Original Article

AN APPROACH TO USING DATA MINING TO SUPPORT EARLY IDENTIFICATION OF ACROMEGALY

Michael S. Broder, MD, MSHS[1]; Eunice Chang, PhD[1]; Sheila R. Reddy, PhD, RPh[1]; Maureen P. Neary, PhD, MS[2]

ABSTRACT

Objective: Data mining using insurance claims presents an opportunity to incorporate new analytic techniques in identifying rare conditions. This study aims to identify dyads of clinical conditions associated with acromegaly that may, with further validation and testing, be used to initially identify and diagnose this rare disease more accurately and efficiently.

Methods: This case-control study used two claims databases to identify acromegaly patients (cases) (International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM]: 253.0) from 2008-2013. Each case was assigned two nonacromegaly controls (same age, gender, and region). Matched patients were randomly split into development and validation datasets. With expert clinician input, we isolated common associated conditions using ICD-9-CM codes. We identified all 2-way combinations of these conditions (dyads) and calculated the rate and risk relative (RR) to controls. Dyads meeting certain criteria (case rate ≥5% [or ≥1% if RR ≥5] or observed RR > expected) were replicated in the validation dataset to confirm results.

Results: We identified 3,731 cases and 7,462 controls: mean age 41.8 (SD, 16.1) years, 51.8% female. A total of 32 and 38 dyads, reduced from 630, met study criteria. Among replicated dyads, case rates varied –15.9% (hypertension and metabolic disorder) to 0.6% (arthritis and menstrual abnormalities). The highest RRs (e.g., valvular insufficiency and colon polyps [RR, 13.5; rate, 0.7%]) also exceeded expected values. Replication showed similar RR direction and size.

Conclusion: This novel analytic approach revealed several dyads that were significantly associated with an acromegaly diagnosis. Presence of high-risk condition pairs, if verified by a detailed data source (e.g., medical charts), may be incorporated into screening tools or serve as potential markers for physicians to consider an acromegaly diagnosis. **(Endocr Pract. 2017;23:422-431)**

Abbreviations:
ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification; **ID** = identification; **RR** = relative risk

INTRODUCTION

Data mining using insurance claims may present an opportunity to incorporate new techniques in identifying rare conditions. Using coded data, it may be possible to supplement the traditional physician-patient interaction, which has been the basis for diagnosing disease since the beginning of modern medicine. Data mining using predictive analytics allows computers to match patterns of a very large number of comorbidities, which may not be possible using standard modeling methodologies.

These powerful analytic techniques may be used to identify multiple unique combinations of clinical conditions that may be associated with a rare disease. In a prior study, such methods have been successfully used to uncover previously unknown or unusual combinations of otherwise common symptoms associated with a rare disease of inter-

Submitted for publication September 4, 2016
Accepted for publication December 14, 2016
From the [1]Partnership for Health Analytic Research, LLC, Beverly Hills, California, and [2]Novartis Pharmaceuticals Corporation, East Hanover, New Jersey.
Address correspondence to Dr. Maureen Neary, One Health Plaza, East Hanover, NJ 07936-1080.
E-mail: maureen.neary@novartis.com
Published as a Rapid Electronic Article in Press at http://www.endocrinepractice.org on January 17, 2017. DOI: 10.4158/EP161575.OR

est (1,2). The future may hold the possibility of combining physician free-text notes, laboratory results, and other detailed information into such algorithms. At present, the most prevalent data useful for data mining can be derived from the use of the standard International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes.

Acromegaly results from pituitary hypersecretion of growth hormone and pituitary adenoma (3,4) and affects up to 130 individuals per million population, or as many as 42,000 people in the United States (5-7). Due to the rarity of the condition and its slow onset (8,9), making the diagnosis can be a challenge (8-11). The average time from the onset of first symptoms to diagnosis is 6 to 10 years (8,10,12,13). Delay in diagnosis and proper clinical management of this disease may lead to multiple chronic comorbidities, many of which may be irreversible (8,14).

Using the example of acromegaly, we designed a framework using conditions associated with acromegaly to test, using coded data, the theory that combinations of conditions that are highly associated with the disease can be identified. If validated, such combinations could be used in an automated process which could function as a screening tool and thereby assist physicians in the early and efficient diagnosis of acromegaly (1).

METHODS

Study Population and Data Source

We conducted a retrospective, matched case-control study using data from two large claims databases in the U.S. Truven Health MarketScan® Commercial Claims Database and IMS Health PharMetrics databases are Health Insurance Portability and Accountability Act compliant. These databases contain de-identified adjudicated pharmacy claims such as outpatient prescriptions and also include medical claims (e.g., inpatient and outpatient services) submitted for payment by providers, healthcare facilities, and pharmacies. Information on member enrollment and benefits as well as on patient, provider, and hospital demographic data are also available in both databases. There is no linkage between these databases and clinical data, and therefore, the claims could not be verified.

This study included patients with claims for acromegaly during a 5-year identification (ID) period from January 1, 2008, to December 31, 2013, and matched controls. Patients were considered to have acromegaly if they had at least two medical claims with a diagnosis of acromegaly (ICD-9-CM: 253.0) on separate days in the ID period or one medical claim with a diagnosis of acromegaly in combination with one other claim for a pituitary tumor (ICD-9-CM: 237.0x), pituitary surgery (hypophysectomy), or cranial stereotactic radiosurgery in any diagnosis field in the ID period. For acromegaly cases, January 1 of the year of the date of the first acromegaly-associated claim in the ID period was defined as the index date, and the year following the index date was the measurement year. Patients who were not continuously enrolled in the measurement year were excluded.

