Data Science for Health Datapalooza

Story

Story

Slides

1. Background
2. Key data sources
3. Population
4. Aggregation
5. Data Contents
6. Data Limitations

APPENDIX A – File Attributes
APPENDIX B – Place of Service Descriptions

Story

Health Datapalooza V is Enough: Graduate to Health Data Science for Medicine

My tweets were:

http://semanticommunity.info/Data_Science/Data_Science_for_Health_Datapalooza

Updated: Sat, 19 Sep 2015 00:45:50 GMT
Powered by mindtouch
• @hdpalooza HDP5 is enough—Graduate to Health Data Science for Medicine—practice of medicine to science of
medicine with health data science.
• @hdpalooza This was an excellent conference for the following reasons (see below)

My question to all is: What was the greatest take home message of HDP5? To me it was from entrepreneur, inventor,
and technology fan, Vinod Khosia, who said "the practice of medicine needs to evolve to the science of medicine", "data
science can do more for the science of medicine than all of the biomedical research put together", and "a Kaiser data
scientist looked at the statin data for Kaiser patients and suggested a change in the protocol for its use that was adopted
and there was a 40% drop in strokes for those patients."

I have asked Dr. Nirav Shah, Senior Vice President and Chief Operating Officer for Clinical Operations, Kaiser
Permanente, for more information of this last statement.

This was an excellent conference for the following reasons:
• Bryan Sivak was a good moderator on the first day and Susannah Fox was an even better moderator on the second
day
• The Workshops looked interesting, but not interesting enough to give up a wonderful Sunday with my family.
• Dr. Eliot Fisher talked about the Dartmouth Atlas of Health Care (I have worked with these data) and Todd Park
paid tribute to George Thomas, (I worked with him on Federal CIO Council activities).
• The Right Honorable Jeremy Hunt and Dr. Atul Gawande gave the most thought provoking talks of the entire
conference.
• The Leading New Data Sources was the most boring session and the DataLab was the most interesting session, I
attended. I told Damon Davis I liked the format of data developer-data provider pairings in the DataLab session and
their should be more of that more frequently like our Federal Big Data Working Group Meetup.
• The HHS Initiatives to Accelerate Innovative Data Applications and the Data Scientist-Extracting Data Forcefully
From Bureaucracies drew my strongest criticism about the need for data science ecosystems (data dictionaries and
ontologies) to focus ONC standards work and data science products (data publications) to provide context for app
developers and the public to effectively use the deluge of open health data being released.
• Dr. Francis Collins gave the talk that should be the theme for next year: Big Data and Biomedical Research, with his
new Assistant Director for Data Science, Phil Bourne, leading the charge of Health Data Science for Medicine. See
my Data Culture at the NIH. I look forward to hearing more about the NIH Data Commons.

The key announcements were:
• Congratulations to ProPublica, the 2014 winners of the Health Data Liberators Award!
• Congratulations to the first-place winners of the Code-a-Palooza, Lyfechannell! Check out all the winners on the
healthdata.gov blog Did any of the the winners actually use the 2012 Medicare Provider Payment Data that was
released for the first time by the Centers for Medicare and Medicaid Services this past April?
• Health Datapalooza 2015, May 31–June 3 in Washington, D.C.
• US CTO @todd_park announces #openFDA at #hdpalooza: "We can't wait to see what you build!" - @openFDA
• Attendees at #hdpalooza are not just creating data, they are creating new systems of healthcare, says
@Atul_Gawande pic.twitter.com/WggboXXZ7k - Maneesh Juneja, @ManeeshJuneja

MORE TO FOLLOW

Work with the App Winners Data Sets:
Data Science for Health Datapalooza V

"Health Datapalooza is an incredible gathering of tech innovators, clinicians, patient advocates, entrepreneurs, health care leaders, policymakers, and researchers. It's helping to fuel the creation of a rising tide of new tools and services that are making a real difference in people's lives." Todd Park, US CTO

The Health Data Consortium (HDC), the group organizing Health Datapalooza IV, encourages the creation of a health data "ecosystem" that promotes and accelerates the innovative use of health data. The Consortium is a collaboration among government, non-profit, and private sector organizations working to foster the availability and use of health data to drive innovations that improve health and health care. The Consortium advocates data best practices and information sharing with data providers, and works with businesses, entrepreneurs, and academia to help them understand how to use data to develop new products, services, apps, and research insights.

I have been to all four previous Health Datapalooza's and this one really interested me because of the following sessions and activities:

• **Open Health Data as a Tool for Investigative Journalism**: Innovative journalists and their media organizations have been taking state and federal health data, crunching it and sometimes building news applications that allow the public to customize big data in a very personal way. ProPublica, for instance, took data on more than 1 billion prescriptions dispensed in Medicare Part D and created Prescriber Checkup, an app that allows users to compare their personal physicians to peers in the same specialty and state. The Center for Investigative Reporting built a tool that allows users to compare painkiller prescribing at VA medical centers nationwide over several years. ClearHealthCosts.com is finding and posting cash and self-pay prices, and is also using crowdsourcing to discover and reveal charges, contracted rates and a full picture of how people are paying for health care. And Kaiser Health News has made it easy to find data on hospital readmission penalties and geographic variations on spending for post-acute care. Reporters from each of these organizations will talk about how they think about using health data, how they customize it for a wider audience and ways in which the government should continue making useable information available.

• **Data Scientist- Extracting Data Forcefully from Bureaucracies**: This session will highlight ways to effectively extract health data from various data platforms, going from data collection all the way through use of the data (e.g., visualization, analysis, and modification of outcomes).

• **How Research Using Big Data is Making a Difference (Or Not)**: Research using big data coupled with new data sources and new analytics is already changing health research. Examples include drawing inferences using the semantics of research publications; cardiac outcomes from data mining; and social network analysis to guide health behavior, to name a few. This session will explore the implications of some of these advances and what they portend for future research.

