Artificial Intelligence in Medicine 74 (2016) 1–8

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Non-obvious correlations to disease management unraveled by Bayesian artificial intelligence analyses of CMS data

Vijetha Vemulapalli[a,*], Jiaqi Qu[b], Jeonifer M. Garren[a], Leonardo O. Rodrigues[a], Michael A. Kiebish[a], Rangaprasad Sarangarajan[a], Niven R. Narain[a], Viatcheslav R. Akmaev[a]

[a] Berg Analytics Division, Berg. 500 Old Connecticut Path, Building B, Framingham, MA 01701, USA
[b] BNP Paribas, 525 Washington Boulevard, Jersey City, NJ 07310, USA

ARTICLE INFO

Article history:
Received 25 July 2016
Received in revised form 1 November 2016
Accepted 7 November 2016

ABSTRACT

Objective: Given the availability of extensive digitized healthcare data from medical records, claims and prescription information, it is now possible to use hypothesis-free, data-driven approaches to mine medical databases for novel insight. The goal of this analysis was to demonstrate the use of artificial intelligence based methods such as Bayesian networks to open up opportunities for creation of new knowledge in management of chronic conditions.
Materials and methods: Hospital level Medicare claims data containing discharge numbers for most common diagnoses were analyzed in a hypothesis-free manner using Bayesian networks learning methodology.
Results: While many interactions identified between discharge rates of diagnoses using this data set are supported by current medical knowledge, a novel interaction linking asthma and renal failure was discovered. This interaction is non-obvious and had not been looked at by the research and clinical communities in epidemiological or clinical data. A plausible pharmacological explanation of this link is proposed together with a verification of the risk significance by conventional statistical analysis.
Conclusion: Potential clinical and molecular pathways defining the relationship between commonly used asthma medications and renal disease are discussed. The study underscores the need for further epidemiological research to validate this novel hypothesis. Validation will lead to advancement in clinical treatment of asthma & bronchitis, thereby, improving patient outcomes and leading to long term cost savings. In summary, this study demonstrates that application of advanced artificial intelligence methods in healthcare has the potential to enhance the quality of care by discovering non-obvious, clinically relevant relationships and enabling timely care intervention.

1. Introduction

In 1960 the United States spent 5.2% of gross domestic product (GDP) on healthcare and by 2004 that number has risen to 16%. Current predictions have healthcare expenditures exceeding 18.7% of GDP today and forecast to increase to over 20% by 2025. There is no doubt of a healthcare crisis unraveling in the United States and in other parts of the world, driven by the rapid increase in the prevalence of chronic diseases such as diabetes and kidney disease. In addition to increasing prevalence of certain diseases, this increase in health care costs can be attributed to new and prohibitively expensive therapeutics for cancer and other disorders, and ever escalating costs of patient mismanagement and malpractice. It is apparent that the returns from the current paradigm of innovation in healthcare have diminished dramatically. Traditional approaches of hypothesis driven clinical research are becoming less effective in translating empirical data to successful health outcomes. Given the rapid growth in the rates of accumulation of healthcare data, every vertical of the global healthcare industry from therapeutics research and development to epidemiology and precision medicine would benefit from data driven, unbiased approaches. Mathematical and statistical tools developed in the field of artificial intelligence (AI) and machine learning are well poised to assist clinical researchers in deciphering complex predictive patterns in healthcare data.

* Corresponding author.
E-mail address: vijetha.vemulapalli@berghealth.com (V. Vemulapalli).

http://dx.doi.org/10.1016/j.artmed.2016.11.001

With government incentives offered to clinical organizations to transition from paper based patient information to well-structured and managed digital form, there has been a tremendous explosion in the availability of patient-centric healthcare data. Such data can be leveraged to open up new avenues in advancing healthcare by improving patient care and creating new efficiencies in delivering care [1–4]. Understanding variation in treatment outcomes due to patient specific molecular and clinical factors [5–7] is essential to the practical implementation of and adoption of precision medicine in clinical practice.

The predominant focus of the analytical efforts in health care has been structured management and compilation of disparate data modalities and data mining with the objective of testing hypotheses generated through reasoning. These efforts are limited by the current thinking in medicine and biology, and often ignore unbiased, observational and "unknown biological" evidence. A data-driven hypothesis generation approach creates a paradigm shift in healthcare by leading to discovery of new and often surprising trends in clinical outcomes. Identification of non-obvious correlations and causal relationships may lead to a significant mitigation of side effects, need for additional medications, reduction of hospitalizations, decrease in unnecessary care, loss of income for patients, and eventually in the overall change in clinical practice.

This study demonstrates how high-level, publicly available healthcare data, in particular, data coming from the United States Centers for Medicare and Medicaid Services (CMS) can be leveraged to generate profound and surprisingly powerful hypotheses in patient management and disease outcome by employing advanced learning methods such as Bayesian networks [8–13] (BNs). BNs or Bayesian artificial intelligence is a mathematical framework for learning probabilistic cause and effect relationships directly from data. The factor or variable relationships learned as directed acyclic graphs. Learning BNs in large data sets is typically an NP-hard optimization problem where an enormous number of possible graphs are attempted in a heuristically driven process of finding locally optimal structures. In this case study, the Diagnosis Related Group (DRG) codes are the factors and the provider specific discharge numbers are the observations.

The approach presented here is a data analytics pipeline developed to work with population claims data on the *"big data"* scale. A network of cause and effect relationships between medical diagnoses was learned in a purely data-driven manner from previously published billing data from CMS. A follow up network analysis revealed novel comorbidities, disease progression factors and the sequence of disease diagnoses for a number of chronic diseases. The findings were confirmed by subsequent association-based statistical analysis using logistic regression models. BNs are a powerful tool that can be readily used in the analysis of big data and presents a dynamic approach to the hypothesis driven epidemiological research. These newly discovered non-trivial relationships might have a significant impact on clinical disease management. The increased utilization of technologies such as AI, in healthcare research, is likely to have a significant impact on clinical disease management and global healthcare efficiency.

