Utilization of Positive and Negative Controls to Examine Comorbid Associations in Observational Database Studies

Jigar R. Desai, PhD, Craig L. Hyde, PhD, Shaum Kabadi, PhD, Matthew St Louis, MS, Vinicius Bonato, PhD, A. Katrina Loomis, Aaron Galaznik, MD, and Marc L. Berger, MD

Background: Opportunities to leverage observational data for precision medicine research are hampered by underlying sources of bias and paucity of methods to handle resulting uncertainty. We outline an approach to account for bias in identifying comorbid associations between 2 rare genetic disorders and type 2 diabetes (T2D) by applying a positive and negative control disease paradigm.

Research Design: Association between 10 common and 2 rare genetic disorders [Hereditary Fructose Intolerance (HFI) and α -1 antitrypsin deficiency] and T2D was compared with the association between T2D and 7 negative control diseases with no established relationship with T2D in 4 observational databases. Negative controls were used to estimate how much bias and variance existed in datasets when no effect should be observed.

Results: Unadjusted association for common and rare genetic disorders and T2D was positive and variable in magnitude and distribution in all 4 databases. However, association between negative controls and T2D was 200% greater than expected indicating the magnitude and confidence intervals for comorbid associations are sensitive to systematic bias. A meta-analysis using this method demonstrated a significant association between HFI and T2D but not for α -1 antitrypsin deficiency.

Conclusions: For observational studies, when covariate data are limited or ambiguous, positive and negative controls provide a method to account for the broadest level of systematic bias, heterogeneity, and uncertainty. This provides greater confidence in assessing associations between diseases and comorbidities. Using this approach we were able to demonstrate an association between HFI and T2D. Leveraging real-world databases is a promising approach to identify and corroborate potential targets for precision medicine therapies.

Key Words: positive negative controls, observational database studies, comorbidity associations

(*Med Care* 2017;55: 244–251)

Within health care, recent years have seen an increasing proliferation and richness of electronic sources of clinical and observational data, whether from electronic health records (EHRs), administrative claims, or other sources of behavioral data. This proliferation, coupled with advances in computation, has enabled easy access and exponential increase in utilization of these databases. This study examines an approach for harnessing real-world data to complement precision medicine research.

Precision medicine is an emerging approach to disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle. (<http://www.nih.gov/precisionmedicine>).¹ Use of real-world data for precision medicine is currently limited by availability of genetic information that can be readily linked to available real-world data sources such as EHRs and administrative claims data. There are, however, cases where phenotypic indicators of genetic information in these observational data sources can be leveraged. One example is the use of diseases identifiable by diagnosis code that have a strong causal association with a known genetic mutation. A recent analysis using this approach, in over 110 million electronic medical records, identified a nondegenerate phenotypic code that links each complex disorder to a unique set of Mendelian diseases/loci.² Moreover, their observations of widespread comorbidity among Mendelian and complex diseases indicate that rare highly penetrant variants play a significant role in complex disease risk.² Thus, observational data sources can be leveraged to explore associations between genetic mutations and disease pathophysiology. This has tremendous implications for the use of real-world data to supplement preclinical information in identifying and corroborating potential targets for developing precision medicine therapies.

Use of observational databases, however, brings with it the need for methods to address limitations of such databases. To consistently address these biases, some methodologists have advocated for routine use of negative controls in observational studies to detect the presence, extent, and direction of uncontrolled bias and systematic error.^{3,4} These recommendations were prompted by recent studies that showed nearly half of significant associations between individual drugs

From Pfizer Inc., New York, NY.

Supported by Pfizer World Wide Research and Development.

The authors are employees of Pfizer Inc. and declare no conflict of interest. Reprints: Craig L. Hyde, PhD, Pfizer Inc., Eastern Point Road, Groton, CT 06340. E-mail: Craig.L.Hyde@pfizer.com.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medicalcare.com.

Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 0025-7079/17/5503-0244

and particular outcomes whether examined with a case-control, cohort, or self-controlled case series design, would be rendered not statistically significant using negative controls to estimate the empirical null distribution (eg, how much bias and variance exists when no association should be observed).⁵ Another study examined 53 drug-outcome pairs utilizing 10 different observational databases and demonstrated sensitivity of conclusions to value choices and selection of database, when holding study methods constant.⁶ Interestingly, this analysis found 21% of studies with a cohort design and 36% of studies from a self-control design can have ratios that range from statistically significant decreases to statistically significant increases depending on the database utilized.⁶ This is compounded by the known, but poorly quantified, bias in the coding process, as assigned ICD-9 (International Disease Classification) codes are often tied to financial incentives.

