Paper 3620-2016

Health Care's 'One-Percenters': Hot-Spotting to Identify Areas of Need and Opportunity

Paul A. LaBrec, 3M Health Information Services

ABSTRACT

Since Atul Gawande popularized the term in describing the work of Dr. Jeffrey Brenner in a New Yorker article, "hot-spotting" has been used in health care to describe the process of identifying "super-utilizers" of health care services, then defining intervention programs to coordinate and improve their care. According to Brenner's data from Camden, New Jersey, 1% of patients generate 30% of payments to hospitals, while 5% of patients generate 50% of payments. Analyzing administrative health care claims data, which contains information about diagnoses, treatments, costs, charges, and patient sociodemographic data, can be a useful way to identify super-users, as well as those who may be receiving inappropriate care. Both groups can be targeted for care management interventions. In this paper, techniques for patient outlier identification and prioritization are discussed using examples from the CMS linkable 2008–2010 Medicare Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF).

INTRODUCTION

Since Atul Gawande popularized the term in describing the work of Dr. Jeffrey Brenner in a New Yorker article, 'Hot Spotting' has been used in health care to describe the process of identifying 'super-utilizers' of health care services, then defining intervention programs to coordinate their care (Gawande, 2011). According to Brenner's data from Camden New Jersey, 1% of patients generate 30% of payments to hospitals, while 5% of patients generate 50% of payments (Dubner, 2015). More recent reports on larger datasets have corroborated these metrics (Health Care Transformation Task Force, 2105).

Hot-spotting methods use claims data to identify the super-utilizers for whom targeted interventions are designed to optimize health care utilization and cost. These interventions are generally multifaceted in scope, targeting not only clinical problems, but social determinants of health, as well. Patients are monitored intensively and a multi-disciplinary team addresses their needs for medical care, pharmaceuticals, health education, transportation, behavioral health, social support, and other factors impacting their health.

In this analysis, I will use claims from the CMS linkable 2008–2010 Medicare Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF) to demonstrate hot-spotting techniques. I will demonstrate how both data analysis and mapping can be useful in identifying and locating high utilizers of health services.

METHODS

The data used in the analysis presented in this paper were downloaded from the CMS linkable 2008–2010 Medicare Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF) (Centers for Medicare and Medicaid Services (CMS), 2013b). These data files and documentation, along with SAS® program files to read and compile data can be found at the following location on the CMS website.

https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF.html

The DE-SynPUF was designed to create a new set of data files that are not actual Medicare claims data, yet are based on such data and resemble the structure and content of true Medicare data files to the degree that they are useful for software and application development and training purposes. According to the CMS documentation the DE-SynPUF files …

> preserve the detailed data structure and metadata of key variables at both the beneficiary and claim levels. However, the data are fully "synthetic," meaning no beneficiary in the DE-SynPUF is an actual Medicare beneficiary. They are all synthetic beneficiaries meant to represent actual beneficiaries. In order to protect the privacy of beneficiaries and to greatly reduce the risk of re-identification, a significant amount of interdependence and

co-variation among variables has been altered in the synthetic process. The synthetic process used significantly diminishes the analytic utility of the file to produce reliable inferences about the actual Medicare beneficiary population (i.e., univariate statistics and regression coefficients produced with the DE-SynPUF will be biased) (ibid, p. 1).

The DE-SynPUF contains beneficiary enrollment and utilization claims data from a total of 2.25MM synthetic beneficiaries contained in 20 samples. Each sample represents approximately 112-116,000 synthetic beneficiaries and contains demographic and plan enrollment data, as well as inpatient, outpatient, physician supplier, and pharmacy claims data (CMS, 2013a).

The analytic utility of the data file differs based on the type and level of analysis being conducted (CMS, 2013b):

- Demographic: The DE-SynPUF estimates of demographic characteristics (date of birth, date of death, sex, race, state, county) of the beneficiary population match the univariate frequency of the full population of beneficiaries enrolled in Medicare at any time during the 2008 year.

- Clinical: The DE-SynPUF estimates for clinical variables such as chronic conditions can provide researchers with bounds on how many cases with a specific condition are likely to be in the Medicare claims, which could be used to generate power calculations for a grant application.

- Economic/financial: The DE-SynPUF estimates for the economic and financial variables provide a lower bound for the true estimate of cost for the full population of beneficiaries enrolled in Medicare at any time during the 2008 year and costs for 2009 and 2010 for this 2008 beneficiary example.

- Multivariate modeling: The dynamic relationships between variables (demographic, health plan enrollment, clinical, economic/financial, and provider information) were altered, to limit re-identification risk. Therefore, analyses from multivariate modeling should be interpreted with caution.

Record types and counts for the full DE-SynPUF are displayed in Table 1. The approximately 2.3MM beneficiary synthetic records represent approximately 5% of total Medicare beneficiaries (CMS, 2013a). Records include a beneficiary summary with demographic and residence data, benefit plan information, and summary costs. Claims data are organized in separate databases as inpatient and outpatient facility claims, provider claims, and pharmacy events. These databases can be linked to one another using the synthetic beneficiary ID (DESYNPUF_ID) as pictured in Figure 1. Some beneficiaries will have multiple claims, while others will have none.