• **Unleashing Data to Enable Health System Change— An Update on CMS Data Tools and Products**: CMS will present new and updated CMS data and information products for researchers and health policymakers. Panelists will present on: CMS public use data, provider charge data, new and updated information visualization tools, access to CMS data via the Virtual Research Data Center, and recent CMS data analysis.

Activities:

• We are delighted to announce that Code-a-Palooza at Health Datapalooza 2014 will utilize newly released Centers for Medicare & Medicaid Services (CMS) data to create visualization tools that help patients and consumers
improve their health care decision-making. Developers can compete for a total of $35,000 in prizes. All participants are eligible for a discount Health Datapalooza registration rate.

• **App Demos & Code-a-Palooza Finalists Preview:** Last month, we put out a call for proposals to create data visualization tools that leverage the newly-released Centers for Medicare & Medicare Services (CMS) claims data in the Code-a-Palooza challenge, run in collaboration with the Office of the National Coordinator for Health IT (ONC). Challenge participants were tasked with creating a proposal for a tool that would help patients make sense of the raw CMS data through visualizations and customizable functions that would support personalized health decision-making.

We're delighted to announce the Code-a-Palooza Finalist teams whose proposals have been chosen:

- Arcadia Solutions, Team Leader: Nicholas Stepro
- DocGraph, Team Leader: Fred Trotter
- DocSpot, Team Leader: Jerry Lin
- karmadata, Team Leader: Sean Power
- Lyfechannel, Team Leader: Dave Vockell
- Medecision Limelight, Team Leader: Nicole DeMicco
- Saltare Systems LLC, Team Leader: Joyce Ho
- Team FloriDUH, Team Leader: Mandi Bishop
- University of Wisconsin-Madison, Team Leader: Edward Zhong
- Zynx Health, Team Leader: Victor Lee

Congratulations to these teams who will move onto Phase II! This month, they will work to build their proposed tools and then present their creations for judging at Health Datapalooza on Monday, June 2, from 1:00-2:30 pm. The winner of Code-a-Palooza will be announced the morning of Tuesday, June 3, at the event.

• **App Demos:** These five organizations will showcase the promise and potential of health data by demonstrating their apps on the main stage. Below are descriptions they’ve provided of their apps:
  
  ◦ **ActualMeds Corp:** ActualMeds provides “medication reconciliation on demand” at point of care. Seamlessly combining medication data from the EHR or HIE, Rx claim/fills, and a patient structured interview into actionable information. ActualMeds produces the most accurate and current medication history. For Payers and Providers with high risk patients, operating efficiencies, quality metrics and outcomes are improved.
  
  ◦ **Social Health Insights:** The State of Indiana wanted to show the excellent and ongoing progress Indiana has made in the adoption of EHRs and the economic impact it has had on the state. By visualizing the data, we let everyone take a look at how the state is doing.
  
  ◦ **Maxwell Health:** At Maxwell Health, our mission is to transform health care in America. To do that, we’ve built an operating system for employee benefits. Maxwell’s mobile experience includes a time- and money-saving health care concierge, a turnkey wellness program that connects with best-in-class health products, and access to all benefits information.
  
  ◦ **Purple Binder:** Purple Binder makes it easy to match people with services that keep them healthy in their community, extending care beyond the clinic. Get actionable data on services ranging from food pantries to low-cost preschools, from homeless shelters to park district programs.
  
  ◦ **CareSync:** CareSync gives people access to meaningful, useful, and simply shareable medical info. Med reminders, daily dashboards, Care Manager, and the Health Timeline™ give users the power to effectively communicate with family, caregivers and providers. Convenient services, including records retrieval & transcription, booking appointments and more put the power of health data into the hands of the patient.
• In addition to the main stage presentation, we will have 2 other areas at Health Datapalooza to showcase amazing apps:
  ◦ Live App Demos: Featured at a stage near the Exhibit Hall, 35 companies will demo their apps in snappy 12-minute time slots.
  ◦ Apps Expo: All 58 app demos will have the opportunity to use a scheduled time to interact with HDP participants in dedicated Apps Expo booths in the Exhibit Hall.

• DataLab: Your opportunity to meet some of the federal government data experts is here. This year we’ll tell some of the inspiring stories of how innovators are using the data complimented by the background information on how and why the data was collected and curated as told by the data stewards. Come learn how to use assets from the Departments of Health & Human Services, other Federal agencies and state governments from the individuals who champion action in improving public access to information to catalyze innovation. Each agency in the federal government is staffed by experts who are well versed in the division’s information resources available on data.gov (administrative data, survey data, research data, medical/scientific content, etc.) The DataLab will also feature opportunities for focused discussions with data experts for “deep dives” into agency’s resources. There will be live demonstrations and announcements of new data resources. The goal of the session is to give innovators and entrepreneurs an overview of new, updated and emerging datasets that can be used to support new applications and services. Facilitators: Justine Carr, M.D., Chair, National Committee on Vital and Health Statistics Workgroup on Data Access and Use and Damon Davis, HHS Director of the Health Data Initiative

MORE TO FOLLOW FROM ATTENDANCE AT CONFERENCE

In preparation for my attendance at the conference and my presentation at the Federal Big Data Working Group Meetup, I data mined the Health Datapalooza 2014 Web pages and their links (see Research Notes below), and answered the three essential questions:

How was the data collected?: See Centers for Medicare & Medicare Services (CMS) claims data and Methodology

Where is the data stored: See Excel and Spotfire Dashboard

What are the data results?: See Spotfire Dashboard and Slides of Screen Captures

Some specific results are as follows:

• The Compressed ZIP package containing the tab delimited data file (Medicare-Physician-and-Other-Supplier-PUF-CY2012.txt) which is 1.7GB uncompressed and contains 27 columns and more than 9 million records, was easily downloaded, uncompressed and imported into Spotfire resulting in a 381 MB sized file.