In this study, we harness the inherent value of observational data to make inferences about clinical and comorbid phenotypes for 2 rare genetic disorders, α -1 antitrypsin deficiency (A1AD) and hereditary fructose intolerance (HFI); we utilize positive and negative controls to address potential sources of systematic bias. Prior preclinical research in animals and small human studies have shown lower levels of α -1 antitrypsin in both type 1 and type 2 diabetes (T2D), suggesting a possible associations between A1AD and metabolic disorders.^{7–11} Recent observations of an unexpectedly high number of comorbidities between pairs of Mendelian diseases² coupled with known propensity for liver abnormalities in both A1AD and HFI led us to hypothesize that these Mendelian disorders are relatively common in patients with T2D. As mutations in both α -1 antitrypsin and aldolase B are associated with diagnosable and highly penetrant medical conditions—A1AD and HFI, respectively—patients with target mutations are identifiable by ICD-9 coding in administrative claims and electronic medical record data.

We evaluated the association between HFI and A1AD and T2D in 4 commonly used observational databases, 2 EHRs and 2 administrative claims. To quantify and adjust for heterogeneity and systematic bias, we compared the associations of these 2 rare genetic disorders with a spectrum of 17 other diseases—5 of which are known to be associated with T2D, 12 diseases (7 chronic, 5 acute) with no known association with T2D.

METHODS

We define comorbidity as the presence of an index disease plus additional disease in a person or group. This is different than instruments that measure global comorbidity or multimorbidity, which is the totality of disease in a person or group, with no index disorder (eg, Charlson Index, Cumulative Illness Rating Scale, Index of Coexisting Diseases, and Kaplan Index). Understanding the impact of comorbidity is confounded by choices regarding the number and types of diseases investigated and population studied.

Positive and Negative Controls

A positive control is defined as a condition that is understood to be positively associated with risk of T2D based on the scientific literature (eg, Fourth Nerve Palsy). A negative

control is defined as a condition that is understood to have no known association with T2D. A literature search (January 2014) was performed using mesh terms for each individual disease and T2D and manually curated with physician's review for clinical presence or absence of association. Thus, negative control diseases chosen show an absence of association in the literature and absence of a clinical explanation of association. Negative control diseases included were Epilepsy, Cervical Dystonia, Blepharospasm, Multiple Sclerosis, Acute Leukemia in Adults, Amyotrophic Lateral Sclerosis, and Complex Regional Pain Syndrome. Positive control diseases included were Sixth Nerve Palsy, Abdominal Aortic Aneurysm, Rheumatoid Arthritis, Peyronie Disease, and Chronic Obstructive Pulmonary Disease.

In addition to positive and negative chronic control diseases, 5 acute disorders (with no reason to expect any association) were examined in our analysis: Rash, Swimmer Ear, Spasmodic Dysphonia, Dislocated Shoulder, and Malaria. We examined all these disease controls to see whether in fact a subset of them may have been biased differently than the rest, and we found that we should not use the nonchronic acute disease among the negative controls for T2D because (a) they behave differently as a group (Supplemental Fig. 1, Supplemental Digital Content 1, <http://links.lww.com/MLR/B294>), than the 7 chronic negative control diseases used for T2D; (b) this difference was nonconservative (the acute disease controls had less correlation to T2D than the chronic negative controls); and (c) T2D categorically matches the “chronic” category, hence there are is an a priori reason to expect a systematic difference in the overall bias between chronic diseases and nonchronic acute diseases.

