Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
J Fam Violence. Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:
J Fam Violence. 2017 January ; 32(1): 79–87. doi:10.1007/s10896-016-9872-5.

Mining Electronic Health Records Data: Domestic Violence and Adverse Health Effects

Gunnur Karakurt,
Department of Psychiatry, Case Western Reserve University

Vishal Patel,
Center for Proteomics and Bioinformatics, Case Western Reserve University

Kathleen Whiting, and
Department of Psychiatry, Case Western Reserve University

Mehmet Koyutürk
Department of Electrical Engineering and Computer Science, Case Western Reserve University

Abstract

Intimate partner violence (IPV) often culminates in acute physical injury, sexual assault, and mental health issues. It is crucial to understand the healthcare habits of victims to develop interventions that can drastically improve a victim's quality of life and prevent future abuse. The objective of this study is to mine de-identified and aggregated Electronic Health Record data to identify women's health issues that are potentially associated with IPV. In this study we compared health issues of female domestic abuse victims to female non-domestic abuse victims. The Domestic abuse population contained 5870 patients, while the Non-Domestic Abuse population contained 14,315,140 patients. Explorys provides National Big Data from the entire USA. Statistical analysis identified 2429 terms as significantly more prevalent among victims of domestic abuse, compared to the general population. These terms were classified into broad categories, including acute injury, chronic conditions, substance abuse, mental health, disorders, gynecological and pregnancy related problems.

Keywords

intimate partner violence; domestic violence; health consequences; big data mining; epidemiology

---

Intimate partner violence (IPV) is a toxic public health problem that influences many layers of society in the US and around the world. The National Intimate Partner and Sexual Violence Survey (Black, Basile, Breiding, Smith, Walters & Merrick, 2011) found that about 30% of women experience physical violence from their intimate partners over their lifetimes. Further investigations of this national data also indicated that one in four women experience severe physical violence, meaning 17.2% of women have been slammed against something, 14.2% have been hit with a fist or something hard, and 11.2% have been beaten

Correspondence concerning this article should be addressed to Gunnur Karakurt, Department of Psychiatry, Case Western Reserve University, 10900 Euclid Ave, Cleveland, OH 44106. gkk6@case.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

by an intimate partner. Women are more likely to be the victim in the relationship and to sustain severe physical injury from the assault than men (Black et al., 2011; Corrigan, Wolfe, Mysiv, Jackson & Bogner, 2003).

Intimate partner violence can adversely influence health, from minor injuries and cuts, to death (Corrigan et al., 2003; Lawrence, Oringo & Brock, 2012). Physical and sexual assault are found to be associated with sexually transmitted infections, including human immunodeficiancy virus (Wingood, DiClemente, McCree, Harrington & Davies, 2001); pelvic inflammatory disease (Letourneau, Holmes, & Chasedunn-Roark, 1999); unintended pregnancy (Hathaway, Mucci, Silverman, Brooks, Mathews & Pavlos, 2000); and psychological distress (Karakurt, Smith & Whiting, 2014). Violence during pregnancy can adversely affect the mother and the newborn's health (El Kady, Gilbert, Xing & Smith, 2005; Huth-Bocks, Levendosky & Bogat, 2002). Medical professionals are beginning to suspect that there may be many other health consequences of IPV less visible and more complex than previously thought. Up to this point, they have focused on individual symptoms, giving little attention to more holistic approaches that could reveal and address the source of the problem. Past research investigating the long-term effects of IPV indicated that chronic pain, neurologic disorders, gastrointestinal disorders, migraine, headaches, and other physical disabilities, as well as posttraumatic stress disorder (PTSD), depression, anxiety disorders, substance abuse, and suicide are more common among victims of violence (Karakurt et al., 2014; Nelson, Bougatsos & Blazina, 2012).

Women who have been the victims of violence reported significantly poorer overall health than women who were the victims of other crimes, and they presented physical symptoms such as rapid heart rate, tension headaches, stomach aches, skin problems, menstrual symptoms, among others (Campbell, 2002). Those who have suffered sexual assault are much more likely to need mental and physical health services within six months. In the year following the assault, medical services increased 15-24%, which means that the cost of treatment of these victims is much higher than the health care costs of non-victims. Often, health problems resulting from violent assault are overlooked by physicians, or attributed to incorrect causes. This can result in more pain and suffering on the patients' part, and wasted time and money for both the patient and the provider. In order to most effectively help these victims, it is imperative that doctors can recognize the signs of IPV and make appropriate treatment intervention decisions.

Although negative health effects of intimate partner violence can be serious and diverse, such as acute, long term, and more systemic health issues, current knowledge of those adverse effects is limited. Factors contributing to this lack of knowledge involve challenges in data collection due to the sensitive nature of the topic and lack of systemic investigations of the symptoms due to difficulties with data analysis. Additionally, most investigations are limited to self-report data, which can be incomplete and inaccurate. Mining "big data" could potentially be a very effective way to identify these subtle symptoms that so far have gone uncorrelated with IPV. Data mining is made possible by the availability of large collections of electronic health records (EHR), data that could be instrumental in the much needed detailed investigation of the consequences of violence on human health. The aim of this study is to utilize large-scale EHR data to investigate the adverse health effects following

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

violent episodes. To our knowledge, this is the first study to attempt investigating the use of mining EHR data to analyze and identify possible correlations between symptoms and population cohorts. Although the statistical analysis used in this study is somewhat limited, we feel its purpose at this point is to verify that these methods of data mining can reproduce symptom associations of IPV currently acknowledged in the research literature.