2. Methods

The methods overview and the data analysis workflow are shown in Fig. 1. Data was obtained from publicly released files from the CMS website [14]. Top 100 most frequently billed diagnosis codes in the year of 2011 for a majority of healthcare providers in the United States were assessed in the data along with the high level provider information, amount charged and amount paid to the provider by the Medicare insurance plan. The wide discrepancy of the provider charges across the United States had been analyzed, presented, and discussed by the media [15–18]. The focus of this analysis was the relationship between the annual incidences of diagnoses across the largest hospitals in the United States. In this data set, the diagnosis of the discharged patients was represented by the DRG codes. Fields containing information about the DRG codes, total number of discharges in 2011 for each diagnosis and the provider code were extracted from the dataset.

The initial step in the data processing workflow was to organize the discharge count information as a data matrix of DRG codes across provider codes. Next, the DRG codes missing in more than 70% of the providers were removed from further analysis along with the provider codes not reporting any patient discharges for more than 25% of the DRG codes in the data set. After the filtering procedure, the dataset consisted of 100 DRG code discharges across 1618 hospitals. The total number of discharges for each provider was used to normalize the number of discharges per a DRG code. In the follow up analysis, it was assumed in the case of a missing data point that no patients were discharged with that diagnosis code from the facility in 2011. Therefore, the remaining missing data were imputed as zeros. Median polish normalization was performed on the data matrix and the data was transformed to the logarithmic scale. All data processing steps up to this point were performed using R (http://www.r-project.org/) [19]. Details of data processing are discussed in Supplementary file 1.

The processed and normalized data matrix was analyzed using the BN learner, bAIcis^{TM} (Berg LLC, Framingham, Massachusetts, U.S.A.). An ensemble of 20 BNs was learned directly from the data using stochastic optimization methods: Each network in the ensemble was created by optimizing the global network structure with respect to the network Bayesian information criterion score from graph families represented by a single linear regression model with up to four regressors. Each DRG code was modeled as a node in the network with the normalized number of discharges across the provider data treated as the variable observations. A final summarized BN was created by performing in silico interventions in the ensemble for all learned node relationships [20]. Such simulations are helpful in determining the strength of the cause and effect link and reducing the number of spurious connections and cross-correlations. In the final step, a confidence metric was assigned to each reported structural relationship in the learned BN ensemble based on the c-statistic measuring the difference in the posterior distributions of the baseline and intervention data. Namely, each driver node in a structural relationship was perturbed 10 fold relative to its baseline mean a 1000 times for each of the ensemble networks. The two posterior distributions for the dependent node (the baseline and the perturbed) were compared using the c-statistics or the area under the receiver operating characteristic curve. Relationships in the resulting DRG code network ensemble were retained if the strength of the connection measured by the c-statistic was larger than or equal to 0.8. Nodes representing related DRG codes were combined. For example: DRG codes that only vary based on presence of complicating conditions. The resulting network ensemble was visualized using Cytoscape (http://www.cytoscape.org/) [21–23]. Table with results of simulation for the selected sub network are shown in supplementary table A.

3. Results and discussion

The DRG code network learned from the CMS data (Fig. 2a) connected 61 DRG codes in a network of 178 connections. If two nodes are connected in the network, it means that a change in the number of discharges for one DRG code affects the number of discharges for the second DRG code with the arrows indicating the directionality of the cause and effect relationship. Interpretation of the meaning of these connections depends on specific diagnoses involved. For

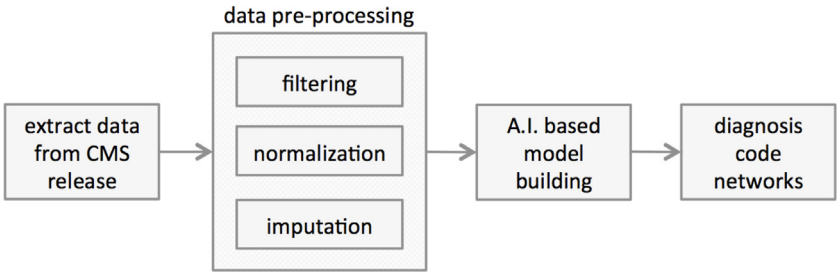

Fig. 1. Data processing workflow.

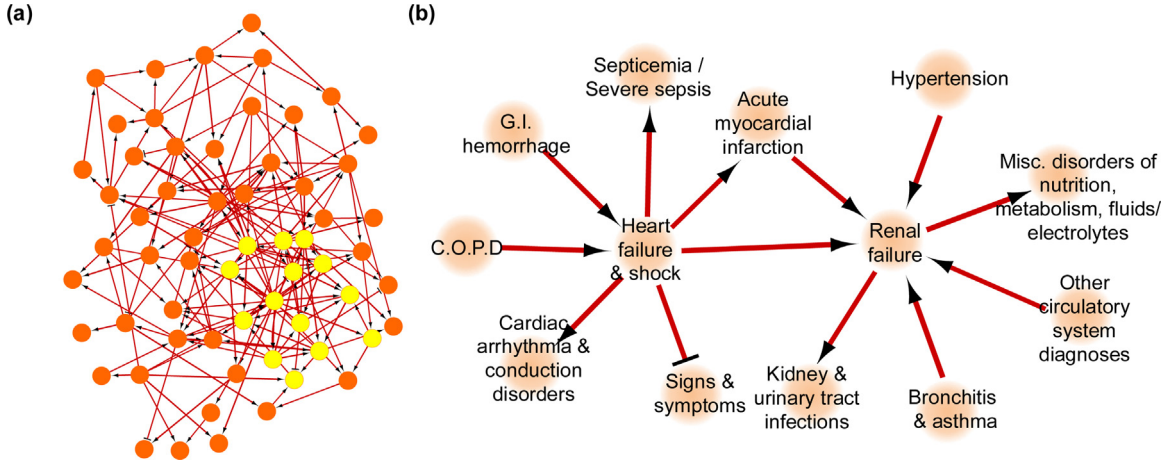

Fig. 2. (a) The panel shows the entire diagnosis code network with the selected subnetwork highlighted. (b) The panel shows the heart failure and renal failure subnetwork. Each node corresponds to a DRG code and edges show interactions between the DRG discharge totals. The thickness of the arrow indicates the strength of the connection. "T" shaped connections indicate that number of diagnosis for the nodes are inversely correlated.

example, when the source node is diabetes and the target node is hypertension, the connection likely represents comorbidities at a population level. When the source is diabetes and the target is neuropathy, the connection likely represents complications from disease.