Databases

This study was performed using 4 large observational data sources, 2 administrative claims [Truven MarketScan Claims Database (2008–2012) and Optum Claims Database (2002–2012)] and 2 EHR databases [GE Centricity EHR Database (1995–2013) and Humedica (2008–2012) EHR Database] (Table 1). The Truven MarketScan database provides data on fully adjudicated inpatient, outpatient, and prescription drug claims for insured employees and their dependents, early retirees, and Medicare-eligible retirees. The commercial data are sourced from almost 100 payers, including government and commercial insurance companies. The Optum Claims Database contains deidentified administrative data from inpatient, outpatient, and prescription drug claims, which also include outpatient lab results; however, data are from just 1 insurer: United Health Care. The Humedica EHR database (when accessed) contained data from general practitioners, specialty care, and hospitalizations for approximately 15 million patients, 8 million of whom have integrated outpatient and hospital records from hospital chains, medical practices, and integrated delivery networks. The GE Centricity database is a pooled EHR database that captures patient-level clinical and demographic data derived from ambulatory care setting from >20,000 clinicians managing ~30 million patients in 49 states.

The final total analytical sample consisted of 17,258,087 subjects (Truven), 9,171,582 subjects (Optum),

TABLE 1. Number of Subjects for Each Disease From 4 Observational Databases

Outcomes	Truven	Optum	Humedica	GE Centricity
Unweighted sample size	96,561,490	47,848,331	14,513,153	33,356,691
Type 2 diabetes patients	6,368,229	2,636,698	998,307	2,225,098
Rare inherited diseases				
Hereditary fructose intolerance	3166	2584	557	635
α -1 antitrypsin deficiency	4507	3279	1584	3027
Negative controls				
Epilepsy	480,818	244,435	100,930	99,869
Cervical dystonia	39,805	21,603	4418	11,357
Blephrospasm	34,438	20,335	3647	4877
Multiple sclerosis	188,947	85,258	32,788	69,737
Acute leukemia in adults	22,064	10,905	1634	3197
Amyotrophic lateral sclerosis	10,910	4439	2519	4477
Complex regional pain syndrome	58,487	30,087	8293	19,698
Positive control diseases				
COPD	9,123,643	5,433,005	1,449,519	3,036,914
Peyronie disease	43,863	20,855	7310	15,436
Rheumatoid arthritis	644,975	325,330	97,562	214,760
Abdominal aortic aneurysm	89,736	289,151	48,183	93,754
Sixth nerve palsy	20,755	10,481	2667	2994
Nonchronic diseases				
Dislocated shoulder	193,884	120,845	24,0842	32,323
Spasmodic dysphonia	123,744	33,137	36,895	109,463
Malaria	3980	3128	725	2487
Rash	2,812,768	1,867,318	421,019	616,991
Swimmer ear	152,682	108,414	11,677	97,144

COPD indicates chronic obstructive pulmonary disease.

2,796,813 subjects (Humedica), and 5,745,757 subjects from GE Centricity. All data in these databases are deidentified to comply with Health Insurance Portability and Accountability Act (HIPAA) regulations, thus this study qualifies as non-human subjects research and was exempt from IRB approval.

Study Population and Endpoint

Each distinct patient and record was identified and extracted and subsequently duplicate diagnoses were removed as to not affect total counts of the outcomes of interest. Although each database contained a cohort of unique patients, there likely was some small overlap such that some patients may have been represented in more than one database. However, the way that data were collected for each dataset—sourcing from different covered populations—would suggest that these datasets are largely mutually exclusive and the amount of overlap is likely a small proportion. The magnitude of the potential overlap cannot be estimated by the research community as datasets are deidentified according to local privacy and HIPAA laws. The outcome endpoint for this analysis was a recorded diagnosis of T2D. Diabetic status was determined by the ICD-9 code 250.xx (ICD-9: 250.x0 or 250.x2). For rare genetic diseases, patients with HFI diagnosis were identified using ICD-9: 271.2 or A1AD diagnosis ICD-9: 273.4. For positive controls: Sixth Nerve Palsy (ICD-9: 378.54), AAA (ICD-9: 441.3, 441.4), RA (ICD-9: 714.0), Peyronie Disease (ICD-9: 607.85), and COPD (ICD-9: 490, 491, 492, 493, 494, 495, 496). For negative controls, the following ICD-9 codes were used: Epilepsy (ICD-9: 345), Cervical Dystonia (ICD-9: 333.83), Blepharospasm (ICD-9: 333.81), Multiple Sclerosis (ICD-9: 340), Acute Lymphoid Leukemia (ICD-9: 204.00,

204.01, 204.02), Amyotrophic Lateral Sclerosis (ICD-9: 335.20), and Complex Regional Pain Syndrome (ICD-9: 337.20, 337.21, 337.22). For the acute nonchronic diseases: Rash (ICD-9: 782.1), Swimmer Ear (ICD-9: 380.12), Spasmodic Dysphonia (ICD-9: 784.42), Dislocated Shoulder (ICD-9: 831), and Malaria (ICD-9: 084.6).

