

Title: A phenome-wide association study identifying risk factors for pediatric post-concussion syndrome

Authors:

Aaron M. Yengo-Kahn, MD^{1,2}; Natalie Hibshman, BA^{2,3}; Christopher M. Bonfield, MD^{1,2}; Eric S. Torstenson, BS^{4,5,6}; Katherine A. Gifford, PsyD^{2,7}; Daniil Belikau⁸; Lea K. Davis, PhD^{6,9}; Scott L. Zuckerman, MD, MPH^{1,2}; Jessica K. Dennis, MSc, PhD^{2,6,8,9,10}

Affiliations:

¹Department of Neurological Surgery, Vanderbilt University Medical Center, Nashville, TN

²Vanderbilt Sport Concussion Center, Vanderbilt University Medical Center, Nashville, TN

³Vanderbilt University School of Medicine, Nashville, TN

⁴Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

⁵Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN

⁶Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN

⁷Department of Neurology, Vanderbilt University Medical Center, Nashville, TN

⁸British Columbia Children's Hospital Research Institute, Vancouver, Canada

⁹Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

¹⁰Department of Medical Genetics, University of British Columbia, Vancouver, Canada

Corresponding Authors and Reprint Requests:

Jessica Dennis, MSc, PhD

British Columbia Children's Hospital

Research Institute

950 West 28th Ave, Variety Building Room 371A

Vancouver, BC, V5Z 4H4 Canada

Email: jessica.dennis@bcchr.ca

Phone: 604-875-2000 x7871

Financial Disclosures and Competing Interests

Aaron Yengo-Kahn – Reports no disclosures or competing interests.

Natalie Hibshman - Reports no disclosures or competing interests.

Chris Bonfield - Reports no disclosures or competing interests.

Eric Torstenson - Reports no disclosures or competing interests.

Katherine Gifford - Reports no disclosures or competing interests.

Daniil Belikau - Reports no disclosures or competing interests.

Lea Davis - Reports no disclosures or competing interests.

Scott Zuckerman - Reports no disclosures or competing interests.

Jessica Dennis - Reports no disclosures or competing interests.

Study Funding

There was no direct funding for this work, however several authors acknowledge they are supported as follows:

Dr. Jessica K. Dennis is supported by the Canadian Institutes of Health Research (award MFE- 142936).

Dr. Katherine Gifford is supported by the National Institute on Aging (award K23045966)

This project was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN. The datasets used for this project were obtained from Vanderbilt University Medical Center's Synthetic Derivative, which is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH funded Shared Instrumentation Grant S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975.

Abstract

Objective: To identify risk factors and generate hypotheses for pediatric post-concussion syndrome (PCS) using a phenome-wide association study (PheWAS).

Methods: A PheWAS (case-control) was conducted following the development and validation of a novel electronic health record-based algorithm that identified PCS cases and controls from an institutional database of >2.8 million patients. Cases were patients ages 5-18 with PCS codes or keywords identified by natural language processing of clinical notes. Controls were patients with mild traumatic brain injury (mTBI) codes only. Patients with moderate or severe brain injury were excluded. All patients used our healthcare system at least three times 180 days before their injury. Exposures included all pre-injury medical diagnoses assigned at least 180 days prior.

Results: The algorithm identified 274 pediatric PCS cases (156 females) and 1,096 controls that were age and sex matched to cases. Cases and controls both had a mean of >8 years of healthcare system use pre-injury. Of 202 pre-injury medical, four were associated with PCS after controlling for multiple testing: headache disorders (OR=5.3; 95% CI 2.8-10.1; $P=3.8e-7$), sleep disorders (OR=3.1; 95% CI 1.8-5.2; $P=2.6e-5$), gastritis/duodenitis (OR=3.6, 95% CI 1.8-7.0; $P=2.1e-4$), and chronic pharyngitis (OR=3.3; 95% CI 1.8-6.3; $P=2.2e-4$).

Conclusions: These results confirm the strong association of pre-injury headache disorders with PCS and provides evidence for the association of pre-injury sleep disorders with PCS. An association of PCS with prior chronic gastritis/duodenitis and pharyngitis was seen that suggests a role for chronic inflammation in PCS pathophysiology and risk. These factors should be considered during the management of pediatric mTBI cases.

INTRODUCTION

The majority of traumatic brain injury (TBI) affecting children are mild traumatic brain injuries (mTBI).^{1,2} While the definition of concussion stipulates negative neuroimaging findings, the symptoms and recovery of concussions and mTBIs are indistinguishable.³ Although most concussions self-resolve in one to three weeks,⁴ approximately 10-30% of patients experience persistent cognitive, somatic, and emotional symptoms known as post-concussion syndrome (PCS).^{3,5-7}

Despite significant efforts to understand which patients are at risk for PCS,⁸ PCS has no clear unifying pathophysiology.⁹ Growing evidence suggests a higher risk of PCS is associated with younger age,^{10,11} female sex,^{10,12} prior mTBI,^{13,14} and certain pre-existing conditions such as headaches,^{15,16 17} and psychiatric diagnoses.^{13,17-20} Nonetheless, models of mTBI recovery remain only modestly predictive,^{17,21} highlighting the need for further research into PCS and inclusion of previously understudied risk factors. Moreover, most previous research has been restricted to a small or moderate number of mTBI patients drawn from specialty concussion clinics^{5,13,22} or from a single institution emergency department,²³ with drop-out rates as high as 40%.²¹

Electronic health record (EHR)-wide investigations utilizing algorithm-based case selection offer at least two major advantages in the search for PCS risk factors. First, large numbers of patients from across the health care system can be ascertained at a fraction of the time and cost of a manually or prospectively ascertained sample,²⁴ with drop-outs minimized by studying health system use patterns in the EHR, as opposed to requiring active patient engagement with the study team. Second, the entire EHR of mTBI patients can be mined using a

Phenome-wide Association Study (PheWAS) to reveal pre-injury medical diagnoses that increase the risk of poor recovery.²⁵

The objective of this study was to identify risk factors and generate hypotheses for pediatric PCS from across the EHR using a PheWAS, harnessing a novel EHR-based algorithm for PCS cases and controls.