Subnetworks centered on 'heart failure & shock' and 'renal failure' were selected (Fig. 2b) for further exploration. These diagnoses were selected based on the cause of death index compiled by the Centers for Disease Control and Prevention [24]. According to this report, heart disease was the top cause of death in 2011. Kidney related conditions were ranked 9, 12, and 13. All renal conditions combined were noted as major causes of death. Therefore, the 'heart failure' and the 'renal failure' subnetworks were analyzed in detail. The goal of the subnetwork analysis was to classify the connections in the subnetwork as known or predicated on information in support of the cause and effect already well established in the clinic, or novel with little to no reference in the scientific or medical literature. Presence of known connections in the networks demonstrates that relationships inferred statistically in a BN structure are valid and can be confirmed by clinical or epidemiological research. Thus, novel connections can be considered to be the basis for formulation of new hypotheses generated in a strictly data driven manner. Based on the analysis it was found that many connections in the 'heart failure' and 'renal failure' subnetworks were known while novel interactions were also identified, in particularly, in the 'renal failure' subnetwork.

3.1. Associations correlated by medical literature

Heart failure is a plurality of conditions that reduce the ability of the heart to pump blood efficiently. These conditions include congenital heart defects, arrhythmias, coronary artery disease that narrow arteries over time and high blood pressure that in combination with heart muscle unable to effectively pump blood can compromise the blood flow in the body. The interaction between 'heart failure and shock' and 'simple pneumonia and pleurisy' was removed from the subnetwork analysis because of inconsistencies in the interaction between these diagnosis codes in the ensemble of networks. Diagnosis relationships discovered from data in this study within the 'heart failure' subnetwork are explored below:

3.1.1. Chronic obstructive pulmonary disease

Pulmonary hypertension is a well-established complication observed in patients with chronic obstructive pulmonary disease. In situations when the lungs try to compensate for low oxygen concentration in blood due to inadequate lung ventilation, the compensatory effect is the observed increase in blood pressure inside the lungs. Increase in blood pressure inside the lungs (pulmonary hypertension) leads to increased stress on the right ventricle that can eventually cause heart failure.

3.1.2. Cardiac arrhythmia and conduction disorders

These disorders could result in heart failure. Therefore diagnosis of heart failure correlates to diagnosis of causal conditions discovered through the network. Based on output from the data analysis, it can be surmised that potential increase in the number of heart failure diagnoses will correspond to the increase in the number of diagnoses of cardiac arrhythmias and conduction disorders in the same patient population.

3.1.3. Gastrointestinal hemorrhage

Heart failure can be a result of hypovolemic shock. Hypovolemic shock occurs due to rapid loss of large blood volume that can result in the failure of multiple organs including the heart. The most

frequent causes of hemorrhagic (hypovolemic) shock are trauma, gastrointestinal hemorrhage and organ injury.

3.1.4. Renal failure

The link between anemia, cardiac problems and renal disease is well known[25–27] and the challenges of treating patients with cardio-renal insufficiencies are well documented [28]. About one-fourth of the patients with renal disease have congestive heart problems and as renal disease worsens, the fraction of patients with heart disease increases to about 65–70% [25]. Large studies have shown that worsening kidney disease is associated with higher mortality and hospitalization rates in patients with previous diagnosis of heart failure [29]. Thus, there is significant evidence supporting the relationship between heart failure and renal failure.

3.1.5. Acute myocardial infarction

Heart failure is known to be a frequent complication of myocardial infarction [30] and hence is a well-established connection. This interaction has been well studied through epidemiological studies and it is also known that the rate of heart failure after myocardial infarction decreases with time [31].

The diagnosis code corresponding to 'renal failure' in this analysis encompassed chronic/acute kidney failure and other renal disorders. Renal failure/insufficiency refers to reduction in the kidney's ability to remove waste products from blood. More than 10% of adults 20 years or older have chronic kidney disease (CKD) [32] The cost of treating CKD is very high because of the wide range of comorbidities and quality of life factors [33]. One study showed that the cost for treating end stage renal disease is continually increasing and CMS costs for this condition have reached 30 billion dollars for 2009 [34]. Known medical relationships within the renal failure subnetwork are explored below:

3.1.6. Kidney and urinary tract infections

This diagnosis code encompasses many types of renal infections including cystitis, abscess, lower and upper urinary tract infections. Recurrent or untreated urinary tract and kidney infections are known to be linked with kidney scarring which contributes to lowered kidney function and renal failure [35–37].

3.1.7. Disorders of nutrition, metabolism, and fluids/electrolytes

Kidneys play an important role in maintaining the fluid and electrolyte balance. Therefore it is possible that a diagnosis of renal failure would lead to follow up tests and diagnosis of nutrition, metabolism, fluid and electrolyte balance problems.

3.1.8. Simple pneumonia and pleurisy

It is established that chronic kidney disease increases susceptibility to infections and that incidence of pneumonia in chronic kidney disease is a well documented infectious complication associated with kidney disease [38,39].

3.1.9. Hypertension

Hypertension or high blood pressure is known to be the second leading cause of kidney failure. It is thought that high blood pressure damages blood vessels in the kidneys hence reducing the ability of the kidneys to remove waste products from the blood [40].

The above investigation of the medical literature vis-à-vis the cause and effect relationships discovered by the BN learner confirms the validity of application of the AI based mathematical methodology to data driven discovery of medical outcome patterns from high level billing data.

3.2. Discovery of a novel relationship in a patient population: linking asthma to renal failure

The 'renal failure' DRG code was inferred to interact with the 'bronchitis and asthma' diagnoses. There is limited information in the medical literature that supports a direct link between incidence of bronchitis/asthma and renal failure. Thus, the asthma and renal failure relationship inferred from this study may be considered a novel discovery that can be linked to medication effects, disease comorbidities or the order of diagnosis based on the dataset used in the analysis. Specific diagnosis codes for this interaction were identified and a linear regression model was built for these variables to confirm the BN finding and identify the significance of the interaction. Significant statistical association between the number of renal failure diagnoses and the number of bronchitis/asthma diagnoses was demonstrated by conventional statistical analysis in the same data set with p-value of less than 1e-5 and the adjusted r^2 of 25%. Moreover, the bronchitis/asthma factor was significant in the multivariate linear regression model controlling for the other predictors in Fig. 2b with p-value 0.0014.