Statistical Analysis

Each disease was evaluated to see whether it had a statistically significant relationship against T2D by Fisher exact test on the 2×2 contingency table of T2D status versus each other disease. Several chronic diseases that have no theoretical basis on which to expect an association with diabetes were selected as negative controls to control for ascertainment bias in the data. The (natural log) odds ratio (OR) of each disease versus T2D was evaluated across 4 different databases. Using inverse-variance weighted meta-analysis, the association for each comorbid disease with T2D was reduced to a single value (with SE). During this process, Cochran Q test was used to determine heterogeneity (ie, variation of log ORs across databases greater than would be attributable to the observed SEs) for each disease against T2D. Because most diseases exhibited extremely significant heterogeneity across databases, an extra variance term representing between-database heterogeneity was added to each variance, a process referred to as a random-effects meta-analysis (as opposed to a fixed-effects meta-analysis, which does not add a variance term for heterogeneity). The negative controls were then further combined by meta-analysis. After each negative control disease had been assessed across databases by random-effects meta-analysis, the negative controls themselves exhibited an insignificant amount of heterogeneity between them. However, as a conservative

measure (and to establish a more general process), they were also combined by random-effects meta-analysis, although results using a fixed-effects meta-analysis for combining the controls are nearly identical, due primarily to the negligible amount of heterogeneity variance added.

Hence, negative controls against T2D were first estimated by meta-analysis across databases, and then were further combined by meta-analysis among themselves to obtain a single negative control estimate versus T2D with SE. The difference of (natural log) OR between each disease and the combined controls was then assessed, using the final random-effects variance for each disease and for the combined controls. This was then converted into a ratio of odds ratios (ROR) with a commensurate 95% confidence interval (CI) and reported as in Figure 3.

RESULTS

Although the genetic basis of A1AD and HFI have been known for over 2 decades,^{12,13} population-based studies investigating associated comorbidities are lacking. Our cohort from 4 large observational databases included 6942 individuals with HFI and 12,397 with A1AD; it also included 12,228,332 individuals diagnosed with T2D, 20,970,893 individuals with positive control diseases, and 1,450,436 individuals with neg-

ative control diseases (Table 1). The prevalence of A1AD and HFI and T2D was similar across all 4 databases and closely match published and Centers for Disease Control estimates, respectively, indicating these databases do not overestimate their prevalence in this investigation. The median follow-up time was 5 years in Truven, 5 years in Optum, 3 years in Humedica, and 7 years in GE Centricity.

As expected, we observed a positive but variable association (both in magnitude and distribution) for each positive control disease for cooccurrence with T2D in each database tested as assessed by individual unadjusted raw ORs (Supplemental Fig. 2, Supplemental Digital Content 2, <http://links.lww.com/MLR/B295>). The magnitude and distribution of the raw OR for each disease varied across databases. We could not detect a discernable pattern to the variability among or between Claims versus EHR databases. For some diseases the pattern of association varied noticeably depending on database investigated.

Unexpectedly, the negative controls tested for association with T2D also show a strong positive association with T2D that was variable in magnitude and distribution in each database for each of the negative control diseases tested (Figs. 1A–D and Supplemental Fig. 3, Supplemental Digital Content 3, <http://links.lww.com/MLR/B296>). A meta-analysis of the negative controls in all 4 databases provided an estimated OR of 1.8–2.3 (Fig. 2). These results did

FIGURE 1. Individual forest plots for each negative control disease from all 4 observational databases. Heterogeneity across diseases and databases for negative controls tested. Unadjusted odds ratio with 95% confidence intervals for association with type 2 diabetes (T2D) and (A) epilepsy, (B) blepharospasm (bleph), (C) multiple sclerosis (MS), and (D) cervical dystonia (CD) in each respective database (DB). The meta-odds ratio for each negative control disease is under the dotted line in all 4 plots. Vertical line at 1 marks the theoretical null. Individual forest plots for additional negative controls can be found in the supplemental content (Supplemental Fig. 3, Supplemental Digital Content 3, <http://links.lww.com/MLR/B296>). het.pval indicates heterogenous *P*-value.