METHODS

Setting and Study Design

A case-control study of pediatric PCS patients was undertaken with age- and sex-matched mTBI controls receiving care at Vanderbilt University Medical Center (VUMC). VUMC is a tertiary care center that provides inpatient and outpatient care in Nashville, TN. The VUMC EHR was established in 1989 and includes data on billing codes (hereafter referred to simply as codes) from the International Classification of Diseases, 9th and 10th editions (ICD9 and ICD10), Current Procedural Terminology (CPT) codes, laboratory values, reports, and clinical notes. The de-identified mirror of the EHR used for research includes patient records on more than 2.8 million individuals. Data analyzed in this study spanned May, 1989 to March, 2019.

Standard Protocol Approvals, Registrations and Patient Consents

The VUMC Institutional Review Board oversees the VUMC EHR research database and approved this study (IRB 151116). The research was performed without patient involvement and relied upon deidentified data for which patient consent was waived.

Case Selection Algorithm Development and Validation

The present study of pediatric PCS patients is part of a broader research program on using EHRs to study PCS, in which we designed and evaluated four different strategies for identifying PCS cases from the EHR (Figure 1). First, we excluded patients with evidence of moderate or severe TBI, which we defined as a codes mapping to the categories of skull fracture,

contusion, or hemorrhage, or a neurosurgical procedure code (ICD9: 01, 02; ICD10: 00, ON, R40.242, R40.243; CPT: 61000-62258).²⁶ Next, we specified four case definitions that selected patients based on the presence of PCS codes (ICD9: 310.2; ICD10: F07.81) and/or keywords in outpatient clinical notes ("postconcuss" or "post-concuss" or "post concuss"). Case Definition 1 required patients to have two PCS codes separated by at least 14 days, an ascertainment strategy that was successful in previous analyses.²⁴

Case Definition 2 required patients to have one PCS code, as well as one PCS keyword, which was identified by applying natural language processing to the outpatient clinical notes in three stages. First, outpatient clinical notes containing PCS keywords using the crude regular expression "post[-]?concuss", which covers the keywords "postconcuss", "post-concuss", and "post concuss," were queried and aggregated. Second, the records were trimmed down to sentences containing PCS keywords using the natural language processing (NLP) library spaCY.²⁷ Third, these sentences were evaluated for negation terms (e.g., "no evidence of postconcussion syndrome"; "postconcussion syndrome not suspected"). For this, the library pyConTextNLP²⁸ was utilized and assigned weights to each sentence corresponding to the evidence in support of PCS. Weights were: (0) sentences contained conflicting or no definitive evidence such as entries in a problem list; (1) sentences were suggestive of positive evidence; (2) sentences had definitively positive evidence; (-1) sentences were suggestive of negative evidence; (-2) sentences had definitively negative evidence. Weights were summed across PCS containing sentences in the clinical note and a final score was assigned to the note itself. Clinical notes with scores of 0 or more were considered to have evidence of PCS.

Case Definition 3 required patients to have only one PCS code, while Case Definition 4 required patients to have only one PCS keyword. Across all case groups, we only included cases

whose last PCS mention (code or keyword) occurred when they were five years of age or older, since symptoms reported in children younger than five years old are difficult to confirm.²⁹

We evaluated each case definition by manually reviewing the EHRs of 50 randomly selected patients from each group, operationally defining PCS as persistent symptoms for longer than two weeks.^{4,30} We abstracted data on the mTBI preceding the PCS diagnosis, including information on care sought, brain imaging, and injury cause, as well as persistence of >1 PCS symptom, as defined as any of the 22 symptoms listed on the PCS symptom scale contained within the 5th edition of the sport concussion assessment tool (SCAT5).³ Data were abstracted into a REDCap database by one reviewer (NH), and difficult-to-classify data elements were resolved by consensus (NH, JKD, and AYK). Patients were classified as true positives if they had both evidence of a mTBI preceding the PCS diagnosis, and if PCS symptoms persisted beyond 14 days of the mTBI.⁴ We calculated the positive predictive value (PPV) of each case definition as the number of true cases divided by the number of cases meeting the case definition.

Manual Review identified 44 and 37 true positives, yielding PPVs of 88% and 74% respectively for case definitions 1 and 2 (Figure 1). The PPVs of case definitions 1 and 2 were 100% and 64% respectively in the subset of manually reviewed pediatric patients, and descriptive characteristics of the true positive set are presented in Table 1.

Selection of PheWAS Participants

Analyses proceeded with cases identified with case definitions 1 and 2 given superior PPV (Figure 1). Each PCS case was matched to four controls on sex and age (within one-year intervals). Controls had no PCS codes or keywords in their EHR, but evidence of an mTBI, i.e.

codes that mapped to the categories of concussion / loss of consciousness, or other and unspecified head injury.²⁶ Patients with evidence of a moderate or severe TBI were excluded from both groups.

In both cases and controls, an index event was defined to delineate pre-injury risk factors for PCS. In cases, the index event was the first mTBI code, or the first PCS code if the patient had no mTBI code or if the first PCS code preceded the first mTBI code. In controls, the index event was the first mTBI code. To ensure adequate data for PheWAS analyses, patients had to have visited VUMC at least three times at least 180 days before the index event. Additionally, controls had to have visited VUMC at least once 30 days after the index event, to ensure that they continued to use VUMC post-injury and had not sought PCS care elsewhere. Finally, we excluded patients whose index event occurred over age 18 or later. In the primary analysis, ICD codes were restricted to those assigned at least 180 days before the index event, identifying diagnoses long preceding the injury.

Exposure Assignment

Phecodes are used in PheWAS to collapse multiple ICD9 and ICD10 codes into a smaller number of higher-order disease categories. The Phecode Map v1.2(beta), as implemented in the PheWAS R package v0.99.5-2,³¹ was used to map ICD9 and ICD10 codes in this >180 day window to phecodes. Patients were assigned to the affected group for a given phecode if they had at least one ICD code that mapped to a given phecode, otherwise they were assigned to the unaffected group. Although most PheWAS analyses in adults require at least two ICD codes

before assigning a patient to an affected phecode group,²⁵ the threshold was lowered to one ICD code to account for the lighter health system usage among children.

Statistical Analyses

Descriptive characteristics of the matched cases and controls were compared using a conditional logistic regression model to account for matching. Statistically significant differences existed in the pre-injury time spent in the EHR and number of pre-injury ICD codes, thus these variables were included as covariates in the PheWAS analysis. PCS case/control status was modeled as the outcome variable and each phecode was considered a predictor variable in separate conditional logistic regression models (clogit function in the R survival package). To ensure sufficient cell counts for interpretable odds ratios, only phecodes with at least 25 affected patients were included. In the primary analysis, 201 phecodes met this threshold and the Bonferroni correction for phenome-wide statistical significance was $P < 2.48e-4$ ($0.05/201$). Sex-stratified analysis on the four phecodes that passed our threshold for statistical significance was performed and the effect size of these phecodes between sexes was compared.