For this analysis I downloaded the first ten samples and selected beneficiaries residing in San Diego County, California. This sample represents three years of claims and eligibility data for a total of 8,543 unique beneficiaries. Beneficiaries who died during a given study year were excluded from the analysis for that year. A total of 375 (1.5%) of the initial study population were lost due to death during the three year study period. The number of beneficiaries included in the study by year is presented in Table 2. The 8,400 beneficiaries included in this analysis represents about 2.5% of the 350,000 seniors recorded as San Diego County residents on the 2010 US Census (CensusViewer, 2012). The analysis population can be viewed as a study cohort with claims available from 2008 through 2010. All analyses were performed using SAS® for PC v9.4.

Summary annual Medicare fee-for-service reimbursement amounts for inpatient, outpatient, and physician supplier claims were taken from the beneficiary data files. There is no summary prescription drug data in the beneficiary file, so annual prescription drug costs were calculated using the gross drug cost by event from the prescription drug events database, summed by beneficiary and year, and merged with the beneficiary data.

DE-SynPUF	Unit of record	Number of Records 2008	Number of Records 2009	Number of Records 2010
Beneficiary Summary	Beneficiary	2,326,856	2,291,320	2,255,098
Inpatient Claims	claim	547,800	504,941	280,081
Outpatient Claims	claim	5,673,808	6,519,340	3,633,839
Carrier Claims	claim	34,276,324	37,304,993	23,282,135
Prescription Drug Events (PDE)	event	39,927,827	43,379,293	27,778,849

Table 1 Data Entrepreneur's Synthetic Public Use File Record Counts

Note: Claim counts for 2010 are lower due to attrition from death, and some effects of disclosure treatment

Source: CMS, 2013b

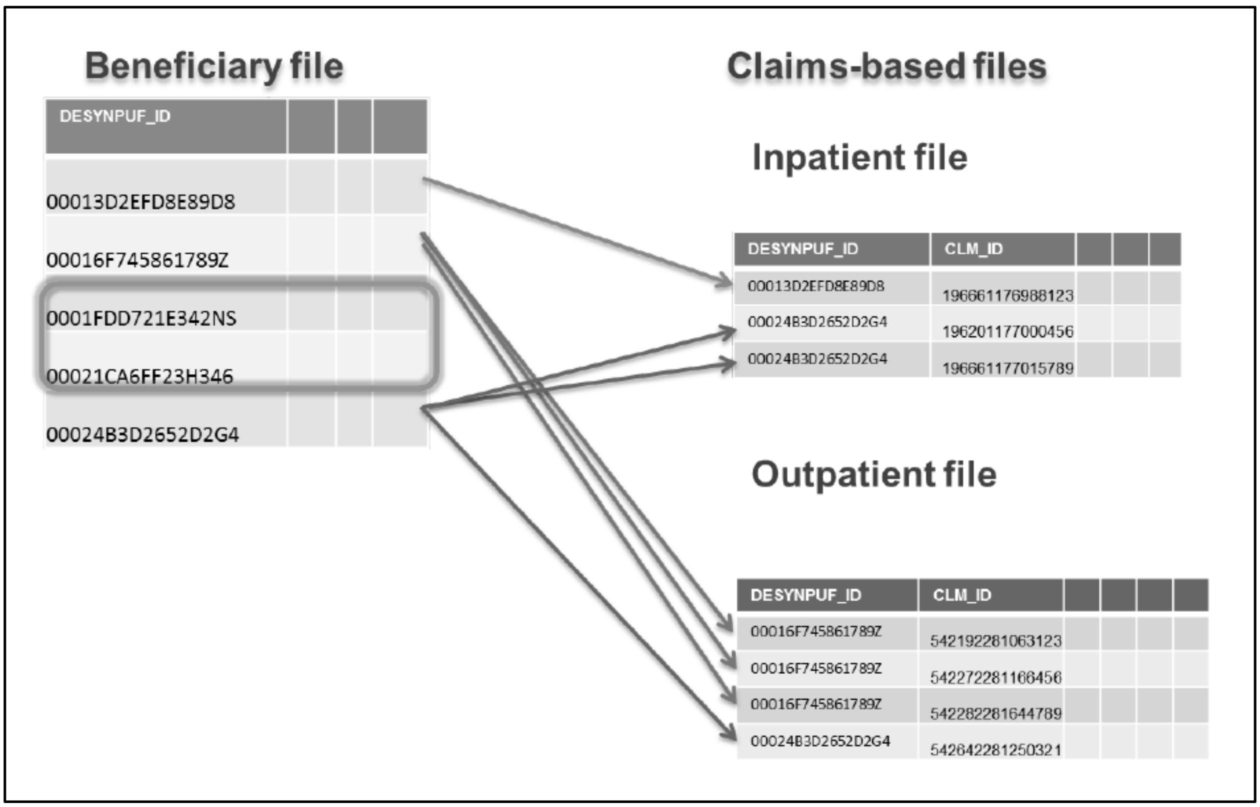

Figure 1 Claims Data File Structure

Source: CMS, 2013b

TOTAL COST OF CARE

Much of the analysis in this paper uses the outcome "Total Cost of Care." Total Cost of Care (TCC) is a measure that has gained increasing presence in discussions of health care payment reform and achievement of the "triple aim" in health care. The triple aim (Berwick, 2001) is defined as:

- Improving the patient experience of care (including quality and satisfaction)
- Improving the health of populations
- Reducing the per capita cost of health care

Health care costs include numerous items. Some of these are obvious and easier to collect (e.g., hospital, physician, pharmacy, laboratory and radiology services), while others are less obvious and/or more difficult to measure or collect (e.g., behavioral health services, durable medical equipment, over-the-counter medication, alternative health services utilization). Most measures of TCC as operationalized in health services research or health care payment reform are derived from administrative claims data. The National Quality Forum has endorsed one such measure as a Total Cost Index (TCI), which includes "all costs associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services" (HealthPartners, 2010). For this analysis.

