International Journal of Management, IT & Engineering
Vol. 11 Issue 4, April 2021
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

A CASE STUDY IN MACHINE - LEARNING AND PROVIDERS/PAYORS DATABASES

Mukesh Kumar Saini
Research Scholar OPJS University

Dr. Jaibir Singh
Asst. Professor

ABSTRACT

The healthcare organization produces enormous collections of data with the potential to expose insights into improving costs and outcomes if analyzed with the appropriate tools. Machine learning and predictive algorithms are formerly in common use in other fields.

In healthcare domain, the largest datasets with the broadest relevance to the US population may reside in payers and provider databases. Analyzing such databases with the modern tools may find exceptional or hard-to-diagnose diseases that pointlessly consume healthcare cost before proper diagnoses are made.
.
The objective to focus the power of modern analytics on transmissible angioedema (HAE), a single exceptional disease, because it demonstrate features of diseases related with high costs: rare, hard to diagnose, progressive, and takes a long time from diagnosis to applicable treatment. Despite the availability of effective therapies, misdiagnoses and under diagnosis of angioedema result in significant burden to the healthcare system.

KEYWORDS: Machine Learning, Health and Health Care, human performance, software engineering, big data, Cases – Patients, Disease - Conditions

INTRODUCTION

Large pools of multitudinous data that are collected, stored, and anatomized to reveal unanticipated patterns and relationships, has speedily evolved to shape and inform nearly all sectors of the global economy.[1] eventually, big data seeks to play a useful economic role by revealing the implicit value hidden in this information. The development of tools able of rooting value from these massive collections of information has made big data applicable to all sectors of the market. Consumers and providers of products and services, as well as governments and controllers, all stand to advantage from the insights emerging from the new learning of big-data. [1] – [3].
Healthcare stands out as a sector with a great deal to gain from the prospective of big data. A 2011 McKinsey Global Institute report estimated that if the US healthcare system could successfully apply big data to drive effectiveness and quality, the periodic potential realized value could be further than $ 300 billion, two- thirds of which would arise from an 8% reduction in expenditures. [1] Experts have recommended that healthcare accept big- data approaches, and similar US groups as the Department of Veterans Affairs and Kaiser Permanente, the integrated managed- care

institute, have enforced innovative pilot programs, numerous of which use clinical data from electronic health records(EHRs) to identify cost- savings opportunities from clinical practice patterns.[1],[4],[5] Broader acceptance of big- data analytics will continue to grow as healthcare associations strive to comply with the accountability requirements of the Patient Protection and Affordable Care Act.[4] – [5]

The lagging rate of acceptance of EHRs in the United States has been a hedge to fully leveraging the potential of this data.

Still, indeed as EHR implementation has progressed, access to this information continues to stand in the way of high- quality exploration regarding treatment effectiveness and cost efficiency. [6] -[8]

Data- sharing and translucency guidelines have only lately been put forward, and programs regarding ethics and regulation still need to be defined by participating institutions.6 One possible way to circumvent the issues raised by using and participating EHRs is to use insurance claims data composed of de-identified diagnosis related details and payment information.

The size and national-wide scope of a claims database are also advantages over the local or regional nature of EHRs when trying to identify opportunities within the data. One similar claims database added up data for further than 170 million US cases from 2014 to2020.9 Claims datasets allow for in- depth assessment of health and quality results when evaluated with tools able of handling datasets of this size.

Applicable prophetic algorithms and machine- learning techniques designed to handle enormous datasets have been available for times, but their connection to healthcare has not been honored until relatively recently.

