

Sex Differences in Health Conditions Associated with Sexual Assault in a Large Hospital Population

Allison M. Lake^{a, b} Slavina B. Goleva^c Lauren R. Samuels^d Laura M. Carpenter^e
Lea K. Davis^{a, b, f, g}

^aVanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; ^bDivision of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; ^cNational Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; ^dDepartment of Biostatistics, Vanderbilt University, Nashville, TN, USA; ^eDepartment of Sociology, Vanderbilt University Medical Center, Nashville, TN, USA; ^fDepartment of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA; ^gDepartment of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

Keywords

Clinical informatics · Women's health · Sex differences ·
Electronic health records

Abstract

Introduction: Sexual assault is an urgent public health concern with both immediate and long-lasting health consequences, affecting 44% of women and 25% of men during their lifetimes. Large studies are needed to understand the unique healthcare needs of this patient population. **Methods:** We mined clinical notes to identify patients with a history of sexual assault in the electronic health record (EHR) at Vanderbilt University Medical Center (VUMC), a large university hospital in the Southeastern USA, from 1989 to 2021 ($N = 3,376,424$). Using a phenome-wide case-control study, we identified diagnoses co-occurring with disclosures of sexual assault. We performed interaction tests to examine whether sex modified any of these associations. Association analyses were restricted to a subset of patients receiving regular care at VUMC ($N = 833,185$). **Results:** The phenotyping approach identified 14,496 individuals (0.43%) across the VUMC-EHR with documentation of sexual assault and achieved a posi-

tive predictive value of 93.0% (95% confidence interval = 85.6–97.0%), determined by manual patient chart review. Out of 1,703 clinical diagnoses tested across all subgroup analyses, 465 were associated with sexual assault. Sex-by-trauma interaction analysis revealed 55 sex-differential associations and demonstrated increased odds of psychiatric diagnoses in male survivors. **Discussion:** This case-control study identified associations between disclosures of sexual assault and hundreds of health conditions, many of which demonstrated sex-differential effects. The findings of this study suggest that patients who have experienced sexual assault are at risk for developing wide-ranging medical and psychiatric comorbidities and that male survivors may be particularly vulnerable to developing mental illness.

© 2022 The Author(s).
Published by S. Karger AG, Basel

Introduction

Sexual assault, which includes rape, forced penetration of another person, sexual coercion, and unwanted sexual contact, is a major public health and human rights concern, estimated to affect 44% of women and 25% of men

in the USA during their lifetimes [1]. Rape alone has an astonishing lifetime prevalence of 20% in women [1]. Sexual assault has immediate and wide-ranging health consequences, including direct physical injuries and adverse effects on reproductive and mental health [2]. Sexual assault is also associated with debilitating long-term health outcomes. Survivors of sexual assault exhibit increased rates of psychiatric conditions including post-traumatic stress disorder (PTSD), anxiety, depression, sleep disorders, eating disorders, and suicide attempts [3]. Furthermore, sexual assault is associated with lifetime diagnoses of multiple functional and somatic disorders, including gastrointestinal symptoms, chronic pain, and functional seizures [4].

The ability to identify patients with a history of sexual assault is crucial for understanding the health consequences related to sexual assault. Large studies are needed to understand the unique healthcare needs of this population. However, the challenges inherent in sexual assault research make this a daunting task. The highly sensitive and often stigmatized nature of sexual assault can leave many patients and research participants unwilling to disclose these experiences. Stigmatization and other cultural factors may also lead survivors to reject the classification of their experiences as sexual assault [5] (cited in [6]). Furthermore, research on male survivors is relatively limited, likely in part due to differential reporting of sexual assault between men and women [7]. Given the sex differences widely observed for common mental health conditions, it is plausible that the effects of sexual assault on mental health differ between men and women. Despite the importance of this line of research, sex differences in sexual assault-related health effects are understudied, and few robust sex differences have been identified [3, 4, 7].

Research utilizing electronic health records (EHRs) has the potential to overcome some limitations of traditional epidemiological approaches. However, few EHR-based studies of sexual assault and the associated health sequelae exist [8, 9]. In the present study, we utilized both billing codes and key phrases to identify patients with a history of sexual assault in a large hospital setting. Our novel key phrase approach, which mines clinical note text for matches to certain relevant phrases, identified over 14,000 patients who have experienced sexual assault out of over 3 million patients, a twofold increase over the 6,998 patients identified using billing codes alone. We then characterized the medical phenome of a subset of this patient population and identified hundreds of associated clinical phenotypes, including many related to psy-

chiatric conditions. Finally, we demonstrated the sex-differentiated nature of dozens of clinical associations, addressing a key gap in sexual assault research.

Materials and Methods

Sample and Data Description

This study included 3,376,424 deidentified patients who received care at Vanderbilt University Medical Center (VUMC), an academic medical center in Nashville, TN, between the years 1989 and 2021. VUMC is a large tertiary care center encompassing a wide variety of specialties and a large catchment area with patients living in Tennessee, Kentucky, and Alabama. The medical records include inpatient and outpatient encounters and span both specialty and primary care. Demographic information, ICD-9 and ICD-10 codes, and clinical notes were obtained from the medical record for use in subsequent analyses. No age restrictions were applied. For all association analyses, we retained only individuals meeting “medical home” criteria as described [10] to limit the study population to patients receiving regular care at VUMC and reduce nonrandom missingness between cases and controls. This was accomplished by restricting our study population to individuals with at least 5 ICD-9 or ICD-10 codes of any type, on different days over a period of at least 3 years. We limited the study population to individuals with EHR-reported male or female sex, resulting in the removal of 11 individuals with “undetermined” sex. We further required at least one recorded body mass index (BMI) for inclusion in association analyses, as BMI is shown to be correlated with approximately two-thirds of the medical phenome and is measured in the majority (84.3%) of patients in the medical home sample. This study was reviewed and approved by the VUMC institutional review board (IRB #212285) and exempted from informed consent requirements as nonhuman subject research using deidentified medical data.

