COMMONWEALTH OF MASSACHUSETTS

HEALTH POLICY COMMISSION

TECHNICAL APPENDIX B6
HIGH-COST PATIENTS

ADDENDUM TO 2014 COST TRENDS REPORT

Table of Contents

1 Summary

This section describes the Health Policy Commission's (HPC) approach for examining the small subgroup of the population that represents a large proportion of medical expenditures in both commercial and Medicare markets using logistic regression and cluster analysis.

2 Logit regression analysis

This first section of technical appendix lays out HPC's methodology for examining (a) persistently high-cost patients (HCP) in total spending, (b) persistent HCP in Emergency Department (ED) spending, (c) intermittently HCP in total spending, and (d) intermittently HCP in ED spending. This is done for both commercially insured and Medicare patients separately.

The purpose of this research is to undertake a systematic study of the demographic, clinical, and regional predictors and characteristics of these patients.

2.1 Sample

We used Massachusetts' All Payer Claims Database for the calendar years 2010-2012 for our analysis. For technical specifications on MA's APCD please refer to:
http://www.mass.gov/anf/docs/hpc/apcd-almanac-technical-notes.pdf

Our sample included data from the three major commercial payers, Blue Cross Blue Shield, Harvard Pilgrim Health Plan, and Tufts Health Plan. Enrollees with missing age and gender were dropped from our dataset. Members with negative expenditures and negative income levels were not included in our study. Expenditures do not capture pharmacy costs or payments outside the claims system.

The sample was screened by the following criteria: 12 months of enrollment in all three data years, spending greater than or equal to zero dollars in total claims, and spending greater than or equal to zero dollars for each category of service. Patients with missing age, gender, diagnosis, and/or cost data in any given year were also dropped from the sample. Patients with the coding errors of neonatal diagnoses (0.8% of commercial data and 0.1% of Medicare data) and Medicare patients with pregnancy diagnoses (0.1%) were also removed from the study. Finally, age limits were set at between 19-64 years old in 2010 for commercial populations and 65 years old or older in 2010 for Medicare population. The final sample size of commercially insured adults was 964,525 and 476,711 for Medicare.

2.2 Definitions of high-cost patients

Persistently high-cost patients in total spending are defined as patients who are in the 95th percentile and above in total medical claims expenditures for three consecutive years (2010-2012). The dollar thresholds to be classified as a total HCP averaged $16,449 and $43,687 in total spending for commercial and Medicare populations respectively. Similarly persistently

HCP in ED spending are those patients who are in the top five percent of ED expenditures for three consecutive years.

Intermittently high-cost patients in total spending are defined as patients who are in the 95th percentile and above in total medical claims expenditures in Year 1 (2010) and Year 3 (2012), but not in Year 2 (2011). Similarly intermittently HCP in ED spending are those patients who are in the top five percent of ED expenditures in Years 1 and 3, but not in Year 2.

2.3 Definitions of categories of service

Categories of Service (COS) were classified according to Optum's Symmetry version 8.2. Specifically, claims were sorted into categories of service using the revenue codes and procedure codes reported on a claim. All lines on a claim are assigned to a single category. This study used only the category of "emergency department" spending. It is possible for a given patient to have spending in all or none of these categories.

2.4 Definitions of medical conditions

Medical conditions were grouped into 34 binary categories according to Optum's Symmetry Episode Risk Groups (ERG) risk adjustment software version 8.2. The ERG software evaluated diagnosis codes in 2010 medical claims to identify the conditions present for each patient. These were then grouped up to 34 medical conditions. It is possible for a given individual to have all or none of these 34 conditions. To briefly define a few medical conditions in particular:

Acute conditions- low-cost conditions across the disease categories of infectious disease, endocrinology, psychiatry, neurology, ENT, pulmonology, gastroenterology, nephrology, urology, dermatology, and orthopedics.

Blood conditions- neoplastic blood conditions (leukemia), hemophilia, sickle cell disease, along with other high cost hematological conditions.