• All of this data can be stored in memory and essentially analyzed in real-time in Spotfire.

• The average Medicare payment amount versus the average Medicare allowed amount is well-correlated (R squared =0.99), and the slope of the best fit straight line is = 0.8 indicating the allowed amount is consistently slightly larger than the paid amount.

• The average submitted charge amount versus the average Medicare allowed amount is less well-correlated (R-squared=0.56), and the slope of the best fit line is = 3.0 indicating the submitted amount is considerably larger than the allowed amount.

MORE TO FOLLOW FROM ADDITIONAL ANALYSES

Slides
Slide 1

http://semanticommunity.info/Data_Science/Data_Science_for_Health_Datapalooza

Slide 2

http://semanticommunity.info/@api/deki/files/29343/MedicareHDPV.xlsx?origin=mt-web

Slide 3

Web Player
Slide 4

Web Player

Slide 5

Slide 6

Spotfire Dashboard

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Media, iframe, embed and object tags are not supported inside of a PDF.

Research Notes

Code-a-Palooza Challenge at Health Datapalooza 2014: http://healthdatapalooza.org/announc...apalooza-2014/

Code-a-Palooza Call for Proposals: http://healthdatapalooza.org/get-inv...apalooza-2014/

How Four Doctors Won a Codeathon - Recap of the Code-a-Palooza: http://healthdata.gov/blog/how-four-...p-code-palooza


Also PDF Download and in ZIP file: https://www.cms.gov/Research-Statist...ethodology.pdf

Code-a-Palooza and the CMS data, please view this webinar. Held?
139 Datasets: [https://data.cms.gov/](https://data.cms.gov/)

Medicare Provider Utilization and Payment Data: Physician and Other Supplier: [https://www.cms.gov/Research-Statist---Supplier.html](https://www.cms.gov/Research-Statist---Supplier.html)


This Compressed ZIP package contains the tab delimited data file (Medicare-Physician-and-Other-Supplier-PUF-CY2012.txt) which is 1.7GB uncompressed and contains more than 9 million records, thus importing this file into Microsoft Excel will result in an incomplete loading of data. Use of database or statistical software is required; a SAS® read-in statement is supplied. Additionally, this ZIP package contains the following supporting documents: CMS-AMA-CPT-2011-license-agreement.pdf, Medicare-Physician-and-Other-Supplier-PUF-SAS-Infile.sas, Medicare-Physician-and-Other-Supplier-PUF-Methodology.pdf

**Summary Tables:**

- Medicare Physician and Other Supplier Aggregate table, CY2012, Microsoft Excel (.xlsx): 144 MB Downloaded
- Medicare National and State/HCPCS Aggregate table, CY2012, Microsoft Excel (.xlsx)


Additional information can be found at our [Frequently Asked Questions](https://www.cms.gov/app/) page

By Topic: Medicare Provider Utilization and Payment Data Subtopic: Physician and Other Supplier

Also of interest:

Steven Randazzo on the [HealthData.gov blog](http://www.healthdata.gov/)

Objectives of four workshops: [http://healthdatapalooza.org/sunday-workshops/#Wk4](http://healthdatapalooza.org/sunday-workshops/#Wk4)

Press Pass and Entry: [http://healthdatapalooza.org/about/contact-us/](http://healthdatapalooza.org/about/contact-us/)

Members of the press are invited to attend Health Datapalooza 2014. Please send an email with your full contact information to datapalooza@courtesyassoc.com to receive the press registration code.

Thank you to our 2014 sponsors, who are contributing to help us to provide an excellent agenda and experience for the 2,200 anticipated attendees at the 2014 event! We are grateful to our many generous sponsors who have invested in the continued growth of the health data ecosystem and Health Datapalooza.

The Health Datapalooza is seeking technology entrepreneurs working on the cutting-edge of patient self-monitoring, consumer self-tracking, and/or employee wellness health technology to participate in the first [Health Datapalooza “Shark Tank”](http://semanticommunity.info/Data_Science/Data_Science_for_Health_Datapalooza) with selected Venture Capital and Angel Investors (session to be held: June 3, 2104; Washington, DC). Do you believe you have health technology that will capture the attention of these Investors? Here is your chance to pitch to a panel of 6 VCs/Angel Investors!
To apply: Submit a 1-page description (12 point font) of your company, your product, and your business model. Please be sure to include your contact information (phone number, e-mail, etc.). Applications will be reviewed by the 2014 Health Datapalooza Consumer Session Organizers (who include the Department of Health and Human Services, Chief Technology Officer – Bryan Sivak). All materials will remain confidential and reviewed by the HDP Consumer Session Organizers only.

Three to Five Entrepreneurs will be selected to give their brief pitch at the 2014 HDP, and either sink or swim with the sharks (who may bite!). Do you want to have a chance to swim with the sharks and see if one of them will take you under their fin? Here is your chance!

Those selected will be responsible for their travel to the 2104 Health DataPalooza and conference registration fee (app developer rate).

