Beyond χ_2. Finding significant differences in cohort studies despite wide ranging demographic diversity

Gordon H. Lemmon, PhD[1-5], Sergiusz Wesolowski, PhD[1-5], Alex Henrie[1-5], Mark Yandell, PhD[1-5]. [1]Program in Personalized Health, [2]Utah Center for Genetic Discovery, [3]Department of Human Genetics, [4]School of Medicine, [5]University of Utah, Salt Lake City, UT

What might the attendee be able to do after being in your session?

The clinical researcher will learn about a novel approach for calculating p-values that relate exposures and outcomes within cohort studies. Rather than stratify the cohort or attempt to build paired cohorts, our approach models the effects of confounding variables such as age, gender, race, ethnicity, insurance type and EMR exposure. The clinician will learn how to estimate the probability of each outcome in each patient in the cohort in the context of confounding variables – and apply these priors to calculate high resolution p-values.

Description of the Problem or Gap

Both comorbidity discovery and cohort studies suffer from the same limitation. The statistical tests employed by these approaches include among others χ_2, binomial, and Fisher's exact. Each of these tests are designed to determine if a nonrandom association between two categorical variables exists. The limitation of these tests is that a confounding variable might be driving the association. For instance, the test might find an association between baldness and obesity, even though this association is actually driven by age.

There are two common approaches for dealing with confounding variables. Using the stratification approach, a researcher will examine sub-populations of more uniform consistency. For instance, a χ_2 p-value will be calculated based on data from a stratum of only females in their 50s. Another approach to dealing with confounders is to build paired cohorts. In this approach, the cases and controls are paired – so that ages and genders have approximately equal representation between cases and controls. In the stratification approach, the more we homogenize our study population, the smaller it becomes. Thus, controlling for confounding variables means sacrificing statistical power. In the paired cohort approach, the more confounders we attempt to pair on, the more difficult it becomes to find pairs of enrollees that are appropriately matched. How many pairs will we find if we are matching on age, gender, ancestry, ethnicity, insurance type and EMR exposure? How close of a match will we need before the pair is added to our cohort?

Furthermore, with many clinical trials now being approved for evaluation on very small sample sizes, it is important that we understand the differences in demographics and health between cases and controls and account for these differences in study design.

Methods: What did you do to address the problem or gap?

We apply logistic regression modeling (LRM) to determine a per-patient baseline probability for each outcome and exposure under investigation. The features are confounding demographic variables. Next, under the null hypothesis, per-patient joint probabilities for pairs of outcomes and/or exposures are calculated as the product of LRM probabilities. These per-patient joint probabilities parameterize a Poisson binomial (PB) distribution. PB is a generalization of a binomial distribution wherein each trial (patient) has a distinct probability of success. We estimate the Poisson binomial survival function (p-value), comparing expected to observed counts. We will refer to our approach as LRM+PB.

Results: What was the outcome(s) of what you did to address the problem or gap?

We use our method to search for comorbidities among medical records from the University of Utah Hospitals and Clinics. The medical records include 1,637,750 patients, 25,065,973 visits and over 553,446,085 observations. These observations include diagnosis codes, procedure codes, and medications – collectively, "medical terms". We calculate Bonferroni corrected p-values between all pairs of medical terms.

For each term we fit an L1 penalized LRM. We build LRMs for 750 CCS diagnosis codes, 250 CCS procedure codes, and 2000 RxNorm CUI medication codes. The percentages of models that included each demographic feature are shown in Table 1.

Table 1. Feature selection by L1 regularization

Feature	% of LRMs that including the feature
Sex	42.2%
Ancestry	27.4%
Ethnicity	37%
Insurance	31.2%
EMR Exposure	100%

Per person joint probabilities for all pairs of terms were calculated as the products of LRM probabilities. From these, the expectation and variance under a Poisson binomial distribution were calculated. Finally, Poisson binomial p-values were calculated for each pair of terms.

Table 2. Log10 p-values comparing stratification with our LRM+PB approach. Colored blue are those p-values that pass a Bonferroni corrected alpha threshold of 1e-7.88.