In summary, a novel hypothesis was generated in an entirely data-driven manner: The incidence of renal failure may potentially be associated with asthma or bronchitis, or perhaps the asthma treatment strategies leading to a direct correlation of the number of asthma/bronchitis discharges to the number of renal failure discharges. Below is a summary description of the exploratory findings in the scientific and medical literature potentially linking asthma and renal failure.

3.3. Scientific evidence in support of the novel linkage between asthma and renal failure

A clinical hypothesis and likely corresponding molecular mechanism is summarized in Fig. 3a and b summarizes potential molecular mechanisms linking the effect of drugs used in the treatment of bronchitis/asthma to their ability to potentially cause renal dysfunction and/or renal failure. Below is a detailed description highlighting the direct and indirect effect of β_2-adrenergic agonists on renal function and its potential to cause renal dysfunction and/or renal failure.

Bronchitis and asthma are respiratory diseases where narrowing of airways results in cough, shortness of breath, tightness in the chest and wheezing. There are many choices of medication currently available to treat the symptoms as well as key pathways involved in the etiology of asthma (Table 1). One of the oldest and most widely used drug classes for symptomatic treatment of asthma is β_2-adrenergic agonists that are referred to as bronchodilators [41]. As shown in Table 1, 7 out of top 10 medications used to treat asthma contain either a long acting or a short acting β_2-adrenergic agonists. β_2- adrenergic agonists interact with the β_2 sub-type of the β-adrenoceptors. Most β_2-adrenergic agonist bronchodilators in asthmatic patients are primarily administered via use of a nebulizer to ensure maximal retention of the drug within the lung environment, the primary site of action. However, the rich vasculature within the lungs enables systemic absorption of the β_2-adrenergic agonists [42]. Thus, the physiological and pharmacological effect of β_2-adrenergic agonists can be manifested systemically in multiple organ systems including heart, skeletal muscle, kidneys expressing receptors [42–44].

This study explores direct and indirect effects of β_2-adrenergic agonists that might potentially lead to renal dysfunction. Most of the β_2-adrenergic agonists used in treatment of asthma are primarily removed from the circulation in the kidneys [45]. The kidneys have widespread expression of the β-adrenergic receptors within the glomerulus and tubular epithelium influencing the hemodynamics and electrolyte balance [46]. Electrolyte imbalances have

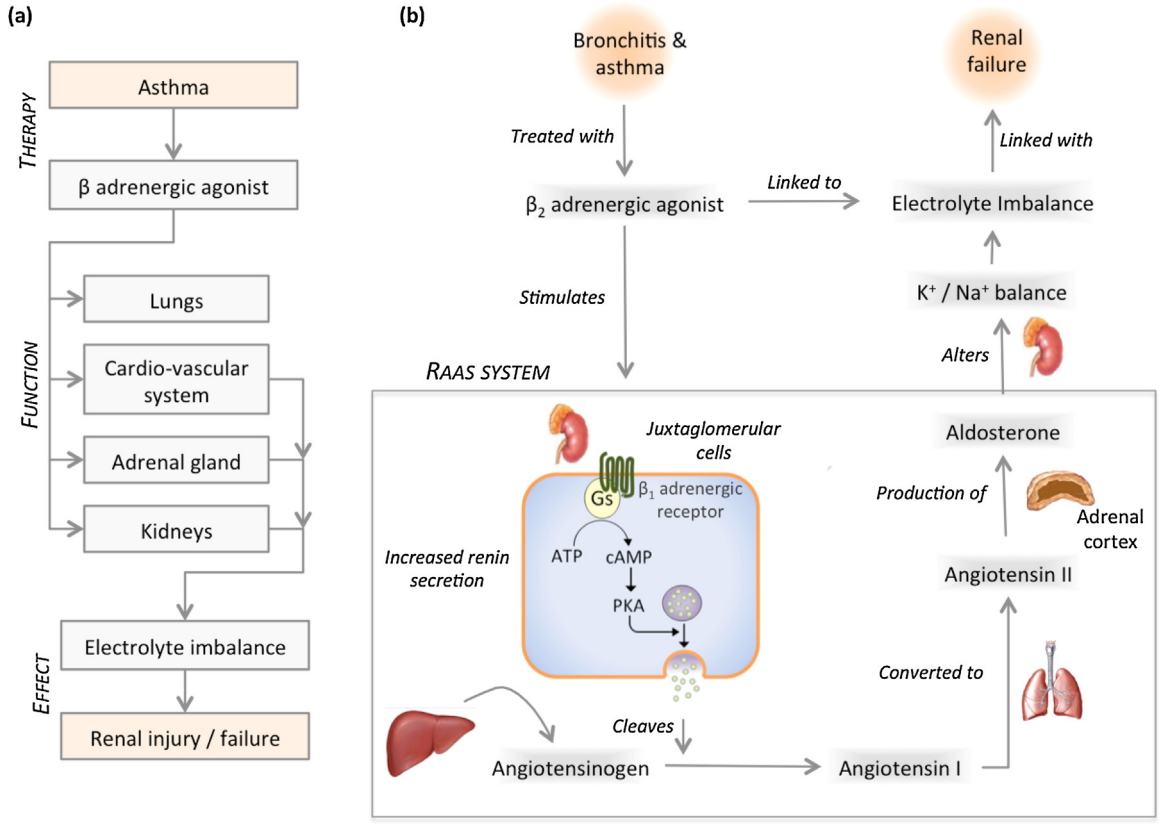

Fig. 3. A mechanistic diagram supporting a novel hypothesis of the association between asthma medications and renal dysfunction. (a) Pathway in this panel was constructed by linking results published clinical research. (b) Potential molecular mechanism supporting the clinical findings. Protein abbreviations: Gs (guanine nucleotide-binding regulatory protein), cAMP (cyclic adenosine monophosphate), PKA (protein kinase A), ACE (angiotensin converting enzyme inhibitor).