Three additional sensitivity analyses were conducted to assess the robustness of the results: (1) restricting PheWAS to the cases and matched controls identified by case definition 1, which had the strongest positive predictive value (PPV); (2) restricting PheWAS to the cases and matched controls with a documented mTBI in their EHR; and (3) broadening the pre-injury window to include ICD codes diagnosed at least 30 days before the index event (as opposed to 180 days in the primary analysis). The pre-injury window was not broadened to include codes assigned immediately before or at the same time as the index event because the focus was on pre-injury risk factors, which have been incompletely documented in previous research. All

statistical analyses were performed in R 3.6.0 (R Foundation, Vienna, Austria). There were no missing data in the primary analysis and no data were imputed.

Data Availability Statement

Full results are available in a public, open access repository (<https://dennislabs.ca/resources-and-software/>). Data may be used with proper citation and crediting.

RESULTS

Participant Characteristics

PheWAS selection criteria identified 274 pediatric cases and 9,069 potential pediatric controls (Figure 2). Cases were older than eligible controls (median age 15.0 vs. 12.1) and were more likely to be female (56.9% vs. 39.1%). Therefore, each case was matched on age and sex to four eligible controls, for a final sample size of 274 cases (156 females and 118 males) and 1096 controls (Table 2). Of these 274 cases, 196 (71.5%) were identified by case definition 1. Over three-quarters of cases had a documented mTBI in their EHR.

Within the matched sample (Table 2), cases and controls had a comparable number of visits at least 180 days before the index event (20.6 vs. 19.2, $P=0.550$). Cases, however, had a higher number of codes assigned at least 180 days before the index event (44.5 vs 34.0, $P=0.058$) and a higher number of days in the EHR pre-index event (3302 days in cases vs. 2929 days in controls, $P=8.6e-4$). Though both groups were in the EHR long before the index injury (>8 years), to safeguard against group differences, the PheWAS analyses included covariates for the number of codes and days pre-index event.

PheWAS

In the primary analysis of codes assigned at least 180 days before the index event, cases were more likely than controls to have reported headache syndromes (OR=5.3; 95% CI 2.8-10.1; $P=3.8e-7$), sleep disorders (OR=3.1; 95% CI 1.8-5.2; $P=2.6e-5$), gastritis and duodenitis (OR=3.6, 95% CI 1.8-7.0; $P=2.1e-4$), and chronic pharyngitis and nasopharyngitis (OR=3.3; 95% CI 1.8-6.3; $P=2.2e-4$)(Figure 3;Table 3). Cases were also more likely than controls to have a pre-injury diagnosis of anxiety disorder (OR=2.4; 95% CI 1.4-4.2; $P=0.001$). This phecode fell below the threshold for phenome-wide significance but was the most strongly associated

psychiatric phecode. Phecode associations were similar in males and females (Table 3). Full sex-stratified PheWAS results are shown graphically in Supplemental Figures 1 and 2.

Restricting analysis to the 196 cases ascertained using case definition 1, results were similar to the primary analysis, with the exception of an attenuated effect estimate for other headache syndromes (OR 1.9; 95% CI 0.68-5.17; $P=0.221$)(Table 4). Restricting analysis to the 208 cases with a documented mTBI in their EHR did not appreciably change any of the top phecode associations (Table 4). Using phecodes diagnosed at least 30 days before the index event, results were largely unchanged compared to those in the primary analysis, except that the migraine phecode now crossed the threshold for phenome-wide significance (OR 2.5; 95% CI 1.6-4.1; $P=1.7e-4$)(Supplemental Figure 3).

DISCUSSION

Our EHR-based PCS algorithm identified 274 pediatric PCS cases and 1096 matched pediatric mTBI controls from a single health care system in the United States. We then leveraged the entire EHR of cases and controls (8-9 years pre-injury, on average) to test over 200 pre-injury diagnoses for association with PCS risk using a PheWAS. Results from the PheWAS recapitulated established PCS risk factors (headache and migraine history) and identified previously underreported associations with pre-injury sleep disorders and inflammatory-related diagnoses.

Pediatric mTBI is a major public health concern, especially for the subset of children who develop PCS. Several governing bodies, including the Centers for Disease Control and Prevention, have called for more research to improve our understanding of mTBI and its sequelae, specifically addressing the small sample sizes and biases in existing research.³² Our EHR-based study directly responds to these research gaps. First, our PCS algorithm identified among the largest samples of PCS cases and controls reported to date,⁸ and is portable to other EHR-based health care systems, facilitating future large-scale studies of pediatric mTBI recovery. Second, our study included a broad mTBI population (all modes of injury recruited from across the healthcare system) for maximum generalizability. Third, we required patients to have used our health care system a minimum number of times before and after the index event to help ensure that cases and controls used VUMC for all their medical care. This requirement meant that we captured mTBI patients with favorable outcomes and included them as controls, whereas prospective studies of mTBI patients often lose these patients to follow-up.²¹ Finally, all objectively measured medical diagnoses before the index event were included as predictor

variables, avoiding our own biases in selecting candidate risk factors, as well as participants' recall biases.

Pre-injury headache is a consistent risk factor for PCS.⁸ In a prospective multi-center analysis of 2,006 pediatric concussion patients, a self-reported physician-diagnosed history of migraines nearly doubled the risk of PCS at 28 days.¹⁷ Another large retrospective cohort study of nearly 2,000 pediatric sport-related concussion patients demonstrated that self-reported pre-morbid headaches statistically increased the risk of prolonged recovery by 15%.¹⁶ In comparison, we found that other headache syndromes and migraines diagnosed by a physician and recorded in the EHR at least 180 days pre-injury increased the risk of PCS by two- to five-fold. Our effect estimates are higher than those previously reported, which could be due to differences in study design. We included a broad pediatric mTBI population as opposed to only sport-concussion patients included in previous studies, and pre-injury medical diagnoses were objectively ascertained instead of by self-report.