**Deadline:** Friday, May 9, 2014; 5:00pm EST
**Acceptance/Rejection Notifications to be sent:** May 19, 2014; 5:00pm EST

Submit your 1-page application to: Dr. Audie Atienza: Audie.Atienza@NIH.gov

---

**Medicare Provider Utilization and Payment Data: Physician and Other Supplier**

Source: [https://www.cms.gov/Research-Statist... Supplier.html](https://www.cms.gov/Research-Statist... Supplier.html)

As part of the Obama Administration’s efforts to make our healthcare system more transparent, affordable, and accountable, the Centers for Medicare & Medicaid Services (CMS) has prepared a public data set, the Medicare Provider Utilization and Payment Data: Physician and Other Supplier Public Use File (Physician and Other Supplier PUF), with information on services and procedures provided to Medicare beneficiaries by physicians and other healthcare professionals. The Physician and Other Supplier PUF contains information on utilization, payment (allowed amount and Medicare payment), and submitted charges organized by National Provider Identifier (NPI), Healthcare Common Procedure Coding System (HCPCS) code, and place of service. This PUF is based on information from CMS’s National Claims History Standard Analytic Files. The data in the Physician and Other Supplier PUF covers calendar year 2012 and contains 100% final-action physician/supplier Part B non-institutional line items for the Medicare fee-for-service population.

While the Physician and Other Supplier PUF has a wealth of information on payment and utilization for Medicare Part B services, the dataset has a number of limitations. Of particular importance is the fact that the data may not be representative of a physician’s entire practice as it only includes information on Medicare fee-for-service beneficiaries. In addition, the data are not intended to indicate the quality of care provided and are not risk-adjusted to account for differences in underlying severity of disease of patient populations. For additional limitations, please review the methodology document available below.

Data are available in two formats:

- Tab delimited file format (requires importing into database or statistical software; SAS® read-in language is included in the download ZIP package)
• Microsoft Excel format (.xlsx), split by provider last name (note: organizational providers with name starting with a numeric are available in the “YZ” file).

CMS has also created two summary tables: 1) aggregated information by physician or other supplier and 2) aggregated information by State and HCPCS code. A detailed methodology document can be found in the Downloads section below which contains important information regarding the limitations of data usage.

**Tab Delimited Format:**

[Note: This Compressed ZIP package contains the tab delimited data file (Medicare-Physician-and-Other-Supplier-PUF-CY2012.txt) which is 1.7GB uncompressed and contains more than 9 million records, thus importing this file into Microsoft Excel will result in an incomplete loading of data. Use of database or statistical software is required; a SAS® read-in statement is supplied. Additionally, this ZIP package contains the following supporting documents: CMS-AMA-CPT-2011-license-agreement.pdf, Medicare-Physician-and-Other-Supplier-PUF-SAS-Infile.sas, Medicare-Physician-and-Other-Supplier-PUF-Methodology.pdf]

**Microsoft Excel Format:**

CMS has also created two summary tables: 1) aggregated information by physician or other supplier and 2) aggregated information by State and HCPCS code. A detailed methodology document can be found in the Downloads section below which contains important information regarding the limitations of data usage.

**Summary Tables:**

Additional information can be found at our [Frequently Asked Questions](#) page.

Inquiries regarding this data can be sent to MedicareProviderData@cms.hhs.gov.

To receive email notifications, please sign up for the Medicare Provider Data GovDelivery subscription here.

**Downloads**

- Medicare Physician and Other Supplier PUF Methodology [PDF, 260KB] My Note: Downloaded this
Related Links

- Interactive Physician and Other Supplier Dataset on data.cms.gov
- Medicare Physician and Other Supplier Look-up Tool
- Page last Modified: 04/23/2014 1:14 PM


Source: [PDF]

CPT codes, descriptions and other data only are copyright 2011 American Medical Association. All rights reserved. CPT is a registered trademark of the American Medical Association (AMA).

You, your employees and agents are authorized to use CPT only as contained in the following authorized materials of Centers for Medicare and Medicaid Services (CMS) internally within your organization within the United States for the sole use by yourself, employees and agents. Use is limited to use in Medicare, Medicaid or other programs administered by CMS. You agree to take all necessary steps to insure that your employees and agents abide by the terms of this agreement.

Any use not authorized herein is prohibited, including by way of illustration and not by way of limitation, making copies of CPT for resale and/or license, transferring copies of CPT to any party not bound by this agreement, creating any modified or derivative work of CPT, or making any commercial use of CPT. License to use CPT for any use not authorized herein must be obtained through the AMA, CPT Intellectual Property Services, AMA Plaza, 330 N. Wabash Ave., Suite 39300, Chicago, IL 60611-5885. Applications are available at the AMA Web site, http://www.ama-assn.org/go/cpt.

Applicable FARS\DFARS Restrictions Apply to Government Use

This product includes CPT which is commercial technical data and/or computer data bases and/or commercial computer software and/or commercial computer software documentation, as applicable which were developed exclusively at private expense by the American Medical Association, 515 North State Street, Chicago, Illinois, 60654. U.S. Government rights to use, modify, reproduce, release, perform, display, or disclose these technical data and/or computer data bases and/or computer software and/or computer software documentation are subject to the limited rights restrictions of DFARS 252.227-7015(b)(2) (November 1995) and/or subject to the restrictions of DFARS 227.7202-1(a) (June 1995) and DFARS 227.7202-3(a) (June 1995), as applicable for U.S. Department of Defense procurements and the limited rights restrictions of FAR 52.227-14 (June 1987) and/or subject to the restricted rights provisions of FAR 52.227-14 (June 1987) and FAR 52.227-19 (June 1987), as applicable, and any applicable agency FAR Supplements, for non-Department of Defense Federal procurements.

AMA Disclaimer of Warranties and Liabilities
CPT is provided “as is” without warranty of any kind, either expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a particular purpose. No fee schedules, basic unit, relative values or related listings are included in CPT. The AMA does not directly or indirectly practice medicine or dispense medical services. The responsibility for the content of this file/product is with CMS and no endorsement by the AMA is intended or implied. The AMA disclaims responsibility for any consequences or liability attributable to or related to any use, non-use, or interpretation of information contained or not contained in this file/product. This agreement will terminate upon notice if you violate its terms. The AMA is a third party beneficiary to this agreement.