FIGURE 2. Meta-analysis of all negative controls across all 4 databases. The combined meta-odds ratio of all negative control diseases, or the empirical null, across all databases with 95% CI is shown under the dotted line. ALS indicates amyotrophic lateral sclerosis; CI, indicates confidence interval; CRPS, complex regional pain syndrome; Het.p, heterogenous *P*-value; T2D, type 2 diabetes.

not differ whether they were calculated using either a fixed-effects or random-effects model. Thus, the magnitude and CIs for both unadjusted and adjusted associations between the 19 diseases tested and T2D was heterogenous in all 4 databases. For 4 out of 5 acute diseases (Rash, Swimmer Ear,

Spasmodic Dysphonia, Dislocated Shoulder, and Malaria), the amount of bias detected was generally lower than for chronic diseases (Supplemental Fig. 1, Supplemental Digital Content 1, <http://links.lww.com/MLR/B294>). This result is not surprising and probably reflects ascertainment and surveillance bias expected to be greater for chronic diseases. Thus, within-disease and between-database variability, and ascertainment and surveillance bias detected provide insight into the extent and direction of systematic bias in these observational databases in relation to T2D diagnosis (Figs. 1–3). Taken together, these data quantify and indicate large systematic bias and potential for overestimation of comorbid associations with T2D in these databases.

HFI and A1AD are rare autosomal recessive disorders where individuals are affected from birth or early childhood, respectively, thus clearly establishing the temporality of HFI or A1AD incidence before the onset of T2D. For these genetic disorders the prevalence rate closely approximates their incidence rates. We observed a 5-year period prevalence of 1/50,000 or 0.002% for HFI and 1/20,000 or 0.005% for A1AD in the 4 databases studied (Table 1). These data are in agreement with published incidence rates for HFI (1/22,000–1/50,000) and slightly underestimate A1AD prevalence (1/8000–1/20,000) in the general population.^{14,15} The unadjusted association between HFI and A1AD and T2D was

FIGURE 3. Random-effects meta-analysis adjusted for negative control diseases across all 4 databases. Forrest plot of meta-analysis of all diseases, across all databases and negative controls adjusted association between 12 chronic and acute conditions, 2 rare genetic diseases HFI and A1AD. The vertical line represents the theoretical null of 1. The empirical null of 1.9 with 95% CI is plotted below dotted line. Corresponding ratio represent ratio of the odds ratio or ROR (ROR=odds ratio of test/odds ratio negative control) for each disease with 95% CI and 2-sided *P*-values are labeled. Negative controls include epilepsy, cervical dystonia, blepharospasm, multiple sclerosis, acute leukemia in adults, amyotrophic lateral sclerosis, and complex regional pain syndrome. Analysis did not control for whether or not subjects were on therapy. AAA indicates abdominal aortic aneurysm; A1AD, α -1 antitrypsin deficiency; CI, confidence interval; COPD, chronic obstructive pulmonary disease; HFI, hereditary fructose intolerance; Het.p, heterogenous *P*-value; Pval, *P*-value; RA, rheumatoid arthritis; T2D, type 2 diabetes.

TABLE 2. Adjusted ROR From Random-effects Meta-analysis Based on Empirically Derived Null From All 4 Databases

Disease	OR	95% LB	OR	95% UB	ROR	95% LB	ROR	95% UB	ROR	<i>P</i> * (vs. Negative Controls ROR)
Sixth nerve palsy	6.35	5.78	6.98	3.15	2.72	3.65	6.21E-53			
AAA	5.78	4.15	8.05	2.87	2.02	4.07	3.51E-09			
RA	3.65	2.92	4.58	1.81	1.41	2.33	3.37E-06			
HFI	3.47	2.21	5.45	1.72	1.08	2.75	0.021			
Peyronie disease	3.02	2.20	4.13	1.50	1.07	2.09	0.017			
A1AD	2.71	1.74	4.20	1.34	0.85	2.11	0.196			
Spasmodic dysphonia	2.71	2.20	3.32	1.34	1.06	1.70	0.012			
COPD	2.66	2.45	2.88	1.32	1.14	1.51	8.69E-05			
Rash	1.57	1.46	1.70	0.78	0.68	0.89	0.0004			
Malaria	1.39	1.05	1.83	0.69	0.51	0.93	0.015			
Dislocated shoulder	1.23	1.07	1.43	0.61	0.51	0.73	2.01E-07			
Swimmer ear	1.02	0.82	1.28	0.51	0.39	0.65	1.76E-07			
Chronic negative control diseases†	2.01	1.79	2.25	N/A	N/A	N/A	N/A			

Random-effects meta-analysis of association between HFI and A1AD and type 2 diabetes across all 4 databases.