Sexual Assault Case/Control Algorithms

Although ICD codes do exist for rape, sexual assault, and sexual abuse, these are typically reserved for care immediately after an assault and are infrequently used to indicate a history of sexual abuse or assault. Across the entire cohort, only 6,998 patient charts included ICD codes for sexual assault (0.21%), likely an extreme underestimate given sexual assault prevalence estimates [1]. To improve detection of patients with a history of sexual assault, we developed an algorithm to search the unstructured free text from clinical notes associated with patient charts for matches to specific phrases. These clinical notes consisted of admission notes, ancillary reports, discharge summaries, emergency department notes, inpatient notes, notes not otherwise classified, nursing reports, outpatient notes, pathology reports, and problem lists.

In developing the algorithm, we first identified charts containing relevant ICD codes and/or matches to an initial set of keywords with obvious relevance (e.g., “rape” and “sexual assault”). We then performed an exploratory chart review and identified an expanded set of relevant keywords and exclusion phrases. This subsequent “phase 1” algorithm identified individuals meeting criteria for disclosure of sexual assault using the chart-review-derived set of key phrases as well as relevant ICD codes (online suppl. Table S1; see www.karger.com/doi/10.1159/000527363 for all online suppl. ma-

terial). We then performed a comprehensive manual review of 25 charts meeting these criteria to further refine our keyword search terms accordingly. This included identifying commonly used phrasing to indicate the presence of a history of sexual assault (i.e., inclusion) or the absence of a history of sexual assault (i.e., exclusion).

Our final “phase 2” keyword-based algorithm consisted of 18 inclusion phrases (e.g., “history of sexual assault,” “hx of rape”) and 12 exclusion phrases (i.e., “denies history of sexual abuse,” “no hx of sexual assault”) to identify patients who disclosed sexual assault (online suppl. Table S1). To allow comparisons between keyword-based and ICD-based approaches, ICD codes were not used in this final keyword-based algorithm. A patient was considered a case if their chart contained a match to at least one inclusion phrase and contained no matches to any exclusion phrases. Otherwise, the patient was considered a control.

To benchmark our keyword-based algorithm, we performed additional analyses using only ICD codes to identify patients with positive and negative histories of sexual assault. In the ICD-code-based approach, charts were defined as cases if they contained at least 1 of 25 ICD-9 and ICD-10 codes pertaining to sexual assault (online suppl. Table S2), and as controls otherwise. We did not exclude individuals meeting both ICD-based and keyword-based criteria.

Chart Review

To evaluate our final “phase 2” keyword-based algorithm, AML conducted a manual review of 100 randomly selected charts – a standard number in the field for validating phenotyping algorithms [11] – identified by the sexual assault algorithm and calculated the positive predictive value (PPV) of the algorithm. The chart reviews were conducted within a subset of patients who are also enrolled in the VUMC biobank, BioVU. These patients tend to have longer health records (median 9.99 years among biobank participants compared to a median of 1.15 years for the entire VUMC-EHR) and thus present an optimal setting for validating a phenotyping algorithm that relies on the presence of key phrases in clinical notes. Charts were identified as true positives if they clearly contained a description of the patient’s reported history of sexual assault.

Statistical Analyses

The presence or absence of a history of sexual assault was used as the primary independent variable in a PheWAS. The medical phenome was categorized by phecodes – single codes representing clinical phenotypes – that were generated by aggregating related billing codes into hierarchical code families [12]. Using the PheWAS package [13] (version 0.99) in R [14], we mapped ICD-9 and ICD-10 codes to 1,817 phecodes using Phecode Map version 1.2 (pewascatalog.org) [15, 16], requiring at least two occurrences of at least one relevant ICD code to define a case. We conducted a sex-combined PheWAS and sex-stratified (i.e., separately in males and females) PheWAS. For each phenotype tested in the sex-combined analysis, we required a minimum of 200 records, with at least 100 records for both males and females. Accordingly, for the sex-stratified analyses, we required a minimum of 100 records per phecode. For each phecode analyzed ($n = 1,482, 1,658, \text{ and } 1,527$ for sex-combined, female-only, and male-only, respectively), we fitted a multivariable logistic regression model with sexual assault history (binary: positive or negative) as the primary inde-

pendent variable and EHR-reported race, ethnicity, median age of all visits in an individual’s medical record, median BMI of all visits in an individual’s medical record, and the log-transformed mean number of ICD codes per day in the record as covariates. Sex-combined analyses additionally included sex as a covariate.

For the 354 phenotypes that were significantly associated with sexual assault in at least one sex in the sex-stratified analyses and that had at least 100 records for each sex, we conducted interaction analyses to formally examine the moderating effect of sex on phenotypic associations with sexual assault. These analyses used models identical to those used in the sex-combined analyses, with the addition of a sex-by-sexual-assault interaction term.

For each association analysis, we applied a Bonferroni correction ($p = 0.05/\text{total number of phecodes tested}$) to correct for the number of phecodes tested, using a threshold of corrected $p < 0.05$ for an association to be considered phenome-wide-significant. Although Bonferroni correction is likely to be overly conservative given the correlation inherent in related diagnostic codes, the large sample size counteracts this loss of power.

