Sex Differences in Reported Pain Across 11,000 Patients Captured in Electronic Medical Records

David Ruau,*†‡ Linda Y. Liu,* J. David Clark,†‡ Martin S. Angst,‡ and Atul J. Butte*

*Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, California.
†Department of Anesthesia, Stanford University School of Medicine, Stanford, California.
‡Anesthesiology Service, Veterans Affairs Palo Alto Health Care System, Palo Alto, California.

Abstract: Clinically recorded pain scores are abundant in patient health records but are rarely used in research. The use of this information could help improve clinical outcomes. For example, a recent report by the Institute of Medicine stated that ineffective use of clinical information contributes to undertreatment of patient subpopulations—especially women. This study used diagnosis-associated pain scores from a large hospital database to document sex differences in reported pain. We used de-identified electronic medical records from Stanford Hospital and Clinics for more than 72,000 patients. Each record contained at least 1 disease-associated pain score. We found over 160,000 pain scores in more than 250 primary diagnoses, and analyzed differences in disease-specific pain reported by men and women. After filtering for diagnoses with minimum encounter numbers, we found diagnosis-specific sex differences in reported pain. The most significant differences occurred in patients with disorders of the musculoskeletal, circulatory, respiratory and digestive systems, followed by infectious diseases, and injury and poisoning. We also discovered sex-specific differences in pain intensity in previously unreported diseases, including disorders of the cervical region, and acute sinusitis (P = .01, .017, respectively). Pain scores were collected during hospital encounters. No information about the use of pre-encounter over-the-counter medications was available. To our knowledge, this is the largest data-driven study documenting sex differences of disease-associated pain. It highlights the utility of electronic medical record data to corroborate and expand on results of smaller clinical studies. Our findings emphasize the need for future research examining the mechanisms underlying differences in pain.

Perspective: This article highlights the potential of electronic medical records to conduct large-scale pain studies. Our results are consistent with previous studies reporting pain differences between sexes and also suggest that clinicians should pay increased attention to this idea.

© 2012 by the American Pain Society. Published by Elsevier Inc. All rights reserved

Key words: Electronic medical records, sex differences, pain intensity, data mining.
A majority of clinical studies suggest that women suffer pain more often (higher prevalence) for musculoskeletal, neuropathic, abdominal, and migraine-related conditions. However, results of these studies are not consistent. Clinical pain research of sex differences has mainly focused on studying pain prevalence, contrasting with the vast majority of experimental pain studies documenting sex-related differences in the perceived intensity of pain. Interestingly, genetic polymorphisms associated in human and animal models have identified a striking number of sexual dimorphisms with either male- or female-specific genetic effects or a significant difference between sexes. One meta-analysis of experimental pain studies of sex differences found that the mixed conclusions among various studies could be explained by insufficient study sizes. The authors determined that a minimum sample size of 41 people per group is essential to obtain adequate power (.70) for detecting a sex difference in pain threshold or tolerance.

Thus, although clinical pain studies point toward existing sex-specific differences in pain intensity for some disease classes, the limited number, the size of many of them, and the mixed results are a challenge to drawing conclusions. In this respect, EMR data represent an extremely valuable resource for clinical research. To our knowledge, data from EMRs have not yet been used systematically to study pain associated with diseases. The aim of this study was to use EMR data from Stanford Hospital and Clinics to investigate sex differences in reported pain intensities in a disease-specific manner. We compared reported pain levels in men and women in 47 different diseases, each having at least 41 encounters for both men and women. Our sample contained data for more than 11,000 patients whose pain scores were recorded as part of routine medical care. We identified significant sex-specific patterns in reported pain intensities. Our findings are consistent with results of experimental pain studies and provide robust evidence that women report increased clinical pain compared with men across an array of diseases.

**Methods**

**Ethics Statement**

This study was a retrospective single-center study approved by the IRB at the Stanford Hospital and Clinics and Lucile Packard Children’s Hospital. The study did not require an informed consent. Collected data were de-identified prior to analysis.

**Design Overview**

We used the Stanford Translational Research Integrated Database Environment (STRIDE) as our source of information. We collected de-identified patient EMR data containing a pain score that had been created between January 2007 and September 2010, and retrieved information on 72,773 unique patients. For each patient, we obtained demographic and encounter-specific information (Table 1). No information that could identify a patient was used. We built a MySQL database on a secure server to house the data and perform queries.