CMS Disclaimer
The scope of this license is determined by the AMA, the copyright holder. Any questions pertaining to the license or use of the CPT should be addressed to the AMA. End Users do not act for or on behalf of the CMS. CMS DISCLAIMS RESPONSIBILITY FOR ANY LIABILITY ATTRIBUTABLE TO END USER USE OF THE CPT. CMS WILL NOT BE LIABLE FOR ANY CLAIMS ATTRIBUTABLE TO ANY ERRORS, OMISSIONS, OR OTHER INACCURACIES IN THE INFORMATION OR MATERIAL CONTAINED ON THIS PAGE. In no event shall CMS be liable for direct, indirect, special, incidental, or consequential

Medicare Physician and Other Supplier PUF Methodology

Source: (PDF)

Medicare Fee-For Service
Provider Utilization & Payment Data
Physician and Other Supplier
Public Use File:
A Methodological Overview
April 7, 2014
Prepared by:
The Centers for Medicare and Medicaid Services,
Office of Information Products and Data Analytics

1. Background

As part of the Obama Administration’s efforts to make our healthcare system more transparent, affordable, and accountable, the Centers for Medicare & Medicaid Services (CMS) has prepared a public data set, the Provider Utilization and Payment Data Physician and Other Supplier Public Use File (herein referred to as “Physician and Other Supplier PUF”), with information on services and procedures provided to Medicare beneficiaries by physicians and other healthcare professionals. The Physician and Other Supplier PUF contains information on utilization, payment (allowed amount and Medicare payment), and submitted charges organized by National Provider Identifier (NPI), Healthcare Common Procedure Coding System (HCPCS) code, and place of service. This PUF is based on information from CMS’s National Claims History (NCH) Standard Analytic Files (SAFs). The data in the Physician and Other Supplier PUF covers calendar year 2012 and contains 100% final-action (i.e., all claim adjustments have been resolved) physician/supplier Part B non-institutional line items (excluding durable medical equipment) for the Medicare fee-for-service (FFS) population.
2. Key data sources

The primary data source for these data is CMS’s CY2012 National Claims History (NCH) Standard Analytic Files (SAFs) which include claims as of 6/30/2013. The NCH SAFs contain 100 percent of Medicare final action claims for beneficiaries who are enrolled in the FFS program. The NCH contains a SAF for each type of Medicare claim type including institutional (i.e., hospital inpatient, hospital outpatient, skilled nursing, home health and hospice) and non-institutional (i.e., physician/supplier Part B and durable medical equipment). Specifically, the Physician/Supplier Part B SAF was used to create the Physician and Other Supplier PUF, which includes services from physicians, non-physician practitioners, laboratories, imaging, ambulances, etc. (does not include claims from the durable medical equipment SAF). Beneficiary and service counts, provider charges, Medicare allowed amounts and payments, place of service, provider type, and Medicare participation indicator were summarized from this SAF.

Provider demographics are also incorporated in the Physician and Other Supplier PUF including name, credentials, gender, complete address and entity type from the National Plan & Provider Enumeration System (NPPES), which CMS developed to assign unique identifiers, known as National Provider Identifiers (NPIs), to health care providers. The health care provider’s demographic information is collected at time of enrollment and updated periodically by CMS approved Electronic File Interchange Organizations (EFIO) that submit information on behalf of the provider. The provider must approve of the updates to NPPES. The demographics information provided in the Physician and Other Supplier PUF was extracted from NPPES at the end of calendar year 2012. For additional information on NPPES, please visit https://nppes.cms.hhs.gov/NPPES/Welcome.do.

3. Population

The Physician and Other Supplier PUF includes data for providers that had a valid NPI and submitted Medicare Part B non-institutional claims (excluding DME) during the 2012 calendar year. To protect the privacy of Medicare beneficiaries, any aggregated records which are derived from 10 or fewer beneficiaries are excluded from the Physician and Other Supplier PUF.

4. Aggregation

The spending and utilization data in the Physician and Other Supplier PUF is aggregated to the following:

a) the NPI for the performing provider,
b) the Healthcare Common Procedure Coding System (HCPCS) code, and
c) the place of service (either facility or non-facility).

There can be multiple records for a given NPI based on the number of distinct HCPCS codes that were billed and where the services were provided. Data has been aggregated based on the place of service because separate fee schedules apply based on whether the place of service submitted on the claim is facility or non-facility.

The provider NPI is the numeric identifier registered in NPPES. HCPCS codes are used to identify medical services and procedures furnished by physicians and other health care professionals and include two levels. Level I codes are the Current Procedural Terminology (CPT) codes that are maintained by the American Medical Association and Level II codes are created by CMS to identify products, supplies and services not covered by the CPT codes (such as...
5. Data Contents

npi – National Provider Identifier (NPI) for the performing provider on the claim.

nppes_provider_last_org_name – When the provider is registered in NPPES as an individual (entity type code='I'), this is the provider’s last name. When the provider is registered as an organization (entity type code = ‘O’), this is the organization name.

nppes_provider_first_name – When the provider is registered in NPPES as an individual (entity type code='I'), this is the provider’s first name. When the provider is registered as an organization (entity type code = ‘O’), this will be blank.

nppes_provider_mi – When the provider is registered in NPPES as an individual (entity type code='I'), this is the provider’s middle initial. When the provider is registered as an organization (entity type code = ‘O’), this will be blank.

nppes_credentials – When the provider is registered in NPPES as an individual (entity type code='I'), these are the provider’s credentials. When the provider is registered as an organization (entity type code = ‘O’), this will be blank.

nppes_provider_gender – When the provider is registered in NPPES as an individual (entity type code='I'), this is the provider’s gender. When the provider is registered as an organization (entity type code = ‘O’), this will be blank.

nppes_entity_code – Type of entity reported in NPPES. An entity code of 'I' identifies providers registered as individuals and an entity type code of ‘O’ identifies providers registered as organizations.