Table 1
Select clinical pharmacology of top 10 asthma drugs in 2011–12. Drug information was obtained from a drug index for prescription drugs [72]. Prescription count information was obtained from a survey by IMS heath [73].

	Drug	Prescriptions dispensed (thousands)	Active ingredient(s)	Active ingredient includes β2-adrenergic agonist	Hypokalemia
1.	Singular	28,110	Montelukast sodium	No	No
2.	Proair HFA	23,931	Albuterol sulfate	Yes (Short acting)	Yes
3.	Advair Diskus	17,534	Fluticasone propionate and salmeterol	Yes (Long acting)	Yes
4.	Ventolin HFA	16,272	Albuterol sulfate	Yes (Short acting)	Yes
5.	Albuterol sulfate	13,978	Albuterol sulfate	Yes (Short acting)	Yes
6.	Spiriva Handihaler	9416	Tiotropium bromide monohydrate	No	No
7.	Flovent HFA	6211	Fluticasone propionate	No	No
8.	Symbicort	4962	Budesonide and formoterol fumarate dihydrate	Yes (Long acting)	Yes
9.	Combivent	4251	Ipratropium bromide and albuterol sulfate	Yes (Short acting)	Yes
10.	Proventil HFA	4084	Albuterol sulfate	Yes (Short acting)	Yes

been noted in asthma patients treated with β_2-adrenergic agonists [47,48]. The β_2-adrenergic receptor has been demonstrated to regulate the sodium linked transporter expression [49] and it has been suggested that it is involved in salt sensitive hypertension [50]. Furthermore, β-adrenergic receptors are known to influence the synthesis and secretion of renin [51,52]. The fact that β_2-adrenergic agonists do not undergo significant biotransformation prior to excretion suggests that the drugs could exert pharmacological effects in the kidneys during the process of excretion.

One of the side effects associated with circulating β_2-adrenergic agonists such as albuterol is its ability to reduce circulating potassium levels and lead to hypokalemia [53–56]. Indeed, hypokalemia is a well-defined and often reported electrolyte imbalance observed in asthmatic patients on β_2-adrenergic agonists [53,56,57]. Persistent hypokalemia can lead to renal dysfunction and tubulointerstitial diease. It has been shown that in rats, hypokalemia induces renal injury [58,59] and, in humans, causes renal failure [60]. In a study of 55 patients, it was shown that chronic hypokalemia was accompanied by renal cystogenesis that resulted in scarring and other kidney damage leading to renal insufficiency [61]. Other studies have also shown that hypokalemia in patients with renal disease increased the rate of progression to end stage renal disease and also increased the rate of mortality [62–64].

Other β-adrenergic agonist effects on renal function include their ability to influence renin release. Studies have demonstrated that β_2-adrenergic agonists tend to activate the renin-angiotensin system (RAS) leading to increases in both the circulating plasma renin and angiotensin II [51,52,65]. Literature documenting the role of RAS inhibitor in conferring protection against chronic kidney disease and renal failure is extensive [66,67]. In addition to its ability to

induce hypokalemia, β_2-agonists mediate increase in aldosterone, and consequently angiotensin II. Aldosterone is a key component of RAS and leads to the adrenal gland secretion of angiotensin II generated within the RAS pathway. Activation of RAS and its components such as aldosterone and angiotensin II has been linked to renal dysfunction [46,68]. This hypothesis is further substantiated by the observation that blocking of aldosterone function is associated with improvement in kidney function [69].

The above evidence supports a role for β_2 adrenergic agonists in the potential etiology of renal failure in patients with asthma. Indeed, there is recent evidence in the literature identifying higher incidence of chronic kidney disease in patients with asthma although the cause is unknown [70,71].

3.4. Impact on healthcare economics and challenges

If the novel link between asthma medication and renal dysfunction identified from this study is confirmed through rigorous epidemiological research, approximately 25% of the diagnoses of asthma/bronchitis may be associated with potentially undesirable drug side effects in certain individuals. Further investigation and modification of treatment guidelines for asthma will not only lead to better patient care but also to considerable cost savings given the cost of annual treatment for a patient with end stage renal disease is estimated at around United States Dollar (USD) 70,000 [34]. Furthermore, clinical intervention will result in significant saving to the CMS that reimbursed USD 2.3 billion for the renal disease treatment shown in the studied data across 3000 hospitals in the United States enrolled in the Inpatient Prospective Payment System.

Given that healthcare costs have been skyrocketing in the recent past and containing costs is one of the major concerns in public health, new knowledge generated using methods such as those presented in this paper are invaluable. Novel insights into treatment complications using AI based methods will contribute to rapid expansion of medical knowledge and improvements in patient care.

3.5. Future direction

The correlation identified in this study would not be readily identified in the current clinical research settings or analyses from literature. This is because the interaction between a treatment of asthma and renal failure is non-obvious and is not currently subject of widespread studies; there has been little evidence until recently [70,71].

Herein, we propose some future directions for this research. The first step would be to use patient level diagnosis and pharmacy data to further test this hypothesis. With patient level diagnosis information, it will be possible to strengthen the link between diagnosis of asthma and diagnosis of renal disease. Use of pharmacy data in addition to diagnosis information will enable gathering more evidence that β_2-adrenergic agonist use is linked to loss of renal function. A combination of diagnosis and pharmacy data should be to study the effect of blood pressure medication such as non-specific β-blockers and angiotensin converting enzyme (ACE) inhibitors. This interaction will provide another pathway to strengthen the link between asthma drugs and renal complications since non-specific β-blockers and ACE inhibitors have the reverse effect of β_2-adrenergic agonists.

Validation of this hypothesis provides the foundation for incorporating routine checks on kidney function into the clinical practice for the treatment of asthmatic patients. Using longitudinal patient level data and genetic predisposition, one could develop a risk model to identify patients at high risk of renal damage and divert those individuals to other treatment options such as corticosteroids, leukotriene modifiers, mast cell stabilizers, anticholinergics, and others. Such treatment strategies can be incorporated into the clinic, through clinical decision support systems, by integrating patient electronic health records with a knowledge base, to provide truly personalized patient care.