In order to achieve this aim, we analyzed the national EHR data provided by the *Explorys* platform (Explorys Inc., Cleveland, OH). *Explorys* is comprised of EHR, electronic medical records (EMR), billing, and claims data sources. The data provided by *Explorys* can help us to identify associations between health problems directly resulting from acute violence and its related physical injuries, and chronic health problems that result from a number of causes and persist (Lorenzi, Kouroubali, Detmer & Bloomrosen, 2009).

The *Explorys* platform was selected to access the national big data in healthcare. *Explorys'* data were collected from various affiliated providers, plans, care settings, and electronic medical systems. This platform ties data together over 315 billion clinical, financial, and operational data elements, about 50 million unique patients from 360 hospitals, and over 317,000 providers. Furthermore, 23 major integrated healthcare systems including Cleveland Clinic, Trinity Health, St. Joseph Health System, Mercy Health, Adventist Health System, and others with patients in all 50 states of the USA participating as a data sources. The data has been used to identify patterns in diseases, treatments, and outcomes by researchers in diverse disciplines.

After completing the search query in *Explorys*, we identified 3458 symptom terms that co-occurred with domestic abuse (we queried for records containing the finding 'domestic abuse'), with statistical analysis indicating 2429 of those terms were significantly more prevalent among victims of IPV. Categorization of the symptoms into broader categories with the assistance of a medical dictionary revealed that the symptoms could be organized into four main classes: acute injuries, chronic symptoms and disorders, gynecological related problems, and mental and behavioral health issues. This data may help healthcare professionals improve screening procedures and treatment plans in the future, not only for victims of domestic violence but for patients in all medical disciplines. The knowledge gained through this analysis provides evidence for the concept that analyzing EHR information can demonstrate significant correlations for utilization in clinical and research applications.

Method

In this section, we describe the data collection and analysis procedures we applied to identify symptoms that are potentially associated with domestic violence. An important consideration in mining EHR data is that these data are incomplete, noisy, and potentially biased by nature. Although data is de-identified in the system in compliance with HIPAA regulations, it is conceivable that a patient's identity could be extrapolated from the data if all records were provided completely and in bulk. To eliminate this risk, commercial platforms providing this information prevent the user from accessing records in their entirety. In particular, due to privacy and intellectual property concerns, the data is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

accessible in summaries and in the form of response to queries, as opposed to being available in bulk. Despite all these concerns, these data are "big", hence they potentially contain information that may be out of reach through any other means. Since the data is very large and provided in the form of summaries, we are not concerned about any gaps in accessible records. For this reason, we tailor existing computational and statistical techniques to mining summaries of large scale EHR data.

Data Collection

All of the data used in this study was obtained from *Explorys,* a commercial platform that provides access to EHR data, as well as EMR, billing, and claims data. *Explorys* utilizes the Systematized Nomenclature of Medicine (SNOMED) for all queries. *Explorys* provides access to data through a query interface that collapses data across patients, thereby preventing identification of patient-specific information for purposes of protecting privacy and intellectual property rights. For this reason, the data can be queried only for frequencies; however, it is possible to specify the variables to be used for the selection of subsets of patients. For example, a query with the selection of "domestic violence" in the "diagnosis" field will return the frequencies of all variables among the patients that have been diagnosed with domestic violence. However, the interface does not allow access to the values of each variable for each patient; therefore the analysis has to be done at the level of frequencies.

The variables in the *Explorys* database include demographic information (age, language, race, religion, gender, zip code), diagnosis, finding, observation, drug pharmaceutical class, drug ingredients, procedure, and vitals. Since variables diagnosis, finding, and observation are highly related to each other and our purpose is to identify health concerns associated with domestic violence, we decided to focus on "diagnosis", since the term diagnosis reflects a specific meaning. Furthermore, the patients are clustered into three groups according to their activity status: Active ever (has a record in the database regardless of time), last three years (the patient has a record that is as recent as in the last three years), and last one year (the patient has a record that is as recent as in the last year). In order to keep database size at a maximum and identify all patterns present in the data with sufficient statistical power, we decided to consider all patients without any restriction on their activity status and selected "active ever" in all of our queries. Finally, since the diagnosis of "domestic violence" was significantly more common among female patients and we wanted to focus on adult victims of domestic violence, we restricted all our queries to female patients between the ages of 18 and 65.

Patient Identification

We utilized *Explorys* query system to identify the potential patients with domestic abuse issues. Similar to SNOMED, the International Classification of Diseases (ICD) is a classification system that provides a system of diagnostic codes for classifying diseases. Various signs, symptoms, social circumstances, and external causes of injury or diseases were designed to map health conditions to corresponding generic categories, and each of these generic categories is assigned a specific code up to six characters long, designed to group a set of similar diseases together. By going through the underlying SNOMED data in *Explorys*, it is possible for researchers to match relevant SNOMED information to specific

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ICD classifications (in the *Explorys* platform, data may originate as ICD coded diagnoses, but it is standardized to SNOMED to facilitate processing). For this study, we were specifically interested in using SNOMED queries for data corresponding to ICD-9-CM v6.11 to identify victims of domestic abuse in the national health database.

Queries

In order to extract valuable information from EHRs in the presence of data access constraints, we ran two queries on the *Explorys* database on May 8, 2014:

1. All patients containing the term "domestic violence" in the "Diagnosis" field (henceforth referred to as the DA group),
2. All patients not containing the term "domestic violence" in the "Diagnosis" field (henceforth referred to as the NDA group).