	Concussion and migraine(1)		**Multiple myeloma and Multiple sclerosis(2)**		**Cancer of pancreas and hypertension(3)**		
Binomial Patients	*Log10 p-value* Binomial	LRM+PB	*Log10 p-value* Binomial	LRM+PB	*Log10 p-value* Binomial	LRM+PB	LRM Features
all_patients	-933.01	-933.01	-121.29	-121.29	-405.45	-405.45	no features
female	-859.67	-936.5	-70.4	-135.02	-191.28	-399.97	gender
+age 50-59	-126.08	-1031.33	-4.8	-64.61	-6.13	-238.81	+age
+caucasian	-92.21	-911.5	-4.83	-86.66	-6.31	-159.45	+ancestry
+nonhispanic	-91.29	-885.72	-6.32	-85.71	-10.19	-60.93	+ethnicity
+commercial	-4.14	-858.85	-0.49	-79.49	-3.1	-60.29	+insurance
+3yr history	-1.06	-70.38	-0.39	-10.14	-1.48	-54.64	+span

Table two investigates 3 different comorbidities. Each of these comorbidities has been described in the literature. Without some way of accounting for confounding variables, all p-values are inflated (first row). As we add more criteria to our stratification strategy, our p-values weaken. This weakening of p-value is largely a byproduct of decreasing sample size rather than accounting for the effects of confounding variables. In contrast, the deflation in p-value seen as features are added to the LRM+PB model, is the result of removing the effects of the confounding variable rather than decreasing the sample size. Notwithstanding PB+LRM retains statistical power to discover true positive associations, it nevertheless still removes more false positives than stratification by age and gender.

Discussion of Results / Conclusion

Coupling logistic regression modeling with the Poisson binomial distribution allows us to retain statistical power while limiting false positive associations. This approach should be preferred to stratification methods or pairing of cases and controls. Similar to the issue with stratification demonstrated in Table 2, the tighter the pairing parameters, the smaller the number of matching pairs and the weaker the p-values will be. Another way to frame this conclusion is that LRM+PB allows us to discover true positive associations with smaller sample sizes.

Attendee's Take-away Tool

We are developing CoDE – comorbidity discovery engine – a website for querying medical terms – diagnoses, procedures, and medications. We provide cooccurrence p-values, and directional p-values – e.g. term A tends to precede term B within a 90 day window; or term C tends to precede term D by greater than 90 days. These directional p-values are used to build a flow diagram of disease progression. The investigator can filter comorbidities by cooccurrence p-value, directional p-values, or by flow rate (percent of patients with term A, who later are coded with term B). See Figure 1.

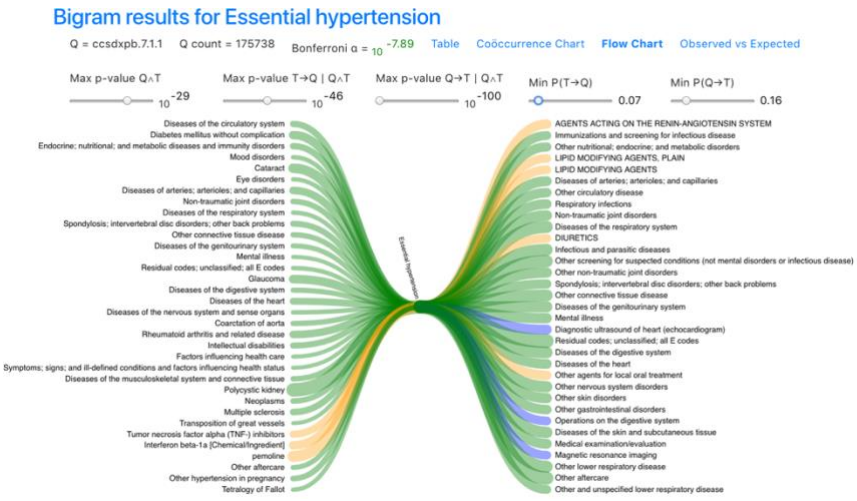

Figure 1. Screenshot from CoDe. Flow chart view shown.

URL: <<PUBLIC URL COMING SOON>>

References

1. Seifert T. The relationship of migraine and other headache disorders to concussion. Handb Clin Neurol. 2018;158:119-26.
2. Shimanovsky A, Alvarez Argote J, Murali S, Dasanu CA. Autoimmune manifestations in patients with multiple myeloma and monoclonal gammopathy of undetermined significance. BBA Clin. 2016;6:12-8.
3. Saif MW, Kaley K, Lamb L. Pancreatic Adenocarcinoma Complicated by Sinistral Portal Hypertension. Cureus. 2016;8(7):e689.