[3] For illustration, predictive analytics designed to assess pitfalls and to model likely results from distant data types (geospatial, text reports, equipment supplies, etc.) are used by the military to enhance operational efficiencies and have applicability to many other fields, similar as felonious disquisition, business, and healthcare.[10] – [13] Predictive systems, driven by machine learning techniques that advance based on empirical data, are ideal tools for identifying the patterns obscured by the volume of insurance claims data. [14]

Identification of patients with rare conditions (diseases) within the claims database may offer a prospective to uncover significant value. Rare conditions, any one of which affects smaller than persons, presently affect about 10% of the US population, or further than 32 million individuals. [1] For illustration, hereditary angioedema (HAE) is a potentially fatal rare disease with an occurrence of 1 in 10000 to 50000 in the United States. Individualities with the disease may be misdiagnosed for as long as 8 years. [15], [16] This inheritable, autosomal- dominant disease causes intermittent, painful attacks of subcutaneous and submucosal swelling of the skin, gastrointestinal tract, and larynx.[17] Although not associated with hives, the skin swelling of HAE frequently leads to misdiagnosis as antipathetic response.18 Swelling of the gastrointestinal tract produces pain, distension, nausea, puking, and diarrhea.[15] Because patients may witness abdominal symptoms for numerous times before manifesting the characteristic subcutaneous swelling of HAE, patients undergo inappropriate surgical and medical treatment for any of a wide range of incorrect judgments , including acute abdomen, biliary bellyache, hepatitis, indigenous enteritis, colic, cholecystitis, nephrolithiasis, pyelonephritis, ruptured ovarian tubercle, intestinal inhibition, duodenal ulcer, and ulcerative colitis.[15] Edema of the larynx may lead to suffocation and death.[18]

Despite the accessibility of effective remedial approaches that address acute treatment, short- term prophylaxis, and conservation remedy, misdiagnoses and under- diagnosis of HAE affect in significant burden to the healthcare system Between 2015 and 2016, US patients with HAE who were misdiagnosed accounted for 5040 emergency department visits, 41% of which caused in hospitalization.[15] At an average cost of$ 1479 per ED visit and an average $22728 for a 5- day hospitalization, significant cost savings might be realized through further prompt and effective diagnosis and management of these patients.[15]

PRACTICAL IMPLICATIONS

A logical tool with the flexibility to be applied to a variety of data sources and specifically identify very small patient subpopulations has the implicit to be an important force in the evolving healthcare landscape.

- Using such a tool, payers may realize cost savings from identifying patients with expensive conditions as early as possible and insure that their care is managed appropriately.
- Physicians would be better informed about how to manage these patients, and patients would receive the care appropriate for their needs.
- These analytical techniques could apply to open and closed healthcare systems, both large and small

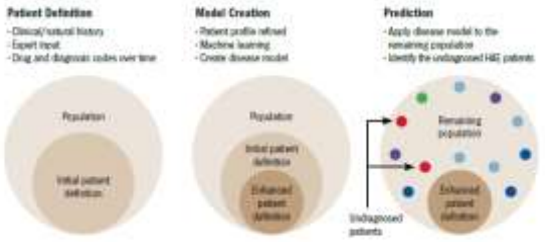
The study presented then was designed to demonstrate the capability of an ultramodern, automated machine learning system to discover undiagnosed rare diseases cases in a claims database. The main contributions of this exploration are to show how the adaption of state- of- the- art technologies in big- data analytics, information proposition, and machine learning can produce a seamlessly integrated frame for database analysis and to demonstrate how this technology could be put to practical use using insurance claims, rather than EHR data, to find undiagnosed rare- disease cases. The case presented then concentrated on HAE.[16]

METHODS

The population for this analysis was uprooted from a database of de-identified patient claims data acquired from Truven Health Analytics (MarketScan claims data).

This claims database contains health insurance claims data for further than 170 million unique lives covering 2014 through 2020. A 3- stage process was employed to discover cases with HAE within this database who hadn't been diagnosed and applied to a claims database to

- Define the characteristics (diagnoses, procedures, treatments, and providers) of patients in the database formerly being treated for HAE
- Use those characteristics to produce a model of patients with HAE
- Use the model to identify patients with HAE in the database who weren't yet diagnosed.