Numerous investigations over the past decade have demonstrated a relationship between post-mTBI sleep disturbances and prolonged recovery,³³ increased symptom reporting,³⁴ and worsened neurocognitive test scores.^{34,35} Another study recently demonstrated that 78% of adult patients diagnosed with PCS who underwent a sleep study after the index head injury were diagnosed with sleep apnea.³⁶ However, only one study³⁷ has evaluated the effect of pre-mTBI sleep difficulties in an older group of 348 adolescent and adult athletes and found that the 34 athletes with self-reported pre-injury sleeping difficulties demonstrated worse reaction times and higher symptom scores at 2, 5-7 and 10-14 days post-injury. Unfortunately, the study did not determine whether those with sleep difficulties went on to develop PCS. We systematically

assessed physician-diagnosed pre-morbid sleep disorders and show these are a risk factor for poor recovery and specifically, PCS.

The link between pre-injury sleep disorders and mTBI recovery is likely multifactorial, including decreased cognitive reserve in sleep-deprived individuals,³⁸ especially those with obstructive sleep apnea.³⁹ The link, however, may also be attributable to poor psychomotor performance, reaction time, and postural control in sleep-deprived individuals, leading to increased mTBI incidence and severity.^{40,41}

The association between chronic inflammation and PCS is presently underreported. We identified an over three-fold increased risk of PCS associated with both previously diagnosed gastritis/duodenitis and chronic pharyngitis/nasopharyngitis. This finding may reflect frequent health system utilization among cases, but cases and controls had a comparable number of visits before the index event. Alternatively, this finding could represent underlying chronic inflammatory processes conferring risk for prolonged or poor recovery after mTBI.

Gastritis, duodenitis, pharyngitis, and nasopharyngitis are diagnoses that involve activation of peripheral immune cells such as monocytes and neutrophils.⁴² These same immune cells also cross the blood brain barrier after an TBI,^{43,44} and animal models have demonstrated improved functional recovery and reduced edema, microglial activation and TBI lesion size after TBI by blocking the migration of monocytes into the central nervous system,⁴⁵ or by depleting systemic neutrophils.⁴⁶ We therefore hypothesize that a baseline state of chronic inflammation in patients with gastritis, duodenitis, pharyngitis, and nasopharyngitis, could produce a more profound inflammatory response to mTBI and consequently, a prolonged recovery. These findings warrant investigations of inflammatory markers and cytokine profiles in mTBI recovery.

A major strength of our study is the large number of pediatric PCS cases and controls from across the health care system and with dense EHR data, which allowed testing of over 200 pre-injury medical diagnoses for association with PCS. A limitation is the imperfect PPVs of our case selection strategies. Our algorithm's PPVs, however, are in line with the PPVs of other EHR-based algorithms,^{47,48} and most top phecode associations were similar in the PheWAS restricted to the cases identified by Case Definition 1, which had the higher PPV. The association with other headache syndromes was attenuated, although the number of patients with this phecode was greatly reduced (18 in the sensitivity analysis vs. 41 in the primary analysis). The attenuated effect estimate could also be because patients reporting chronic headache are eventually misdiagnosed as having PCS. A second potential study limitation is our definition of PCS based on symptom presence at 14 days post-injury. Some have argued that typical recovery takes longer than 14 days in children,³ but the definition of PCS is inconsistent across clinicians^{9,30} and diagnostic manuals,^{49,50} and a large study of 1,953 pediatric sports concussion patients reported a median time to symptom resolution of 18 days (range 1-353).¹⁶ A third potential limitation is heterogeneity in mTBI severity. In manual review, we found that nearly three-quarters of PCS cases sought care for the mTBI, whereas by design, all controls sought care for their injury. Nonetheless, injury severity characteristics (e.g., loss of consciousness and post-traumatic amnesia) are inconsistent predictors of PCS risk,⁸ and sensitivity analyses of the PCS cases with a documented mTBI in their EHR yielded results similar to those in the primary analysis. Fourth, our PheWAS relied on pre-injury billing code diagnoses, which can be imprecise. For example, other headache syndromes and migraine are captured by mutually exclusive billing codes and phecodes. Clinically, however, these diagnoses may overlap, and there could be some misattribution of other headache syndromes to migraines, and vice-versa.

Finally, the patients utilized for the study were cared for within a single health system in the United States, which may limit the generalizability of the findings.

CONCLUSION

An EHR-based PCS algorithm can be used to identify a broad group of pediatric mTBI patients with and without PCS, and to study pre-injury medical diagnoses associated with PCS risk in a PheWAS. The results confirm the association of pre-injury headache disorders with PCS and provide evidence for the association of pre-injury sleep disorders with PCS. An association of PCS with prior chronic gastritis/duodenitis and pharyngitis was seen, suggesting a role for chronic inflammation in PCS pathophysiology and risk. Evaluation of these factors should be considered during the clinical management of pediatric mTBI cases.

References

1. Dewan MC, Mummareddy N, Wellons JC 3rd, Bonfield CM. Epidemiology of Global Pediatric Traumatic Brain Injury: Qualitative Review. *World Neurosurg.* 2016;91:497–509.e1.
2. Yue JK, Winkler EA, Burke JF, et al. Pediatric sports-related traumatic brain injury in United States trauma centers. *Neurosurg Focus.* 2016;40:E3.
3. McCrory P, Meeuwisse W, Dvorak J, et al. Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in Berlin, October 2016. *Br J Sports Med.* BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine; 2017;51:838–847.
4. Haider MN, Leddy JJ, Pavlesen S, et al. A systematic review of criteria used to define recovery from sport-related concussion in youth athletes. *Br J Sports Med.* 2018;52:1179–1190.
5. Tator CH, Davis H. The postconcussion syndrome in sports and recreation: clinical features and demography in 138 athletes. *Neurosurgery.* 2014;75 Suppl 4:S106–S112.
6. Zuckerman SL, Brett BL, Jeckell AS, Yengo-Kahn AM, Solomon GS. Prognostic Factors in Pediatric Sport-Related Concussion. *Curr Neurol Neurosci Rep.* 2018;18:104.
7. Tator CH, Davis HS, Dufort PA, et al. Postconcussion syndrome: demographics and predictors in 221 patients. *J Neurosurg.* 2016;125:1206–1216.
8. Iverson GL, Gardner AJ, Terry DP, et al. Predictors of clinical recovery from concussion: a systematic review. *Br J Sports Med.* 2017;51:941–948.
9. Iverson GL. Network Analysis and Precision Rehabilitation for the Post-concussion Syndrome. *Front Neurol.* 2019;10:489.
10. Zuckerman SL, Apple RP, Odom MJ, Lee YM, Solomon GS, Sills AK. Effect of sex on symptoms and return to baseline in sport-related concussion. *J Neurosurg Pediatr.*

2014;13:72–81.