**Setting and Participants**

We obtained a total of 161,827 patient-reported pain scores. Scores were recorded during in- or out-patient encounters (hospital or clinic visits). They were obtained from 2 numerical pain scale types: verbal and nonverbal external observer-based. In adults, pain was quantified by asking patients to rate their pain intensity subjectively on an 11-point numerical pain rating scale (scores between 0 and 10) anchored by the words “no pain” and “worst pain imaginable.” We excluded children from our dataset. Pain in children at Stanford Hospital and Clinics is assessed with the 6-point subjective Wong-Baker Faces Pain Rating Scale (scores 0–5). If patients were unable to communicate, pain was assessed by trained personnel using the 11-point face, Legs, Activity, Cry, Consolability (FLACC) scale. Additionally, some patient pain scores were recorded but the information about the pain scale used was missing or indicated both a verbal and a nonverbal assessment, referred to in Table 1 as “mixed” pain scale type.

Stanford Hospital and Clinics uses the International Statistical Classification of Diseases and Related Health Problems codes (ICD-9) to identify diagnoses. The ICD-9 classification is a hierarchical classification of diseases and symptoms, divided into 18 major categories called chapters that are themselves subdivided into sections.
To generate a comprehensive overview of the wide spectrum of diagnoses (318), we clustered individual diagnoses into sections. Sections are defined by grouping diagnoses sharing the first 3 digits of an ICD-9 code. This method reduced the original number of diagnosis codes from 318 to 249 and has been used previously.\(^3\) This number was further reduced to 47 when filtered for a minimum of 41 encounters (Table 2 and Supplementary Table 1). Because our focus was on diseases, we excluded procedural and external codes (V and E codes) and codes representing signs and symptoms (780–799). For example, we removed codes for “headache,” but kept those for “migraine.”

Many patient files contained multiple diagnosis codes; 1 code was usually designated as the primary diagnosis. We included data with a primary diagnosis and at least 1 recorded pain score. A patient may be represented in our data multiple times if s/he came in for multiple visits and had pain scores associated with a primary diagnosis. To obtain pain estimates that were not confounded by subsequent medications or procedures targeting pain alleviation, we used only the first pain score associated with an encounter. We determined the first pain score using the time stamp attached to it. To filter out erroneous conclusions from undersized cohorts, we considered only diagnoses with 41 or more patient encounters per sex, following the guideline noted in the introduction regarding the minimum sample size to use when assessing pain thresholds.\(^3\)

**Statistical Analysis**

We averaged pain scores for males and females for each section with at least 41 encounters per sex. We analyzed sex differences in pain scores by performing 2-sample t-tests with unequal sample size and unequal variances (similar to the Welch 2-sample t-test). We used the Benjamini-Hochberg method to correct for multiple hypothesis testing.\(^2\) Alongside the \(P\) values, we computed the Cohen’s \(d\) effect sizes as follows
\[
d = \frac{\bar{x}_m - \bar{x}_f}{s_{pooled}}\]

where \(\bar{x}_m\) and \(\bar{x}_f\) are the average pain scores for male and female, respectively.\(^2\) Positive values indicate higher male average pain, and negative values indicate higher female average pain. The result is unit free and Cohen proposed benchmark values for what is to be considered small, medium, and large effects (\(d = .2, .5, .8\), respectively).\(^5\) To increase our confidence and to identify trends in the disease-specific sex differences in pain, we used incremental minimum encounter thresholds for each sex. These thresholds were 41–68 and >68 encounters per disease, per sex; 68 being the median of the number of encounters per diagnosis for male patients. We performed linear