nppes_provider_street1 – The first line of the provider’s street address, as reported in NPPES.

nppes_provider_street2 – The second line of the provider’s street address, as reported in NPPES.

nppes_provider_city – The city where the provider is located, as reported in NPPES.

nppes_provider_zip – The provider’s zip code, as reported in NPPES.

nppes_provider_state – The state where the provider is located, as reported in NPPES. The fifty U.S. states and the District of Columbia are reported by the state postal abbreviation. The following values are used for other areas:

'XX' = 'Unknown'
'AA' = 'Armed Forces Central/South America'
'AE' = 'Armed Forces Europe'
'AP' = 'Armed Forces Pacific'
'AS' = 'American Samoa'
'GU' = 'Guam'
'MP' = 'North Mariana Islands'

http://semanticommunity.info/Data_Science/Data_Science_for_Health_Datapalooza

Updated: Sat, 19 Sep 2015 00:45:50 GMT
Powered by mindtouch
nppes_provider_country – The country where the provider is located, as reported in NPPES. The country code will be ‘US’ for any state or U.S possession. For foreign countries (i.e., state values of ‘ZZ’), the provider country values include the following:

‘AE’ = ‘United Arab Emirates’ ‘IL’= Israel
‘AR’= ‘Argentina’ ‘IN’= India
‘AU’= ‘Australia’ ‘IS’= Iceland
‘BR’= ‘Brazil’ ‘IT’= Italy
‘CA’= ‘Canada’ ‘JP’= Japan
‘CH’= Switzerland ‘KR’= ‘Korea’
‘CN’= China ‘NL’= ‘Netherlands’
‘CO’= Colombia ‘PK’= ‘Pakistan’
‘DE’= ‘Germany’ ‘SA’= ‘Saudi Arabia’
‘ES’= ‘Spain’ ‘SY’= ‘Syria’
‘FR’= France ‘TR’= ‘Turkey’
‘GB’= Great Britain ‘VE’= ‘Venezuela’
‘HU’= Hungary

provider_type – Derived from the provider specialty code reported on the claim. For providers that reported more than one specialty code on their claims, this is the specialty code associated with the largest number of services.

medicare_participation_indicator – Identifies whether the provider participates in Medicare and/or accepts assignment of Medicare allowed amounts. The value will be ‘Y’ for any provider that had at least one claim identifying the provider as participating in Medicare or accepting assignment of Medicare allowed amounts.

place_of_service – Identifies whether the place of service submitted on the claims is a facility (value of ‘F’) or non-facility (value of ‘O’). Non-facility is generally an office setting; however other entities are included in non-facility. See “Appendix B – Place of Service Descriptions” for the types of entities included in facility and non-facility.

hcpcs_code – HCPCS code for the specific medical service furnished by the provider.

hcpcs_description – Description of the HCPCS code for the specific medical service furnished by the provider.

line_srvc_cnt – Number of services provided; note that the metrics used to count the number provided can vary from service to service.

bene_unique_cnt – Number of distinct Medicare beneficiaries receiving the service.

bene_day_srvc_cnt – Number of distinct Medicare beneficiary/per day services. Since a given beneficiary may receive multiple services of the same type (e.g., single vs. multiple cardiac stents) on a single day, this metric removes double-counting from the line service count to identify whether a unique service occurred.
average_Medicare_allowed_amt – Average of the Medicare allowed amount for the service; this figure is the sum of the amount Medicare pays, the deductible and coinsurance amounts that the beneficiary is responsible for paying, and any amounts that a third party is responsible for paying.

stdev_Medicare_allowed_amt – Standard deviation of the Medicare allowed amounts. The standard deviation indicates the amount of variation from the average Medicare allowed amount that exists within a single provider, HCPCS service, and place of service.

average_submitted_chrg_amt – Average of the charges that the provider submitted for the service.

stdev_submitted_chrg_amt – Standard deviation of the charge amounts submitted by the provider. The standard deviation indicates the amount of variation from the average submitted charge amount that exists within a single provider, HCPCS service, and place of service.

average_Medicare_payment_amt – Average amount that Medicare paid after deductible and coinsurance amounts have been deducted for the line item service.

stdev_Medicare_payment_amt – Standard deviation of the Medicare payment amount. The standard deviation indicates the amount of variation from the average Medicare payment amount that exists within a single provider, HCPCS service, and place of service.

6. Data Limitations

Although the Physician and Other Supplier PUF has a wealth of payment and utilization information about many Medicare Part B services, the dataset also has a number of limitations that are worth noting.

First, the data in the Physician and Other Supplier PUF may not be representative of a physician’s entire practice. The data in the file only has information for Medicare beneficiaries with Part B FFS coverage, but physicians typically treat many other patients who do not have that form of coverage. The Physician and Other Supplier PUF does not have any information on patients who are not covered by Medicare, such as those with coverage from other federal programs (like the Federal Employees Health Benefits Program or Tricare), those with private health insurance (such as an individual policy or employer-sponsored coverage), or those who are uninsured. Even within Medicare, the Physician and Other Supplier PUF does not include information for patients who are enrolled in any form of Medicare Advantage plan.

The information presented in this file also does not indicate the quality of care provided by individual physicians. The file only contains cost and utilization information, and for the reasons described in the preceding paragraph, the volume of procedures presented may not be fully inclusive of all procedures performed by the provider.