Results

Sexual Assault Prevalence in the EHR

Across our entire cohort (3,376,424 patients), the keyword-based approach identified 14,496 individuals (0.43%) with documentation of sexual assault, whereas only 6,998 cases (0.21%) (with an overlap of 2,928 individuals) received ICD codes for sexual assault. After applying medical-home criteria to our cohort (see Methods), we removed 11 individuals with unknown sex and 154,804 individuals with missing BMI, resulting in a total sample of 833,185 individuals. Among these individuals, 9,333 were designated as cases by the keyword-based algorithm (0.94%, Table 1) and 4,422 by the ICD-code-based algorithm (0.45%, online suppl. Table S3), with an overlap of 2,014 individuals. Patients with keyword- or ICD-based documentation of sexual assault were predominantly female and younger on average than the remainder of the medical-home population (Table 1, online suppl. Table S3).

Sexual Assault Case-Control Algorithm Performance

Of the 100 patient charts identified as cases by the sexual assault case-control algorithm and randomly selected to assess the PPV of the algorithm, 93 were found to be true positives (PPV = 93.0%, 95% confidence interval [CI]: 85.6–97.0%). In the 7/100 charts identified as false positives, inclusion phrases were present either as negation of sexual assault history (e.g., “denies a h/o sexual abuse”) or in reference to someone other than the patient (e.g., “[sibling] was sexually abused”).

Table 1. Demographics of patients classified as sexual assault cases or controls by our keyword-based phenotyping algorithm^{a,b}

	Case (N = 9,333)	Control (N = 823,852)	Test statistic	p value
Sex, n (%)				
Female	7,610 (81.5)	467,617 (56.8)	$\chi^2 = 2311.26$	<0.001
Male	1,723 (18.5)	356,235 (43.2)		
Record-median age				
Median [Min, Max] – female	29.0 [0, 86.0]	38.0 [0, 89.0]	$t = -39.47$	<0.001
Median [Min, Max] – male	23.0 [0, 86.0]	37.0 [0, 89.0]		
Record-median BMI				
Median [Min, Max] – female	27.0 [11.3, 59.0]	25.7 [10.0, 60.0]	$t = 18.04$	<0.001
Median [Min, Max] – male	24.4 [11.3, 59.2]	25.8 [10.0, 60.0]		
White and non-Hispanic, n (%)				
No	2436 (26.1)	207,663 (25.2)	$\chi^2 = 3.87$	0.0492
Yes	6,897 (73.9)	616,189 (74.8)		
Mean records per day				
Median [Min, Max]	0.0319 [0.000775, 2.01]	0.0168 [0.000462, 3.92]	$t = 53.07$	<0.001

^aTo assess case-control differences in demographic variables, two-sided *t* tests were performed for continuous variables (record-median age, record-median BMI, log-transformed mean records per day), and χ^2 tests were performed for binary variables (sex, white/non-Hispanic). ^bMean records per day is a measure of the density of diagnostic codes across an individual's health record. Values were log-transformed prior to the two-sided *t* test.

History of Sexual Assault Is Associated with Hundreds of Clinical Phenotypes

Our sex-combined PheWAS identified 386 out of 1,482 medical and psychiatric conditions as significantly overrepresented (Bonferroni-adjusted $\alpha = 3.34e-05$) among patients disclosing sexual assault. Of the 50 associations with the lowest *p* values, 40 were psychiatric disorders, including schizophrenia (odds ratio (OR) = 86.36, 95% CI = 76.31, 97.75), suicide or self-inflicted injury (OR = 86.33, 95% CI = 74.82, 99.6), depression (OR = 27.5, 95% CI = 25.57, 29.57), and PTSD (OR = 122.81, 95% CI = 113.47, 132.91, Fig. 1, online suppl. extended data Table 1 “sex_combined” sheet). These top associations also included phecodes relating to personality disorders: “personality disorders” (OR = 107.35, 95% CI = 94.51, 121.92) and “antisocial/borderline personality disorder” (OR = 129.17, 95% CI = 111.73, 149.32). In total, 64 of the 71 tested phecodes pertaining to psychiatric disorders were significantly associated with sexual assault.

The 386 statistically significant associations also included many nonpsychiatric health conditions. Some of these associations plausibly represent immediate physical consequences of sexual assault, including urinary tract infection (OR = 1.79, 95% CI = 1.68, 1.91), sexually transmitted infection (OR = 4.6, 95% CI = 3.85, 5.5), and concussion (OR = 2.19, 95% CI = 1.9, 2.53), Figure 2, online suppl. extended data Table 1 “sex_combined” sheet).

Many of the nonpsychiatric clinical associations with the largest effect sizes pertained to poisonings, including poisoning by psychotropic agents (OR = 18.13, 95% CI = 15.55, 21.13) and poisoning by analgesics, antipyretics, or antirheumatics (OR = 5.35, 95% CI = 4.69, 6.09).

Survivors of sexual assault frequently suffer from disorders involving functional and somatic symptoms including chronic pelvic pain and chronic gastrointestinal symptoms [4]. In our analysis, the broad “somatoform disorder” phecode is strongly associated with sexual assault (OR = 53.84, 95% CI = 46.35, 62.55). We additionally note multiple clinical associations related to functional and somatic symptom disorders, including abdominal pain (OR = 2.18, 95% CI = 2.07, 2.29), chronic (nonspecific) pain (OR = 2.24, 95% CI = 2.09, 2.41), unspecified myalgia and myositis (OR = 2.31, 95% CI = 2.13, 2.5), and cervicalgia (OR = 1.75, 95% CI = 1.62, 1.88). We also identify a range of urinary symptoms associated with sexual assault, consistent with prior studies linking sexual assault and genitourinary symptoms [17, 18], including dysuria (OR = 2.11, 95% CI = 1.95, 2.29), urinary frequency (OR = 2.18, 95% CI = 1.95, 2.43), urinary retention (OR = 2.67, 95% CI = 2.33, 3.06), urinary incontinence (OR = 1.97, 95% CI = 1.78, 2.19), and cystitis (OR = 2.35, 95% CI = 2.06, 2.68).