Nonacromegaly controls were selected from a 5% random sample of all patients in the claims databases. Patients were selected as matched controls if they had no claims with acromegaly or pituitary tumor diagnosis, no claims with pituitary surgery or cranial stereotactic radiosurgery in the ID period, and were continuously enrolled in the calendar year. Patients could not have been included in both the case and control cohorts. For each acromegaly patient, two controls with the same age, gender, and region in the same measurement year were randomly selected into the final study cohort. Patients could be enrolled in both databases; therefore, possible duplicates were identified and excluded from the study. The entire dataset of cases and controls was then randomly and equally split into either development or validation datasets. In order to avoid spurious associations and to confirm our findings, the analysis was conducted in the development dataset then repeated in the validation dataset.

Study Measures

We determined the sample characteristics, including age, gender, and geographic region. Published literature was reviewed to identify the most common clinical conditions linked with acromegaly (4,6,9,11,15). Our primary outcome of interest was a subset of the most common combinations of clinical conditions associated with patients with acromegaly. With expert endocrinologist input, we isolated key characteristics and conditions of clinical significance. The conditions were broadly grouped as musculoskeletal (e.g., carpal tunnel syndrome, arthralgia), cardiovascular (e.g., valvular insufficiency, arrhythmia), metabolic (e.g., metabolic disorder, including diabetes), reproductive (e.g., menstrual abnormalities, infertility), gastrointestinal (e.g., colon polyps), skin and soft tissue (e.g., hyperhidrosis, macroglossia), and others (see Supplementary Table 1 for full list of conditions). The final list comprised 36 conditions identifiable using ICD-9-CM codes, which included signs, symptoms, and comorbidities of acromegaly. We reviewed all claims for both cases and controls to find evidence of any of these 36 conditions and calculated their prevalence within our sample.

Statistical Analysis

We divided prevalence in cases by prevalence in controls to calculate the relative risk (RR) of each condition. A similar procedure was carried out for all possible 2-way combinations (dyads) of the 36 conditions (630 dyads in total). In a 2-step item reduction, among these 630 combinations, we identified (*1*) dyads with prevalence ≥5% among cases or prevalence ≥1% in cases but a RR ≥5 in cases compared to controls; and (*2*) dyads with a higher RR than expected based on the prevalence of each condition alone. We then selected for validation the 10 most

prevalent (in cases) dyads meeting each of these criteria. Means, SDs, medians, and percentages were reported as appropriate. Statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC).

RESULTS

Using the algorithm described above, we identified 9,349 cases of acromegaly during the 2008-2013 ID period, 7,462 of whom were continuously enrolled in the measurement year. From a population of over 8 million continuously enrolled patients without acromegaly, we selected controls in a 2:1 ratio by exactly matching each case on age, gender, region, and measurement year. Cases (n = 7,462) and controls (n = 14,924) were randomly divided into a development and a validation dataset, each with 3,731 cases and 7,462 controls (Fig. 1). The mean age of cases and controls was 41.8 (SD, 16.1) years, and 51.8% were female. All regions of the U.S. were represented: Midwest (24%), Northeast (25%), South (36%), and West (16%) (Table 1). Results did not differ between development and validation data sets (not shown).

The frequency and RR of the initial 36 conditions in the acromegaly cases compared to controls varied widely. The highest frequency condition was hypertension at 34.9%; the lowest was goiter at 0.1%. The RRs varied from infinite for conditions identified only in patients with acromegaly (increased aldosterone levels and macroglossia), to 1.7 (hypertension). Sleep apnea was associated with the highest RR among acromegaly patients (3.9), and hypertension had the lowest RR of 1.7 (Supplementary Table 1).

In the first step of item reduction, conditions with either (*1*) prevalence ≥5% in cases or (*2*) prevalence ≥1% in cases and RR ≥5 for cases compared to controls were subjected to pairwise comparisons. The rate and RR of each of the 32 dyads meeting these criteria were calculated. The highest frequency dyad (in acromegaly) was hypertension and metabolic disorder (which included type 2 diabetes mellitus, insulin resistance, hyperinsulinemia, dysmetabolic syndrome, and hyperglycemia) at 15.9%; the lowest was colon diverticula and sleep apnea at 1.0%. The RRs varied from 13.3 for cardiomegaly/cardiomyopathy and sleep apnea to 2.8 for arthralgia and hypertension (Supplementary Table 2).

In the next step of item reduction, of the 630 dyads, 38 with a higher RR than expected based on the prevalence of each condition alone were reviewed. In this group, the highest RR was for cardiomegaly/cardiomyopathy and colon diverticula, which had a RR of 26.0 and frequency of 0.348% in cases. The dyad with the lowest RR was arthralgia and prostatic hypertrophy with a RR of 4.0 and frequency of 0.322% in cases (Supplementary Table 3).

The 10 dyads with the highest prevalence in cases were selected from each set (e.g., the 32 dyads selected using prevalence/RR rules, and the 38 selected using RR alone), leaving a final item set of 20 dyads. In the final item list, the highest RRs were for valvular insufficiency and colon polyps (RR, 13.5); menstrual abnormalities and sleep apnea (RR, 12.5); metabolic disorder and hyperhidrosis (RR, 12.0); colonic diverticula and sleep apnea (RR, 11.1); and arthropathy and sleep apnea (RR, 10.0). The frequencies for these dyads were: valvular insufficiency and colon polyps (0.7%); menstrual abnormalities and sleep apnea (0.7%); metabolic disorder and hyperhidrosis (0.6%); colonic diverticula and sleep apnea (1.0%); and arthropathy and sleep apnea (0.8%) (Table 2, Fig. 2). Because of the small sample size, the confidence intervals for RR overlapped substantially. The replication showed similar RR direction and size between the development and validation datasets.