**P*-value represents difference in disease compared with “negative control diseases.”

†“Negative control diseases” include epilepsy, cervical dystonia, blepharospasm, multiple sclerosis, acute leukemia in adults, amyotrophic lateral sclerosis, and complex regional pain syndrome.

AAA indicates abdominal aortic aneurysm; A1AD, α -1 antitrypsin deficiency; COPD, chronic obstructive pulmonary disease; HFI, hereditary fructose intolerance; LB, lower bound; OR, odds ratios; RA, rheumatoid arthritis; ROR, ratio of odds ratios; UB, upper bound.

positive and heterogenous ($P < 0.001$) in all 4 databases. The unadjusted pooled OR calculated using a random-effects meta-analysis model was 3.48 (95% CI, 2.21–5.46) for HFI and 2.71 for A1AD (95% CI, 1.75–4.20) (Supplemental Fig. 4, Supplemental Digital Content 4, <http://links.lww.com/MLR/B297>).

To address the broadest level of uncertainty and systematic error, we conducted a meta-analysis of the negative controls for each disease (separately for each database and combined together). These results were compared with those found for individual negative controls to T2D, as well as for HFI and A1AD to T2D. Using a random-effects model, we observed the adjusted association between HFI and T2D was positive and significant compared with negative controls (ROR = 1.73; 95% CI, 1.08–2.75) (Fig. 3 and Table 2). However, the association between A1AD and T2D (ROR = 1.35; 95% CI, 0.86–2.12) was not significant ($P = 0.2$) (Fig. 3 and Table 2). Using a fixed-effects model to perform the meta-analysis, we observed that all diseases tested, including HFI (ROR = 2.19; 95% CI, 2.07–2.31) and A1AD (ROR = 1.33; 95% CI, 1.27–1.40), show a positive and statistically significant ROR ($P < 0.00001$) compared with negative control diseases (Supplemental Table, Supplemental Digital Content 5, <http://links.lww.com/MLR/B298>). In general, the strength of significance for association of each disease and T2D was higher and with reduced CIs in the fixed-effects meta-analysis model, which does not account for heterogeneity between databases when compared with random-effects model.

Thus, we advocate a proper assessment of true variability in OR is attained by utilizing multiple databases and by using a random-effects meta-analysis to account for heterogeneity across them for each disease. We further recommend doing this for a multitude of positive and negative controls and comparing a further meta-analysis of these controls against the test disease in the final step.

DISCUSSION

Although biases in observational databases have been well documented, great strides have been made in handling

and quantifying some of these biases.^{16–18} Similar to the application of positive/negative controls to assessment of drug-outcome pairs and adverse drug event analyses in observational databases,^{5,6,19} we show comorbidity analyses require empirical *P*-value calibration, by analysis of multiple positive and negative controls in multiple observational databases, to account for systematic bias and clinical heterogeneity. Using this approach, we were able to demonstrate an association between HFI and T2D. To our knowledge, this is the largest population-based study providing context around comorbid association of rare genetic disorders and T2D simultaneously with a spectrum of common diseases. Leveraging real-world databases is a promising approach to identify and corroborate potential targets for precision medicine therapies.