Medicare allowed amounts and Medicare payments for a given HCPCS code/place of service can vary based on a number of factors, including modifiers, geography, and other services performed during the same day/visit. For example, modifiers (which are two-character designators that signal a change in how the HCPCS code for the procedure or service should be applied) may be included on the claim line when the service intensity was increased or decreased, when an additional physician administered services, or when the service provided differs from the procedure definition. In some cases, modifiers impact allowed amounts and payments. In addition, allowed amounts and payments vary geographically because Medicare makes adjustments for most services based on an area’s cost of living. Allowed
amounts and payments can also be adjusted when a physician renders multiple services to a beneficiary on the same day, which is referred to as a multiple procedure payment reduction. For standard payment and allowed amount rates by CPT/HCPCS code, please go to http://www.cms.gov/apps/physician-fe.../overview.aspx.

Additionally, the data are not risk adjusted and thus do not account for difference in the underlying severity of disease of patient populations treated by providers. Also, since the data presented are summarized from actual claims received from providers and no attempts were made to modify any data (i.e., no statistical outliers were removed or truncated), in rare instances the average submitted charge amount may reflect errors included on claims submitted by providers.

As noted earlier, the file does not include data for services that were performed on 10 or fewer beneficiaries, so users should be aware that summing the data in the file may underestimate the true Part B FFS totals. In addition, some providers bill under both an individual NPI and an organizational NPI. In this case, users cannot determine a provider’s actual total because there is no way to identify the individual’s portion when billed under their organization.

Medicare pays differently when services are provided in a facility setting versus a freestanding physicians’ office (or other non-facility setting). When services are delivered in a facility setting, Medicare makes two payments, one for the physician’s professional fee and one for the facility. For services delivered in a facility (place_of_Service =“F”), the data in the Physician and Other Supplier PUF only represents the physician’s professional fee and does not include the facility payment. On the other hand, for services delivered in a non-facility setting, such as a physician’s office (place_of_Service =“O”), the Physician and Other Supplier PUF represents the complete payment for the service.

If users try to link data from this file to other public datasets, please be aware of the particular Medicare populations included and timeframes used in each file that will be merged. For example, efforts to link the Physician and Other Supplier PUF data to Part D prescription drug data would need to account for the fact that some beneficiaries who have FFS Part B coverage (and are thus included in the Physician and Other Supplier PUF) do not have Part D drug coverage (and thus not represented in Part D data files). At the same time, some beneficiaries that have Part D coverage (and are thus included in the Part D data) do not have FFS Part B coverage (and thus not included in the Physician and Other Supplier PUF). Another example would be linking to data constructed from different or non-aligning time periods, such as publically available data on physician referral patterns, which is based on an 18-month period.

Finally, users should be aware that payments from some CMS demonstration programs are included in the Physician and Other Supplier PUF. Since some CMS demonstration programs utilize the Medicare claims submission process, payments for services under these demonstrations are included in the data file and may be grouped under specific demonstration HCPCS codes or aggregated under non-demonstration specific HCPCS codes. Demonstration programs that are paid outside of the Medicare claims submission process are not included in the Physician and Other Supplier PUF.

APPENDIX A – File Attributes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Format</th>
<th>Length</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>npi</td>
<td>Char</td>
<td>10</td>
<td>National Provider Identifier</td>
</tr>
<tr>
<td>Field</td>
<td>Type</td>
<td>Length</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>---------</td>
<td>--------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>nppes_provider_last_org_name</td>
<td>Char</td>
<td>70</td>
<td>Last Name/Organization Name</td>
</tr>
<tr>
<td>nppes_provider_first_name</td>
<td>Char</td>
<td>20</td>
<td>First Name</td>
</tr>
<tr>
<td>nppes_provider_mi</td>
<td>Char</td>
<td>1</td>
<td>Middle Initial</td>
</tr>
<tr>
<td>nppes_credentials</td>
<td>Char</td>
<td>20</td>
<td>Credentials</td>
</tr>
<tr>
<td>nppes_provider_gender</td>
<td>Char</td>
<td>1</td>
<td>Gender</td>
</tr>
<tr>
<td>nppes_entity_code</td>
<td>Char</td>
<td>1</td>
<td>Entity Code</td>
</tr>
<tr>
<td>nppes_provider_street1</td>
<td>Char</td>
<td>55</td>
<td>Street Address 1</td>
</tr>
<tr>
<td>nppes_provider_street2</td>
<td>Char</td>
<td>55</td>
<td>Street Address 2</td>
</tr>
<tr>
<td>nppes_provider_city</td>
<td>Char</td>
<td>40</td>
<td>City</td>
</tr>
<tr>
<td>nppes_provider_zip</td>
<td>Char</td>
<td>20</td>
<td>Zip Code</td>
</tr>
<tr>
<td>nppes_provider_state</td>
<td>Char</td>
<td>2</td>
<td>State Code</td>
</tr>
<tr>
<td>nppes_provider_country</td>
<td>Char</td>
<td>2</td>
<td>Country Code</td>
</tr>
<tr>
<td>provider_type</td>
<td>Char</td>
<td>43</td>
<td>Provider Type</td>
</tr>
<tr>
<td>medicare_participation_indicator</td>
<td>Char</td>
<td>1</td>
<td>Medicare Participation Indicator</td>
</tr>
<tr>
<td>place_of_Service</td>
<td>Char</td>
<td>1</td>
<td>Place of Service</td>
</tr>
<tr>
<td>hcpcs_code</td>
<td>Char</td>
<td>5</td>
<td>HCPCS Code</td>
</tr>
<tr>
<td>hcpcs_description</td>
<td>Char</td>
<td>30</td>
<td>HCPCS Description</td>
</tr>
<tr>
<td>line_srvc_cnt</td>
<td>Num</td>
<td>8</td>
<td>Number of Services</td>
</tr>
<tr>
<td>bene_unique_cnt</td>
<td>Num</td>
<td>8</td>
<td>Number of Medicare Beneficiaries</td>
</tr>
<tr>
<td>bene_day_srvc_cnt</td>
<td>Num</td>
<td>8</td>
<td>Number of Medicare Beneficiary/Day Services</td>
</tr>
<tr>
<td>average_Medicare_allowed_amt</td>
<td>Num</td>
<td>8</td>
<td>Average Medicare Allowed Amount</td>
</tr>
<tr>
<td>stdev_Medicare_allowed_amt</td>
<td>Num</td>
<td>8</td>
<td>Standard Deviation Medicare Allowed Amount</td>
</tr>
<tr>
<td>average_submitted_chrg_amt</td>
<td>Num</td>
<td>8</td>
<td>Average Submitted Charge</td>
</tr>
<tr>
<td>stdev_submitted_chrg_amt</td>
<td>Num</td>
<td>8</td>
<td>Standard Deviation Submitted Charge Amount</td>
</tr>
</tbody>
</table>
### APPENDIX B – Place of Service Descriptions