11. Lee YM, Odom MJ, Zuckerman SL, Solomon GS, Sills AK. Does age affect symptom recovery after sports-related concussion? A study of high school and college athletes. *J Neurosurg Pediatr.* 2013;12:537–544.
12. Kostyun RO, Hafeez I. Protracted recovery from a concussion: a focus on gender and treatment interventions in an adolescent population. *Sports Health.* 2015;7:52–57.
13. Morgan CD, Zuckerman SL, Lee YM, et al. Predictors of postconcussion syndrome after sports-related concussion in young athletes: a matched case-control study. *J Neurosurg Pediatr.* 2015;15:589–598.
14. Miller JH, Gill C, Kuhn EN, et al. Predictors of delayed recovery following pediatric sports-related concussion: a case-control study. *J Neurosurg Pediatr.* 2016;17:491–496.
15. Register-Mihalik J, Guskiewicz KM, Mann JD, Shields EW. The effects of headache on clinical measures of neurocognitive function. *Clin J Sport Med.* 2007;17:282–288.
16. Heyer GL, Schaffer CE, Rose SC, Young JA, McNally KA, Fischer AN. Specific Factors Influence Postconcussion Symptom Duration among Youth Referred to a Sports Concussion Clinic. *J Pediatr.* 2016;174:33–38.e2.
17. Zemek R, Barrowman N, Freedman SB, et al. Clinical Risk Score for Persistent Postconcussion Symptoms Among Children With Acute Concussion in the ED. *JAMA.* 2016;315:1014–1025.
18. Ellis MJ, Ritchie LJ, Koltek M, et al. Psychiatric outcomes after pediatric sports-related concussion. *J Neurosurg Pediatr.* 2015;16:709–718.
19. Legarreta AD, Brett BL, Solomon GS, Zuckerman SL. The role of family and personal psychiatric history in postconcussion syndrome following sport-related concussion: a story of compounding risk. *J Neurosurg Pediatr.* 2018;22:238–243.
20. Biederman J, Feinberg L, Chan J, et al. Mild Traumatic Brain Injury and Attention-Deficit

- Hyperactivity Disorder in Young Student Athletes. *J Nerv Ment Dis.* 2015;203:813–819.
21. Cnossen MC, Winkler EA, Yue JK, et al. Development of a Prediction Model for Post-Concussive Symptoms following Mild Traumatic Brain Injury: A TRACK-TBI Pilot Study. *J Neurotrauma* [online serial]. Epub 2017 Mar 27. Accessed at: <http://dx.doi.org/10.1089/neu.2016.4819>.
 22. Howell DR, Potter MN, Kirkwood MW, Wilson PE, Provance AJ, Wilson JC. Clinical predictors of symptom resolution for children and adolescents with sport-related concussion. *J Neurosurg Pediatr.* Epub 2019 Apr 16.:1–8.
 23. Babcock L, Byczkowski T, Wade SL, Ho M, Mookerjee S, Bazarian JJ. Predicting postconcussion syndrome after mild traumatic brain injury in children and adolescents who present to the emergency department. *JAMA Pediatr.* 2013;167:156–161.
 24. Dennis J, Yengo-Kahn AM, Kirby P, Solomon GS, Cox NJ, Zuckerman SL. Diagnostic Algorithms to Study Post-Concussion Syndrome using Electronic Health Records: Validating a Method to Capture an Important Patient Population. *J Neurotrauma* [online serial]. Epub 2019 Mar 28. Accessed at: <http://dx.doi.org/10.1089/neu.2018.5916>.
 25. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. *Annu Rev Genomics Hum Genet.* 2016;17:353–373.
 26. Carroll CP, Cochran JA, Guse CE, Wang MC. Are we underestimating the burden of traumatic brain injury? Surveillance of severe traumatic brain injury using centers for disease control International classification of disease, ninth revision, clinical modification, traumatic brain injury codes. *Neurosurgery.* 2012;71:1064–1070; discussion 1070.
 27. Honnibal M. Install spaCy · spaCy Usage Documentation [online]. Install spaCy. Accessed at: <https://spacy.io/usage>. Accessed July 15, 2020.
 28. Chapman B. pyConTextNLP [online]. Github; Accessed at: <https://github.com/chapmanbe/pyConTextNLP>. Accessed July 15, 2020.
 29. Davis GA, Purcell L, Schneider KJ, et al. The Child Sport Concussion Assessment Tool 5th

- Edition (Child SCAT5): Background and rationale. *Br J Sports Med.* 2017;51:859–861.
30. Rose SC, Fischer AN, Heyer GL. How long is too long? The lack of consensus regarding the post-concussion syndrome diagnosis. *Brain Inj.* 2015;29:798–803.
 31. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics.* 2014;30:2375–2376.
 32. Suskauer SJ, Yeates KO, Sarmiento K, et al. Strengthening the Evidence Base: Recommendations for Future Research Identified Through the Development of CDC’s Pediatric Mild TBI Guideline. *J Head Trauma Rehabil.* 2019;34:215–223.
 33. Hoffman NL, O’Connor PJ, Schmidt MD, Lynall RC, Schmidt JD. Relationships between Post-Concussion Sleep and Symptom Recovery: A Preliminary Study. *J Neurotrauma* [online serial]. Epub 2019 Dec 26. Accessed at: <http://dx.doi.org/10.1089/neu.2019.6761>.
 34. Kostyun RO, Milewski MD, Hafeez I. Sleep disturbance and neurocognitive function during the recovery from a sport-related concussion in adolescents. *Am J Sports Med.* 2015;43:633–640.
 35. Hoffman NL, Weber ML, Broglio SP, et al. Influence of Postconcussion Sleep Duration on Concussion Recovery in Collegiate Athletes. *Clin J Sport Med* [online serial]. 2019;Publish Ahead of Print. Accessed at: <https://journals.lww.com/cjsportsmed/Pages/articleviewer.aspx?year=9000&issue=00000&article=99221&type=Abstract>. Accessed July 29, 2019.
 36. Santos A, Walsh H, Anssari N, Ferreira I, Tartaglia MC. Post-Concussion Syndrome and Sleep Apnea: A Retrospective Study. *J Clin Med Res* [online serial]. 2020;9. Accessed at: <http://dx.doi.org/10.3390/jcm9030691>.
 37. Sufrinko A, Pearce K, Elbin RJ, et al. The effect of preinjury sleep difficulties on neurocognitive impairment and symptoms after sport-related concussion. *Am J Sports Med.* 2015;43:830–838.

