The AAAI-17 Joint Workshop on
Health Intelligence
WS-17-09

Multidimensional and Longitudinal Indicators in Population Health

Guido Antonio Powell,[a,*] Yu T Luo,[a, b] Aman Verma,[a], David A Stephens,[b] David L Buckeridge[a]

[a] Surveillance Lab, McGill Clinical and Health Informatics, [b] Mathematics and Statistics, McGill University, Montreal, Quebec, Canada

* Corresponding author: guido.powell@mail.mcgill.ca

Abstract

Within population health information systems, indicators are commonly presented as independent, cross-sectional measures, neglecting the multivariate, longitudinal nature of disease progression and health care use. We use administrative claims data for patients with a previous diagnosis of chronic obstructive pulmonary disease in Montreal, Canada to explore two approaches to facilitating the discovery and interpretation of patterns across indicators and over time. The first approach identifies regional clusters based on patterns across four health service indicators. Our second approach uses a hidden Markov model to analyze individual-level trajectories based on the same four indicators. Both approaches offer additional insights, such as a dual interpretation of low use of general practitioner services. These approaches to the analysis and visualization of health indicators can provide a foundation for information displays that will help decision makers identify areas of concern, predict future disease burden, and implement appropriate policies.

Introduction

Decision makers within health systems depend on indicators to guide their actions. The increasing adoption of information systems has done much to improve the availability of indicators, but many existing systems present indicators independently, providing little contextual information. This context is often needed to interpret indicators. Take, for example, an indicator of health service use, such as the proportion of patients visiting a general practitioner (GP) in the past year. Such an indicator is typically presented in a stratified manner to facilitate comparison across regions, by patient demographics (age, sex), or over time as a series of cross-sectional estimates. Although many patterns can be identified through indicators stratified in this manner, making sense of the patterns often requires additional information, such as the specialist, or emergency department consultation rates in the same regions.

In most population health information systems, the basic methods for data manipulation do not support the efficient exploration of hypotheses that span multiple regional indicators or refer to temporal patterns within individuals. In attempting to explore these types of questions, a user can become overwhelmed in trying to understand how individual cross-sectional estimates, such as yearly and regional distributions of GP visits, relate to similarly stratified estimates of other relevant indicators. For example, the patterns of GP use may be related to patterns of specialist services, emergency department visits, and hospitalizations. Further complicating the interpretation of these associations across multiple health regions is the heterogeneity of associations, which is exponentially greater at the individual level, when the temporal ordering or timing of health care use is considered. New strategies for the analysis and visualization of health indicators are clearly needed to support the discovery and interpretation of multidimensional and longitudinal patterns in health indicators.

Our objective was to explore two approaches for analyzing and visualizing indicators of health service use that take into account their multivariate nature, their inherent heterogeneity, and their complex evolution over the temporal progression of an individual disease course. The first approach offers a visualization of multiple indicators of health service use. The approach deals with the complexity of identifying patterns both within and across regions by clustering the data into latent regional profiles. The second approach provides a model of how patients transition through different latent patterns of health service use, based on the same set of indicators, to illustrate variation in patient trajectories over time.

Ultimately, these approaches should help decision-makers to identify and interpret patterns within the health system and facilitate data-driven and evidence-based decision making.

Methods

This research, to develop new approaches to analyzing and visualizing population health indicators, was conducted as part of the Population Health Record (PopHR) project (Shaban-Nejad et al. 2016). PopHR is a semantic web application that helps measure and monitor population health and health system performance, integrating administrative data on health service use with data on behavioral health determinants from surveys and other sources. Health service use data were obtained from the public health insurance provider in Quebec, Canada. The current implementation of PopHR uses an open cohort of approximately 1 million people created by taking a 25% random sample of the population of the Montreal census metropolitan area (CMA). Analyses were performed using data for 81,580 individuals over 35, who were identified as having chronic obstructive pulmonary disease (COPD), based on diagnostic codes for hospitalization or medical billing (e.g. ICD9 491x, 492x, 496x; ICD10 J41-J44).

The indicators of health service used in these analyses were based on visits with a GP or visits with a specialist (respirologist or internist) in an outpatient setting, visits to an emergency department, and all-cause hospitalizations. For the first analysis, each indicator was calculated as the proportion of all COPD patients who used the service in the past year. These estimates were aggregated by the regions associated with the 57 local community service centres (CLSC) within the Montreal CMA, standardized for age and sex distributions of the region, and averaged over the period of 2012-2014. The second analysis modeled yearly counts of each service use from 1998-2014, for each of the 81,580 COPD patients.

The first analysis performed a hierarchical cluster analysis on a set of Euclidean distances between CLSCs based on their four indicator values (proportion of COPD patients visiting a GP, a specialist, an ED, or hospitalized). Each indicator value was centered on the mean across regions. Clustering was performed using the complete linkage (hclust in R). Visualization of intergroup dissimilarity of clusters was used to decide on the number of clusters.