These two queries provided us with a "test" population composed of patients diagnosed with domestic violence and a "background" population composed of all other patients. The DA population contained F_T=5870 patients, while the NDA population contained F_B=14,315,140 patients. The age distribution of each population is shown in Figure 1. Age distribution of the DA and NDA groups are consistent with prior research that younger women are at higher risk of intimate partner violence, peaking at ages 25 to 34.

Statistical Analysis

For this analysis we considered both log odds ratio and χ^2 statistics since they provide information at different levels. More specifically, log odds ratio provides an interpretable measure of how likely a victim of domestic violence is to be diagnosed with a given disease, as compared to any other patient in the population. However, log odds ratio does not reward terms that are more frequent in the general population, and can therefore be over-estimated for terms that are less frequent in the population. While this can be accounted for by estimating the standard error, this statistic does not provide sufficient resolution for our application, since the background population is much larger as compared to the test population. For this reason, we also considered the χ^2 statistic for purposes of ranking terms according to the difference in test and background populations. We corrected the p-values estimated for the χ^2 tests for multiple hypotheses testing using Bonferroni correction. Using these techniques to identify diagnoses that are associated with domestic abuse, we compared the frequencies of each term in the DA and NDA population. Namely, for diagnosis "d" (henceforth referred to as a "term"), let $f_T(d)$ and $f_B(d)$ respectively denote the frequency of diagnosis d in the DA and NDA populations. Our aim is to quantify the significance of $f_T(d)$ given $f_B(d)$. For this purpose, we used two well-established statistics: 1) log odds ratio, computed as log-odds(d)=log(($f_T(d)/F_T)(F_B/ f_B(d)$)), 2) χ^2 statistic, as computed based on the 2×2 table populated by $f_T(d)$, $F_{T,}$ $f_B(d)$, F_B, and associated p-value.

Furthermore, in *Explorys*, the frequencies are rounded to the nearest 10 for privacy purposes. This is done so that outside parties cannot extrapolate patient information by analyzing large datasets. However, since this may cause rounding errors, we also computed confidence intervals for the log-odds ratio of each term using Monte Carlo simulations. Namely, we generated a possible original instance (i.e., with unrounded frequencies) of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

data by randomly assigning integer values in the interval [f-5, f+5] to each frequency f in the data (where f is always a multiple of 10) and computed the test statistics for each term using these randomly generated values. We repeated this procedure 1000 times and, for each term, recorded the 5th smallest and 5th largest of the test statistics among these 1000 runs estimate a 99% confidence interval for the respective test statistic.

The terms were ranked based on both log odds ratio and χ^2 statistic as follows: We obtained two rankings, one based on log odds ratio, the other based on χ^2 statistic. Subsequently, we computed the maximum (worse) of the two ranks assigned to each term. We then ranked the terms according to this maximum rank. This method ensured that a term is ranked high only if it is ranked high according to both criteria.

Results

Key Findings

We identified 3458 terms to be related with DA based on the search query conducted in Explorys to find all records in the system with the diagnosis finding ‘domestic abuse’. Explorys returned 3458 comorbid diagnosis terms found within the records of the DA cohort. Our statistical analysis indicated that 2429 terms were identified as significantly ($p < 0.05$, after correction for multiple hypothesis testing) more prevalent among victims of domestic abuse. Two raters then classified these terms into broad categories with the assistance of a medical dictionary (Venes & Taber, 2013). These categories include: acute injury; chronic, substance abuse; mental health; disorders; gynecological issues; pregnancy related problems; allergy; procedures; congenital/hereditary; nutrition; neoplasm; personal history; family history; neuropathy; diabetes; gastrointestinal; cardiovascular; nervous system; respiratory; musculoskeletal; eyes, ears, nose & throat, excretory; endocrine; immune system; skin related problem. The distribution of the 2429 terms into these categories is shown in Figure 2.

One term can potentially get classified to more than one category. For example, the term ‘psychoactive substance-induced organic mental disorder’ is coded as acute condition, substance abuse and mental health. Another example is the term ‘anemia due to substance’, which was coded with acute, chronic, substance use and cardiovascular. There are 39 terms which the two coders disagreed on; consensus was reached through discussion. Interrater reliability was 99%. The percentage of terms that fall into each category is shown in Figure 2. As seen in the table, approximately one in three significant term represents an acute condition, one in four significant term represents an acute injury, and one in six significant term represents a chronic disease.

The statistics for the most significant 10 terms are shown in Table 1. These 10 terms were selected as the most significant based on ranking according to both log odds ratio and χ^2 statistic, as described in the last paragraph of the Methods section. Our results indicated that contusion of soft tissue, superficial bruising, contusion, head and neck injury, wound of skin, injury of head, injury of integument, contusion of multiple sites, superficial injury and injury of face are the 10 most commonly reported terms. These results show that the terms that are most significantly associated with domestic violence correspond to acute injuries.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Discussion

As the implementation of EHR in health practices nationwide continues to increase, and as more research has been performed relating to intimate partner violence and associated symptoms, several changes have occurred in the healthcare system that have increased the visibility of IPV among healthcare providers. There are a number of direct and indirect negative health outcomes associated with violent assault. Results of the study indicate that over 2429 symptoms were identified as significantly more prevalent among victims of domestic abuse. These symptoms covered a wide variety of pathologies. However, detailed investigation of these symptoms indicated that they fall under four categories:

1. Acute symptoms mostly due to injury inflicted from the abuser such as fractures, contusion, and wounds.
2. Chronic symptoms, disorders and cardiovascular problems.
3. Gynecological and pregnancy related problems.
4. Mental and behavioral health issues.