DISCUSSION

Rare diseases can present a diagnostic challenge: physician experience may only very infrequently lead to enough knowledge to quickly identify signs of the condition. Using the example of acromegaly, our study demonstrates the utility of analyzing information from large administrative databases to identify associations between specific clinical conditions and a rare disease. We used a data mining technique to identify combinations of clinical conditions that are associated with acromegaly, which may be a first step in developing some potentially useful aids for diagnosis. Application of this powerful analytic technique might also be able to be used to characterize conditions associated with other rare diseases.

Our exploratory analysis of claims data identified several dyad combinations of clinical conditions that were many times more likely to occur among acromegaly patients than among their matched controls. For example, the risk of having sleep apnea in the presence of either

Table 1
Patient Characteristics

	Development dataset	
	Acromegaly	**Nonacromegaly**
n	3,731	7,462
Age, years, mean (SD)	41.8 (16.1)	41.8 (16.1)
≤40	1,408 (37.7)	2,816 (37.7)
41-50	943 (25.3)	1,886 (25.3)
51-64	1,380 (37.0)	2,760 (37.0)
Female, n (%)	1,934 (51.8)	3,868 (51.8)
Region, n (%)	896 (24.0)	1,792 (24.0)
Midwest		
Northeast	919 (24.6)	1,838 (24.6)
South	1,332 (35.7)	2,664 (35.7)
West	584 (15.7)	1,168 (15.7)

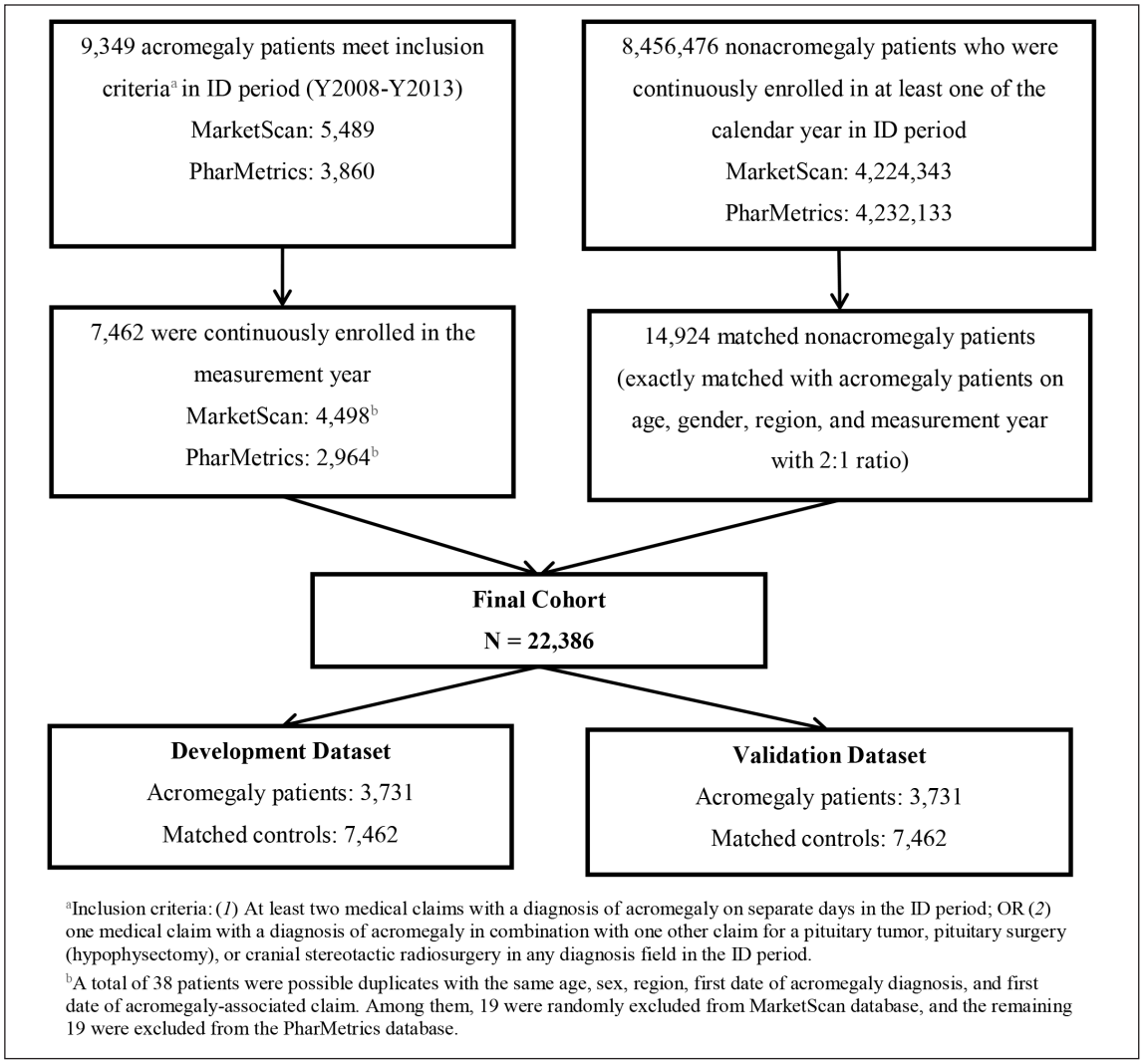

Fig. 1. Patient selection. *ID* = identification.

menstrual abnormalities, colon diverticula, or arthropathy, or the risk of having metabolic disorder in the presence of hyperhidrosis, were at least 10-times greater in acromegaly patients than nonacromegaly patients (RRs from 10 to 13.5). Interestingly, sleep apnea was found to be a common clinical condition in three of the five highest-risk dyad combinations.

