CI)	IRR	(95%CI)
GP cluster usage						
No usage	1	(1; 1)	1	(1; 1)	1	(1; 1)
Moderate usage	0.62***	(0.57; 0.66)	0.67***	(0.62; 0.72)	0.41***	(0.33; 0.50)
High usage	0.67***	(0.64; 0.71)	0.70***	(0.66; 0.73)	0.40***	(0.35; 0.46)
Very high usage	0.76***	(0.72; 0.79)	0.76***	(0.72; 0.80)	0.39***	(0.34; 0.45)
Gender						
Males vs. females	1.06***	(1.03; 1.10)	1.07***	(1.04; 1.11)	1.24***	(1.13; 1.36)
Age (years)						
18/44	1	(1; 1)	1	(1; 1)	1	(1; 1)
45/59	1.20***	(1.12; 1.28)	1.21***	(1.14; 1.29)	1.10	(0.91; 1.32)
60/74	1.74***	(1.64; 1.86)	1.73***	(1.62; 1.84)	1.44***	(1.20; 1.73)
75+	2.30***	(2.15; 2.46)	2.31***	(2.16; 2.47)	1.42***	(1.18; 1.71)
Indigenous status						
Yes vs. No	1.47***	(1.37; 1.59)	1.50***	(1.39; 1.61)	2.18***	(1.79; 2.67)
SEIFA						
Highest Disadvantage	1	(1; 1)	1	(1; 1)	1	(1; 1)
High disadvantaged	0.95*	(0.91; 1.00)	0.95*	(0.91; 0.99)	0.96	(0.84; 1.09)
Moderate disadvantage	0.95	(0.90; 1.00)	0.94*	(0.89; 0.99)	0.86*	(0.76; 0.97)
Less disadvantage	0.98	(0.93; 1.03)	0.97	(0.92; 1.02)	0.95	(0.82; 1.10)
Least disadvantage	0.93**	(0.88; 0.98)	0.90***	(0.86; 0.95)	0.94	(0.81; 1.09)
Accessibility						
Very remote	1	(1; 1)	1	(1; 1)	1	(1; 1)
Remote	1.00	(0.87; 1.13)	1.00	(0.88; 1.13)	0.76*	(0.59; 0.96)
Moderate	0.97	(0.88; 1.08)	0.98	(0.88; 1.08)	0.84	(0.64; 1.09)
Accessible	0.92	(0.83; 1.03)	0.92	(0.82; 1.02)	0.73*	(0.57; 0.95)
Highly accessible	0.89*	(0.82; 0.98)	0.90*	(0.83; 0.99)	0.97	(0.78; 1.21)
Duration of diabetes (years)	1.03***	(1.03; 1.04)	1.04***	(1.03; 1.04)	1.05***	(1.04; 1.06)
Complication severity level						
No complication	1	(1; 1)	1	(1; 1)	1	(1; 1)
1 complication	1.33***	(1.27; 1.40)	1.27***	(1.21; 1.33)	1.05	(0.94; 1.18)
2 complications	1.68***	(1.60; 1.77)	1.58***	(1.51; 1.66)	1.57***	(1.37; 1.80)
3+ complications	2.12***	(2.02; 2.22)	1.90***	(1.81; 2.00)	2.72***	(2.34; 3.15)
Number of comorbidities	1.07***	(1.06; 1.07)	1.04***	(1.03; 1.04)	1.07***	(1.05; 1.09)
Number of specialist services	1.01***	(1.01; 1.01)	0.99***	(0.98; 0.99)	0.97***	(0.96; 0.98)
Non-diabetes related hospitalisation	1.05***	(1.02; 1.09)	0.99	(0.96; 1.02)	0.99	(0.90; 1.10)
Diabetes related hospitalisation lag1			1.36***	(1.31; 1.40)	4.65***	(3.94; 5.49)
Diabetes related hospitalisation baseline			1.11***	(1.07; 1.14)	1.14*	(1.02; 1.27)
Group mean number of specialist visits			1.04***	(1.04; 1.05)	1.06***	(1.05; 1.08)
Group mean non-diabetes related hospitalisations			1.60***	(1.50; 1.72)	1.89***	(1.52; 2.36)
AIC	191782.6		190686.5		202182.5	
BIC	192075.6		191019.9		202515.9	

Exponentiated coefficients

="* p<0.05 ** p<0.01 *** p<0.001"

Table 5 Margin incident rate of diabetes related PPH

GP usage	Incidence rate	95% CI	
No GP usage	0.38	0.36	0.40
Moderate GP usage	0.25	0.24	0.27
High GP usage	0.26	0.26	0.27
Very high GP usage	0.29	0.28	0.30