Several phenome-wide-significant phecodes identified in our analysis pertained to seizures and epilepsy, in-

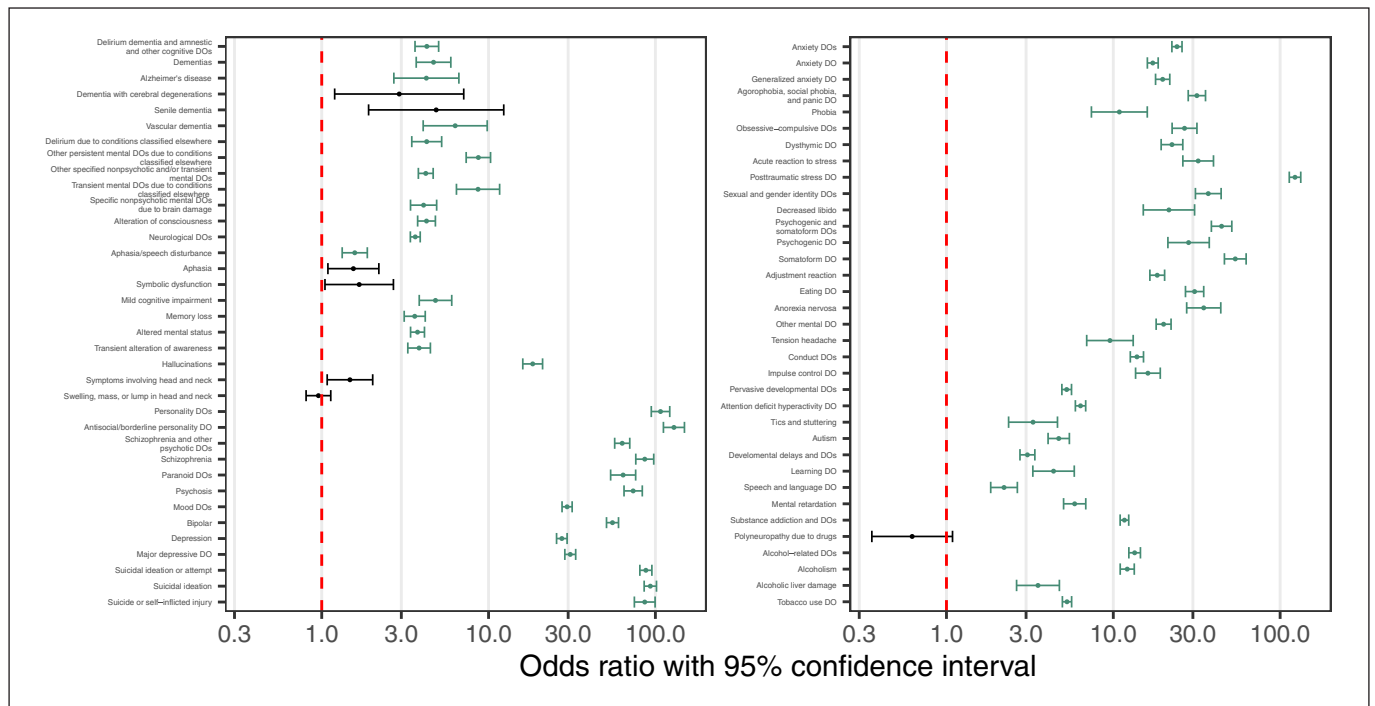


Fig. 1. Psychiatric phenotypic associations with sexual assault. Log₁₀-scale forest plots displaying ORs with 95% CI for adjusted associations between sexual assault and all tested psychiatric phecodes in sex-combined analyses. Associations achieving statistical significance are colored green. Related phecodes are grouped together (for example, the “Antisocial/borderline personality DO” phecode [301.2] is related to its parent phecode “Personality DOs” [301]). Red dashed line indicates an OR of 1. DO, disorder.

cluding the broad “epilepsy, recurrent seizures, convulsions” phecode and all phecodes classified under it. Given the previously identified relationship between sexual assault and the development of functional seizures, we investigated whether these associations were partially explained by patients experiencing functional seizures as classified by our previously published phenotyping algorithm [10]. Indeed, we observed a high level of overlap between seizure-related phecodes and functional seizure status (online suppl. Table S4). Furthermore, when we included functional seizure status as a covariate in the multivariable logistic regression models, all associations between sexual assault and seizure-related phecodes were at least partially attenuated (online suppl. Table S5).

Sex-Differential Phenotypic Association Analyses

Sex differences in prevalence, age of onset, and presentation of many neuropsychiatric disorders are well characterized in the literature [19, 20]. We analyzed our cohort for sex-differential phenotypic associations to determine whether a history of sexual assault is modified by

sex. First, we first performed sex-stratified PheWAS of sexual assault disclosure across all phecodes with sufficient case counts (online suppl. extended data Table 1 “male_only” and “female_only” sheets). We identified 410 phecodes significantly associated with sexual assault in at least one of the analyses, 332 of which were also significant in the sex-combined analysis.