Main Categories Relating to IPV

Acute symptoms—Intimate partner violence by its very nature is incredibly damaging to the body. Physical force that leads to injuries can range from minor abrasions and bruises to multisystem trauma and even death. Past research indicates that victims are most commonly injured by being struck with a hand or an object, and/or crushing impact (Sheridan & Nash, 2007). These blunt force injuries can involve abrasions, contusions, lacerations, and fractures (Rooms & Shapiro, 2006). When IPV involves using weapons, there is a higher risk of mortality. The most common type of injury suffered by IPV victims is soft tissue injury (Sheridan & Nash, 2007). Indeed, our findings also indicated that among 5870 patients, 1650 of them visited medical centers for contusion of soft tissue. Trauma to the body's soft tissues is also commonly reported among victims, with frequent reports of cuts and bruises to the limbs and abdomen (Wu, Huff & Bhandari, 2010). Unless treated, these injuries can lead to infection, necessitating treatment by medical providers from a variety of backgrounds.

Injuries are most likely to occur in the head, face, and neck areas (Sheridan & Nash, 2007). Related injuries with high levels of sensitivity and moderate levels of specificity are linked with these locations (Sheridan & Nash, 2007). Thus it is not surprising that traumatic brain injury is a very real risk for victims of IPV, with reported prevalence rates ranging from 30%-74% (Kwako, Glass, Campbell, Melvin, Barr & Gill, 2011). Emergency department records show that victims describe loss of consciousness, dizziness, and disorientation, frequently leading to memory problems and other neurological sequelae.

The personal nature of the relationship between victim and perpetrator causes many perpetrators to pursue injury to the victim's face over the rest of the body. This can result in significant damage to the rest of the body as the victim attempts to protect his/her face with other limbs. In addition to the previously mentioned acute symptoms, bone fractures and musculoskeletal problems, including dislocation, are a serious risk. Unfortunately, many of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the medical professionals who treat these acute injuries and conditions either have misperceptions about IPV, not enough education to identify the signs of IPV, or feel uncomfortable confronting patients about IPV because they do not know how to approach the situation (Sprague, Madden, Dosanjh, Godin, Goslings, Schemitsch, & Bhandari, 2013). A thorough analysis of symptoms correlated with IPV will enable healthcare providers to determine how best to screen for and intervene in the event of suspected IPV. Treatment of acute symptoms provides medical professionals with many opportunities to address situations involving domestic violence if proper training can be provided. It should be noted that all of the top 10 commonly observed symptom terms based on odds ratio and χ^2 were coded into the 'acute injury' category for this study (Table 1). These 10 terms were 'contusion of soft tissue', 'superficial bruising', 'contusion', 'head and neck injury', 'wound of skin', 'injury of head', 'injury of integument', 'contusion of multiple sites', 'superficial injury', and 'injury of face'. It is interesting to note that three of these terms specifically relate to injuries involving the head/neck/face area, which is a well-known characteristic of domestic abuse victims' injury patterns (Kwako et al., 2011).

Chronic symptoms, disorders, and cardiovascular problems.—There are a variety of underlying causes that may have contributed to long-term changes in health perceptions relating to IPV and violent assault. Physiological changes have been shown to correspond to traumatic violent assault and the subsequent psychological issues associated with it. The traumatic exposure of violent assault results in repeated physiological arousal from the central nervous system (classical conditioning); this chronic physiological arousal may be associated with PTSD (Koss & Heslet, 1992). Individuals with PTSD may experience higher rates of cardiovascular, respiratory, musculoskeletal, and neurological symptoms due to the persistent physiological arousal that results from the traumatic event. Following the assault, victims also experience changes in the endocrine system response which can suppress immune system functioning and may indirectly cause the onset of illness or injuries related to poor health. Alterations of the hypothalamic-pituitary-adrenal axis (HPA) have been connected to PTSD, and immune functioning changes may negatively influence the victims' susceptibility to fatigue, infections, and malignant disease.

There is some evidence that learning theory mechanisms like conditioning and anticipatory anxiety play an important role in the development of physiological symptoms that are linked to the experience of emotional and psychological distress among victims of violent assault and IPV. It has been hypothesized that the classical conditioning of emotional responses along with the post-trauma avoidance of reminders may lead to PTSD (Koss & Heslet, 1992). The unconditioned stimulus of a life-threatening or extremely distressing assault situation elicits an automatic response with physiological components (including functional alterations in breathing and heart rate), cognitive components (including fear of having one's life threatened or one's body broken or defiled), and behavioral components (including screaming, running, freezing, and other "fight or flight" responses). This unconditioned emergency-survival response can be paired with various cues related to the assault situation or environment, but not necessarily dangerous in and of themselves (such as the weather during the assault, time of day, etc). The learned fear response resulting from the traumatic event cannot be unlearned, extinguished, or desensitized without intervention. Physical

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

reactions that occurred during the assault, such as shaking, rapid heart rate, nausea, and shortness of breath (Resnick, Acierno & Kilpatrick, 1997), may all be re-lived when the fear response is triggered, to the intense discomfort of the victim. A cycle of avoidance may occur in which physiological symptoms associated with or relating to the attack set off the cognitive experience of danger, leading to typical behavioral responses involving avoidance of those triggers (Koss & Heslet, 1992).