The approach used in our study could serve as a platform for gathering information from large datasets wherein the use of automated systems would assist the physicians in identifying combinations of symptoms—or potential markers—for various diseases. An analogous approach has been successfully developed for other conditions in the inpatient setting, specifically using electronic alerts to improve diagnosis of acute kidney injury during hospitalization and to predict chronic obstructive pulmonary disease exacerbations using telehealth monitoring (16,17). These studies used, or proposed the use of, alerts for clinicians, but neither reported clinical outcomes. Ours was an exploratory study, and future studies to extend and validate these findings would be necessary before they could be used in clinical settings to assist with case identification. Incorporating data from electronic health records (EHRs) and from longer observation periods are two potentially useful initial steps (18). In addition, if tools derived from a data mining method such as this are incorporated in clinical practice, they should be carefully studied to evaluate the possible impact of false positives (and negatives) on resource use and the flow of clinical care. With this study, we aimed to take a first step in the development of a system that might one day aid clinicians by identifying cases that might have been missed. A limitation of this method is that it may lead to a large number of false positives, where these comorbidities may be common in a number of other conditions. Further refinement of the case identification algorithm would be needed to attempt to limit the rate of false positives. In addition, one doubts that any such system, no matter how well developed, would be much

Table 2
Actual and Expected Risk of Final Selected Dyads

Condition 1	Condition 2	RR[a] (95% CI)	Expected RR[b]
Valvular insufficiency	Colon polyps	13.5 (4.7-38.6)	8.1
Menstrual abnormalities	Sleep apnea	12.5 (4.4-35.9)	9.0
Metabolic disorder[c]	Hyperhidrosis	12.0 (4.2-34.6)	11.9
Colonic diverticula	Sleep apnea	11.1 (5.0-24.9)	9.3
Arthropathy	Sleep apnea	10.0 (4.2-24.0)	9.1
Valvular insufficiency	Colonic diverticula	9.0 (3.7-21.8)	7.8
Metabolic disorder	Menstrual abnormalities	7.8 (4.7-12.9)	6.3
Arthralgia	Valvular insufficiency	6.6 (4.2-10.2)	6.1
Arthritis	Menstrual abnormalities	6.0 (2.6-14.1)	5.8
Metabolic disorder	Fatigue	5.8 (4.6-7.5)	7.0
Metabolic disorder	Sleep apnea	5.0 (3.9-6.3)	10.8
Arthralgia	Colonic diverticula	4.8 (3.0-7.6)	4.4
Hypertension	Sleep apnea	4.0 (3.3-4.8)	6.7
Arthralgia	Metabolic disorder	3.6 (2.9-4.3)	5.1
Hypertension	Fatigue	3.4 (2.8-4.0)	4.3
Arthralgia	Fatigue	3.4 (2.7-4.1)	4.7
Arthritis	Arthralgia	3.0 (2.5-3.5)	4.7
Arthritis	Hypertension	3.0 (2.5-3.6)	4.3
Hypertension	Metabolic disorder	2.9 (2.5-3.2)	4.7
Arthralgia	Hypertension	2.8 (2.4-3.2)	3.2

Abbreviations: CI = confidence interval; RR = relative risk.
[a]RR of the selected combination of conditions for acromegaly vs. nonacromegaly patients.
[b]The calculation of expected RR for conditions 1 & 2 assumes independence between risks of individual conditions.
[c]Type 2 diabetes mellitus, insulin resistance, hyperinsulinemia, dysmetabolic syndrome, hyperglycemia.

help to a specialist accustomed to seeing acromegaly in all its stages. On the other hand, an inexperienced clinician with respect to the specialty area of endocrinology, or one who never or rarely sees this disease, might be alerted to the possibility, and thus shorten the period from onset of symptoms to referral to a specialist and correct diagnosis.

A major strength of our study includes the use of two large U.S. commercial claims databases together representing over 100 million covered lives. Further, to improve the robustness of our findings, the selected signals observed in the development set were replicated in the validation set. Our study has several limitations. Even with a very large underlying population, sample sizes for individual condition dyads were small, leading to wide confidence intervals around the results. With a condition as rare as acromegaly—with an incidence of between 3 and 10 per million person-years—this outcome was not unexpected (19,20). Insurance claims are not collected for research purposes, and inaccuracies in assigning ICD-9 codes are common. In addition, patients with acromegaly may be more likely to receive diagnoses for other conditions on the basis of having frequent interaction with the healthcare system. Although control patients were not necessarily healthier than cases (they simply did not have acromegaly), if controls were systematically assigned ICD-9-CM codes less often than cases, this could potentially have affected our findings. This study was limited to patients with commercial insurance, and the results therefore may not be generalizable to the rest of the population. Furthermore, the data used in our study consisted of claims submitted for billing, not medical records. We used the presence of certain ICD-9-CM codes to identify patients with acromegaly. These cases could not be confirmed with a review of medical records, as a link to medical records was not available. Similarly, pathology and laboratory test data were not available, and their inclusion in future studies (e.g., by incorporating data from EHRs) might increase the clinical utility of the findings. We used only one of many potential methods and data sources for this study. We encourage other researchers to use other data sets and novel methods to confirm or refute