4. Conclusions

AI mathematical methodology is a powerful set of tools in data science. The application of BNs to high level billing data from CMS has demonstrated its utility in uncovering non-obvious relationships in the data, in particular, a potentially critical interaction between β_2-adrenergic agonist asthma medications and renal dysfunction. This hypothesis generated purely from data gained more validity after a thorough review of the relevant medical literature. The review showed it is likely that the newly discovered causal link is real. A confirmatory epidemiological study is necessary to implement this finding in clinical practice. Results presented in this study lend strong support for the use of hypothesis free, data-driven methodology in “big data” approaches to healthcare research and management. Such methods are complementary to classical epidemiological research and may be able to provide the researchers with non-obvious and unconventional leads that are challenging to focus on in a hypothesis driven experimental design and analysis. The new perspective presented here creates an opportunity to make significant progress in personalization of the treatment pathways for some of the most prevalent diseases such as diabetes, hypertension, congestive heart failure, and cancer. In conclusion, the work presented here provides a rationale for employing use of Bayesian artificial intelligence algorithms for the analyses of disparate healthcare, socioeconomic, demographic, genetic, and even data from wearables to advance medical research thereby improving patient outcomes and reducing treatment costs.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.artmed.2016.11.001.

References

[1] Sun J, McNaughton CD, Zhang P, Perer A, Gkoulalas-Divanis A, Denny JC, et al. Predicting changes in hypertension control using electronic health records from a chronic disease management program. J Am Med Inform Assoc 2013;21:337–44, http://dx.doi.org/10.1136/amiajnl-2013-002033.

[2] Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform 2008;77:81–97, http://dx.doi.org/10.1016/j.ijmedinf.2006.11.006.

[3] Stewart W, Roy J, Sun J, Ebadollahi S. Clinical utility of machine learning and longitudinal EHR data. In: Dua S, Acharya UR, Dua P, editors. Mach. Learn. Healthc. Informatics SE – 10, vol. 56. Berlin, Heidelberg: Springer; 2014. p. 209–27, http://dx.doi.org/10.1007/978-3-642-40017-9_10.

[4] Sim JJ, Handler J, Jacobsen SJ, Kanter MH. Systemic implementation strategies to improve hypertension: the kaiser permanente southern california experience. Can J Cardiol 2014, http://dx.doi.org/10.1016/j.cjca.2014.01.003.

[5] Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. J Am Med Inform Assoc n.d.;16:371–9. 10.1197/jamia.M2846.

[6] Amarasingham R, Patel PC, Toto K, Nelson LL, Swanson TS, Moore BJ, et al. Allocating scarce resources in real-time to reduce heart failure readmissions: a prospective, controlled study. BMJ Qual Saf 2013;22:998–1005, http://dx.doi.org/10.1136/bmjqs-2013-001901.

[7] Oetjens M, Bush WS, Birdwell KA, Dilks HH, Bowton EA, Denny JC, et al. Utilization of an EMR-biorepository to identify the genetic predictors of calcineurin-inhibitor toxicity in heart transplant recipients. Pac Symp Biocomput 2014;25:253–64.

[8] Pe'er D, Regev A, Elidan G, Friedman N. Inferring subnetworks from perturbed expression profiles. Bioinformatics 2001;17(Suppl. 1):S215–24, http://dx.doi.org/10.1093/bioinformatics/17.suppl_1.S215.

[9] Pe'er D. Bayesian network analysis of signaling networks: a primer. Sci STKE 2005;2005:pl4, http://dx.doi.org/10.1126/stke.2812005pl4.

[10] Heckerman D. A tutorial on learning with bayesian networks. In: Holmes DE, Jain LC, editors. Innovations in Bayesian Networks: Theory and Applications. Berlin, Heidelberg: Springer; 2008. p. 33–82, http://dx.doi.org/10.1007/978-3-540-85066-3_3.

[11] Woolf PJ, Prudhomme W, Daheron L, Daley GQ, Lauffenburger DA. Bayesian analysis of signaling networks governing embryonic stem cell fate decisions. Bioinformatics 2005;21:741–53, bti056 [pii]\r10.1093/bioinformatics/bti056.

[12] Pearl J. Causality. New York, NY, USA ©2000: Cambridge University Press; 2009. p. 21–6.

[13] Korb KB, Ann Nicholson E. Introducing bayesian networks. Bayesian Artif Intell 2003:29–54, http://dx.doi.org/10.1201/b10391-4.

[14] Medicare Provider Utilization and Payment Data: Inpatient n.d. http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient.html (Accessed July 02, 2014).

[15] KHN Morning Briefing – Summaries of health policy coverage from major news organizations. 2013. http://kaiserhealthnews.org/morning-breakout/hospital-spending/ (Accessed January 14, 2015).

[16] Wayne A. Hospital Charges Vary Across U.S for Same Procedures. Bloomberg.com; 2013. http://www.bloomberg.com/news/2013-05-08/hospital-charges-vary-across-u-s-for-same-procedures.html (accessed January 14, 2015).

[17] Frost P, Dizikes C. Hospital fees vary for same treatment, U.S. data show. Chicago Trib 2013. http://articles.chicagotribune.com/2013-05-09/business/ct-biz-0509-hospital-prices-20130509_1_high-cost-hospitals-u-s-data-health-care (Accessed January 14, 2015).

[18] Meier B, McGinty JC, Creswell J. Hospital billing varies wildly, government data shows. NY Times 2013. http://www.nytimes.com/2013/05/08/business/hospital-billing-varies-wildly-us-data-shows.html?pagewanted=all&_r=0 (Accessed January 14, 2015).

[19] R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013 http://www.R-project.org/.

[20] Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, et al. Comparing families of dynamic causal models. PLoS Comput Biol 2010;6, http://dx.doi.org/10.1371/journal.pcbi.1000709.

[21] The cytoscape consortium. GNU Lesser Gen Public Licens Version 21, Febr 1999 n.d. http://www.cytoscapeconsortium.org.