Gynecological and pregnancy related problems—Gynecological problems associated with intimate partner violence are extensive. Rape and sexual assaults can be devastating to the female reproductive system, causing a wide range of ailments including uterine ruptures, vaginal tears, numerous sexually transmitted diseases, unplanned pregnancy, miscarriage or stillbirth, and in the worst cases even maternal and/or fetal death (American College of Obstetricians and Gynecologists, 2012; El Kady et al., 2005; Kwako et al., 2011). These gynecological issues resulting from assault occur in addition to the other substantial mental and physical effects, not only from sexual assault but any form of domestic violence. Depression and anxiety, a frequent correlate of IPV and 2.5 times more likely among women who are abused during pregnancy, can cause dangerously low weight gain during pregnancy as well as low birth weight (Parker, McFarlane & Soeken, 1994; Taillieu & Brownridge, 2010). Unfortunately, abuse during pregnancy often results in late entry to prenatal care, as well as missed visits to the doctor (Taillieu & Brownridge, 2010), which can compromise the health of both mother and child. For women suffering spousal abuse, regardless of whether or not they are pregnant, the likelihood of having gynecological problems is three times greater than average (Campbell, 2002). Many of these are chronic or acute conditions.

Mental and behavioral health issues—The mental health of violent assault victims also suffers following the attack. Ninety percent of domestic abuse victims exhibited PTSD (Resnick et al., 1997) symptoms within weeks of having been attacked, and it has been shown that they are more likely to develop PTSD. Depression is also prevalent among victims of intimate partner violence. These detrimental psychological symptoms following violent assault contribute not only to mental health problems like depression, PTSD, and panic attacks, but also to physical health problems like substance abuse, heart attacks, and higher rates of sexually transmitted diseases.

Increased rates of substance abuse after violent assault (particularly cigarette smoking) is often considered a coping mechanism that the victim perceives as a way to forget the traumatic event and reduce current fear, anxiety, or depression (Resnick et al., 1997). Victims frequently exhibit risky sexual behavior and often turn to alcohol abuse. Although it is used for self-treatment by victims, alcohol also leads to increased rates of physical violence. Over 60% of all homicides and 49% of non-intentional injury fatalities are related to alcohol. Of the individuals seeking hospital treatment for serious violent assault, 70% were under the influence of alcohol or drugs, and 60% said their attackers were under the influence (Resnick et al., 1997). Almost one third of women who report drug use have also reported being victimized in their lifetime.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Limitations

As with any research, the results of this study should be viewed through its limitations. One major limitation is the difficulty we encountered when trying to accurately and thoroughly identify domestic abuse cases. In order for us to access and identify the domestic abuse cases, we had to utilize the ICD codes through the underlying SNOMED data. However, the ICD code for domestic abuse is not widely used by medical professionals, which likely means that our data query missed many instances of domestic abuse. Our data also suggests that the cases we identified using the SNOMED equivalence of the ICD code are skewed towards the most severe occurrences of domestic violence, probably because the severe injuries made it more obvious that domestic abuse was the cause and thus more likely that the ICD code was used. The corollary of this, of course, is that the data neglects less severe instances of abuse, missing more subtle symptom correlations of IPV.

Another limitation we encountered during this study was actually the quality of the data itself. The data is very unstandardized; it's noisy, raw, and does not include any temporal information. Even the query system poses problems for analysis, because it was designed for clinical research to provide only frequency data. This makes it nearly impossible to control for confounding variables. Nevertheless, the data provides valuable information on potentially severe cases of domestic abuse, and provides a wide range of symptoms that are more prevalent among victims. With proper application, this information can be used to improve screening and diagnosis.

Future Research

While the results obtained through our first attempt to mine EHR data add modest information to the investigation of adverse health effects relating to IPV, these results demonstrate the validity of using this methodology to investigate IPV-related health effects. We envision these results as the first stage of an extended research project harnessing big data with more sophisticated computational approaches for research into IPV screening, diagnosis, and treatment strategies. Our results show that through data analysis of EHR, we can identify categories of adverse health effects that are more significantly related to victims of IPV than the general population. Furthermore, our results corroborate the findings of previously conducted research that did not utilize EHR, although our study does not involve any direct patient contact or extensive data collection. We were able to obtain a much greater volume of data by expending a drastically lower amount of time, effort, and funds than observational studies. Our future research will develop a network map of the associations between these 28 identified symptom categories, and if these methods continue to replicate accepted research findings regarding health effects related to IPV, we will further develop a network analysis of EHR to map IPV comorbidities. This will enable us to identify what symptoms are risk factors for IPV and allow us to develop improved screening tools for clinical use.

Despite the limitations of using data mining to determine symptom relationships and associations, there are several promising avenues for further research. Through creative analysis of query results, it may be possible to identify some of the confounding factors. It will be useful in the future to assess the relative frequency of all factors, such as

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

demographics, in both the DA and NDA groups, as well as dividing the population into groups based on factors with significant differences. It would also be very useful to attempt to distinguish risk factors from consequences by running temporal queries. For instance, given the results from a “domestic abuse” query at one time point, we could then go on to do multiple queries later to analyze what diagnoses are frequent in the following one, five, even 10 years. All of these efforts could advance the specificity of identifying symptoms and conditions correlated to IPV victimization.

Unfortunately, all of these advancements can only help if the healthcare system advances as well. Victims often neglect their health, and doctors frequently misdiagnose conditions because they are not always fully aware of the patient's medical history or the latest findings regarding which symptoms have been associated with certain conditions. By examining the data from EHR and analyzing what health factors are associated with IPV, it may be possible to improve the chances of successful interventions and diagnoses, improve victims’ quality of life, and possibly even prevent future health problems.