By utilizing positive and negative control diseases and subsequent meta-analysis for T2D comorbidity across databases and diseases, we could assess biases in 4 databases (both in magnitude and distribution). This enabled us to assess within-disease and between-database heterogeneity, as well as ascertainment bias, which could result in the potential for overestimation (inflated OR and *P*-values) of comorbid associations. We observed a nearly 200% increase in the empirical null (estimated by using negative controls and a more accurate estimate of what should be observed if there is no association to be found). This suggests that any observed comorbid association with T2D with an OR ratio of 1.8–2.3 is within systematic error and should be considered a false positive. We observed that fixed-effects and random-effects meta-analysis models yield different results regarding the association of A1AD with T2D. We present the very different and inflated results from a fixed-effects meta-analysis (Supplemental Table, Supplemental Digital Content 5, <http://links.lww.com/MLR/B298>), which does not account for heterogeneity between databases. Together our findings confirm that disease comorbidity analyses are sensitive to uncontrolled systematic bias, selection of data source, and meta-analysis models. Given the impact of comorbidity on design and interpretation of research studies and clinical care, we advocate the approach of

accounting for heterogeneity between databases and comparing to positive and negative controls, by random-effects meta-analysis, to protect against false positives in association testing and inflated *P*-values due to extremely large sample sizes.

Our study also shows that for comorbidity analyses, clinical heterogeneity, in addition to database selection, introduces significant systematic bias that must be addressed. Comparing difficult-to-diagnose with easily diagnosed diseases introduces considerable systematic biases due to increased likelihood of capturing data in 1 set of variables if other variables are available in that individual or cohort. Moreover, diseases that can be observed in routine clinical care and do not require specialized testing (eg, Blepharospasm) can have different covariates and differential variability compared with difficult-to-diagnose disorders. Established methods used to account for covariate imbalance (eg, Multiple Regression, Propensity Score Matching, and Instrumental Variables) perform optimally when rich covariate data are available and the population is normally ascertained. Positive and negative controls analyses, in conjunction with established covariate analysis, seems to be a valuable approach in situations where covariate imbalances are unknown or covariate data are limited (eg, clinical surveillance and ascertainment bias).

Although we used 4 US observational databases in this study, it is likely that other databases, in the United States and elsewhere, will have varying degrees of systematic bias due to collection methods and health care delivery in their respective regions. Therefore, new sets of positive and negative controls should be selected for each new outcome interrogated for that study. Subject matter expertise is required in the selection of positive and negative controls, as incorrect selection of controls can skew results and introduce more error. It can be argued that some databases may have a common overall bias, or that database source may explain a significant amount of variability among ascertainment and/or surveillance biases. This is a dangerous assumption to apply to each disease, and in general it should be expected that ascertainment and surveillance bias between any pair of diseases may differ by both database and disease. Further, even if 1 database experiences a particular “type” of bias, that bias might likely still have a different numerical effect across different disease pairs. Moreover, it has been previously reported that observed ratios for 53 drug-outcome pairs, assessed in 10 different databases, can range from statistically significant decreases to statistically significant increases depending on the database utilized, even when holding study methods constant.⁶

There are limitations to this study. One potential limitation is that this study did not fully adjust for various standard or nonstandard risk factors. These databases do not have complete information for all covariates of interest, nor record this information with the same frequency and fidelity. Particular covariates may be sparsely available in claims database and/or missing at a high rate in EHRs. The intent of our analytic approach is *not* to supplant established methods for covariate analysis, but rather to augment the ability to detect and account for the presence and direction of uncontrolled bias. Interestingly, our data suggest that duration of follow-up does not explain the majority of heterogeneity and enrichment observed in these databases. Rather, ascertainment and surveillance bias and

clinical heterogeneity of diseases is more important. Consistent with this is our finding that the database with longest follow-up time (GE centrality) had a consistently lower enrichment for negative controls tested. Longer duration of follow-up would generally be expected to increase the odds of T2D diagnosis and show greater, not lower, enrichment as observed.

Another limitation of this study is the use of ICD-9 codes rather than phenotypic algorithms. Although it has been shown that complete phenotypic algorithms can outperform single codes for some diseases, validated phenotypic algorithms are not available for all diseases evaluated (<http://www.phekb.org>).²⁰ However, rare diseases are commonly assigned 1 ICD-9 code. This reductionist method allows for quantification of biases introduced by ICD-9 coding to be more directly compared, notwithstanding the well-appreciated underestimation of rare disease prevalence using single ICD-9 codes.^{21,22}

The challenges in leveraging observational health care data are akin to those faced by gene expression and network pathway analysis a decade ago.^{23,24} The practitioners conducting observational database studies should take note of the tremendous advances utilizing gene expression and network pathway data have made despite similar challenges in data quality, reproducibility, and advent of new, more sensitive, data capture (eg, RNAseq).^{24,25} Although observational studies have limitations, important insights can be gained despite the presence of bias and data limitations by developing methods that account for these biases. Continued focus and effort to develop new methods is a critical step in realizing these insights.