This analysis measures Total Cost of Care using data available in the DE-SynPUF as the sum of Medicare fee-for-service payments for inpatient, outpatient, professional (physician supplier) services, plus gross drug cost from pharmacy utilization data.

SYNTHETIC ZIP CODE ASSIGNMENT FOR MAPPING

The DE-SynPuf beneficiary data records contain variables for beneficiary state and county of residence. In order to demonstrate an exercise in geographic mapping of patients, San Diego County ZIP Codes were assigned to beneficiaries in the dataset. A set of 102 San Diego ZIP Codes with population estimates of seniors (age 65+ years) was extracted from a 3M internal demographic database. ZIPs comprising less than 1% of the county senior population were dropped from the list, leaving 70 ZIP codes. These ZIPs were weighted to approximate the relative senior population (e.g., if a ZIP contained 3% of the county's seniors, it would be weighted 3x that of a ZIP containing 1% of the county's seniors.) This list was randomly matched to the list of 8,543 Medicare beneficiaries in our 2008 sample. In this manner a fictitious ZIP Code was assigned to each beneficiary in order to create the map presented in the results section.

RETROSPECTIVE COHORT ANALYSIS

The DE-SynPUF contains longitudinal claims data, and thus affords the opportunity to follow beneficiaries over time in a cohort analysis. I created two cohorts for this analysis: 1) a "high-cost" cohort defined as those beneficiaries having an average total cost of care greater than five standard deviations above the mean for the study population in 2008; and 2) a "non-high-cost" cohort which included the remaining study beneficiaries.

Although I examined average costs for all beneficiaries by year (Table 5), most of the analysis concerns cohorts defined in the base year (2008) and followed for the two subsequent years (2009, 2010). The cohort approach in a hot-spotting analysis is particularly useful as it allows the analyst to assess the persistence of high-cost events in specific beneficiaries.

RESULTS

After removing beneficiary deaths from the analysis as described in the Methods, the study population began at 8,432 in 2008 and ended at 8,168 in 2010 (Table 2).

Claim Year	Beneficiaries on 1-JAN	Deaths (%)	Beneficiaries on 31-DEC
2008	8,543	111 (1.3)	8,432
2009	8,432	137 (1.6)	8,295
2010	8,295	127 (1.5)	8,168

Table 2 Final Beneficiary Count from the DE-SynPUF San Diego County, CA Sample

Select demographic characteristics of the beneficiaries are presented in Table 3. Nearly all beneficiaries were enrolled for a full year in each analysis year, most (57%) were female, white (73%), and the mean age was 73 years in the first study year. In the study base year (2008), 1,206 (14.3%) of beneficiaries in the study sample were under age 65 years. Persons under age 65 can receive Medicare benefits for certain disabilities. They were not excluded from this analysis, as the focus is identifying high-cost beneficiaries regardless of age or other demographic characteristics.

Claim Year	Benes	Mean Enrollment Months (sd)	Mean Age Years (sd)	% Female	% White	% Black	% Others	% Hispanic
2008	8,432	11.5 (2.2)	72.7 (12.4)[1]	56.8	73.2	8.0	12.6	6.3
2009	8,295	11.7 (1.9)	73.7 (12.4)	56.8	73.3	8.0	12.6	6.3
2010	8,168	11.6 (2.0)	74.7 (12.5)	56.8	73.1	8.0	12.6	6.3

Table 3 Select Demographic Characteristics - DE-SynPUF San Diego County, CA Sample

[1] 14.3% of the sample were under age 65 years in 2008.

Chronic Condition	2008		2009		2010	
	Cases	Rate (%)	Cases	Rate (%)	Cases	Rate (%)
End-Stage Renal Disease	620	7.4	759	9.2	546	6.7
Alzheimer’s or related disorder	1,544	18.3	1,796	21.7	1,332	16.3
Heart Failure	2,276	27.0	2,752	33.2	1,989	24.4
Chronic Kidney Disease	1,307	15.5	1,656	20.0	1,030	12.6
Cancer	501	5.9	595	7.2	406	5.0
COPD	1,029	12.2	1,219	14.7	687	8.4
Depression	1,775	21.0	2,040	24.6	1,370	16.8
Diabetes	3,114	36.9	3,337	40.2	2,313	28.3
Ischemic Heart Disease	3,446	40.9	3,725	44.9	2,876	35.2
Osteoporosis	1,436	17.0	1,509	18.2	1,006	12.3
Rheumatoid/Osteo Arthritis	1,253	14.9	1,382	16.7	772	9.5
Stroke/Transient Ischemic Attack	362	4.3	420	5.1	225	2.8

Table 4 Prevalence of Select Chronic Disease Indicators - DE-SynPUF San Diego County, CA Sample

The prevalence of select chronic conditions reported in the DE-SynPUF Beneficiary data is presented in Table 4. Common chronic conditions in this population are diabetes, ischemic heart disease, heart failure, depression—all affecting at least 20% of the study population in the baseline year (2008).