The second analysis fit a hidden Markov model to individual-level aggregates of yearly health service use, using the depmix command "depmixS4" package from R (Visser and Speekenbrink 2010). The model assumed a Markov process between four discrete latent states of health service use, each defined as five dimensions of service uses (GPs, specialists, ED, hospitalization, or no use). The probability of each service use within a state and probabilities of moving from one state to another were parameters learned by the model.

Results

The unscaled mean proportions for the four indicators across the 57 regions were 0.57 [95% CI: 0.561-0.77] for GP visits, 0.31 [0.311-0.314] for specialist visits, 0.30 [0.288-0.302] for ED visits, and 0.11 [0.112-0.114] for

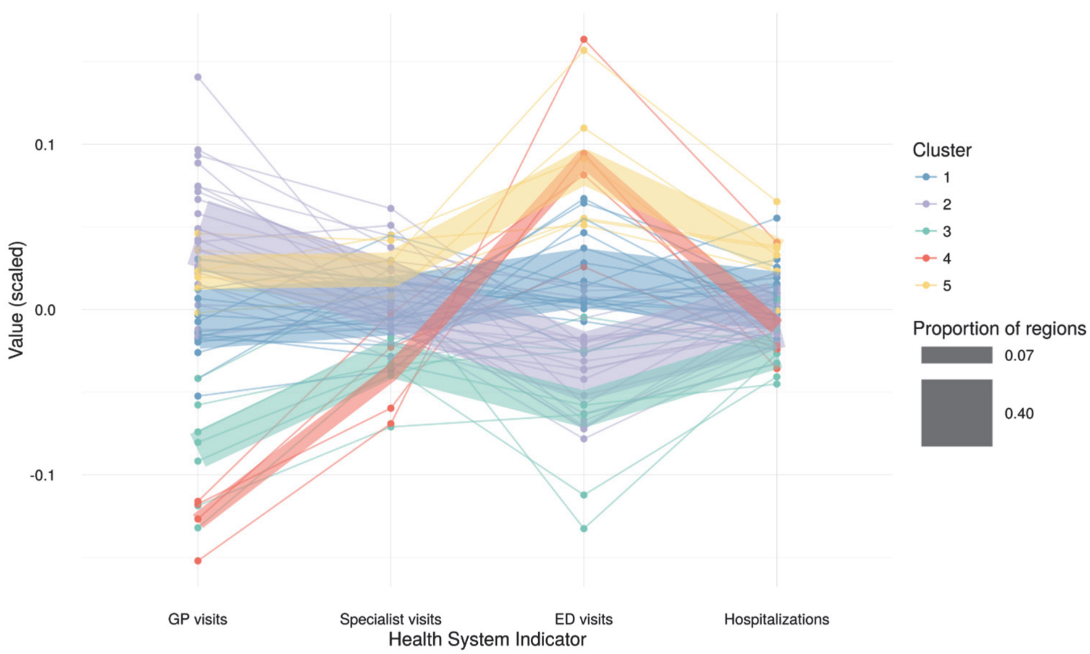

Figure 1. Clusters of 57 health regions in Montreal, Quebec based on the patterns across health system indicators within each region. Color represents the five clusters and the bolder lines, weighted by the size of the cluster, follow the cluster centroids over the indicators

hospitalizations. The first cluster analysis using scaled values of each indicator produced visually informative results, identifying five subsets of regions with different indicator patterns (Figure 1).

Most regions were fit into two clusters where the use of all services was close to average (17 in cluster 1 and 23 in cluster 2). Of 11 regions characterized by low GP use, clustering disambiguated seven 'low overall' regions (cluster 3) from the four 'high ED' regions (cluster 4). Six regions were characterized as 'high overall' (cluster 5).

In the second analysis, the HMM identified four latent health service use states with expected probabilities of service use shown in Table 1. State 1 was characterized by higher specialist visits, state 2 by higher GP visits, state 3 by higher ED visits and hospitalizations, and state 4 by overall low service use.

Patients were most likely to start in state 4 (low overall use, 0.64), compared to a 0.18 probability of starting in state 1 (high specialist), 0.08 for state 2 (high GP), and 0.11 for state 3 (high ED and high hospital). On any given year, patients mostly remained in the same state rather than transitioning. Otherwise, patients were most likely to transition to state 4 (low overall). State 3 (high ED and high hospitalization) was the least stable state (i.e. where patients were least likely to remain).

State	GP Visits	Specialist visits	ED visits	Hospitalizations
1	.008	**.010**	.002	.001
2	**.031**	.002	.004	.001
3	.009	.003	**.017**	**.005**
4	.008	.000	.001	.001

Table 1. Probabilities of observable service use characterizing four latent health service use states in COPD