**Table 2. Diagnoses With Significant Differences in Reported Pain Between Sexes**

<table>
<thead>
<tr>
<th>Diagnosis Section (ICD Section Code)</th>
<th>ICD-9 Chapter</th>
<th>FEMALE</th>
<th>MALE</th>
<th>CORRECTED P VALUE (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other and unspecified disorders of back (724)</td>
<td>Musculoskeletal system/ connective tissue</td>
<td>6.03 (2.34)</td>
<td>5.53 (2.51)</td>
<td>2.62E-05 (−.21)</td>
</tr>
<tr>
<td>Osteoarthritis and allied disorders (715)</td>
<td>Musculoskeletal system/ connective tissue</td>
<td>5.37 (2.37)</td>
<td>4.56 (2.33)</td>
<td>2.62E-05 (−.34)</td>
</tr>
<tr>
<td>Other and unspecified disorders of joint (719)</td>
<td>Musculoskeletal system/ connective tissue</td>
<td>5.41 (2.31)</td>
<td>4.88 (2.35)</td>
<td>3.74E-05 (−.22)</td>
</tr>
<tr>
<td>Other disorders of soft tissues (729)</td>
<td>Musculoskeletal system/ connective tissue</td>
<td>5.73 (2.52)</td>
<td>4.98 (2.42)</td>
<td>3.01E-04 (−.30)</td>
</tr>
<tr>
<td>Other disorders of synovium, tendon, and bursa (727)</td>
<td>Musculoskeletal system/ connective tissue</td>
<td>4.80 (2.37)</td>
<td>3.91 (1.91)</td>
<td>4.43E-03 (−.41)</td>
</tr>
<tr>
<td>Other and unspecified arthropathies (716)</td>
<td>Musculoskeletal system/ connective tissue</td>
<td>6.00 (2.10)</td>
<td>4.93 (2.33)</td>
<td>6.98E-03 (−.49)</td>
</tr>
<tr>
<td>Human immunodeficiency virus [HIV] disease (042)</td>
<td>Infectious and parasitic diseases</td>
<td>6.33 (2.09)</td>
<td>5.1 (2.48)</td>
<td>7.08E-03 (−.51)</td>
</tr>
<tr>
<td>Other disorders of cervical region (723)</td>
<td>Musculoskeletal system/ connective tissue</td>
<td>5.78 (2.26)</td>
<td>5.22 (2.26)</td>
<td>.010 (−.25)</td>
</tr>
<tr>
<td>Essential hypertension (401)</td>
<td>Circulatory system</td>
<td>5.96 (2.30)</td>
<td>4.99 (2.41)</td>
<td>.012 (−.41)</td>
</tr>
<tr>
<td>Acute sinusitis (461)</td>
<td>Respiratory system</td>
<td>5.37 (2.17)</td>
<td>4.46 (2.18)</td>
<td>.017 (−.42)</td>
</tr>
<tr>
<td>Sprains and strains of knee and leg (844)</td>
<td>Injury and poisoning</td>
<td>5.16 (2.55)</td>
<td>4.02 (2.27)</td>
<td>.024 (−.47)</td>
</tr>
<tr>
<td>Other hernia of abdominal cavity w/o obstruction or gangrene (553)</td>
<td>Digestive system</td>
<td>6.09 (1.96)</td>
<td>4.96 (2.47)</td>
<td>.033 (−.50)</td>
</tr>
<tr>
<td>Complications peculiar to certain specified procedures (996)</td>
<td>Injury and poisoning</td>
<td>6.00 (2.37)</td>
<td>4.98 (2.37)</td>
<td>.046 (−.43)</td>
</tr>
<tr>
<td>Disorders of muscle, ligament, and fascia (728)</td>
<td>Musculoskeletal system/ connective tissue</td>
<td>5.58 (2.65)</td>
<td>4.67 (2.21)</td>
<td>.046 (−.37)</td>
</tr>
</tbody>
</table>

**NOTE.** Pain differences were declared significant for corrected T-test \(P\) values <.05. \(P\) values were corrected for multiple hypothesis testing by the Benjamini-Hochberg method. \(d\): Cohen’s \(d\) effect size.
regression analysis using the male and female average pain score for the 2 diagnosis groups defined by the threshold.

In order to facilitate a higher-level understanding of pain differences between the sexes, we grouped our diagnosis sections into 1 of the 16 chapters represented in our disease list. This process allowed us to place our disease sections into broader disease categories, such as diseases of the circulatory system, diseases of the musculoskeletal system, and neoplasms. We assigned each diagnosis section to a chapter and then tested for overrepresentation of particular chapters among diagnoses with and without sex-specific pain differences (Table 2). The enrichment analysis for each chapter was performed using a hypergeometric test. We used the Benjamini-Hochberg method to correct for multiple hypothesis testing across the 16 chapters.