Total Cost of Care for the high-cost cohort (TCC > 5 SDs above the mean) within each study year is compared to the TCC for the total population for each in year in Table 6. In each study year, the high-cost cohort comprises approximately 1% of the population. The high-cost cohorts presented in this table were redefined each year (i.e., a beneficiary could appear as high-cost one year and not the next or vice versa).

Select Financial Measures	2008 (n=8,432)	2009 (n=8,295)	2010 (n=8,168)
	Mean (sd)	**Mean (sd)**	**Mean (sd)**
Total Medical Allowed	$3,700 (9434)	$4,052 (8249)	$2,448 (5702)
Pharmacy Cost	$406 (1083)	$445 (1135)	$300 (812)
Total Cost of Care	$4,107 (9540)	$4,498 (8397)	$2,747 (5798)

Table 5 Select Financial Measures - DE-SynPUF San Diego County, CA Sample

Total Medical Allowed = Inpatient, Outpatient, Professional Claim Reimbursement

Year	Total Benes	% Total	Mean TCC	SD	Min	Max
2008	8,432	100.0%	$4,107	$9,540	$0	$165,130
2009	8,295	100.0%	$4,498	$8,397	$0	$119,830
2010	8,168	100.0%	$2,747	$5,797	$0	$89,820
	Hi-Cost Benes					
2008	76	0.9%	$74,426	$20,719	$51,880	$165,130
2009	67	0.8%	$65,511	$16,098	$46,490	$119,830
2010	73	0.9%	$45,454	$14,720	$31,830	$89,820

Table 6 Total Cost of Care for High Cost Beneficiaries Compared to All Beneficiaries

COHORT ANALYSIS RESULTS

This section presents the results of the retrospective cohort analysis described in the Methods above. For these analyses, the study population cohorts were fixed using data from the baseline year (2008). A beneficiary defined as high-cost or non-high-cost in 2008 remained in that cohort for analysis of subsequent claims in 2009 and 2010. Using 2008 claims data 76 beneficiaries were identified as high-cost (> 5 SDs above mean TCC). The high-cost beneficiaries were similar to the non-high-cost cohort in age and gender, though a higher proportion of high-cost beneficiaries was Hispanic (Table 7).

Cost Category	Claim Year	Benes	Mean Enrolled Months (sd)	Mean Age Years (sd)	% Female	% White	% Black	% Others	% Hispanic
Non-High-Cost	2008	8,356 (99%)	11.5 (2.2)	72.8 (12.4)	56.8	73.2	8.0	12.6	6.2
High-Cost	2008	76 (1%)	12.0 (0.0)	71.0 (14.8)	56.6	72.4	6.6	10.5	10.5

Table 7 Select Demographic Characteristics – High-Cost / Non-High-Cost Cohorts

As expected, the high-cost cohort had much higher rates of chronic disease than the non-high-cost group (Table 8). Nearly all were diabetic and had heart and/or kidney disease; more than half suffered from depression and/or Alzheimer's disease. The most common conditions among the non-high-cost cohort were ischemic heart disease and diabetes.

Select baseline (2008) financial measures for each cohort are presented in Table 9. By definition, average costs for the high-cost cohort far exceed those of the non-high-cost cohort. An interesting finding is that among the high-cost cohort prescription drug costs, although nearly double those of the non-high-cost cohort, amounted to only 1% of TCC. Among the non-high-cost cohort, however, prescription drug costs accounted for 12% of costs.

Chronic Condition	Non-High-Cost (n=8,356)		High-Cost (n=76)	
	Cases	Rate (%)	Cases	Rate (%)
End-Stage Renal Disease	581	7.0	39	51.3
Alzheimer's or related disorder	1,499	17.9	45	59.2
Heart Failure	2,204	26.4	72	94.7
Chronic Kidney Disease	1239	14.8	68	89.5
Cancer	486	5.8	15	19.7
COPD	979	11.7	50	65.8
Depression	1,732	20.7	43	56.6
Diabetes	3,045	36.4	69	90.8
Ischemic Heart Disease	3,371	40.3	75	98.7
Osteoporosis	1,407	16.8	29	38.1
Rheumatoid/Osteo Arthritis	1,223	14.6	30	39.5
Stroke/Transient Ischemic Attack	342	4.1	20	26.3

Table 8 Prevalence of Select Chronic Disease Indicators – High-Cost and Non-High-Cost Cohorts

Measure	Cohort	Benes	Mean	SD	Min	Max
TCC	Non-High-Cost	8,356	$3,467	$6,527	$0	$51,630
	High-Cost	76	$74,426	$20,719	$51,880	$165,130
TMA	Non-High-Cost	8,356	$3,064	$6,401	$0	$51,630
	High-Cost	76	$73,680	$20,819	$45,990	$160,440
RX	Non-High-Cost	8,356	$403	$1,077	$0	$9,760
	High-Cost	76	$747	$1,627	$0	$6,760

Table 9 Select Baseline (2008) Financial Measures – High-Cost and Non-High-Cost Cohorts

Table 10 presents the annual total cost of care for each cohort over the three-year study period. The data display a fair amount of variability in costs over the study period. The average cost of the high-cost cohort dropped from $74,426 in 2008 to $22,676 (-70%) in 2009 and further decreased to $7,185 in 2010, less than 10% of what is was in 2008. Some members of the high-cost cohort in 2008 incurred no costs in 2009 or 2010. The non-high-cost cohort also demonstrated variability, with average TCC increasing by 25% over baseline in 2009, then back down to 22% below baseline in 2010.