The majority of significant sex-stratified associations, including many psychiatric conditions, were observed in both males and females with concordant directions of effect. The majority of these were also assessed in a sex-by-exposure interaction term. However, some associations did not have sufficient case counts to be included in the interaction analysis. These phenotypes fell into two categories: (a) sex-specific and (b) infrequent (<100 cases) in one sex.

We identified two male-specific phecodes associated with sexual assault: testicular hypofunction (OR = 2.25, 95% CI = 1.73, 2.93) and testicular dysfunction (OR = 2.24, 95% CI = 1.72, 2.91). The female-specific associations were diverse and included genitourinary conditions

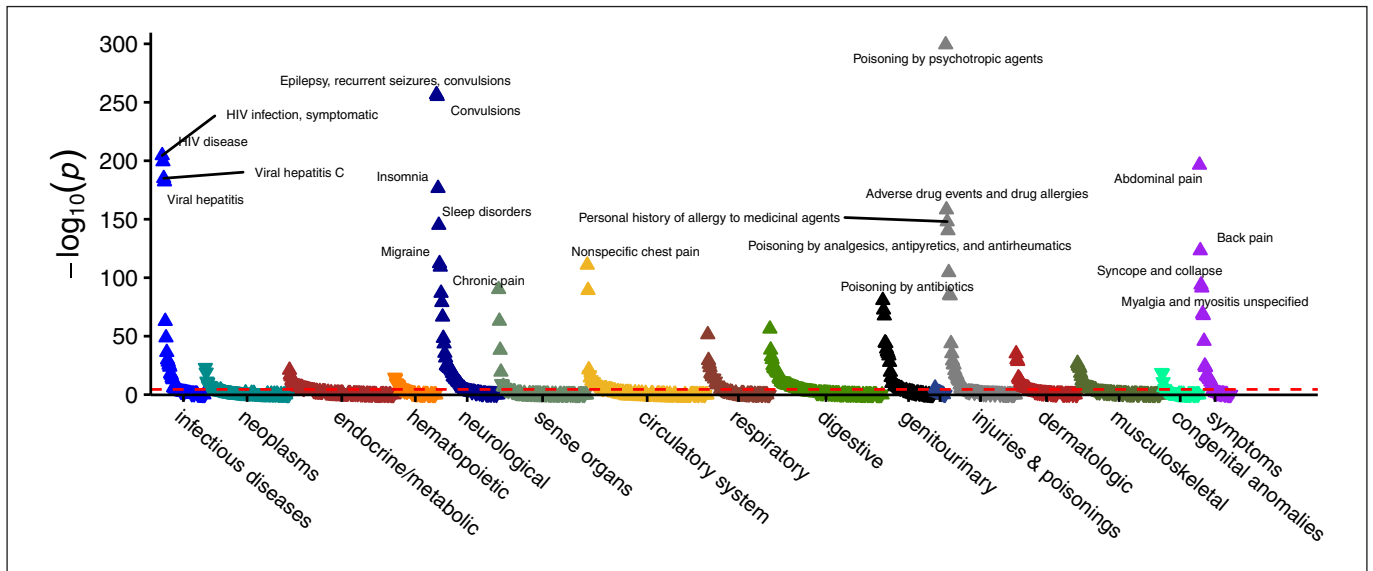


Fig. 2. Nonpsychiatric associations with sexual assault. PhewAS plot from sex-combined analysis displaying $-\log_{10}(p)$ for all nonpsychiatric phecodes analyzed. Dashed red line indicates significance threshold (Bonferroni-corrected $p < 0.05$).