Conclusion

In conclusion, the utilization of EHR systems as valuable data sources is vital to ensuring the proper dissemination of health services to victims of violent assault and intimate partner violence, as well as to the population as a whole. By having access to a patient's complete medical record, including the details of all previous healthcare procedures/visits, all reported symptoms and diagnosed conditions, and all lifestyle/health habits, health care providers can more accurately determine what is actually going on with the patient mentally and physically, and will be better able to prescribe more effective treatment interventions. This type of information also helps researchers to find relationships between conditions that would be considered independent otherwise.

Potential applications of the patterns that can be identified by analyzing EHR data include clinical screening. If we can identify conditions that are significantly linked to IPV, we can use such information to develop screening tools that check for those related conditions to see if a patient is at risk of IPV, thus giving healthcare professionals the opportunity to refer the patient for appropriate resources. Conversely, if a patient is known to be a victim of IPV, physicians and healthcare providers will know to check for the presence of associated conditions and can pursue preventative measures against some adverse health effects. If healthcare providers have a better understanding of the relationship between IPV and comorbid health conditions, it will improve their ability to effectively diagnose illnesses and treat patients.

Data from EHR systems gives doctors and researchers insight into trends across different populations. Knowledge of these trends and statistics will better equip doctors to recognize the telltale signs of IPV and to distinguish between health problems directly resulting from acute violent assault and its associated physical injuries, versus chronic health problems resulting from a number of other unrelated causes. This type of knowledge will reduce misdiagnosis, will assist in developing more effective screening tools and treatments, and will improve the quality of life and healthcare for victims of IPV and violent assault.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgments

This publication was made possible in part by the Clinical and Translational Science Collaborative of Cleveland, NIH/NCRR CTSA KL2TR000440 from the National Center for Advancing Translational Sciences (NCATS) component of the National Institutes of Health and NIH roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Brief Author Biographies

Gunnur Karakurt, Ph.D., LMFT, is an Assistant Professor at the Department of Psychiatry, Case Western Reserve University in Cleveland, OH. She is also a multidisciplinary clinical researcher at the same institution.

Vishal N. Patel, MD Ph.D., is currently a Data Scientist at Pieces Technologies, Inc., in Dallas, TX.

Kathleen Whiting, BS is currently a PhD student in the Neuroscience Program at the Uniformed Services University of the Health Sciences in Bethesda, MD. She also worked as a research assistant in the Psychiatry Department at Case Western Reserve University in Cleveland, OH.

Mehmet Koyuturk, Ph.D is currently T. & A. Schroeder Associate Professor at the Department of Electrical Engineering and Computer Science at Case Western Reserve University in Cleveland, OH.

References

American College of Obstetricians and Gynecologists. Intimate partner violence - committee opinion no. 518. Obstetrics & Gynecology. 2012; 119:412–417. [PubMed: 22270317]

Black MC, Basile KC, Breiding MJ, Smith SG, Walters ML, Merrick MT. National intimate partner and sexual violence survey 2010 summary report and sexual violence survey. 2011

Campbell J. Health consequences of intimate partner violence. The Lancet. 2002

Corrigan J, Wolfe M, Mysiv W, Jackson R, Bogner J. Early identification of mild traumatic brain injury in female victims of domestic violence. American Journal of Obstetrics & Gynecology. 2003; 188(5):S71–S76. [PubMed: 12748454]

El Kady D, Gilbert W, Xing G, Smith L. Maternal and neonatal outcomes of assaults during pregnancy. Obstetrics & Gynecology. 2005; 105(2):357–363. doi: 105:357–63. [PubMed: 15684165]

Hathaway J, Mucci L, Silverman J, Brooks D, Mathews R, Pavlos C. Health status and health care use of massachusetts women reporting partner abuse. American Journal of Preventive Medicine. 2000; 19(4):302–307. [PubMed: 11064235]

Huth-Bocks A, Levendosky A, Bogat G. The effects of domestic violence during pregnancy on maternal and infant health. Violence & Victims. 2002; 17(2):169–185. [PubMed: 12033553]

Karakurt G, Smith D, Whiting J. Impact of intimate partner violence on women's mental health. Journal of Family Violence. 2014; 29(7) doi: 10.1007/s10896-014-9633-2.

Koss M, Heslet L. Somatic consequences of violence agains women. Archives of Family Medicine. 1992; 1:53–59. [PubMed: 1341588]

Kwako L, Glass N, Campbell J, Melvin K, Barr T, Gill J. Traumatic brain injury in intimate partner violence: A critical review of outcomes and mechanisms. Trauma, Violence & Abuse. 2011; 12(3): 115–126. doi: 10.1177/1524838011404251.

Lawrence E, Oringo A, Brock R. The impact of partner abuse on partners. Partner Abuse. 2012; 3(4)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Letourneau E, Holmes M, Chasedunn-Roark J. Gynecological health consequences to victims of interpersonal violence. Women's Health Issues. 1999; 9(2):115–120. [PubMed: 10189822]

Lorenzi N, Kouroubali A, Detmer D, Bloomrosen M. How to successfully select and implement electronic health records (EHR) in small ambulatory practice settings. BMC Medical Informatics and Decision Making. 2009; 9:15–15. doi: 10.1186/1472-6947-9-15. [PubMed: 19236705]

Nelson H, Bougatsos C, Blazina I. Screening women for intimate partner violence: A systematic review to update the 2004 U.S. preventive services task force recommendation. Annals of Internal Medicine. 2012; 156(11):796–808. doi: 10.7326/0003-4819-156-11-201206050-00447. [PubMed: 22565034]

Parker B, McFarlane J, Soeken K. Abuse during pregnancy: Effects on maternal complications and birth weight in adult and teenage women. Obstetrics & Gynecology. 1994; 84:323–328. [PubMed: 8058224]

Resnick HS, Acierno R, Kilpatrick DG. Health impact of InterpersonalViolence. section II: Medical and mental health outcomes. Behavioral Medicine. 1997; 23:79–85. [PubMed: 9309347]

Rooms, RR., Shapiro, PD. Approach for emergency medical personnel.. In: Lynch, VA., Duval, JB., editors. Forensic Nursing. Elsevier Mosby; St. Louis, MO: 2006. p. 341-350.