Year	**Non-High-Cost**	**Mean TCC**	**SD**	**Min**	**Max**
2008	8,356	$3,467	$6,527	$0	$51,630
2009	8,219	$4,329	$7,936	$0	$119,830
2010	8,092	$2,706	$5,697	$0	$83,910
Year	**Hi-Cost**	**Mean TCC**	**SD**	**Min**	**Max**
2008	76	$74,426	$20,719	$51,880	$165,130
2009	76	$22,676	$23,614	$0	$106,150
2010	76	$7,185	$11,704	$0	$89,820

Table 10 Total Cost of Care over Study Period – High-Cost and Non-High-Cost Cohorts

PATIENT PROFILE

In fact, only one individual of the 76 in the baseline 2008 high-cost cohort was also identified in both the 2009 and 2010 high-cost cohorts. This beneficiary was a 60-year old female with multiple chronic conditions. A patient profile is presented in Table 11. Care for this patient cost over $200,000 for the three-year study period—about $6,000 per month. She appears to be a good candidate for care management intervention. This patient is what hot-spotters call a “super-user,” someone who uses a much higher than average amount of health care services. The patient had 10 hospital stays totaling 56 days between 2008 and 2010 receiving care totaling nearly $140,000 in fee-for-service costs (Table 12). She stayed at 4 different hospitals and was treated by 6 different attending physicians. At only half of her stays was she at the same hospital treated by the same attending physician. She also visited 14 unique providers at 10 unique outpatient facilities at a cost of nearly $45,000 (Table 13). Her Medicare carrier (physician) reimbursement data show similarly fragmented care. Furthermore, this same patient had 157 prescription drug events over the three-year study period totaling $13,990 in gross cost.

<table>
<tr><td>Age in 2008</td><td>60</td><td colspan="2" rowspan="3">Annual Medicare fee-for-service reimbursements for this patient averaged over $6,000 per month in 2008-2010.</td></tr>
<tr><td>Gender</td><td>Female</td></tr>
<tr><td>Race</td><td>White</td></tr>
<tr><td>Chronic Conditions</td><td colspan="3">ESRD, Diabetes, Congestive Heart Failure, Chronic Kidney Disease, Ischemic Heart Disease</td></tr>
<tr><td>Year</td><td>TMA</td><td>RX</td><td>TCC</td></tr>
<tr><td>2008</td><td>$53,740</td><td>$3,260</td><td>$57,000</td></tr>
<tr><td>2009</td><td>$67,740</td><td>$4,820</td><td>$72,560</td></tr>
<tr><td>2010</td><td>$83,910</td><td>$5,910</td><td>$89,820</td></tr>
</table>

Table 11 Patient Profile – High-Cost Medicare Beneficiary

Admit Date	Hospital	Length of Stay	Attending Physician	DRG	DRG Description	Payment
19-Apr-08	A	6	A	386	INFLAMMATORY BOWEL DISEASE W CC	$5,000
14-May-08	B	3	B	442	DISORDERS OF LIVER EXCEPT MALIG,CIRR,ALC HEPA W CC	$21,000
8-Jul-08	C	7	C	643	ENDOCRINE DISORDERS W MCC	$11,000
1-May-09	A	5	A	257	UPPER LIMB & TOE AMPUTATION FOR CIRC SYSTEM DISORDERS W/O CC/MCC	$23,000
30-Jun-09	A	1	A	685	ADMIT FOR RENAL DIALYSIS	$14,000
16-Dec-09	D	4	D	698	OTHER KIDNEY & URINARY TRACT DIAGNOSES W MCC	$5,000
13-Jan-10	E	2	E	412	CHOLECYSTECTOMY W C.D.E. W CC	$41,000
24-Feb-10	C	11	F	923	OTHER INJURY, POISONING & TOXIC EFFECT DIAG W/O MCC	$10,000
23-Mar-10	A	3	A	286	CIRCULATORY DISORDERS EXCEPT AMI, W CARD CATH W MCC	$0
24-Aug-10	A	14	A	916	ALLERGIC REACTIONS W/O MCC	$9,000
TOTAL	4	56	6			$139,000

Table 12 Inpatient Claims for High-Cost Medicare Beneficiary

	TOTAL	2008	2009	2010
Facilities	10	5	4	6
Providers	14	6	7	9
Claims	34	9	15	10
Cost	$44,510	$8,030	$17,600	$18,880

Table 13 Outpatient Claims for High-Cost Medicare Beneficiary

MAPPING

A hot-spotting mapping exercise was performed using the 2010 DE-SynPUF beneficiary data randomly tagged with ZIP Codes as described in the Methods. In a true payer dataset, enrollee residence data contained in files would be used to map persons by ZIP Code. The average (mean) total cost of care (TCC) was calculated for beneficiaries of each ZIP Code represented in the data and presented on the map in Figure 2. This type of map is good way to visualize the location of hot-spots of high utilizers. Elements can be added, such as the location of various health care facilities (hospitals, clinics, urgent care centers, etc.), to show their location with respect to high utilizers. Orange and red areas on the map highlight ZIP Codes with the highest average TCC. Land areas displayed as white did not align with any data from the beneficiary records (e.g., Camp Pendleton and Cleveland National Forest).