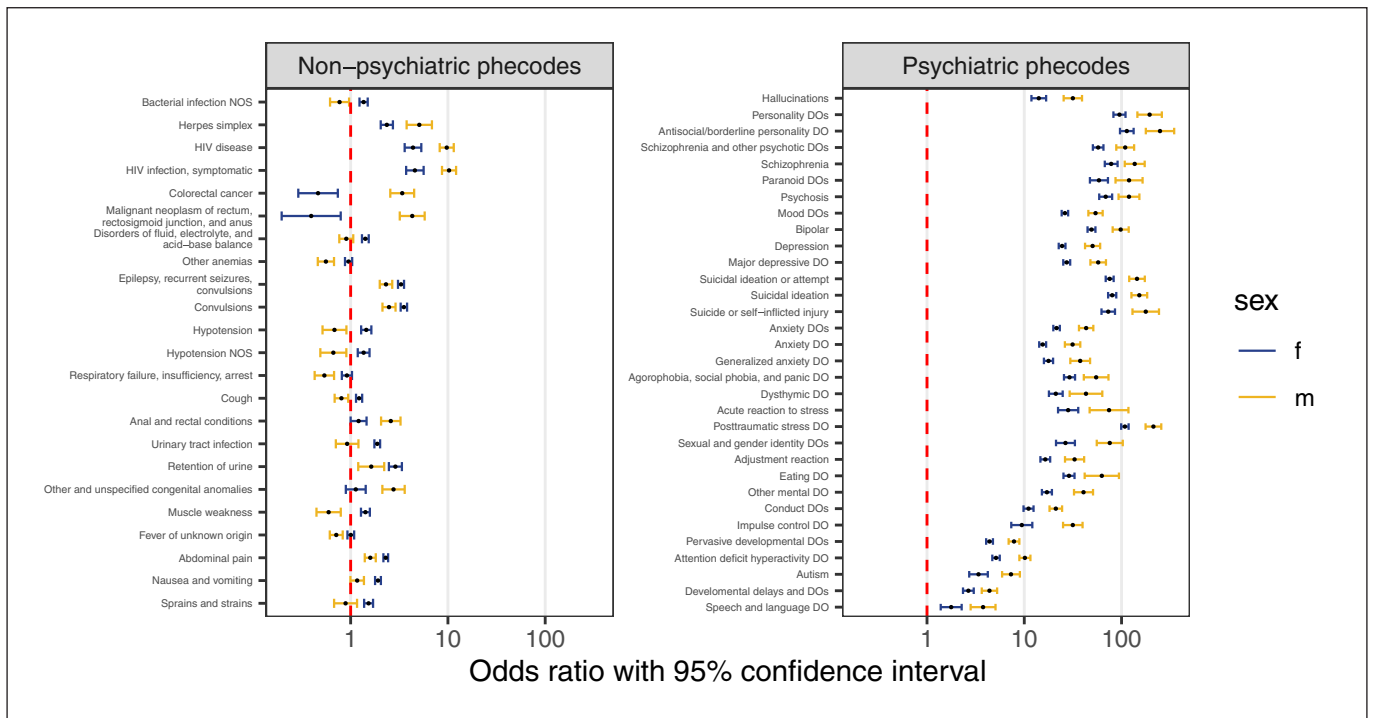


Fig. 3. History of sexual assault increases odds of clinical phenotypes in a sex-differential manner. For each of the 55 phenotypes with significant sexual-assault-by-sex interaction effects, \log_{10} -scale odds ratio (OR) and 95% CI is plotted separately by sex for nonpsychiatric phecodes (left) and psychiatric phecodes (right). The ORs correspond to the sex-stratified associations between sexual assault and each phenotype. Associations are grouped by related phenotypes. Red dashed line indicates an OR of 1. DO, disorder.

relating to female genital organs and psychiatric diagnoses associated with pregnancy. For example, we note that sexual assault is strongly associated with the broad “mental disorders during/after pregnancy” phecode (OR = 28.77, 95% CI = 24.6, 33.66).

Some associations were significant in one sex but could not be included in the sex-by-exposure interaction test because the outcome was infrequent (<100 cases) in the other sex. For example, we observed a strong association between sexual assault and the female-predominant diagnosis of dissociative disorder (OR = 113.77, 95% CI = 86.15, 150.24), which recapitulates previous findings that the majority of patients with this disorder have a history of abuse or neglect [21, 22] (cited in [23]).

Of the 410 significant sex-stratified associations, the 354 phecodes with sufficient case counts in both sexes were tested in an interaction analysis examining the moderating effect of sex on sexual assault comorbidities. Of these, 55 phecodes demonstrated significant differences between males and females (Bonferroni-adjusted $\alpha = 1.41e-04$, Fig. 3, online suppl. extended data Table 1 “sex_interaction” sheet), 32 of which pertained to psychiatric conditions, including depression, anxiety disorders, schizophrenia, and PTSD. Interestingly, in all 32 sex-differential psychiatric associations with sexual assault, the effect sizes were significantly greater in males than in females.

Among the nonpsychiatric associations, those stronger in females than in males included “sprains and strains,” gastrointestinal symptoms (“abdominal pain”; “nausea and vomiting”), urinary retention, urinary tract infection, and seizure-related phecodes (“epilepsy, recurrent seizures, convulsions”; “convulsions”). Nonpsychiatric associations that were stronger in males included anal and rectal conditions, HIV-related phecodes, herpes simplex infection, and unspecified congenital anomalies. Several sex-differential associations with nonpsychiatric phecodes exhibit opposite effects in men and women. For example, hypotension, unspecified bacterial infection, muscle weakness, and cough were observed more frequently in females affected by sexual assault than in females not affected, but the reverse was true for males. Conversely, colorectal cancer was observed more frequently in males with a sexual assault history than in those without, but the reverse was true for females.

Benchmarking of Keyword Algorithm Using ICD-Code-Based Case-Selection

To determine whether the associations identified by a keyword-based case-control definition are consistent

with those identified by a simpler ICD-code-based approach, we repeated the PheWAS using ICD codes to classify patients as positive or negative for a history of sexual assault (see Methods). Among phecodes identified as phenome-wide-significant by the keyword-based approach, log ORs from the two approaches were strongly correlated (Pearson $R \geq 0.84$, online suppl. Fig. S1; online suppl. extended data Table 2).

Discussion

Epidemiological studies have established that individuals with a history of sexual assault have an increased risk of developing both immediate and long-term physical and mental health conditions [2–4, 7, 8, 23–25]. We aimed to replicate and extend these findings using a novel clinical informatics-based phenotyping approach in a large hospital setting. Our approach ascertained substantially more patients than did diagnostic codes alone, yielding improved power to study the medical phenomes of sexual assault survivors. This study not only confirmed the utility of our approach but also produced novel insights into the consequences of sexual assault for males and females.

Our phenotyping approach identified hundreds of clinical phenotypes significantly associated with sexual assault, including a preponderance of associations with psychiatric conditions such as schizophrenia, depression, PTSD, and suicidal behaviors. We observed several associations between sexual assault and phenotypes related to functional and somatic symptom disorders including gastrointestinal symptoms, chronic pain, and seizures. These results are consistent with findings from the few epidemiological studies examining the impact of sexual violence on the development of functional and somatic symptom disorders [4, 8]. Nevertheless, they remain largely understudied and warrant further investigation.