ACKNOWLEDGMENT

The authors thank Rumya Raghunath for her comments and help in preparing this manuscript.

REFERENCES

1. Precision medicine initiative. 2015. Available at: <http://www.nih.gov/precisionmedicine>. Accessed October 15, 2015.
2. Blair DR, Lyttle CS, Mortensen JM, et al. A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. *Cell*. 2013;155:70–80.
3. Lipsitch M, Tchetgen Tchetgen E, Cohen T, et al. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21:383–388.
4. Gustafson P, McCandless LC. Probabilistic approaches to better quantifying the results of epidemiologic studies. *Int J Environ Res Public Health*. 2010;7:1520–1539.
5. Schuemie MJ, Ryan PB, DuMouchel W, et al. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med*. 2014;33:209–218.
6. Madigan D, Ryan PB, Schuemie M, et al. Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol*. 2013;178:645–651.
7. Zhang B, Lu Y, Campbell-Thompson M, et al. Alpha1-antitrypsin protects beta-cells from apoptosis. *Diabetes*. 2007;56:1316–1323.
8. Sandström CS, Ohlsson B, Melander O, et al. An association between type 2 diabetes and alpha-antitrypsin deficiency. *Diabet Med*. 2008;25:1370–1373.
9. Lu Y, Choi YK, Campbell-Thompson M, et al. Therapeutic level of functional human alpha 1 antitrypsin (hAAT) secreted from murine muscle transduced by adeno-associated virus (rAAV1) vector. *J Gene Med*. 2006;8:730–736.
10. Sandler M, Gemperli BM, Hanekom C, et al. Serum alpha 1-protease inhibitor in diabetes mellitus: reduced concentration and impaired activity. *Diabetes Res Clin Pract*. 1988;5:249–255.
11. Lisowska-Myjak B, Pachecka J, Kaczyńska B, et al. Serum protease inhibitor concentrations and total antitrypsin activity in diabetic and

- non-diabetic children during adolescence. *Acta Diabetol.* 2006;43:88–92.
12. Ali M, Rellos P, Cox TM. Hereditary fructose intolerance. *J Med Genet.* 1998;35:353–365.
 13. Fregonese L, Stolk J. Hereditary alpha-1-antitrypsin deficiency and its clinical consequences. *Orphanet J Rare Dis.* 2008;3:16.
 14. Cox TM. The genetic consequences of our sweet tooth. *Nat Rev Genet.* 2002;3:481–487.
 15. de Serres FJ, Blanco I, Fernández-Bustillo E, PI S and PI Z alpha-1 antitrypsin deficiency worldwide. A review of existing genetic epidemiological data. *Monaldi Arch Chest Dis.* 2007;67:184–208.
 16. Berger ML, Mamdami M, Atkins D, et al. Good research practices for comparative effectiveness research: defining, reporting, and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force—Part I. *Value Health.* 2009;12:1044–1052.
 17. Cox E, Martin BC, Van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force—Part II. *Value Health.* 2009;12:1053–1061.
 18. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol.* 2012;65:126–131.
 19. Schuemie MJ, Trifirò G, Coloma PM, et al. Detecting adverse drug reactions following long-term exposure in longitudinal observational data: the exposure-adjusted self-controlled case series. *Stat Methods Med Res.* 2014;31:31–37.
 20. The phenotype knowledge base, PheKB. 2014. Available at: <http://www.phekb.org>. Accessed October 15, 2015.
 21. Rath A, Olry A, Dhombres F. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Hum Mutat.* 2012;33:803–808.
 22. Sollie A, Sijmons RH, Lindhout D, et al. A new coding system for metabolic disorders demonstrates gaps in the International Disease Classifications ICD-10 and SNOMED-CT, which can be barriers to genotype-phenotype data sharing. *Hum Mutat.* 2013;34:967–973.
 23. Ramasamy A, Mondry A, Holmes CC, et al. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.* 2008;5:1320–1332.
 24. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol.* 2012;8:1–10.
 25. Rung J, Brazma A. Reuse of public genome-wide gene expression data. *Nat Rev Genet.* 2013;14:89–99.