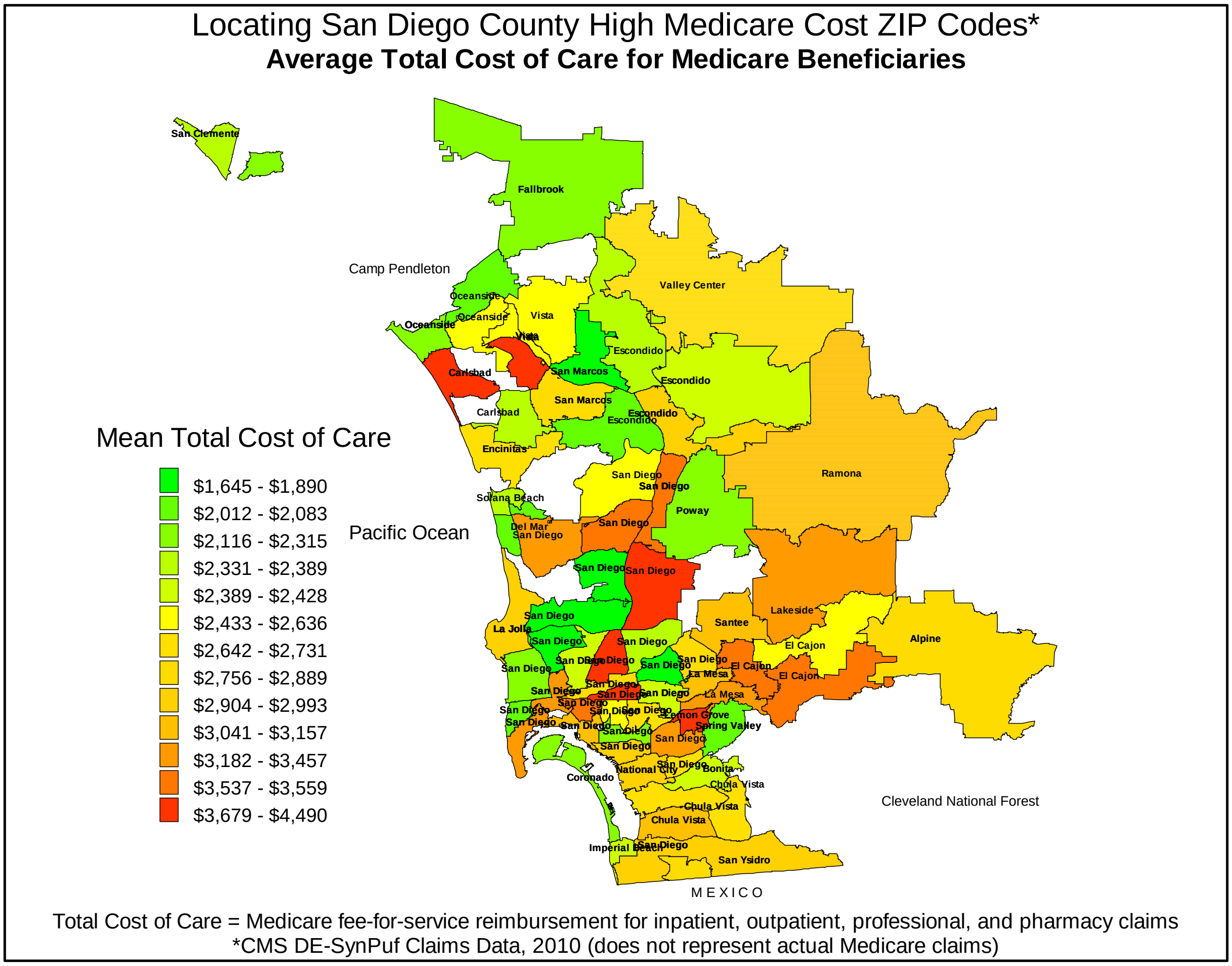

Figure 2 Locating San Diego High Medicare Cost ZIP Codes

PROC MAPIMPORT was used to import the US by ZIP Code Tabulation Area (ZCTA) shape file from the 2010 Census. Using PROC SQL, the current data was merged with the SAS® dataset SASHELP.ZIPCODE which contains boundary information including longitude and latitude coordinates, ZIP Code centroids, city names, county codes, and more. The ANNOTATE dataset was used to add city names to the map, and PROC GMAP with ODS output generated and outputted the finished map. The code used to generate this map is reproduced at the end of the paper.

DISCUSSION

The primary goal of hot-spotting programs is to identify patients, like the woman identified in the Medicare claims data in this exercise, and help them to receive intensive care management. The goal is ultimately to improve the patient's health through better care coordination and more intensive oversight of their care. This care often extends beyond clinical settings into social services in order to improve the environment in which the patient lives while they are not at a health provider's office. In this manner, such a program can help the patient realize their best chance at health.

By profiling patients using hot-spotting techniques, limited health care resources can be targeted to improve the health and reduce the cost of high-risk patients. Administrative claims data are a good source of information to develop prioritized care management patient lists to be used by insurers and providers to more effectively target patients for enhanced care. Enhanced care can take the form of providing additional and/or more regular primary care services, identifying hospitalized patients for more intensive monitoring or discharge follow-up programs, providing prescription management services, or other therapeutic intervention programs. In particular, the cost of nonadherence to prescription medication has been estimated at $100-300MM annually (PhRMA, 2011). Medication management is an important component of overall care management and can be monitored through claims-based adherence measures, such as the medication possession ration and proportion of days covered (Leslie, 2008).