38. Renn RP, Cote KA. Performance monitoring following total sleep deprivation: effects of task type and error rate. *Int J Psychophysiol*. Elsevier; 2013;88:64–73.
39. Rosenzweig I, Glasser M, Polsek D, Leschziner GD, Williams SCR, Morrell MJ. Sleep apnoea and the brain: a complex relationship. *Lancet Respir Med*. 2015;3:404–414.
40. Ma J, Yao Y-J, Ma R-M, et al. Effects of sleep deprivation on human postural control, subjective fatigue assessment and psychomotor performance. *J Int Med Res*. journals.sagepub.com; 2009;37:1311–1320.
41. Raikes AC, Athey A, Alfonso-Miller P, Killgore WDS, Grandner MA. Insomnia and daytime sleepiness: risk factors for sports-related concussion. *Sleep Med*. 2019;58:66–74.
42. Pahwa R, Singh A, Jialal I. *Chronic Inflammation*. StatPearls. Treasure Island (FL): StatPearls Publishing; 2019.
43. Szymdynger-Chodobska J, Strazielle N, Gandy JR, et al. Posttraumatic Invasion of Monocytes across the Blood—Cerebrospinal Fluid Barrier. *J Cereb Blood Flow Metab*. SAGE Publications Ltd STM; 2012;32:93–104.
44. Soares HD, Hicks RR, Smith D, McIntosh TK. Inflammatory leukocytic recruitment and diffuse neuronal degeneration are separate pathological processes resulting from traumatic brain injury. *J Neurosci. Soc Neuroscience*; 1995;15:8223–8233.
45. Corps KN, Roth TL, McGavern DB. Inflammation and neuroprotection in traumatic brain injury. *JAMA Neurol*. 2015;72:355–362.
46. Kenne E, Erlandsson A, Lindbom L, Hillered L, Clausen F. Neutrophil depletion reduces edema formation and tissue loss following traumatic brain injury in mice. *J Neuroinflammation*. Springer; 2012;9:17.
47. Smoller JW. The use of electronic health records for psychiatric phenotyping and genomics. *Am J Med Genet B Neuropsychiatr Genet*. 2018;177:601–612.
48. Lingren T, Chen P, Bochenek J, et al. Electronic Health Record Based Algorithm to Identify

Patients with Autism Spectrum Disorder. PLoS One. 2016;11:e0159621.

49. Wortzel HS, Arciniegas DB. The DSM-5 approach to the evaluation of traumatic brain injury and its neuropsychiatric sequelae. *NeuroRehabilitation*. 2014;34:613–623.
50. Boake C, McCauley SR, Levin HS, et al. Limited agreement between criteria-based diagnoses of postconcussional syndrome. *J Neuropsychiatry Clin Neurosci*. 2004;16:493–499.

Table 1: Characteristics of the true positives identified by manual review of 100 records meeting case definitions 1 and 2.

	True Positives (N=81)	Pediatric True Positives (N=13)
Care sought for index TBI		
Yes – VUMC	39 (48.1)	8 (61.5)
Yes – Not VUMC	21 (25.9)	3 (23.1)
No	1 (1.2)	0 (0)
Unknown	20 (24.7)	2 (15.4)
Hospital length of stay for index TBI		
None – ED visit	41 (50.6)	9 (69.2)
None – outpatient visit	3 (3.7)	1 (7.7)
Hospitalization <3 days	2 (2.5)	0 (0)
Hospitalization >= 3 days	4 (4.9)	1 (7.7)
Unknown	31 (38.3)	2 (15.4)
Imaging for the index TBI		
Positive imaging result	8 (9.9)	1 (7.7)
Cause of index TBI		
Motor vehicle collision	35 (43.2)	4 (30.8)
Fall – not sports-related	29 (35.8)	2 (15.4)
Assault	5 (6.2)	1 (7.7)

Sports-related	6 (7.4)	5 (38.5)
Other / unknown	6 (7.4)	1 (7.7)

Table 2: Descriptive characteristics of the matched sample used in the PheWAS.

Participant characteristics ^a	Controls (N=1096)	Cases (N=274)	P value ^b
Age at index event in years	14.4 (3.1)	14.4 (3.1)	NA
Female, N (%)	624 (56.9)	156 (56.9)	NA
Number of days in EHR before index event	2929 (1671.2)	3302 (1637.3)	8.6e-4
Number of ICD codes at least 180 days before index event	34.0 (68.6)	44.5 (101.6)	0.058
Number of visits at least 180 days before index event	19.2 (33.7)	20.6 (31.3)	0.550
Number of ICD codes at least 30 days before index event	37.0 (73.8)	50.2 (114.1)	0.038
Number of visits at least 30 days before index event	20.8 (35.0)	22.5 (33.8)	0.480
Documented TBI in EHR, N (%)	1096 (100)	208 (75.9)	NA ^c
Motor vehicle collision cause code ^d before index event, N (%)	107 (9.8)	6 (2.2)	1.2e-4

^aData are presented as mean (sd) unless noted otherwise, ^bP values were calculated from a conditional logistic regression model, ^cModel failed to converge due to a complete separation of the data, ^dICD9 and ICD10 external cause of injury codes in patients' EHRs were mapped to the motor vehicle cause of injury category as previously published.²⁶

Table 3: Phecodes associated with pediatric PCS at the phenome-wide threshold for statistical significance

Phecode	Overall				Females				Males				Females vs. Males
	N True	N False	OR (95% CI)	P value	N True	N False	OR (95% CI)	P value	N True	N False	OR (95% CI)	P value	P value
Other headache syndromes	41	1329	5.3 (2.8-10.1)	3.85e-7	32	748	4.3 (2.1-8.6)	7.53e-5	9	581	12.4 (2.5-60.6)	0.002	0.84
Sleep disorders	62	1308	3.1 (1.8-5.2)	2.60e-5	32	748	4.1 (1.9-8.7)	2.37e-4	30	560	2.3 (1.1-5.0)	0.028	0.26
Gastritis and duodenitis	44	1326	3.6 (1.8-7.0)	2.08e-4	26	754	3.8 (1.6-9.0)	0.003	18	572	3.4 (1.2-10.0)	0.024	0.46
Chronic pharyngitis and nasopharyngitis	44	1326	3.3 (1.8-6.3)	2.21e-4	36	744	4.3 (2.1-8.8)	9.54e-5	8	582	1.3 (0.3-6.4)	0.756	0.14
Anxiety disorders ^b	71	1299	2.4 (1.4-4.2)	0.001	51	729	1.7 (0.9 - 3.4)	0.131	20	570	4.7 (1.9 - 11.8)	0.001	0.87

^aThe phecode “Other headache syndromes” includes all headaches other than tension and migraine headaches.