Sheridan DJ, Nash KR. Acute injury patterns of intimate partner violence victims. Trauma, Violence & Abuse. 2007; 8:281–289. doi: 10.1177/1524838007303504.

Sprague S, Madden K, Dosanjh S, Godin K, Goslings JC, Schemitsch EH, Bhandari M. Intimate partner violence and musculoskeletal injury: Bridging the knowledge gap in orthopaedic fracture clinics. BMC Musculoskeletal Disorders. 2013; 14(23) doi: 10.1186/1471-2474-14-23.

Taillieu TL, Brownridge DA. Violence against pregnant women: Prevalence, patterns, risk factors, theories, and directions for future research. Aggression & Violent Behavior. 2010; 15:14–35. doi: 10.1016/j.avb.2009.07.013.

Venes, D., Taber, CW. Taber's cyclopedic medical dictionary. 22nd ed.. F.A. Davis; Philadelphia: 2013.

Wingood G, DiClemente R, McCree D, Harrington K, Davies S. Dating violence and the sexual health of black adolescent females. Pediatrics. 2001; 107:E72–E75. doi: 10.1542/peds.107.5.e72. [PubMed: 11331722]

Wu V, Huff H, Bhandari M. Pattern of physical injury associated with intimate partner violence in women presenting to the emergency department: A systematic review and meta-analysis. Trauma, Violence & Abuse. 2010; 11(2):71–82. doi: 10.1177/1524838010367503.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A. Authors Gunnur Karakurt, Vishal Patel, Kathleen Whiting and Mehmet Koyuturk declare that they have no conflicts to report.

B. This research does not consider human subjects.

C. This publication was made possible in part by the Clinical and Translational Science Collaborative of Cleveland, NIH/NCRR CTSA 4UL1TR000439 from the National Center for Advancing Translational Sciences (NCATS) component of the National Institutes of Health and NIH roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH

D. Gunnur Karakurt, Vishal Patel, Kathleen Whiting, & Mehmet Koyutürk, Mining Electronic Health Records Data: Domestic Violence and Adverse Health Effects

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

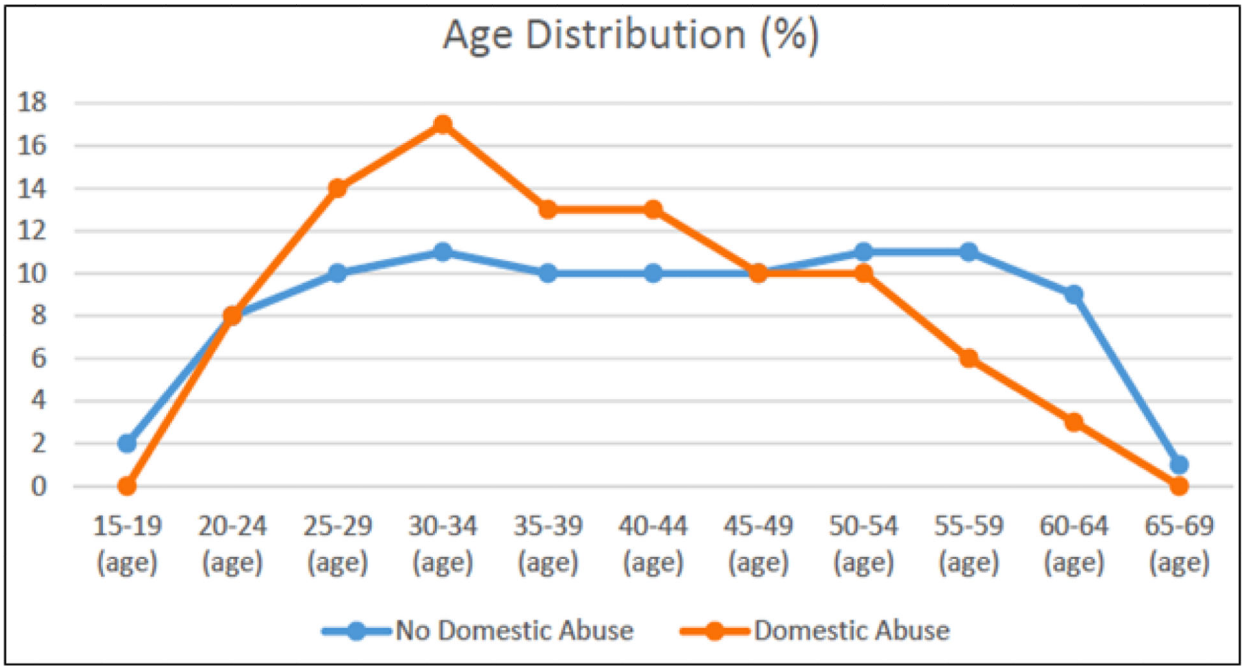

Figure 1.