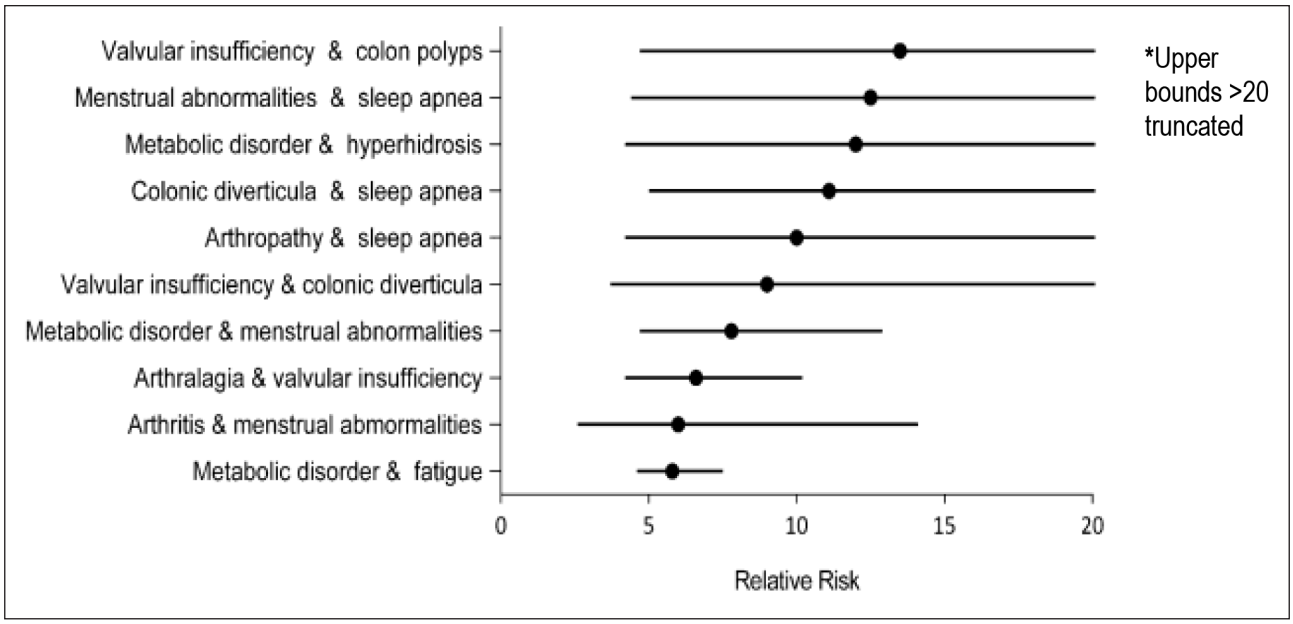

Fig. 2. Ten dyads with highest relative risk, shown with 95% confidence intervals.

our findings. Finally, before these findings are used in a clinical setting, the results should be replicated in a more detailed data source; one in which the conditions identified are verified to actually be present, rather than identified using insurance claims data alone.

CONCLUSION

We present a novel approach to use data mining to identify uncommon pairs of clinical conditions associated with a rare disease—acromegaly in this example. Our study revealed several combinations (e.g., sleep apnea in the presence of menstrual abnormalities, colon diverticula, or arthropathy; and metabolic disorder in the presence of hyperhidrosis) that were many times more likely to occur among acromegaly patients. The presence of high-risk dyad combinations of clinical conditions may serve as a potential marker in the early diagnosis of acromegaly. Broadly, our findings may be useful in developing clinical screening tools to assist physicians in identifying patients for early diagnostic testing for rare diseases. With further validation, possibly including the use of clinical data from medical charts and incorporation of other potentially useful information such as age, the specific high-risk dyad combinations we identified may one day be useful in clinical decision making. This may be used as an aid for the identification of patients at highest risk of acromegaly who should be further evaluated to determine if they may have this condition. Early detection and treatment may provide benefit for long-term clinical outcomes for patients with acromegaly.

ACKNOWLEDGMENT

Dr. William Ludlam, an endocrinologist, contributed in identifying the set of conditions to predict onset of acromegaly.

DISCLOSURE

Funding for this study was provided by Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, NJ 07936-1080. M.P.N. is an employee of Novartis Pharmaceuticals Corporation. M.B., E.C., and S.R.R. are employees of the Partnership for Health Analytic Research, LLC (PHAR, LLC), a health services research company paid by Novartis to conduct this research.