^bThe association between “Anxiety disorders” and PCS fell just below the threshold for phenome-wide significance. “Anxiety disorders”, however, were the fifth-ranked phecode in the PheWAS and results are included in this table for comparison with previous literature.

Table 4: Sensitivity analysis of top phecodes associated with pediatric PCS

Phecode	Including only cases meeting case definition 1				Including only cases with a traumatic brain injury code			
	N True	N False	P value	OR (95% CI)	N True	N False	P value	OR (95% CI)
Other headache syndromes	18	962	0.221	1.9 (0.7 - 5.2)	26	864	3.86E-04	4.3 (1.9 - 9.5)
Sleep disorders	45	935	3.26E-04	3.1 (1.7 - 5.9)	38	852	2.73E-02	2.2 (1.1 - 4.3)
Gastritis and duodenitis	27	953	0.004	3.4 (1.5 - 7.7)	26	864	5.42E-04	4.7 (2.0 - 11.5)
Chronic pharyngitis and nasopharyngitis	30	950	0.046	2.2 (1.0 - 4.9)	23	867	0.010	3.4 (1.3 - 8.5)
Anxiety disorder	47	933	0.011	2.4 (1.2 - 4.5)	52	838	0.002	2.8 (1.5 - 5.3)

Appendix 1: Authors

Name	Location	Contribution
Aaron Yengo-Kahn, MD	Vanderbilt University Medical Center, Nashville, TN	Study conception and design, data acquisition, interpretation of results and manuscript drafting and revision for intellectual content
Natalie Hibshman, BA	Vanderbilt University School of Medicine, Nashville, TN	Major role in data acquisition, interpretation of results, manuscript drafting
Christopher Bonfield, MD	Vanderbilt University Medical Center, Nashville, TN	Study concept and design, interpretation of data and critically revised the manuscript for important intellectual content.
Eric Torstenson, BS	Vanderbilt University Medical Center, Nashville, TN	Analysis and interpretation of data and critically revised the manuscript for important intellectual content
Katherine Gifford, PsyD	Vanderbilt University Medical Center, Nashville, TN	Substantial role in interpretation of data and revised the manuscript for important intellectual content
Daniil Belikau	British Columbia Children's Hospital Research Institute, Vancouver, Canada	Analysis of data and revised the manuscript for important intellectual content
Lea Davis, PhD	Vanderbilt University Medical Center, Nashville, TN	Study design, interpretation of data and revised the manuscript for important intellectual content
Scott Zuckerman, MD, MPH	Vanderbilt University Medical Center, Nashville, TN	Study conceptualization and design, interpretation of results and revised the manuscript for intellectual content
Jessica Dennis, MSc, PhD	University of British Columbia, Vancouver, Canada	Study conception and design, data acquisition, analysis, and interpretation as well as drafting and revision of the manuscript for intellectual content

Figure Legends

Figure 1 - Evaluation of strategies to ascertain post-concussion syndrome (PCS) cases from electronic health records. TBI denotes traumatic brain injury and PCS, post-concussion syndrome.

Figure 2 – Patient Selection Flow Diagram. In controls, the index event was the first mTBI code. In cases, the index event was the first mTBI code, or the first PCS code if there was no mTBI code assigned or if the first mTBI code occurred after the first PCS code. EHR denotes Electronic health records; PCS, post-concussion syndrome, PPV, positive predictive value; (m)TBI, (mild) traumatic brain injury; and VUMC, Vanderbilt University Medical Center.

Figure 3 – PheWAS of 201 Phecodes Assigned ≥ 180 Days Pre-Index Event. To ensure sufficient cell counts for interpretable odds ratios, only phecodes with at least 25 affected patients were included. The red line denotes the Bonferroni threshold for phenome-wide significance ($0.05/201=2.48 \times 10^{-4}$). Phecodes plotted above the x-axis are associated with an increased odds of PCS while those plotted below the x-axis are associated with a reduced odds of PCS. Interactive plots and full results are available at [<https://dennislabs.ca/resources-and-software/>].

Figure 1 - Evaluation of strategies to ascertain post-concussion syndrome (PCS) cases from electronic health records. TBI denotes traumatic brain injury and PCS, post-concussion syndrome.