2.5 Definitions of chronic conditions

From the 34 conditions, 17 chronic conditions of interest were identified: arthritis, asthma, child psychiatric disorders, blood conditions, diabetes, epilepsy, glaucoma, cardiology, HIV/AIDS, hyperlipidemia, hypertension, mental health, multiple sclerosis (MS) and ALS (amyotrophic lateral sclerosis), psychiatric disorders, renal failure, mood disorders, and substance abuse. Five of these conditions were then grouped into a behavioral health category (child psychiatric disorders, mental health, psychiatric disorders, mood disorders, and substance abuse) and the remainder as chronic medical conditions.

2.6 Definitions of catastrophic medical conditions

There is no standardized definition for high-cost "catastrophic" illnesses. Reviewing the literature on high cost illnesses, several are mentioned as particularly catastrophic from a cost perspective: cancer (4), HIV/AIDS (4), transplants (3), stroke (3), myocardial infarction/cardiovascular surgery (3), neonatal conditions (3), renal failure (2), sepsis (2), orthopedic procedures (1), hemophilia (1), severe trauma (1), pulmonary insufficiency/respiratory failure (1), intracranial abscess (1), peritonitis (1), liver disease (1), selected infections (1), aplastic anemia (1), and quadriplegia/ paraplegia (1). [1,2,3,4] The numbers in parenthesis represent the number of times a disease is mentioned as a catastrophic illness out of the reviewed articles.

The current study uses the most commonly cited catastrophic conditions applied within the structure of the HPC's data. This study's definition of "catastrophic" illnesses therefore includes: cancers, HIV/AIDS, transplants, coronary artery disease (CAD)/heart failure, and renal failure.

2.7 Regional and demographic variables

A patient's zip code was used to group him/her into a Hospital Service Area (HSA). HSAs are local health care markets for hospital care as defined by the Dartmouth Atlas Project.[5] Indicators for each HSA were used as controls for hospital/provider mix in the regression analyses, but findings are not discussed in this report. The average spending HSA was calculated for total and ED spending respectively to be used as bases for comparison in logistic regression analyses. For commercial patients, the base region was set at Norfolk for total spending, and Woonsocket for ED spending. For Medicare patients, the base region was set at Cambridge for total spending, and Fitchburg for ED spending.

Using 2010 as the base year, nine age categories were created: 0-18, 19-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85 plus. The lowest spending category was used as a comparison group. In the commercial population, the third age bracket (25-34 years old) was used as the comparison group in the logistic regression analyses. In the Medicare group, the seventh age bracket (65-74 years old) was used as the comparison group in the logistic regression.

The APCD does not include individual patient or family income data. Instead patients were assigned to an income level based on the 2011 median household incomes of their zip code as reported by the U.S. Census American Community Survey. Patients were then classified by quintiles of community income. The first quintile (Q1) is defined as the lowest income. The top (Q5) is defined at the highest income communities. In the commercial population, the range of incomes for each quintile was as follows: Q1 $2,500-26,714; Q2 $26,744-30,694; Q3 $30,708-34,097; Q4 $34,240-40,797; Q5 $40,865-158,153. In the Medicare population, the range of incomes for each quintile was as follows: Q1 $2,500-28,200; Q2 $28,235-32,228; Q3 $32,292-36,771; Q4 $36,786-44,000; Q5 $44,052-158,153. The lowest income (Q1) quintile was set as the base for comparison in the logistic regression analyses.

2.8 Multivariate analysis

The following patient characteristics were selected as independent predictor variables: gender, age group, age group and gender interaction terms, the 34 medical conditions present in 2010 (and separately an indicator for the number of chronic diseases present in 2010), location of residence (HSA), and community income. Using Stata software, logistic regressions were used to predict the binary outcome variables of: (a) persistent total HCP, (b) persistent ED HCP, (c) intermittent total HCP or (d) intermittent ED HCP against the independent variables. Odds ratios were then computed and tested for significance.