Supplementary Table 1
Relative Risks[a] of Individual Conditions

Order	Condition	Acromegaly n (%)	Nonacromegaly n (%)	RR[a]
1	Increased aldosterone levels	10 (0.268)	0 (0.000)	∞
2	Macroglossia	7 (0.188)	0 (0.000)	∞
3	Malocclusion	37 (0.992)	1 (0.013)	74.0
4	Galactorrhea	52 (1.394)	2 (0.027)	52.0
5	Hyperphosphatemia	9 (0.241)	2 (0.027)	9.0
6	Kyphoscoliosis	97 (2.600)	25 (0.335)	7.8
7	Hypercalciuria	21 (0.563)	7 (0.094)	6.0
8	Cardiomyopathy or cardiomegaly	122 (3.270)	45 (0.603)	5.4
9	Goiter	5 (0.134)	2 (0.027)	5.0
10	Nasal polyps	11 (0.295)	5 (0.067)	4.4
11	Hyperhidrosis	43 (1.153)	20 (0.268)	4.3
12	Myopathy	8 (0.214)	4 (0.054)	4.0
13	**Sleep apnea**	**526 (14.098)**	**268 (3.592)**	**3.9**
14	Colon cancer	20 (0.536)	11 (0.147)	3.6
15	Decreased libido	9 (0.241)	5 (0.067)	3.6
16	Peripheral neuropathy	73 (1.957)	43 (0.576)	3.4
17	**Aortic or mitral valve**	**206 (5.521)**	**125 (1.675)**	**3.3**
18	Splenomegaly	22 (0.590)	14 (0.188)	3.1
19	**Arrhythmias**	**303 (8.121)**	**213 (2.854)**	**2.8**
20	Heart failure	76 (2.037)	54 (0.724)	2.8
21	**Metabolic disorder**	**974 (26.106)**	**705 (9.448)**	**2.8**
22	**Fatigue**	**721 (19.325)**	**572 (7.666)**	**2.5**
23	**Arthritis**	**491 (13.160)**	**391 (5.240)**	**2.5**
24	**Colon polyps**	**195 (5.226)**	**158 (2.117)**	**2.5**
25	Carpal tunnel syndrome	101 (2.707)	84 (1.126)	2.4
26	Colonic diverticula	172 (4.610)	145 (1.943)	2.4
27	Arthropathy	134 (3.592)	116 (1.555)	2.3
28	Infertility	38 (1.018)	33 (0.442)	2.3
29	Hypertriglyceridemia	55 (1.474)	48 (0.643)	2.3
30	**Menstrual abnormalities**	**315 (8.443)**	**275 (3.685)**	**2.3**
31	Soft-tissue edema	159 (4.262)	142 (1.903)	2.2
32	Erectile dysfunction	86 (2.305)	86 (1.153)	2.0
33	Hepatomegaly	13 (0.348)	13 (0.174)	2.0
34	**Arthralgia**	**822 (22.032)**	**884 (11.847)**	**1.9**
35	Prostatic hypertrophy	53 (1.421)	57 (0.764)	1.9
36	**Hypertension**	**1,302 (34.897)**	**1,517 (20.330)**	**1.7**

Abbreviation: RR = relative risk.
[a]RR of the individual condition for acromegaly vs. nonacromegaly patients. Conditions with prevalence ≥5% in acromegaly are highlighted in bold.

Supplementary Table 2
Relative Risks[a] of Dyads of Selected Conditions[b]

Order	Condition 1	Condition 2	Acromegaly n (%)	Nonacromegaly n (%)	RR[a]
1	Cardiomyopathy or cardiomegaly	Sleep apnea	40 (1.072)	6 (0.080)	13.3
2	Colonic diverticula	Sleep apnea	39 (1.045)	7 (0.094)	11.1
3	Arrhythmias	Sleep apnea	74 (1.983)	15 (0.201)	9.9
4	Colon polyps	Sleep apnea	50 (1.340)	12 (0.161)	8.3
5	Aortic or mitral valve insufficiency	Metabolic disorder	90 (2.412)	22 (0.295)	8.2
6	Aortic or mitral valve insufficiency	Sleep apnea	48 (1.287)	12 (0.161)	8.0
7	Metabolic disorder	Menstrual abnormalities	74 (1.983)	19 (0.255)	7.8
8	Fatigue	Sleep apnea	173 (4.637)	47 (0.630)	7.4
9	Arthritis	Sleep apnea	121 (3.243)	33 (0.442)	7.3
10	Aortic or mitral valve insufficiency	Cardiomyopathy or cardiomegaly	43 (1.153)	12 (0.161)	7.2
11	Arrhythmias	Fatigue	105 (2.814)	31 (0.415)	6.8
12	Arthralgia	Sleep apnea	166 (4.449)	50 (0.670)	6.6
13	Arthralgia	Aortic or mitral valve insufficiency	82 (2.198)	25 (0.335)	6.6
14	Cardiomyopathy or cardiomegaly	Metabolic disorder	52 (1.394)	16 (0.214)	6.5
15	Carpal tunnel syndrome	Metabolic disorder	50 (1.340)	16 (0.214)	6.3
16	**Metabolic disorder**	**Fatigue**	**239 (6.406)**	**82 (1.099)**	**5.8**
17	Arthritis	Aortic or mitral valve insufficiency	46 (1.233)	16 (0.214)	5.8
18	Arrhythmias	Metabolic disorder	126 (3.377)	45 (0.603)	5.6
19	Aortic or mitral valve insufficiency	Fatigue	70 (1.876)	25 (0.335)	5.6
20	Arrhythmias	Cardiomyopathy or cardiomegaly	42 (1.126)	15 (0.201)	5.6
21	Arthritis	Arrhythmias	68 (1.823)	25 (0.335)	5.4
22	Metabolic disorder	Colon polyps	75 (2.010)	28 (0.375)	5.4
23	Colonic diverticula	Fatigue	48 (1.287)	19 (0.255)	5.1
24	**Metabolic disorder**	**Sleep apnea**	**221 (5.923)**	**89 (1.193)**	**5.0**
25	**Hypertension**	**Sleep apnea**	**304 (8.148)**	**153 (2.050)**	**4.0**
26	**Arthralgia**	**Metabolic disorder**	**261 (6.995)**	**147 (1.970)**	**3.6**
27	**Hypertension**	**Fatigue**	**304 (8.148)**	**181 (2.426)**	**3.4**
28	**Arthralgia**	**Fatigue**	**235 (6.299)**	**140 (1.876)**	**3.4**
29	**Arthritis**	**Hypertension**	**291 (7.800)**	**191 (2.560)**	**3.0**
30	**Arthritis**	**Arthralgia**	**320 (8.577)**	**214 (2.868)**	**3.0**
31	**Hypertension**	**Metabolic disorder**	**593 (15.894)**	**416 (5.575)**	**2.9**
32	**Arthralgia**	**Hypertension**	**393 (10.533)**	**284 (3.806)**	**2.8**

Abbreviation: RR = relative risk.
[a]RR of the selected combination of conditions for acromegaly vs. nonacromegaly patients.
[b]Combinations with either a prevalence rate ≥5% in acromegaly patients or a prevalence rate ≥1% in and a RR ≥5 compared to controls.
The 10 dyads with the highest prevalence in acromegaly are shown in bold.