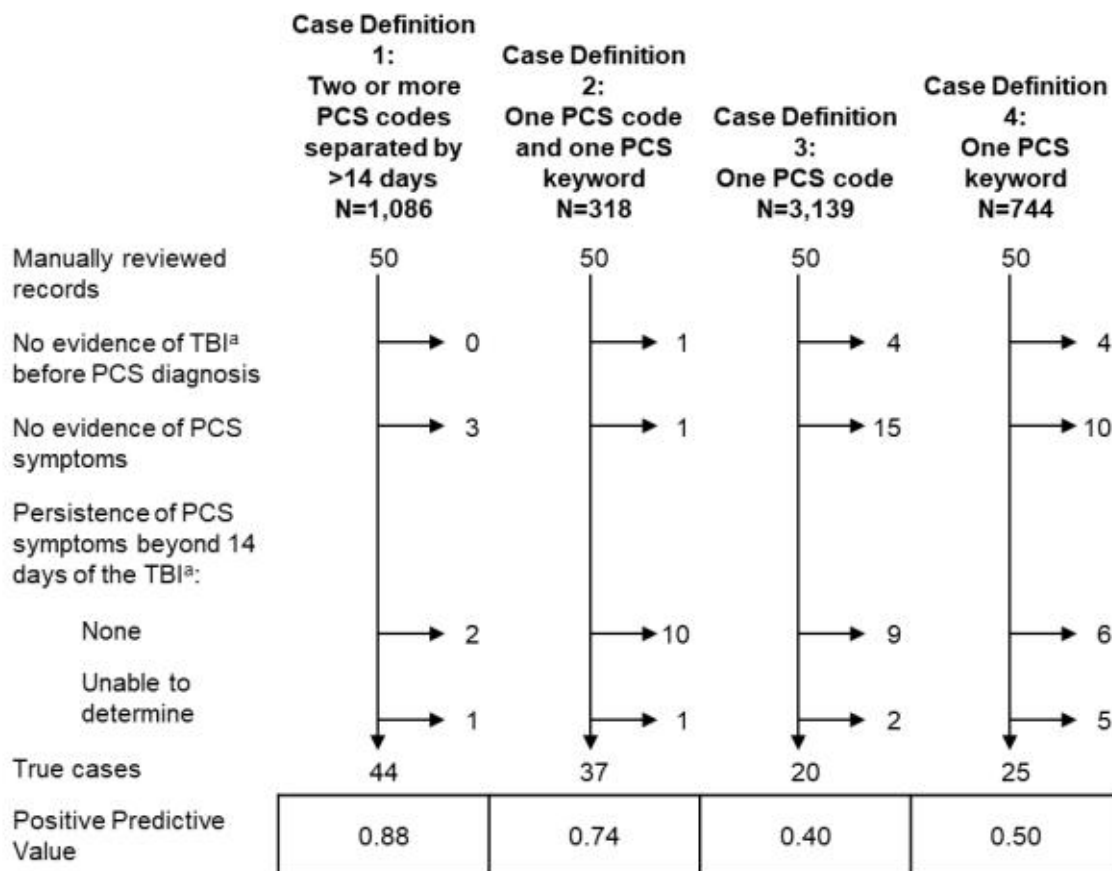


Figure 2 – Patient Selection Flow Diagram. In controls, the index event was the first mTBI code. In cases, the index event was the first mTBI code, or the first PCS code if there was no mTBI code assigned or if the first mTBI code occurred after the first PCS code. EHR denotes Electronic health records; PCS, post-concussion syndrome, PPV, positive predictive value; (m)TBI, (mild) traumatic brain injury; and VUMC, Vanderbilt University Medical Center.

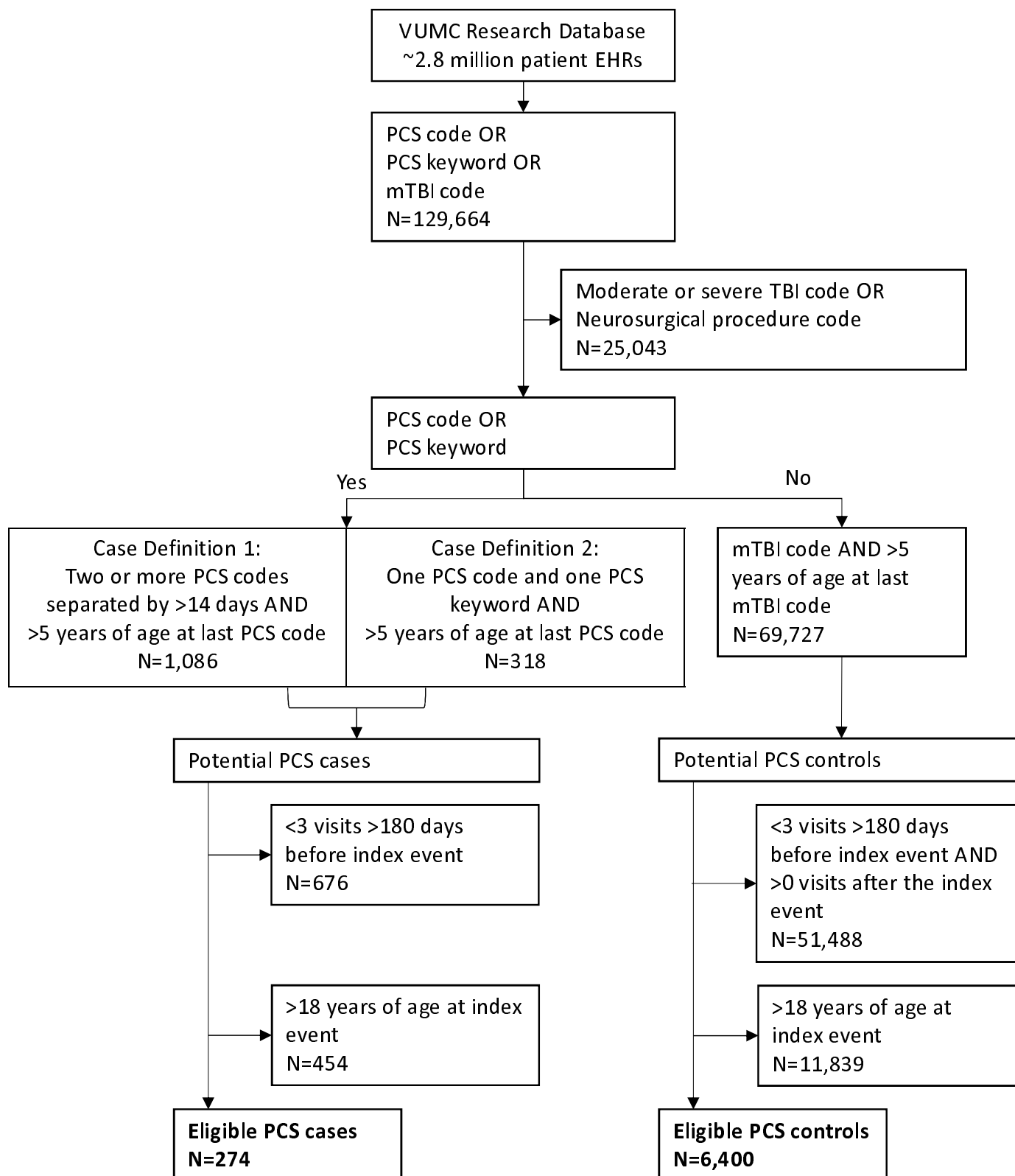


Figure 3 – PheWAS of 201 Phecodes Assigned ≥ 180 Days Pre-Index Event. To ensure sufficient cell counts for interpretable odds ratios, only phecodes with at least 25 affected patients were included. The red line denotes the Bonferroni threshold for phenome-wide significance ($0.05/201=2.48 \times 10^{-4}$). Phecodes plotted above the x-axis are associated with an increased odds of PCS while those plotted below the x-axis are associated with a reduced odds of PCS. Interactive plots and full results are available at [<https://dennislabs.ca/resources-and-software/>].