We used violin plots to display pain score distributions by sex and age. The width of the violin plot was proportional to the number of encounters at each pain level. We also placed boxplots depicting the 25th through 75th percentiles inside the violin plots. Significance P values between matched age-sex groups were computed using the Welch 2-sample t-test with unequal variance. All analyses were performed using R statistical software.15

Results

The cohort represents a broad sampling of patients routinely but not specifically evaluated for pain. Table 1 summarizes the demographics and pain characteristics of these patients. There were 56.2% females and 43.8% males, and all age groups were represented except pediatric patients (which made up <2% of the cohort). The majority of the population was white (51%) and about a quarter of the patients were of unknown ethnicity.

Fig 1 shows average male and female pain scores for the 47 diagnosis sections, plotted as ordered pairs. The graph shows that on average, women reported higher pain scores for the majority of sections—the majority of the points are above the diagonal line representing equal pain levels between sexes. Pain scores for women were higher in 18/25 (72%) sections with 41 to 68 patient encounters per sex group. This proportion reached 21/22 (95%) when the minimum number of encounters per diagnosis was increased to 69 or more. Several diagnosis sections (above the black dashed line) exhibited a pain score difference of greater than 1 unit between the sexes for some of the diseases and disorders listed in Table 2. Thirteen diseases had a small effect size comprised between –.2 and –.5. Two additional sections had a medium effect size, with d >.5. The 2 sections were human immunodeficiency virus, and other hernia of abdominal cavity sections having 41 or more encounters. There were significant differences between male and female pain scores in 14 of 47 diagnoses after multiple hypothesis correction (Table 2). Results for all 47 diagnoses are shown in Supplementary Table 1.

The slopegraph in Fig 2 shows the difference in average pain score for diagnoses with at least 69 encounters per group. Average pain scores were typically higher for women. We also observed remarkable agreement between the sexes of overall ranking of diagnoses by average pain intensity, with a significant Spearman rank correlation of 62.2% (P = 3.1 × 10\(^{-6}\)).

Several studies have reported pain differences between the sexes for some of the diseases and disorders listed in Table 2. Among them, osteoarthritis & allied disorders and arthropathies have been widely studied and both higher prevalence and pain intensity in women have been reported.1,17,34 Similarly, we found that women with unspecified disorders of the back (especially low back pain) or sprain and strain of knee and leg consistently reported higher pain intensity scores than men, in line with other studies.3,33

Effect sizes were all negative for diseases with significant sex differences (range –.51 to –.21) indicating a higher average pain value for women than for men (Table 2). Thirteen diseases had a small effect size comprised between –.2 and –.5. Two additional sections had a medium effect size, with d >.5. The 2 sections were human immunodeficiency virus, and other hernia of abdominal cavity
w/o obstruction or gangrene. Lastly, 8/14 of the diseases from Table 2 belonged to the musculoskeletal system and connective tissue ICD-9 chapter. This category was significantly overrepresented among diseases with pain differences between the sexes ($P = .02$).

Table 2 also shows evidence for sex-specific differences in reported pain intensity for diseases not previously studied in this context. These diseases included disorders of cervical region (such as torticollis or more generally pain in neck), and acute sinusitis ($p = .01$ and $.017$, respectively). To further explore these findings, we stratified pain scores by age, as age is thought to influence experimental pain sensitivity (Supplementary Fig 1). In several cases, male and female average pain scores remained significantly different between age groups.

**Discussion**

We analyzed a large repository of EMR data and found elevated clinical pain scores in women in multiple diseases and disease categories. Our results are consistent with previous findings in the clinic setting, as well as findings in experimental pain studies. Our study also suggests that women report increased pain intensity for acute inflammatory conditions including sinusitis and arthropathies. Sinusitis is a novel finding. Acute inflammatory conditions can be related to postoperative pain studies reporting a tendency toward greater pain intensities among women.

We believe that our broad, inclusive cohort and data-driven approach were instrumental in identifying significant sex differences in the reported intensity of pain. EMR data are useful for studying pain in a large group of individuals and across diseases. Additionally, our use of the ICD-9 hierarchical structure allowed us to group diagnoses into broad classes and better interpret the disease and disorder-associated findings. Enrichment of sex-specific pain differences in general categories may help elucidate common processes among diseases in a single category. This kind of analysis requires large databases and may provide a unique approach to understanding functional differences between the sexes.