Patient Definition
-Clinical/natural history
-Expert input
-Drug and diagnosis codes over time
Population
Initial patient definition
Model Creation
-Patient profile refined
-Machine learning
-Create disease model
Population
Initial patient definition
Enhanced patient definition
Prediction
-Apply disease model to the remaining population
-Identify the undiagnosed HAE patients
Remaining population
Enhanced patient definition
Undiagnosed patients

Three-stage Discovery Process for Identifying Undiagnosed Hereditary Angioedema Patients in an Insurance Claims Database

Stage 1- Patient Definition

To study the characteristics of HAE victims and to identify their statistical 'signature' the first step was to find a group of patients in the database who surely had HAE. Diagnosis codes from International Classification of Diseases, Ninth Revision, Clinical revision(ICD-9-CM) alone may not be completely dependable a code may be used for billing purposes without official diagnosis; old codes may be used indeed after new, more specific codes come available; an ICD-9-CM code occasionally represents a group of diseases and data entry errors could occur.

Consulting physicians and drug experts for this study agreed that patients prescribed 1 of the 4 HAE-specific medicines available in the United States were, without mistrustfulness, patients with HAE. The 4 medicines were Cinryze (C1 esterase inhibitor [human]; Shire), Firazyr (icatibant; Shire), Berinert (C1 esterase inhibitor [human]; CSL Behring), and Kalbitor (ecallantide; Shire). Therefore, patients linked in the database as being prescribed 1 or further of these 4 medicines formed the population of index HAE patients.

Stage 2 - Model Creation

To identify which features or combination of features are most statistically applicable for screening HAE from non-HAE patients, an information-theoretic conception of collective information(MI) was employed to determine the differentiating features. MI is a measure of how important information about one set of data can be determined from another set of data.20 In this analysis, the features with higher MI values were likely to be more instructional for differencing HAE from non-HAE patients. After the MI of individual features or their combinations was computed, the process of point selection began. The thing of feature selection was to define the lowest subset of features that inclusively contain utmost of the mutually participated information and therefore most clearly define the characteristics of the patient with HAE. Machine learning algorithms drove the analysis of feature selection that created a model of HAE. Therefore, the model comported of the smallest possible and contemporaneously utmost screening characteristics of patients with HAE, resulting in an enhanced patient definition.

Stage 3 – Prediction

Once a model of the characteristics of the patients with HAE was determined from the index patients with HAE, the remaining population of patients in the data set was scored by the model to find undiagnosed patients. For every remaining case in the data set, the first step in scoring was to compute the features that didn't appear in the set of index patients with HAE. Each case's features were input to the HAE model, which produced a numerical score. This score represented the prospect that the patient had undiagnosed HAE, and patients were ranked from most probably to least likely to have the condition.

RESULTS

Stage 1 - Patient Definition

Searching the 2014 - 2020 Market-Scan database for all patients recommended two C1 inhibitors (Cinryze, Berinert), incatibant (Firazyr), and ecallantide (Kalbitor) revealed 1002 index patients with HAE.

Stage 2 - Model Creation

The histories of the patients linked in Stage 1 were analyzed to determine the treatment, procedural, diagnostic, and healthcare provider characteristics prior to getting definitive treatment for HAE. By comparing the characteristics of patients with HAE with those of demographically matched non-HAE patients, machine learning algorithms named the characteristics that were most descriptive and predictive of eventual HAE opinion by a practitioner.[22] - [23]

These characteristics were linked and refined, forming the enhanced patient definition characteristics listed in the Table encompassing diagnosis, procedures, treatment, and providers. Note that some nonadjacent lines of descriptive text appear identical within the table. These are associated with different ICD-9-CM codes depending on the provider's position of involvement with the patient; therefore, they appear further than formerly in the table with different ranking. [22] - [23]