Age distribution of the patient groups that have/do not have "domestic violence" in the diagnosis field.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

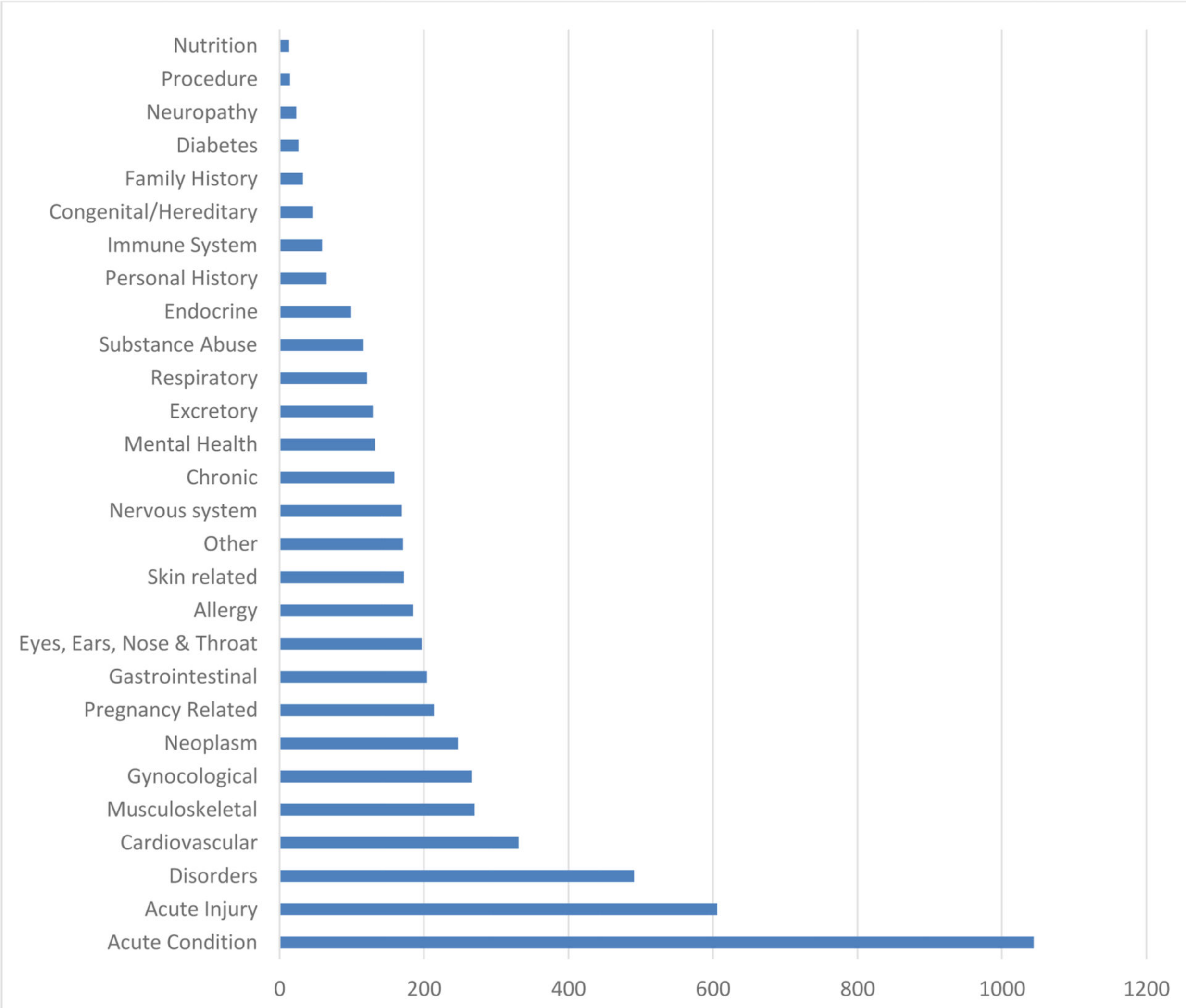

Figure 2.

Distribution of diagnoses that are significantly more frequent at the presence of domestic violence, based on a broad classification of diagnoses. The x-axis represents the number of diagnoses that were classified under the respective disease class on the y-axis. Significance was based on log odds ratio and χ^2 statistic calculations to compare the frequencies of each term in the domestic abuse (DA) versus no domestic abuse (NDA) population, as detailed in the ‘Statistical Analysis’ section.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Ten Most Commonly Observed Terms Based on Odds Ratio and χ^2 Statistics

Term	χ^2	Log Odds Ratio (99% Conf. Int.)	Domestic Abuse Frequency	No Domestic Abuse Frequency
Contusion of soft tissue	31120.92	[3.35, 3.36]	1650	120760
Superficial bruising	30344.49	[3.34, 3.35]	1600	116460
Contusion	24558.15	[2.96, 2.96]	3380	591060
Head and neck injury	21996.67	[2.87, 2.88]	3470	681330
Wound of skin	20349.84	[2.92, 2.93]	1770	203350
Injury of head	19929.99	[2.76, 2.77]	2560	416480
Injury of integument	17637.12	[2.69, 2.70]	2220	353140
Contusion of multiple sites	16649.49	[3.47, 3.49]	620	32360
Superficial injury	16626.38	[2.72, 2.73]	1820	255610
Injury of face	14591.93	[2.72, 2.73]	1470	191350

Note: The reported χ^2 statistics are computed based on the frequencies shown in the table (rounded to the nearest 10 by the database). The intervals shown for log odds ratio represent 99% confidence intervals for the log odds ratio estimated using Monte Carlo simulation to assess the effect of possible bias introduced by the rounding.