To achieve success, hot-spotting programs require numerous highly trained dedicated staff committed to long-term interventions. This success can be difficult to achieve and to sustain among localized provider panels, let alone large scale patient populations. Furthermore, since these interventions involve social programs traditionally outside the domain of health care delivery, greater cooperation and integration of state and federal agencies, and cooperation with private providers and insurers would be necessary to achieve and maintain success at a large scale (Van Horn, 2011; McArdle, 2011)

There are, however, some methodological problems with demonstrating the value of traditional hot-spotting. First, these programs are generally not developed using experimental designs, may have no control group, often target small populations, and may not collect adequate data on potential study confounders. For these reasons and others, comparative results from these programs may suffer from bias and a lack of generalizability. Second, pre/post-intervention studies of extreme cases (super-utilizers) are at risk for the statistical phenomenon 'regression to the mean,' whereby the average utilization volume or cost among the non-random population of super-users is more likely to go down in future measurements than it is to stay high. This will be true for measures over time whether or not any intervention is applied. Regression to the mean occurs because not all the patients identified as super-users at a single point or period in time will remain super-users, causing average measures among that population to 'regress' toward the average of the total population (Linden, 2013). In a recent analysis of Medicare fee-for-service expenditures from 2010---2012, 37% of patients in the top 10% spender tier became "reverters" who later dropped out of this spender tier. Another Institute of Medicine report found about 40% of patients in the top 5% spender tier were also reverters, most of whom tended to be younger with good or excellent self-reported health status (Health Care Transformation Task Force, 2015).

The phenomenon of regression to the mean can be seen in the cohort analysis of De-SynPUF data. The high average costs incurred by the 2008 'high-cost' cohort did not persist over the next two years. In fact, average costs for that cohort in 2010 were only about 10% of what they were in 2008. We found only one patient identified as a super-utilizer for all three study years. This statistical outlier problem underscores the fact that a longitudinal look at patient experience and analysis of the drivers behind high cost are important tools in identifying the truly persistent high cost patients.

Even with the caveats described above, there is value in analyzing administrative health care claims data to identify patients most appropriate for care management intervention. Claims data can be risk-stratified to discriminate between those patients who are likely to use more services and more expensive services than others based on their diagnostic history, utilization experience, and demographic profile. Stratifying patients by risk for high utilization and cost is an important component of care management intervention. Many quantitative and qualitative risk stratification tools are available from both public and private sources.

Finally, the identification of super-utilizers or the inefficient or inappropriate delivery of care is increasingly being recognized as a cost-effective strategy and one worthy of payment. Various payment reform programs initiated by private insurers reward providers for reducing the costs of caring for chronically ill patients. Pay-for-Performance (P4P) programs, bundled payments, global budgeting, and other strategies are being used to increase the proportion of value-based versus direct fee-for-service payments (Miller, 2015). In the public sector, CMS began instituting reimbursement penalties for excess Medicare readmissions through the Hospital Readmissions Reduction Program (HRRP) in 2012, and recently began making separate payments for chronic care management (CCM) services under the Medicare Physician Fee Schedule (CMS, 2015b). The Department of Health and Human Services (DHHS) has set a goal to have 30 percent of Medicare payments in alternative payment models by the end of 2016 and 50 percent by the end of 2018. Overall, HHS seeks to have 85 percent of Medicare fee-for-service payments in value-based purchasing by 2016 and 90 percent by 2018 (CMS, 2015a).

Programs such as these in both the public and private sector are providing incentive to manage excess health services utilization and improve population health.

CONCLUSION

My objective in writing this paper was to provide a brief overview of the topic of hot-spotting in health care. I discussed the origin of the term and the concepts it encapsulates, and introduced the work of its early practitioners. I gave an overview of the CMS DE-SynPUF database and, I hope, whet your appetite to try some Medicare data analysis yourself. This tool provides a good laboratory to become familiar with administrative claims data and explore health services research methods. I presented concepts of cohort analysis cost of care measurement. I presented a patient profile which got into the details of claim record. I demonstrated how hot-spotting can be aided with visualization using mapping tools. Finally, I gave some caveats to the exercise of hot-spotting and reviewed utilization reduction and payment reform. In the end, I hope you remember the patient whose case we reviewed, and know that, although fictitious, she represents many persons who would greatly benefit from better management of their care.