Stage 3 - Prediction

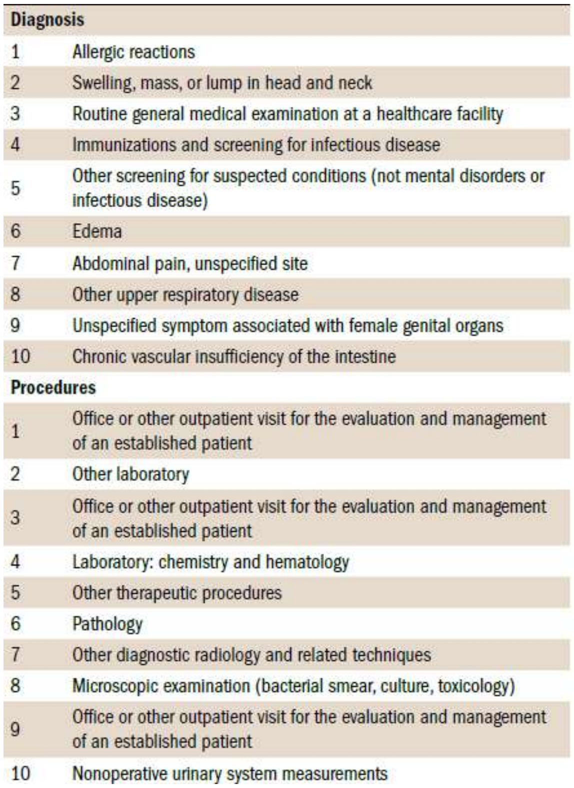
A model of the angioedema patient's history and summary, based on the enhanced patient definition determined in Stage 2, was applied to the remaining population of patients in the database. With the prediction classifier set to a discovery probability>0.8 in Stage 3 of this analysis, applying the model to the remaining population indicated 5511 potentially undiagnosed patients with HAE.

Although the data in the database is de-linked, the patient information in the database is linked to metropolitan statistical areas (MSAs) to understand the geographic distribution of the information. The Office of Management and Budget defines MSAs for use by government statistical agencies. [21] The distribution of the prognosticated HAE patients across the United States is depicted in the chart in Figure 2.

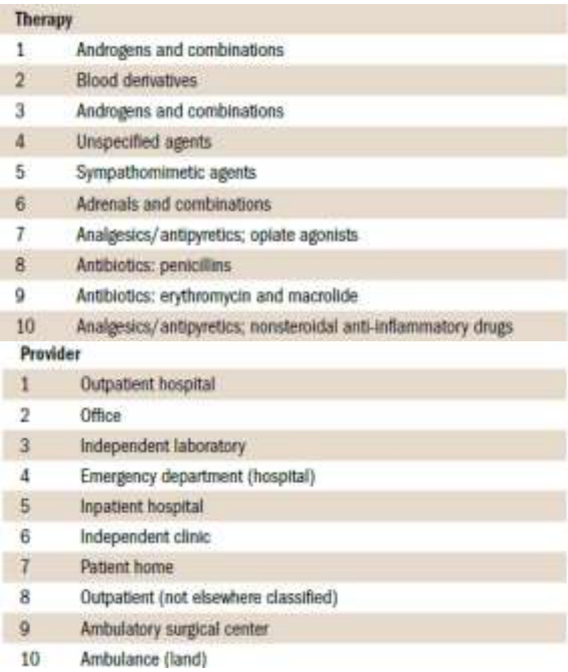
DISCUSSION

The major contributions of this study relate to the identification of patients with a rare disease, the insight method used, and the source of data used. Together, these data elements define a new environment in which payers, practitioner, and ca patients may benefit from the value still locked down in big healthcare data. As healthcare gradationally embraces the value of data analytics, it still struggles with overcoming access and translucency issues with regard to using patient records.[6] – [8] As a result, numerous of the sweats to apply predictive analytics are directed at the EHRs of a single institution or network. This produces results that may have limited applicability to other health provider systems, are based on a limited population size, and frequently concentrate on furnishing rapid-fire feedback to warn the healthcare provider of implicit care issues. The frequency and ubiquity of these cautions, frequently with limited practical value, has been known to produce "alert fatigue," which results in some healthcare providers ignoring these warnings, therefore further diminishing the usefulness of these predictive analytics. As a result, the thing of these analytics to give decision- making mechanisms that maximize the value of medical care aren't completely realized By utilizing a de-identified claims database compliant with the Health Insurance Portability and Accountability Act of 1996, walls to translucency and sharing are overcome. The size of the databases further than 170 million patients in this study and ensures confidence in the applicability of the issues to the US population and in the significance of the results. Analyzing a claims database rather than EHRs may allow some payers to more effectively concentrate an analysis on uncommon disease states to answer questions about the best possible ways to cover costs while maximizing care options for the population they serve.