Although the use of EMR data is powerful in certain ways, it is limited in others. Limitations of this study

---

**Figure 2.** Slopegraph representing the average pain score change between men and women for 22 diagnoses with >68 encounters. Abbreviations: dis, disorder; dx, disease.
include incomplete documentation. For example, 1 common problem we encountered was missing diagnoses or no primary diagnosis. The vast majority of records also did not identify the pain scale used (subjective versus observer-dependent rating), making it impractical to evaluate the influence of the different scales on the sex differences in pain. Lab results or prescriptions may also not be recorded in a timely fashion or may be missing. Although this problem was not a focus of this study, it represents a major obstacle for associating pain levels with a patient’s physiological state. Additionally, a patient’s full health record may be scattered between multiple medical centers, some of which may not use EMR systems. The implementation of integrated reporting methods will alleviate this problem.\(^2\) While obtaining de-identified data from a single institution is becoming more straightforward than in the past, gaining access to large-scale EMR data from multiple institutions is still a serious challenge.\(^3\) Ideally, accessing national-level EMR data repositories, such as the Veterans Affairs national registry, would enable nationwide integrative studies. However, sex biases within these cohorts, including Veterans Affairs cohorts, would have to be addressed. For example, while the percentage of women in the veteran population is growing, it is still very low, with a projected estimate of 8.3% in 2011.\(^4\)

There are also specific limitations to using EMR data to study pain. Sociocultural and psychological factors are not recorded in the EMR and may confound findings. For example, women have a higher tendency to seek medical care than men,\(^5\) and patients may report different pain levels based on the gender of the physician evaluating them.\(^6,7,12,20\) We also used only the first pain measurement recorded during an encounter to avoid potential bias introduced by physician-prescribed analgesic treatment, but in most cases, we had limited knowledge concerning the use of over-the-counter pain relievers prior to the encounter. Nevertheless, we believe that working with large-scale human data from a broad population is extremely valuable for corroborating the findings of previous studies. It was also useful for identifying sex-related differences for conditions not previously studied in this context.

While overall pain scores were higher in women compared with men on average, we cannot speak towards the causes underlying these differences, which could include hormonal, genetic, or psychological factors. Sex and gender differences in pain have been increasingly studied in the past 20 years. Studies have found that sex hormones may influence pain sensitivity.\(^6,8-10,13\) In this regard, the stage of the menstrual cycle should be considered when recording female pain scores.

Unfortunately, this information was not available for our analysis. Furthermore, polymorphisms with sex-specific effects have been associated with sex-specific differences in reported pain intensity.\(^7,18\) In addition, psychological factors clearly influence the perception of pain. For example, the effects of a patient’s mood and lack of personal coping strategies, such as a tendency to catastrophize, have been well studied.\(^12,20,36\)

Clinical pain studies of sex differences have mainly focused on pain prevalence, and to a lesser extent on pain intensity differences. The latter has been the topic of experimental pain studies. Sex differences in the intensity of disease-related pain among men and women are not as well studied and remain a somewhat controversial issue due to contradictory results. Our results suggest that women report higher clinical pain intensities than men for at least some disease entities. Best documented by our results are differences of the musculoskeletal system. The small-to-medium effect sizes we observed may partially explain conflicting results from previous clinical studies. To our knowledge there has been no report of effect size estimates for sex difference in clinical pain across multiple diseases, while at least 1 meta-analysis was conducted for sex differences in experimental pain. Riley et al\(^32\) reported small and medium summary weighted Cohen’s d effect sizes for threshold pain, similar to our results. However, a larger effect was observed for pain tolerance measures.

Taken together, our results corroborate the general agreement that women report higher pain intensity in musculoskeletal diseases and also suggest differences in diseases of the respiratory system. Our data support the idea that sex differences exist, and they indicate that clinicians should pay increased attention to this idea. Our findings also support the need for further research in this area, such as studies investigating mechanistic and physiological causes of sex-related differences in pain. For example, 1 approach would be to systematically include female subjects in population studies. Diagnosis-associated pain scores are the most widely used pain measure in the biomedical community. This study shows that they can be exploited to shed light on a basic but crucial question: does pain differ between males and females?

**Acknowledgments**

David Ruau and Linda Y. Liu contributed equally.

We thank Susan Weber, Sylvia Tam, Gomathi Krishnan, and Todd Ferris for data extraction; Alex Skrenchuk from Stanford University for computer support; and Valerie Natale for useful comments.

### References


36. The Journal of Pain