We noted several phenotypes associated with sexual assault pertaining to toxic ingestions, such as poisoning by psychotropic agents and analgesics. Psychotropic agents including antidepressants, as well as commonly used analgesics such as paracetamol, are used frequently in suicide attempts [26, 27]. Consistent with this, our analysis demonstrated a strong association between sexual assault and suicide and self-harm. Given the high overlap between toxic ingestions and documentation of suicidal behavior in our data (online suppl. Table S6), the associations between sexual assault and toxic ingestions are likely in part driven by suicidal behavior. Further-

more, when we conditioned on the “suicide or self-inflicted injury” phecode, these associations were diminished (online suppl. Table S7).

Finally, we addressed a gap in the sexual assault literature by leveraging our approach to study sex differences in risk factors and comorbidity patterns among survivors of sexual assault. We found that dissociative disorder, which occurs predominantly in women, is strongly associated with sexual assault in our cohort, recapitulating the observation that most women with the disorder have a history of abuse or neglect [21–23]. Our interaction analysis revealed that abdominal pain, nausea and vomiting, epilepsy, recurrent seizures, and convulsions are more commonly observed in female than male survivors of sexual assault. While we cannot conclusively assert that these conditions represent functional or somatic symptom disorders, prior work linking sexual assault to this class of disorders [4] makes this interpretation plausible. Furthermore, we showed that the seizure-related associations are partially attenuated when conditioning on functional seizures’ case-control status, the phenotyping algorithm for which we recently published [10].

Interestingly, most of the sex-differential psychiatric associations with sexual assault were stronger in men than in women. This finding should be interpreted in light of possible differential reporting and recall bias between males and females who have experienced sexual assault. For example, if males tend to report experiences of sexual assault less frequently than females, it is possible that the men who do disclose such a history in a medical setting are those who are experiencing especially severe health consequences as a result of their trauma. In this scenario, observed clinical associations with sexual assault would appear stronger in males than in females, which is what we observe for many sex-differential phenotypes in our analysis (in particular, psychiatric conditions).

Limitations

The cross-sectional design of our study prohibits us from drawing conclusions about causal or temporal relationships between sexual assault and associated phenotypes. Some of the phenotypic associations we identified, including phecodes relating to congenital anomalies and neurodevelopmental disorders, are likely to have preceded sexual assault and could be interpreted as risk factors for experiencing sexual assault later in life. However, many of the identified associations are consistent with health sequelae of sexual assault identified in longitudinal case-control and cohort studies.

Second, our analysis is subject to the ascertainment bias inherent in EHR-based medical phenome studies. It is difficult to extrapolate our results, which are derived from a cohort of individuals seeking medical care, to the broader population. For example, survivors of sexual assault who frequently interact with the healthcare system may be more likely to receive a greater number of medical diagnoses than those who do not seek or have limited access to care. Nevertheless, it is critical that healthcare systems are prepared to anticipate the care needs of survivors of sexual assault to provide high-quality primary preventative care. Moreover, we have attempted to account for this bias by including a measure of healthcare utilization as a covariate in our analyses.

Third, the small number of cases identified by our keyword-based algorithm relative to sexual assault prevalence estimates [1] suggests that, though it is an improvement over the use of ICD codes alone, our algorithm is still misclassifying many cases as controls and underestimating the prevalence of sexual assault in the VUMC patient population. This underestimation likely stems from the ubiquitous underreporting of sexual assault in healthcare settings. For example, only approximately one third of women with injuries due to rape seek any type of medical care [28]. In the context of this misclassification, our analyses of clinical associations with sexual assault are likely to be overly conservative. While an overly conservative analysis introduces a type II error, it does not reduce confidence in the positive clinical associations we have identified.

Preprint

A preprint version of this article is available on medRxiv [29].

Statement of Ethics

This study was reviewed and approved by the VUMC Institutional Review Board (IRB #212285) and exempted from informed consent requirements as nonhuman subjects research using deidentified medical data.

Conflict of Interest Statement

The authors have declared no conflicts of interest.

Funding Sources

Research reported in this publication was supported by NIGMS of the National Institutes of Health under award No. T32GM007347 and Grant No. R01 HG011405. Grant No. R01 HG011405 was partially funded by the Office of Research on Women's Health, Office of the Director, NIH, and the National Human Genome Research Institute. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Office of Research on Women's Health or the National Human Genome Research Institute.

The synthetic derivative (SD) was supported by the National Center for Research Resources, Grant UL1 RR024975-01, and is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06. The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