The Top 10 Diagnostic, Procedural, Therapeutic, and Healthcare Provider Characteristics Most Predictive of Heredity Angioedema Diagnosis



Diagnosis	
1	Allergic reactions
2	Swelling, mass, or lump in head and neck
3	Routine general medical examination at a healthcare facility
4	Immunizations and screening for infectious disease
5	Other screening for suspected conditions (not mental disorders or infectious disease)
6	Edema
7	Abdominal pain, unspecified site
8	Other upper respiratory disease
9	Unspecified symptom associated with female genital organs
10	Chronic vascular insufficiency of the intestine
Procedures	
1	Office or other outpatient visit for the evaluation and management of an established patient
2	Other laboratory
3	Office or other outpatient visit for the evaluation and management of an established patient
4	Laboratory: chemistry and hematology
5	Other therapeutic procedures
6	Pathology
7	Other diagnostic radiology and related techniques
8	Microscopic examination (bacterial smear, culture, toxicology)
9	Office or other outpatient visit for the evaluation and management of an established patient
10	Nonoperative urinary system measurements



Therapy	
1	Androgens and combinations
2	Blood derivatives
3	Androgens and combinations
4	Unspecified agents
5	Sympathomimetic agents
6	Adrenals and combinations
7	Analgesics/antipyretics; opiate agonists
8	Antibiotics: penicillins
9	Antibiotics: erythromycin and macrolide
10	Analgesics/antipyretics; nonsteroidal anti-inflammatory drugs
Provider	
1	Outpatient hospital
2	Office
3	Independent laboratory
4	Emergency department (hospital)
5	Inpatient hospital
6	Independent clinic
7	Patient home
8	Outpatient (not elsewhere classified)
9	Ambulatory surgical center
10	Ambulance (land)

An opportunity arose to sustain the computed prediction of undiagnosed patients. An apprise to the Market Scan database covering the 11 months from January 2015 through November 2015 was scanned for further information about patients with hypothetically undiagnosed HAE. An aggregate of 888 of the prognosticated 5511 undiagnosed cases were set up to have new healthcare information during those 11 months.

Of those 888 implicit HAE patients, 14 had new claims data canons for HAE, therefore affirming the applicability of the computed predictive model. Although this doesn't constitute statistically rigorous confirmation of the model, evidence of diagnosis in these 14 patients prognosticated to have HAE suggests the implicit power of this model to have an impression on cost and product management for rare and hard-to-diagnose diseases.

This study concentrated the power of state- of- the- art analytics on a single rare disease HAE, because HAE shares numerous disease features associated with high costs rare, hard to diagnose, progressive, and takes a long time from diagnosis to applicable treatment. During the 8 years it may take to diagnose a patient with HAE, patients constantly visit EDs, are admitted for surgical center stays, and receive inappropriate costly procedures.15 Earlier diagnosis and treatment would remove patients from the cycle of expensive procedures, ineffective treatment that drives them back for further of the same disease therefore reducing waste and improving patient treatment results. With3.2 million implicit patients with rare diseases in the United States, predictive analytics applied to insurance claimed databases to classify them could open the door for payers to help practitioners maximize effects and value in the care of these patients population. [14]

Machine learning algorithms used in this study have crossed over from other disciplines, similar as defense and business, that are formerly demonstrating the flexibility and adaptability essential in their design.[10] – [13] This study successfully demonstrated the capability of this state- of- the- art predictive analysis to find rare disease patients in a large and complex insurance database. A logical tool with the flexibility to be applied to a variety of data sources and to specifically identify patient subpopulations of interest to payers or healthcare institutions, has the implicit to be an important power in the evolving healthcare geography. Using such a tool, payers may realize cost savings from relating patients with expensive conditions and from taking way to insure that their care is managed appropriately. Physicians may be better informed about how to manage these patients, and patients would get the care best suited for their requirements.