Table B-1. Non-Facility Based Place of Service (place_of_Service ="O")

<table>
<thead>
<tr>
<th>Place of Service Code</th>
<th>Place of Service Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Pharmacy</td>
</tr>
<tr>
<td>03</td>
<td>School</td>
</tr>
<tr>
<td>04</td>
<td>Homeless Shelter</td>
</tr>
<tr>
<td>05</td>
<td>Indian Health Service Free-standing Facility</td>
</tr>
<tr>
<td>06</td>
<td>Indian Health Service Provider-based Facility</td>
</tr>
<tr>
<td>07</td>
<td>Tribal 638 Free-standing Facility</td>
</tr>
<tr>
<td>08</td>
<td>Tribal 638 Provider-based Facility</td>
</tr>
<tr>
<td>09</td>
<td>Prison/ Correctional Facility</td>
</tr>
<tr>
<td>11</td>
<td>Office</td>
</tr>
<tr>
<td>12</td>
<td>Home</td>
</tr>
<tr>
<td>13</td>
<td>Assisted Living Facility</td>
</tr>
<tr>
<td>14</td>
<td>Group Home</td>
</tr>
<tr>
<td>15</td>
<td>Mobile Unit</td>
</tr>
<tr>
<td>16</td>
<td>Temporary Lodging</td>
</tr>
<tr>
<td>17</td>
<td>Walk-in Retail Health Clinic</td>
</tr>
<tr>
<td>20</td>
<td>Urgent Care Facility</td>
</tr>
<tr>
<td>25</td>
<td>Birthing Center</td>
</tr>
<tr>
<td>32</td>
<td>Nursing Facility</td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>21</td>
<td>Inpatient Hospital</td>
</tr>
<tr>
<td>22</td>
<td>Outpatient Hospital</td>
</tr>
<tr>
<td>23</td>
<td>Emergency Room – Hospital</td>
</tr>
<tr>
<td>24</td>
<td>Ambulatory Surgical Center</td>
</tr>
<tr>
<td>26</td>
<td>Military Treatment Facility</td>
</tr>
<tr>
<td>31</td>
<td>Skilled Nursing Facility</td>
</tr>
<tr>
<td>34</td>
<td>Hospice</td>
</tr>
<tr>
<td>41</td>
<td>Ambulance - Land</td>
</tr>
<tr>
<td>42</td>
<td>Ambulance – Air or Water</td>
</tr>
<tr>
<td>51</td>
<td>Inpatient Psychiatric Facility</td>
</tr>
<tr>
<td>52</td>
<td>Psychiatric Facility-Partial Hospitalization</td>
</tr>
<tr>
<td>50</td>
<td>Federally Qualified Health Center</td>
</tr>
<tr>
<td>54</td>
<td>Intermediate Care Facility/Mentally Retarded</td>
</tr>
<tr>
<td>55</td>
<td>Residential Substance Abuse Treatment Facility</td>
</tr>
<tr>
<td>57</td>
<td>Non-residential Substance Abuse Treatment Facility</td>
</tr>
<tr>
<td>60</td>
<td>Mass Immunization Center</td>
</tr>
<tr>
<td>62</td>
<td>Comprehensive Outpatient Rehabilitation Facility</td>
</tr>
<tr>
<td>65</td>
<td>End-Stage Renal Disease Treatment Facility</td>
</tr>
<tr>
<td>71</td>
<td>Public Health Clinic</td>
</tr>
<tr>
<td>72</td>
<td>Rural Health Clinic</td>
</tr>
<tr>
<td>81</td>
<td>Independent Laboratory</td>
</tr>
<tr>
<td>99</td>
<td>Other Place of Service</td>
</tr>
<tr>
<td>33</td>
<td>Custodial Care Facility</td>
</tr>
<tr>
<td>49</td>
<td>Independent Clinic</td>
</tr>
<tr>
<td>55</td>
<td>Residential Substance Abuse Treatment Facility</td>
</tr>
<tr>
<td>60</td>
<td>Mass Immunization Center</td>
</tr>
<tr>
<td>57</td>
<td>Non-residential Substance Abuse Treatment Facility</td>
</tr>
<tr>
<td>62</td>
<td>Comprehensive Outpatient Rehabilitation Facility</td>
</tr>
<tr>
<td>65</td>
<td>End-Stage Renal Disease Treatment Facility</td>
</tr>
<tr>
<td>71</td>
<td>Public Health Clinic</td>
</tr>
<tr>
<td>72</td>
<td>Rural Health Clinic</td>
</tr>
<tr>
<td>81</td>
<td>Independent Laboratory</td>
</tr>
<tr>
<td>99</td>
<td>Other Place of Service</td>
</tr>
<tr>
<td></td>
<td>Description</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>53</td>
<td>Community Mental Health Center</td>
</tr>
<tr>
<td>56</td>
<td>Psychiatric Residential Treatment Center</td>
</tr>
<tr>
<td>61</td>
<td>Comprehensive Inpatient Rehabilitation Facility</td>
</tr>
</tbody>
</table>