Supplementary Table 3
Dyads With a Relative Risk Higher Than Expected

Order	Condition 1		Condition 2		Conditions 1 & 2			
	Condition	RR[a]	Condition	RR[a]	Acromegaly n (%)	Nonacromegaly n (%)	RR[a] (95% CI)	Expected RR[b]
1	Cardiomyopathy or cardiomegaly	5.4	Colonic diverticula	2.4	13 (0.348)	1 (0.013)	26.0 (3.4-198.7)	12.9
2	Peripheral neuropathy	3.4	Aortic or mitral valve insufficiency	3.3	11 (0.295)	1 (0.013)	22.0 (2.8-170.3)	11.2
3	Cardiomyopathy or cardiomegaly	5.4	Colon polyps	2.5	11 (0.295)	1 (0.013)	22.0 (2.8-170.3)	13.4
4	Metabolic disorder	2.8	Infertility	2.3	9 (0.241)	1 (0.013)	18.0 (2.3-142.0)	6.4
5	Peripheral neuropathy	3.4	Colon polyps	2.5	8 (0.214)	1 (0.013)	16.0 (2.0-127.9)	8.4
6	Arthritis	2.5	Hypertriglyceridemia	2.3	14 (0.375)	2 (0.027)	14.0 (3.2-61.6)	5.8
7	**Aortic or mitral valve insufficiency**	**3.3**	**Colon polyps**	**2.5**	**27 (0.724)**	**4 (0.054)**	**13.5 (4.7-38.6)**	**8.1**
8	**Menstrual abnormalities**	**2.3**	**Sleep apnea**	**3.9**	**25 (0.670)**	**4 (0.054)**	**12.5 (4.4-35.9)**	**9.0**
9	Arrhythmias	2.8	Splenomegaly	3.1	6 (0.161)	1 (0.013)	12.0 (1.4-99.6)	8.9
10	Decreased libido	3.6	Fatigue	2.5	6 (0.161)	1 (0.013)	12.0 (1.4-99.6)	9.1
11	**Metabolic disorder**	**2.8**	**Hyperhidrosis**	**4.3**	**24 (0.643)**	**4 (0.054)**	**12.0 (4.2-34.6)**	**11.9**
12	Heart failure	2.8	Colonic diverticula	2.4	12 (0.322)	2 (0.027)	12.0 (2.7-53.6)	6.7
13	Arthralgia	1.9	Hypertriglyceridemia	2.3	17 (0.456)	3 (0.040)	11.3 (3.3-38.6)	4.3
14	**Colonic diverticula**	**2.4**	**Sleep apnea**	**3.9**	**39 (1.045)**	**7 (0.094)**	**11.1 (5.0-24.9)**	**9.3**
15	Soft-tissue edema	2.2	Menstrual abnormalities	2.3	16 (0.429)	3 (0.040)	10.7 (3.1-36.6)	5.1
16	Soft-tissue edema	2.2	Splenomegaly	3.1	5 (0.134)	1 (0.013)	10.0 (1.2-85.6)	7.0
17	Carpal tunnel syndrome	2.4	Aortic or mitral valve insufficiency	3.3	5 (0.134)	1 (0.013)	10.0 (1.2-85.6)	7.9
18	Arthralgia	1.9	Myopathy	4.0	5 (0.134)	1 (0.013)	10.0 (1.2-85.6)	7.4
19	Peripheral neuropathy	3.4	Heart failure	2.8	5 (0.134)	1 (0.013)	10.0 (1.2-85.6)	9.6
20	Hyperhidrosis	4.3	Menstrual abnormalities	2.3	5 (0.134)	1 (0.013)	10.0 (1.2-85.6)	9.9
21	**Arthropathy**	**2.3**	**Sleep apnea**	**3.9**	**30 (0.804)**	**6 (0.080)**	**10.0 (4.2-24.0)**	**9.1**
22	**Aortic or mitral valve insufficiency**	**3.3**	**Colonic diverticula**	**2.4**	**27 (0.724)**	**6 (0.080)**	**9.0 (3.7-21.8)**	**7.8**
23	Peripheral neuropathy	3.4	Prostatic hypertrophy	1.9	4 (0.107)	1 (0.013)	8.0 (0.9-71.5)	6.3
24	Hypertriglyceridemia	2.3	Soft-tissue edema	2.2	4 (0.107)	1 (0.013)	8.0 (0.9-71.5)	5.1
25	Heart failure	2.8	Colon polyps	2.5	12 (0.322)	3 (0.040)	8.0 (2.3-28.3)	6.9
26	Hypertension	1.7	Infertility	2.3	8 (0.214)	2 (0.027)	8.0 (1.7-37.7)	4.0
27	**Metabolic disorder**	**2.8**	**Menstrual abnormalities**	**2.3**	**74 (1.983)**	**19 (0.255)**	**7.8 (4.7-12.9)**	**6.3**
28	Arthralgia	1.9	Infertility	2.3	11 (0.295)	3 (0.040)	7.3 (2.0-26.3)	4.3
29	**Arthralgia**	**1.9**	**Aortic or mitral valve insufficiency**	**3.3**	**82 (2.198)**	**25 (0.335)**	**6.6 (4.2-10.2)**	**6.1**
30	Menstrual abnormalities	2.3	Colon polyps	2.5	13 (0.348)	4 (0.054)	6.5 (2.1-19.9)	5.7
31	Hepatomegaly	2.0	Fatigue	2.5	3 (0.080)	1 (0.013)	6.0 (0.6-57.7)	5.0
32	Colonic diverticula	2.4	Hepatomegaly	2.0	3 (0.080)	1 (0.013)	6.0 (0.6-57.7)	4.7
33	**Arthritis**	**2.5**	**Menstrual abnormalities**	**2.3**	**21 (0.563)**	**7 (0.094)**	**6.0 (2.6-14.1)**	**5.8**
34	Soft-tissue edema	2.2	Colonic diverticula	2.4	11 (0.295)	4 (0.054)	5.5 (1.8-17.3)	5.3
35	Menstrual abnormalities	2.3	Infertility	2.3	19 (0.509)	7 (0.094)	5.4 (2.3-12.9)	5.3
36	Arthropathy	2.3	Prostatic hypertrophy	1.9	5 (0.134)	2 (0.027)	5.0 (1.0-25.8)	4.3
37	**Arthralgia**	**1.9**	**Colonic diverticula**	**2.4**	**57 (1.528)**	**24 (0.322)**	**4.8 (3.0-7.6)**	**4.4**
38	Arthralgia	1.9	Prostatic hypertrophy	1.9	12 (0.322)	6 (0.080)	4.0 (1.5-10.6)	3.5