The methods used in this analysis could apply to open and closed healthcare systems, both large and small. Large healthcare systems that invest in state- of- the- art predictive analytics would have tools at the ready to answer critical questions about how their patient population requirements are being met and how their costs are apportioned. More importantly, similar tools could give insight into what might be done to meet patients changing requirements and respond efficiently to the demands of the evolving managed care system. There's eventuality for lower indigenous systems and individual health plans or employers to apply lessons learned from published analyses of larger systems to guide examination of their own data. Applying the methods used then to other changing that are rare, hard to diagnose, progressive, and take a long time from diagnosis to applicable treatment has the potential to prospect payers, practitioners , and patients in the responsible care environment.

Limitations

Certain limitations to this analysis should be on considered. Despite being generally representative of the US population, the MarketScan database is composed of data from a subset of the US population and therefore isn't an arbitrary sample. [9], [22], [23]

The data come substantially from large employers, so medium and small organization data aren't represented. [22] Organizational claims data generally contain some rendering inaccuracies and missing data, which might affect in misclassification or other bias. Also, although self-validating cross-checks were incorporated as part of developing the logical model, real- world validation of the identification of these undiagnosed patients isn't complete.

CONCLUSIONS

This analysis successfully predictive the capability of this state- of- the- art prophetic analysis to find implicit rare disease patients in a large and complex database. Machine learning techniques applied to de-identified claims database are easily able of relating these undiagnosed and erroneously treated patients. This information could be precious to claims directors and employers who may realize savings by helping physicians bring these patients to applicable treatment sooner. The potential exists to apply this method to other diseases that are rare, hard to diagnose, progressive, and may take a long time from opinion to applicable treatment. It's a lesson for managed care providers of all types that new data analysis and patient isolation methods are applicable to the patient populations they manage. It's time for healthcare to join other data intensive reveal the value in embracing technologies that reveal the value in the largest asset they manage information.

REFERENCES

[1]. Manyika J, Chui M, Brown B, et al; McKinsey Global Institute. Big data: the next frontier for innovation, competition, and productivity. McKinsey & Company website. http://www.mckinsey.com/business-functions/business-technology/our-insights/bigdata-The-next-frontier-for-innovation. Published May 2011. Accessed June 22, 2016.

[2]. McAfee A, Brynjolfsson E. Big data: the management revolution. *Harvard Business Review* website. https://hbr.org/2012/10/big-data-the-managementrevolution. Published October 2012. Accessed June 22, 2016.

[3]. Naidus E, Celi LA. Big data in healthcare: are we close to it? *Rev Bras Ter Intensiva*.2016;28(1):8-10. doi:10.5935/0103-507X.20160008.

[4]. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)*. 2014;33(7):1123-1131. doi:10.1377/hlthaff.2014.0041.

[5]. Parikh RB, Obermeyer Z, Bates DW. Making predictive analytics a routine part of patient care. *Harvard Business Review* website. https://hbr.org/2016/04/making-predictive-analytics- a-routine-part-of-patient-care. Published April 21, 2016. Accessed June 28, 2016.

[6]. Amarasingham R, Audet AM, Bates DW, et al. Consensus statement on electronic health predictive analytics: a guiding framework to address challenges. *EGEMS (Wash DC)*. 2016;4(1):1163. doi:10.13063/2327-9214.1163.