SAS® Code used to produce Figure 2:

```
*Combine data and ZIP Code coordinate files*;
proc sql;
 create table SGFCMS.hotspot_10 as
  (
   select
         a.*,
         b.*,
         c.ZCTA5CE10
   from SGFCMS.TCC_MEAN a
   left join SASHELP.ZIPCODE b on (a.ZIP = b.ZIP)
   left join maps.gazateer2 c on (a.ZIPCHAR = c.ZCTA5CE10)
   )
   ;
quit;

*Create Annotate Dataset*;
data labels_black;
   length function color $ 8 text $ 22;
   retain function 'label' xsys ysys '2' hsys '3' when 'a';
  set SGFCMS.hotspot_10;
   text=''||CITY;
   style="'Arial/bo'";
   color='black'; size=1.05; position='B';
run;

*Produce Map*;
ods graphics on;
ODS RTF file="<file location>\superuser.rtf" style=statistical ;
goptions ftitle='Arial';
title1 h=1.75 "Figure 1. Locating San Diego County High Medicare Cost
ZIP Codes*";
title2 h=1.5 "Average Total Cost of Care for Medicare Beneficiaries";
footnote1 "Total Cost of Care = Medicare fee-for-service reimbursement for
inpatient, outpatient, professional, and pharmacy claims" justify=left;
footnote2 "*CMS DE-SynPuf Claims Data, 2010 (does not represent actual
Medicare claims)" justify=left;
```

```
/* colors from green->red */
pattern1 v=s c=cx00ff00; pattern2 v=s c=cx65ff00; pattern3 v=s c=cx88ff00;
pattern4 v=s c=cxbaff00; pattern5 v=s c=cxd0ff00; pattern6 v=s c=cxffff00;
pattern7 v=s c=cxffe000; pattern8 v=s c=cxffdc00; pattern9 v=s c=cxffd100;
pattern10 v=s c=cxffc000; pattern11 v=s c=cxff9a00; pattern12 v=s
c=cxff7700; pattern13 v=s c=cxff3400;
  legend1
  origin=(5,10) pct
  offset=(2,2)
  mode=share
  across=1
  label=(justify=center height=1.5 'Mean Total Cost of Care'
        position=top)
  value=(j=r)
  shape=bar(1.5,2.5) pct;
proc gmap data=SGFCMS.hotspot_10 map=cazip;
  id ZCTA5CE10; /* matches values between response and map dataset */
  choro TCC_MEAN /* response variable */
  / levels=13 /* pattern/color levels */
  legend=legend1 /*add legend options*/
  coutline=black  /*boundary lines*/
  annotate=labels_black;
  note move=(2.25,4.5) in h=0.75 "Camp Pendleton"
         move=(2.25,2.75) in h=1.1 "Pacific Ocean"
         move=(5.75,1) in h=0.75 "Cleveland National Forest"
         move=(4.5,0.4) in h=0.75 "M E X I C O"
;
run;
quit;
ODS RTF close;
ods graphics off;
```

REFERENCES

Berwick DM, Nolan TW, Whittington J. May 2008. "The Triple Aim: Care, Health, and Cost." Health Affairs. May 2008 vol. 27 no. 3, pp. 759-769.

CensusViewer. Population of San Diego County, California: Census 2010 and 2000 Interactive Map, Demographics, Statistics, Graphs, Quick Facts. Accessed 3/19/16 at http://censusviewer.com/county/CA/San%20Diego

Centers for Medicare and Medicaid Services. "Better Care. Smarter Spending. Healthier People: Paying Providers for Value, Not Volume." Fact Sheet. January 26, 2015. (https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-01-26-3.html)

Centers for Medicare and Medicaid Services. "Chronic Care Management Services." Fact Sheet. May, 2015. (https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/ChronicCareManagement.pdf)

Centers for Medicare and Medicaid Services. Linkable 2008–2010 Medicare Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF): Codebook. Updated February 20, 2013.

Centers for Medicare and Medicaid Services (CMS) Linkable 2008–2010 Medicare Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF): User Manual. Updated January 15, 2013. (https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/Downloads/SynPUF_Codebook.pdf)

Dubner, S. "How Do We Know What Really Works in Healthcare?" Freakonomics podcast. April 2, 2015. (http://freakonomics.com/2015/04/02/how-do-we-know-what-really-works-in-healthcare-a-new-freakonomics-radio-podcast/)

Gawande, A. "The Hot Spotters: can we lower medical costs by giving the neediest patients better care?" The New Yorker. January 24, 2011. (http://www.newyorker.com/magazine/2011/01/24/the-hot-spotters)

Health Care Transformation Task Force. "Proactively Identifying the High Cost Population." White Paper, July, 2015. (http://www.hcttf.org/resouces-tools/2015/5/14/whitepaper)

HealthPartners. Total Cost of Care Population-based PMPM Index (NQF Measure 1604), 2010. Accesed 18-MAR-16 at http://www.qualityforum.org/Home.aspx

Leslie, S. Calculating Medication Compliance, Adherence, and Persistence in Administrative Pharmacy Claims Databases. Western Users of SAS Software Proceedings. November, 2008. Accessed 3/19/16 at http://wuss.org/proceedings08/08WUSS%20Proceedings/papers/anl/anl09.pdf

Linden, A. "Assessing regression to the mean effects in health care initiatives." BMC Medical Research Methodology 2013, 13:119. (http://www.biomedcentral.com/1471-2288/13/119)

McArdle, M. "The Value of Health Care Experiments." The Atlantic. January 24, 2011. (http://www.theatlantic.com/business/archive/2011/01/the-value-of-health-care-experiments/70106/)

Miller, H. "The Building Blocks of Successful Payment Reform: Designing Payment Systems that Support Higher-Value Health Care." Center for Healthcare Quality and Payment Reform, 2015. (http://www.chqpr.org/downloads/BuildingBlocksofPaymentReform.pdf)

PhARMA. "Improving Prescription Medication Adherence is Key to Better Health." January, 2011. Accessed 3/19/16 at http://www.phrma.org/sites/default/files/pdf/PhRMA_Improving%20Medication%20Adherence_Issue%20Brief.pdf

Van Horn, L. "The Terrible Economics of Health Care 'Hot-Spotting.'" Forbes. March 28, 2011. (http://www.forbes.com/sites/larryvanhorn/2011/03/28/the-terrible-economics-of-health-care-hot-spotting/)

ACKNOWLEDGMENTS

The author acknowledges Melissa Gottschalk for her editorial review of the manuscript.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Paul A. LaBrec
3M Health Information Systems
plabrec@mmm.com
www.3Mhis.com