- 1 Smith SG, Zhang X, Basile KC, Merrick MT, Wang J, Kresnowjo M, et al. [The National Intimate Partner and Sexual Violence Survey \(NISVS\): 2015 data brief – updated release \(Internet\)](#). Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention; 2018 [cited 2020 Aug 30]. Available from: <https://www.cdc.gov/violenceprevention/datasources/nisvs/2015NISVSdatabrief.html>.
- 2 Jina R, Thomas LS. Health consequences of sexual violence against women. *Best Pract Res Clin Obstet Gynaecol*. 2013 Feb 1;27(1):15–26.
- 3 Chen LP, Murad MH, Paras ML, Colbenson KM, Sattler AL, Goranson EN, et al. Sexual abuse and lifetime diagnosis of psychiatric disorders: systematic review and meta-analysis. *Mayo Clin Proc*. 2010 Jul;85(7):618–29.
- 4 Paras ML, Murad MH, Chen LP, Goranson EN, Sattler AL, Colbenson KM, et al. Sexual abuse and lifetime diagnosis of somatic disorders: a systematic review and meta-analysis. *JAMA*. 2009 Aug 5;302(5):550–61.
- 5 Dartnall E, Jewkes R. Sexual violence against women: the scope of the problem. *Best Pract Res Clin Obstet Gynaecol*. 2013 Feb;27(1):3–13.
- 6 Oram S, Khalifeh H, Howard LM. Violence against women and mental health. *Lancet Psychiatr*. 2017 Feb;4(2):159–70.
- 7 Rhodes AE, Boyle MH, Tonmyr L, Wekerle C, Goodman D, Leslie B, et al. Sex differences in childhood sexual abuse and suicide-related behaviors. *Suicide Life Threat Behav*. 2011; 41(3):235–54.
- 8 Young-Wolff KC, Sarovar V, Klebaner D, Chi F, McCaw B. Changes in psychiatric and medical conditions and healthcare utilization following a diagnosis of sexual assault: a retrospective cohort study. *Med Care*. 2018 Aug; 56(8):649–57.
- 9 Brignone E, Gundlapalli AV, Blais RK, Kimerling R, Barrett TS, Nelson RE, et al. Increased health care utilization and costs among veterans with a positive screen for military sexual trauma. *Med Care*. 2017 Sep; 55(Suppl 2):S70–S77.
- 10 Goleva SB, Lake AM, Torstenson ES, Haas KF, Davis LK. Epidemiology of functional seizures among adults treated at a university hospital. *JAMA Netw Open*. 2020 Dec 29; 3(12):e2027920.
- 11 Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc*. 2013 Jun;20(e1):e147–154.
- 12 Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics*. 2010 May 1; 26(9):1205–10.
- 13 Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics*. 2014 Aug 15; 30(16):2375–6.
- 14 R Core Team. [R: a language and environment for statistical computing \(Internet\)](#). Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: <https://www.R-project.org/>.
- 15 Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013 Dec;31(12):1102–11.
- 16 Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, et al. Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR Med Inform*. 2019 Nov 29;7(4):e14325.
- 17 Mayson BE, Teichman JM. The relationship between sexual abuse and interstitial cystitis/painful bladder syndrome. *Curr Urol Rep*. 2009 Nov 1;10(6):441–7.
- 18 Lalchandani P, Lisha N, Gibson C, Huang AJ. Early life sexual trauma and later life genitourinary dysfunction and functional disability in women. *J Gen Intern Med*. 2020 Nov 1; 35(11):3210–7.
- 19 Arnett AB, Pennington BF, Willcutt EG, DeFries JC, Olson RK. Sex differences in ADHD symptom severity. *J Child Psychol Psychiatr*. 2015;56(6):632–9.
- 20 Brody DJ, Pratt LA, Hughes JP. [Prevalence of depression among adults aged 20 and over: United States, 2013–2016 \(Internet\)](#). Hyattsville, MD: National Center for Health Statistics; 2018 [cited 2020 Aug 30]. [NCHS Data Brief]. Report No.: 303. Available from: <https://www.cdc.gov/nchs/products/databriefs/db303.htm>.
- 21 Kluft RP. An update on multiple personality disorder. *Psychiatr Serv*. 1987 Apr 1;38(4): 363–73.
- 22 Ross CA, Miller SD, Bjornson L, Reagor P, Fraser GA, Anderson G. Abuse histories in 102 cases of multiple personality disorder. *Can J Psychiatry*. 1991 Mar;36(2):97–101.
- 23 Teicher MH, Samson JA. Childhood maltreatment and psychopathology: a case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. *Am J Psychiatry*. 2013 Oct 1;170(10):1114–33.

Author Contributions

Lea K. Davis, Allison M. Lake, and Slavina B. Goleva conceived of the study design and aims. Allison M. Lake drafted the manuscript and performed all statistical analyses. All the authors participated in the critical review and interpretation of the results. Allison M. Lake, Lauren R. Samuels, Laura M. Carpenter, and Lea K. Davis contributed to the review and editing of the manuscript. All the authors approved the final version of the manuscript.

Data Availability Statement

All summary-level data produced in the present work are included in this article. Deidentified VUMC EHR data analyzed in the present work are not available for public access due to privacy restrictions. Further inquiries can be directed to the corresponding author.

- 24 Finestone HM, Stenn P, Davies F, Stalker C, Fry R, Koumanis J. Chronic pain and health care utilization in women with a history of childhood sexual abuse. *Child Abuse Negl.* 2000 Apr 1;24(4):547–56.
- 25 Kilpatrick DG, Ruggiero KJ, Acierno R, Saunders BE, Resnick HS, Best CL. Violence and risk of PTSD, major depression, substance abuse/dependence, and comorbidity: results from the National Survey of Adolescents. *J Consult Clin Psychol.* 2003 Aug;71(4):692–700.
- 26 Hawton K, Bergen H, Simkin S, Cooper J, Waters K, Gunnell D, et al. Toxicity of antidepressants: rates of suicide relative to prescribing and non-fatal overdose. *Br J Psychiatry.* 2010 May;196(5):354–8.
- 27 Sarchiapone M, Mandelli L, Iosue M, Andrisano C, Roy A. Controlling access to suicide means. *Int J Environ Res Public Health.* 2011 Dec;8(12):4550–62.
- 28 [Full report of the prevalence, incidence, and consequences of violence against women: findings from the national violence against women survey \(Internet\)](#). National Institute of Justice; 1998 [cited 2021 Nov 10]. Available from: <https://nij.ojp.gov/library/publications/full-report-prevalence-incidence-and-consequences-violence-against-women>.
- 29 Lake AM, Goleva SG, Samuels LR, Carpenter LM, Davis LK. Health conditions associated with sexual assault in a large hospital population. *medRxiv.* 2022.