[22] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504, http://dx.doi.org/10.1101/gr.1239303.

[23] Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 2011;27:431–2, http://dx.doi.org/10.1093/bioinformatics/btq675.

[24] Hoyert DL, Xu J. Deaths Preliminary Data for 2011. Natl VItal Stat Reports 2012; 61:1–52.

[25] Silverberg D, Wexler D, Blum M, Schwartz D, Iaina A. The association between congestive heart failure and chronic renal disease. Curr Opin Nephrol Hypertens 2004;13:163–70.

[26] Stenvinkel P, Carrero JJ, Axelsson J, Lindholm B, Heimbürger O, Massy Z. Emerging biomarkers for evaluating cardiovascular risk in the chronic kidney disease patient: how do new pieces fit into the uremic puzzle? Clin J Am Soc Nephrol 2008;3:505–21, http://dx.doi.org/10.2215/CJN.03670807.

[27] Wexler D, Silverberg D, Blum M, Sheps D, Keren G, Wollman Y, et al. Anaemia as a contributor to morbidity and mortality in congestive heart failure. Nephrol Dial Transplant 2005;20(Suppl 7):vii11–5, http://dx.doi.org/10.1093/ndt/gfh1101.

[28] Gil P, Justo S, Castilla MA, Criado C, Caramelo C. Cardio-renal insufficiency: the search for management strategies. Curr Opin Nephrol Hypertens 2005;14:442–7, http://dx.doi.org/10.1097/01.mnh.0000170753.41279.70.

[29] Damman K, Navis G, Voors AA, Asselbergs FW, Smilde TDJ, Cleland JGF, et al. Worsening renal function and prognosis in heart failure: systematic review and meta-analysis. J Card Fail 2007;13:599–608, http://dx.doi.org/10.1016/j.cardfail.2007.04.008.

[30] Albert NM, Lewis C. Recognizing and managing asymptomatic left ventricular dysfunction: after myocardial infarction. Crit Care Nurse 2008;28:20–37.

[31] Hellermann JP, Goraya TY, Jacobsen SJ, Weston S a, Reeder GS, Gersh BJ, et al. Incidence of heart failure after myocardial infarction: is it changing over time? Am J Epidemiol 2003;157:1101–7, http://dx.doi.org/10.1093/aje/kwg078.

[32] Centers for Disease Control and Precention (CDC). National Chronic Kidney Disease Fact Sheet: General Information and National Estimates on Chronic Kidney Disease in United States, 2010. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2010.

[33] Smith DH, Gullion CM, Nichols G, Keith DS, Brown JB. Cost of medical care for chronic kidney disease and comorbidity among enrollees in a large HMO population. J Am Soc Nephrol 2004;15:1300–6, http://dx.doi.org/10.1097/01.ASN.0000125670.64996.BB.

[34] Kidney Disease Statistics for the United States. NIH Publ No 12–3895 n.d. http://kidney.niddk.nih.gov/KUDiseases/pubs/kustats/ (accessed December 12, 2014).

[35] Montini G, Tullus K, Hewitt I. Febrile urinary tract infections in children. N Engl J Med 2011;365:239–50, http://dx.doi.org/10.1056/NEJMra1007755.

[36] Jakobsson B, Berg U, Svensson L. Renal scarring after acute pyelonephritis. Pediatr Nephrol 1994;8:609, http://dx.doi.org/10.1007/BF00858145.

[37] Lee YJ, Lee JH, Park YS. Risk factors for renal scar formation in infants with first episode of acute pyelonephritis: a prospective clinical study. J Urol 2012;187:1032–6, http://dx.doi.org/10.1016/j.juro.2011.10.164.

[38] Naqvi SB, Collins AJ. Infectious complications in chronic kidney disease. Adv Chronic Kidney Dis 2006;13:199–204, http://dx.doi.org/10.1053/j.ackd.2006.04.004.

[39] Huang S-T, Lin C-L, Chang Y-J, Sher Y-P, Wu M-J, Shu K-H, et al. Pneumococcal pneumonia infection is associated with end-stage renal disease in adult hospitalized patients. Kidney Int 2014, http://dx.doi.org/10.1038/ki.2014.79.

[40] National Center for Health Statistics. Health, United States, 2011. With Special Feature on Socioeconomic Status and Health. Hyattsville, MD, USA: 2012. Library of Congress Catalog Number 76-641496.

[41] Ohar JA, Donohue JF. Mono- and combination therapy of long-acting bronchodilators and inhaled corticosteroids in advanced COPD. Semin Respir Crit Care Med 2010;31:321–33, http://dx.doi.org/10.1055/s-0030-1254072.

[42] Cazzola M, Page CP, Calzetta L, Matera MG. Pharmacology and therapeutics of bronchodilators. Pharmacol Rev 2012;64:450–504, http://dx.doi.org/10.1124/pr.111.004580.

[43] Santulli G, Iaccarino G. Pinpointing beta adrenergic receptor in ageing pathophysiology: victim or executioner? Evidence from crime scenes. Immun Ageing 2013;10:1–13, http://dx.doi.org/10.1186/1742-4933-10-10.

[44] Liggett SB. Update on current concepts of the molecular basis of beta2-adrenergic receptor signaling. J Allergy Clin Immunol 2002;110:S223–7. S0091674902014963.

[45] Morgan DJ. Clinical pharmacokinetics of beta-agonists. Clin Pharmacokinet 1990;18:270–94.

[46] Nakamura A, Niimi R, Imaizumi A, Yanagawa Y. Renal effects of beta2-adrenoceptor agonist and the clinical analysis in children. Pediatr Res 2007;61:129–33, http://dx.doi.org/10.1203/01.pdr.0000249998.24772.3b.

[47] Bodenhamer J, Bergstrom R, Brown D, Gabow P, Marx JA, Lowenstein SR. Frequently nebulized beta-agonists for asthma: effects on serum electrolytes. Ann Emerg Med 1992;21:1337–42.

[48] Alamoudi OS. Electrolyte disturbances in patients with chronic, stable asthma: effect of therapy. Chest 2001;120:431–6.