Abbreviations: CI = confidence interval; RR = relative risk.
[a]RR of the selected combination of conditions for acromegaly vs. nonacromegaly patients.
[b]The expected RR for conditions 1 & 2 was calculated under the assumption that risks of conditions are independent.
The 10 dyads with the highest prevalence in acromegaly are shown in bold.

REFERENCES

1. **Broder MS, Chang E, Sun GH, Neary MP, Ludlam W.** Identification of potential markers for Cushing disease. *Endocr Pract*. 2016;22:567-574.
2. **Munson M, Wrobel J, Holmes C, Hanauer D.** Data mining for identifying novel associations and temporal relationships with charcot foot. *J Diabetes Res*. 2014; 2014:214353.
3. **Miller RE, Learned-Miller EG, Trainer P, Paisley A, Blanz V.** Early diagnosis of acromegaly: computers vs clinicians. *Clin Endocrinol (Oxf)*. 2011;75:226-231.
4. Mayo Foundation for Medical Education and Research. Acromegaly. Available at: http://www.mayoclinic.org/diseases-conditions/acromegaly/symptoms-causes/dxc-20177626. Accessed March 3, 2016.
5. **Ben-Schlomo A, Sheppard MC, Stephens JM, Pulgar S, Melmed S.** Clinical, quality of life, and economic value of acromegaly disease control. *Pituitary*. 2011;14:284-294.
6. **Broder MS, Neary MP, Chang E, Cherepanov D, Katznelson L.** Treatments, complications, and healthcare utilization associated with acromegaly: a study in two large United States databases. *Pituitary*. 2014;17:333-341.
7. **Keyfitz K, Flieger W.** Population Clock. World population: an analysis of vital data. 1968. Available at: http://www.cabdirect.org/abstracts/19702701244.html. Accessed March 9, 2016.
8. **Schneider HJ, Kosilek RP, Günther M, et al.** A novel approach to the detection of acromegaly: accuracy of diagnosis by automatic face classification. *J Clin Endocrinol Metab*. 2011;96:2074-2080.
9. **Reid TJ, Post KD, Bruce JN, Nabi Kanibir M, Reyes-Vidal CM, Freda PU.** Features at diagnosis of 324 patients with acromegaly did not change from 1981 to 2006: acromegaly remains under-recognized and under-diagnosed. *Clin Endocrinol (Oxf)*. 2010;72:203-208.
10. **Melmed S.** Medical progress: acromegaly. *N Engl J Med*. 2006;355:2558-2573.
11. **Melmed S, Casanueva F, Cavagnini F, et al.** Consensus statement: medical management of acromegaly. *Eur J Endocrinol*. 2005;153:737-740.
12. **Lugo G, Pena L, Cordido F.** Clinical manifestations and diagnosis of acromegaly. *Int J Endocrinol*. 2012;2012:540398.
13. **Nachtigall L, Delgado A, Swearingen B, Lee H, Zerikly R, Klibanski A.** Changing patterns in diagnosis and therapy of acromegaly over two decades. *J Clin Endocrinol Metab*. 2008;93:2035-2041.
14. **Ayuk J, Sheppard MC.** Does acromegaly enhance mortality? *Rev Endocr Metab Disord*. 2008;9:33-39.
15. **Nabarro JD.** Acromegaly. *Clin Endocrinol (Oxf)*. 1987;26:481-512.
16. **Porter CJ, Juurlink I, Bisset LH, Bavakunji R, Mehta RL, Devonald MAJ.** A real-time electronic alert to improve detection of acute kidney injury in a large teaching hospital. *Nephrol Dial Transplant*. 2014;29:1888-1893.
17. **Mohktar MS, Redmond SJ, Antoniades NC, et al.** Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data. *Artif Intell Med*. 2015;63:51-59.
18. **Parikh RB, Kakad M, Bates DW.** Integrating predictive analytics into high-value care: the dawn of precision delivery. *JAMA*. 2016;315:651-652.
19. **Holdaway IM, Rajasoorya C.** Epidemiology of acromegaly. *Pituitary*. 1999;2:29-41.
20. **Broder MS, Chang E, Neary MP, Ludlam W.** Incidence of acromegaly in the United States: a claims-based analysis. ENDO 2016; 2016 April 3; Boston, MA. Available at: https://endo.confex.com/endo/2016endo/webprogram/Paper24139.html. Accessed February 6, 2017.