2.9 Results

Tables B6:1-B6:32 (published as a separate document in Excel format) show descriptive statistics and regression results related to commercial and Medicare patients with persistently high costs, intermittently high-costs, persistently high ED costs, and intermittently high ED costs.

3 Cluster analysis

This second section of technical appendix explains the cluster analysis methodology used to identify and describe the five clusters of APCD commercially-insured persistently high cost patients (HCP) highlighted in HPC's 2014 Cost Trends Report.

Broadly speaking, cluster analysis groups patients such that those in the same cluster are relatively similar along the dimensions explored. HPC's analysis was based on similarity of emergency department spending, inpatient spending, clinical conditions, and demographic characteristics between patients.

Our analysis found 17 clusters of which 5 were chosen as particularly relevant for clinical intervention or management.

3.1 Sample

The following filtering steps were performed on the All Payer Claims Database (APCD) to select a set of patients for analysis.

- Remove patients with missing age, gender, diagnosis, and/or cost data in any given year (2010-2012).
- Remove patients who had less than 12 months of enrollment in any of the three years (2010-2012).
- Remove patients whose total spending or spending in a category of service was less than $0 in 2010-2012.
- Remove patients aged 65 years or older in 2011.

- Limit to persistently high cost patients. Randomly select 2,564 patients. This is done to accommodate computing run time limitations.

3.2 Patient characteristics

In this study, four characteristics were used to compare the similarity between patients:

1. *Age*: an integer ranging from 0 to 64 using 2011 as the base year.
2. *Emergency Department Spending*: an integer equal to the dollar amount spent by a patient on emergency services in 2011.
3. *Inpatient Spending*: an integer equal to the dollar amount spent by a patient on inpatient services in 2011.
4. *Medical Condition*: a 237-element binary vector based on the ERG categories established by Optum's Symmetry Episode Risk Groups (ERG) risk adjustment software version 8.2. For example, if the 28th element in a patient's medical condition vector equals 1 and all the other 236 elements equal 0, then that patient has a lower cost hematology condition, and only that. Only clinical data from 2011 was used.

3.3 Creating clusters.

Patients were compared to each other using a kernel function, and Ward's minimum variance method was chosen as a partitioning criterion in order to obtain homogeneous clusters having minimum inter-cluster variance and maximum between-cluster variance.[1] More details on the HPC's clustering methods are available upon request.

We used odds ratios to identify distinct demographic, spending and clinical factors associated with each cluster relative to the broader population of persistently high-cost patients. Table B6:33 (published as a separate document in Excel format) shows spending by category of service for five clusters of particular interest.

REFERENCES

[1] Wyszewianski, L. Financially catastrophic and high-cost cases: definitions, distinctions, and their implications for policy formulation. Inq. J. Med. Care Organ. Provis. Financ. 23, 382–394 (1986).
[2] Joynt, K. E., Gawande, A. A., Orav, E. J. & Jha, A. K. Contribution of preventable acute care spending to total spending for high-cost Medicare patients. JAMA 309, 2572–2578 (2013).

[1] The "ward.D2" option in hclust function in the default {stats} package in R Version 3.1.2 was utilized in this study to cluster patients using the agglomerative clustering strategy of Ward's method.

[3] Fitch, Kate and Bruce Pyenson. 'Benefit Designs for High Cost Medical Conditions.' Milliman Research Report. Milliman Inc, New York, NY (April 2011) at <http://publications.milliman.com/research/health-rr/pdfs/benefit-designs-high-cost.pdf>.

[4] Maguire, A. M. et al. 'Carving out' conditions from global capitation rates: protecting high-cost patients, physicians, and health plans in a managed care environment. Am. J. Manag. Care 4, 797–806 (1998).

[5] The Dartmouth Institute for Health Policy and Clinical Practice. "Data by Region." Dartmouth Atlas Project. Dartmouth College. <http://www.dartmouthatlas.org/data/region/>.