[49] Aschenbach JR, Borau T, Gäbel G. Glucose uptake via SGLT-1 is stimulated by beta(2)-adrenoceptors in the ruminal epithelium of sheep. J Nutr 2002;132:1254–7.

[50] Pojoga L, Kolatkar NS, Williams JS, Perlstein TS, Jeunemaitre X, Brown NJ, et al. Beta-2 adrenergic receptor diplotype defines a subset of salt-sensitive hypertension. Hypertension 2006;48:892–900, http://dx.doi.org/10.1161/01.HYP.0000244688.45472.95.

[51] Weber MA, Stokes GS, Gain JM. Comparison of the effects on renin release of beta adrenergic antagonists with differing properties. J Clin Invest 1974;54:1413–9, http://dx.doi.org/10.1172/JCI107888.

[52] Weinberger MH, Aoi W, Henry DP. Direct effect of beta-adrenergic stimulation on renin release by the rat kidney slice in vitro. Circ Res 1975;37:318–24, http://dx.doi.org/10.1161/01.RES.37.3.318.

[53] Udezue E, D'souza L, Mahajan M. Hypokalemia after normal doses of neubulized albuterol (salbutamol). Am J Emerg Med 1995;13:168–71, http://dx.doi.org/10.1016/0735-6757(95)90086-1.

[54] Assadi FK. Therapy of acute bronchospasm. Complicated by lactic acidosis and hypokalemia. Clin Pediatr (Phila) 1989;28:258–60, http://dx.doi.org/10.1177/000992288902800603.

[55] Epelbaum S, Benhamou PH, Pautard JC, Devoldere C, Kremp O, Piussan C. Respiratory arrest in an asthmatic girl treated with beta-2-mimetics and theophylline: possible role of hypokalemia in sudden death in asthmatic patients. Ann Pediatr (Paris) 1989;36:473–5.

[56] Kemperman CJF, Kuilman M, Njio LKF. A retrospective and explorative study of hypokalemia in psychiatric disorders: a beta2-receptor related phenomenon. Eur Arch Psychiatry Neurol Sci 1988;237:161–5, http://dx.doi.org/10.1007/BF00451284.

[57] Brown MJ, Brown DC, Murphy MB. Hypokalemia from beta2-receptor stimulation by circulating epinephrine. N Engl J Med 1983;309:1414–9, http://dx.doi.org/10.1056/NEJM198312083092303.

[58] Suga SI, Phillips MI, Ray PE, Raleigh JA, Vio CP, Kim YG, et al. Hypokalemia induces renal injury and alterations in vasoactive mediators that favor salt sensitivity. Am J Physiol Renal Physiol 2001;281:F620–9.

[59] Reungjui S, Roncal CA, Sato W, Glushakova OY, Croker BP, Suga S-I, et al. Hypokalemic nephropathy is associated with impaired angiogenesis. J Am Soc Nephrol 2008;19:125–34, http://dx.doi.org/10.1681/ASN.2007030261.

[60] Menahem SA, Perry GJ, Dowling J, Thomson NM. Hypokalemia-induced acute rena failure. Nephrol Dial Transplant 1999;(September 14):2216–8.

[61] Torres VE, Young WF, Offord KP, Hattery RR. Association of hypokalemia, aldosteronism, and renal cysts. N Engl J Med 1990;322:345–51, http://dx.doi.org/10.1056/NEJM199002083220601.

[62] Wang H-H, Hung C-C, Hwang D-Y, Kuo M-C, Chiu Y-W, Chang J-M, et al. Hypokalemia, its contributing factors and renal outcomes in patients with chronic kidney disease. PLoS One 2013;8:e67140, http://dx.doi.org/10.1371/journal.pone.0067140.

[63] Einhorn LM, Zhan M, Hsu VD, Walker LD, Moen MF, Seliger SL, et al. The frequency of hyperkalemia and its significance in chronic kidney disease. Arch

Intern Med 2009;169:1156–62, http://dx.doi.org/10.1001/archinternmed.2009.132.
[64] Hayes J, Kalantar-Zadeh K, Lu JL, Turban S, Anderson JE, Kovesdy CP. Association of hypo- and hyperkalemia with disease progression and mortality in males with chronic kidney disease: the role of race. Nephron Clin Pract 2012;120:c8–16, http://dx.doi.org/10.1159/000329511.
[65] Millar EA, McInnes GT, Thomson NC. Investigation of the mechanism of β2-agonist-induced activation of the Renin—Angiotensin systemtle. Clin Sci 1995;88:433–7.
[66] Taal MW, Brenner BM. Renoprotective benefits of RAS inhibition: from ACEI to angiotensin II antagonists. Kidney Int 2000;57:1803–17, http://dx.doi.org/10.1046/j.1523-1755.2000.00031.x.
[67] Weir MR. Effects of renin-angiotensin system inhibition end-organ protection: can we do better? Clin Ther 2007;29:1803–24, http://dx.doi.org/10.1016/j.clinthera.2007.09.019.
[68] Epstein M. Aldosterone and the hypertensive kidney: its emerging role as a mediator of progressive renal dysfunction: a paradigm shift. J Hypertens 2001;19:829–42, http://dx.doi.org/10.1097/00004872-200105000-00001.
[69] Wenzel U. Aldosterone and progression of renal disease. Curr Opin Nephrol Hypertens 2008;17:44–50, http://dx.doi.org/10.1097/MNH.0b013e3282f29028.
[70] Liu D-W, Zhen X-G, Liang Y, Jing X-G, Zhang T-S, Zhang G-J, et al. Persistent asthma increases the risk of chronic kidney disease: a retrospective cohort study of 2354 patients with asthma. Chin Med J (Engl) 2013;126:4093–9.
[71] Huang H-L, Ho S-Y, Li C-H, Chu F-Y, Ciou L-P, Lee H-C, et al. Bronchial asthma is associated with increased risk of chronic kidney disease. BMC Pulm Med 2014;14:80, http://dx.doi.org/10.1186/1471-2466-14-80.
[72] RxList n.d. www.rxlist.com (Accessed December 03, 2015).
[73] IMS health n.d. http://www.imshealth.com/portal/site/imshealth (Accessed December 05, 2015).