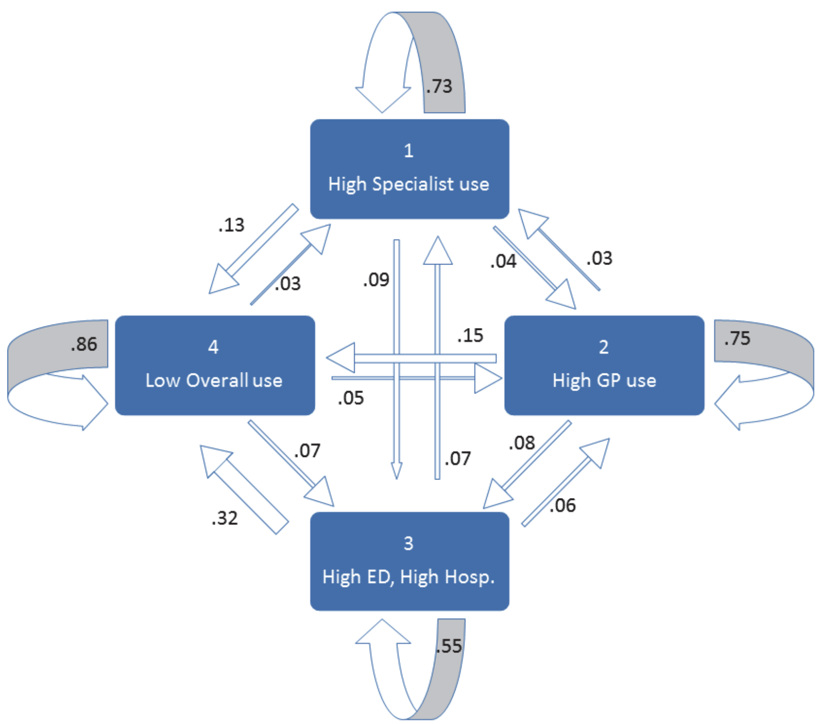

Figure 2. Hidden Markov model of COPD service use. Blue boxes represent hidden service use states , while arrow direction and width represent transitions between states and their probabilities

Discussion

The analysis and visualization methods explored in this study demonstrate the potential for decision-makers to gain insights into health system dynamics through the exploration of regional estimates across multiple indicators as well as by considering patient trajectories of health care use over time. Making such methods available within population health information systems would aid users in identifying and examining regions or patterns of patient trajectories that may benefit from further analysis and intervention. For example, results from the first approach can give meaning to several regions' rates, where estimates of GP use is relatively low. Decision makers can focus interventions on regions where low GP use is associated with corresponding high ED visits, as these estimates are likely more representative of inappropriate care.

The second approach we explored incorporated additional data, summarizing not only the patterns of use of various health service, but also the patient pathways between different types of health services. The tendency for COPD patients to predominantly be in a state of low service use is an insight not easily obtained from single, cross-sectional indicators. Nor could a typical indicator have demonstrated that patients are most likely to remain in the same state over time. The model also allowed for a refined understanding of low GP use, shedding more light on the dynamic relationship between states of low use and states of high use of emergency care and hospitalizations.

Examples of related work in the literature include evaluations of emergency department performance, where cluster analyses explained performance differences in similar hospitals based on physician availability (Anderson et al. 2016). Research has also used Markov models and health service data to model disease progression in various patient groups, including COPD (Wang, Sontag, and Wang 2014), (Jensen et al. 2014). Our study has demonstrated the potential value of similar approaches in the context of population health information systems. The set of indicators we considered could be expanded considerably to include measures of concepts such of comorbid diagnoses, drug dispensations, mortality, and socioeconomic status. However, restricting analyses to four indicators belies the larger complexity in modeling health systems. Expanding these analysis even further presents challenges to the integration in usable interfaces and in guiding decision making.

Analysis of the increasing volume of population health data has the potential to enhance the efficiency and effectiveness of health systems and to ultimately improve population health. However, methods are needed for making sense of indicators within information systems, through multivariate and longitudinal analysis and visualization. Making better use of the available data would allow more comprehensive characterization of health systems and should also improve predictions of disease burden. Such analyses can also eventually guide tailored decisions about policy changes and interventions to better address the current and future needs of specific regions and patient groups – bringing a "precision" perspective to population health (Collins and Varmus 2015).

References

Anderson, D.; Pimentel, L.; Golden, B.; Wasil, E.; and Hirshon, J. M. 2016. Drivers of ED efficiency: a statistical and cluster analysis of volume, staffing, and operations, *Am. J. Emerg. Med.,* vol. 34, no. 2, pp. 155–161, Feb.

Collins, F. S. and H. Varmus, H. 2015. A New Initiative on Precision Medicine, *N. Engl. J. Med.,* vol. 372, no. 9, pp. 793–795, Feb.

Jensen, A. B.; Moseley, P. L.; Oprea, T. I.; Ellesoe, S. G.; Eriksson, R.; Schmock, H.; Jensen, P. B.; Jensen, L. J.; and Brunak, S. 2014. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients," N*at. Commun.,* vol. 5, Jun.

Shaban-Nejad, A, Lavigne, M.; Okhmatovskaia, A.; and Buckeridge, D. L. 2016. PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data, *Ann. N. Y. Acad. Sci.*, Oct.

Visser, I.; Speekenbrink M. 2010. depmixS4: An R Package for Hidden Markov Models. *Journal of Statistical Software*. vol. 36, no. 7, pp. 1-21.

Wang, X.; Sontag, D.; and Wang, F. 2014. Unsupervised Learning of Disease Progression Models. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, pp. 85–94.