[7]. Using real-world evidence to accelerate safe and effective cures: advancing medical innovation for a healthier America. Bipartisan Policy Center website. Published June 2016. Accessed June 27**,** 2016.

[8]. Doshi JA, Hendrick FB, Graff JS, Stuart BC. Data, data everywhere, but access remains a big issue for researchers: a review of access policies for publicly funded patient-level health care data in the United States. *EGEMS (Wash DC)*.

2016;4(2):1204. doi:10.13063/2327-9214.1204.
[9]. MarketScan Research Databases. Truven Health website. http://truvenhealth.com/yourhealthcare-focus/analytic-research/marketscan-research-databases. Accessed June 28, 2016.

[10]. Klein A. Police enlist war tech in crime fight. *Washington Post* website. https://www.washingtonpost.com/local/police-enlist-war-tech-in-crimefight/2013/02/18/0a9e18e2-6bc6-11e2-ada0-5ca5fa7ebe79_story.html. Published February 18, 2013. Accessed June 29, 2016.

[11]. Ward MJ, Marsolo KA, Froehle CM. Applications of business analytics in healthcare. *Bus Horiz*. 2014;57(5):571-582. doi:10.1016/j.bushor.2014.06.003.
[12]. Wood C. How does the military use big data? Emergency Management website. http://www.emergencymgmt.com/safety/Military-Use-Big-Data.html. Published January 6, 2014. Accessed June 29, 2016.

[13]. Custom Strategies/IBM Government Analytics Forum. Putting predictive analytics to work for the army—an executive perspective. Government Executive website. http://www.govexec.com/govexec-sponsored/2015/04/putting-predictive-analytics-work-armyexecutive-perspective/111406/. Published April 30, 2015. Accessed June 29, 2016.

[14]. Rare diseases: facts and statistics. Global Genes website. http://globalgenes.org/rare-diseases-facts-statistics/. Published January 1, 2012. Accessed June 2, 2016.

[15]. Ali MA, Borum ML. Hereditary angioedema: what the gastroenterologist needs to know. *Clin Exp Gastroenterol*. 2014;7:435-445. doi:10.2147/CEG.S50465.
[16]. Lumry WR, Castaldo AJ, Vernon MK, Blaustein MB, Wilson DA, Horn PT. The humanistic burden of hereditary angioedema: impact on health-related quality
of life, productivity, and depression. *Allergy Asthma Proc*. 2010;31(5):407-414.
doi:10.2500/aap.2010.31.3394.

[17]. Riedl M. Recombinant human C1 esterase inhibitor in the management of hereditary angioedema. *Clin Drug Investig*. 2015;35(7):407-417. Review. doi:10.1007/s40261-015-0300-z.

[18]. Agostoni A, Aygören-Pürsün E, Binkley KE, et al. Hereditary and acquired angioedema: problems and progress: proceedings of the third C1 esterase inhibitor deficiency workshop and beyond. *J Allergy Clin Immunol*. 2004;114(suppl 3):S51-S131. doi:10.1016/j.jaci.2004.06.047.

[19]. Gómez-Traseira C, Pérez-Fernández E, López-Serrano MC, et al. Clinical patternand acute and long-term management of hereditary angioedema due to C1-esteraseinhibitor deficiency. *J Investig Allergol Clin Immunol*. 2015;25(5):358-364.

[20]. Ross BC. Mutual information between discrete and continuous data sets. *PloS One*. 2014;9(2):e87357. doi:10.1371/journal.pone.0087357.

[21]. Metropolitan and micropolitan statistical areas main. US Census Bureau website. http://www.census.gov/population/metro/. Updated July 2015. Accessed September 12, 2016.

[22]. Hansen LG, Chang S. Health research data for the real world: The MarketScan Databases. Truven Health website. http://truvenhealth.com/portals/0/assets/PH_11238_0612_TEMP_MarketScan_WP_FINAL.pdf.

[23]. MJH Life Sciences™ , Pharmacy Times – Pharmacy Practice News and Expert Insights. https://www.pharmacytimes.com/