11-6-2021

Development of an algorithm to assess unmeasured symptom severity in gynecologic care

Kemi M. Doll

Annie Green Howard

Till Stürmer

Tim Carey

Wanda K. Nicholson

See next page for additional authors

Recommended Citation

Doll KM, Howard AG, Stürmer T, Carey T, Nicholson WK, Carey E, Myers E, Nerenz D, and Robinson WR. Development of an algorithm to assess unmeasured symptom severity in gynecologic care. Am J Obstet Gynecol 2021.

Authors

Kemi M. Doll, Annie Green Howard, Till Stürmer, Tim Carey, Wanda K. Nicholson, Erin Carey, Evan Myers, David R. Nerenz, and Whitney R. Robinson

GYNECOLOGY

Development of an algorithm to assess unmeasured symptom severity in gynecologic care

Kemi M. Doll, MD, MSCR; Annie Green Howard, PhD; Till Stürmer, MD, MPH, PhD; Tim Carey, MD, MPH; Wanda K. Nicholson, MD, MPH, MBA; Erin Carey, MD, MSCR; Evan Myers, MD; David Nerenz, PhD; Whitney R. Robinson, PhD, MSPH

BACKGROUND: Healthcare disparities research is often limited by incomplete accounting for differences in health status by populations. In the United States, hysterectomy shows marked variation by race and geography, but it is difficult to understand what factors cause these variations without accounting for differences in the severity of gynecologic symptoms that drive the decision-making for hysterectomy.

OBJECTIVE: This study aimed to demonstrate a method for using electronic health record–derived data to create composite symptom severity indices to more fully capture relevant markers that influence the decision for hysterectomy.

STUDY DESIGN: This was a retrospective cohort study of 1993 women who underwent hysterectomy between April 4, 2014, and December 31, 2017, from 10 hospitals and >100 outpatient clinics in North Carolina. Electronic health record data, including billing, pharmacy, laboratory data, and free-text notes, were used to identify markers of 3 common indications for hysterectomy: bulk symptoms (pressure from uterine enlargement), vaginal bleeding, and pelvic pain. To develop weighted symptom indices, we finalized a scoring algorithm based on the relationship of each marker to an objective measure, in combination with clinical expertise, with the goal of composite symptom severity indices that had sufficient variation to be useful in comparing different patient groups and allow discrimination among severe symptoms of bulk, bleeding, or pain.

RESULTS: The ranges of symptom severity scores varied across the 3 indices, including composite bulk score (0–14), vaginal bleeding score (0–44), and pain score (0–30). The mean values of each composite symptom severity index were greater for those who had diagnostic codes for vaginal bleeding, bulk symptoms, or pelvic pain, respectively. However, each index demonstrated a variation across the entire group of hysterectomy cases and identified symptoms that ranged in severity among those with and without the target diagnostic codes.

CONCLUSION: Leveraging multisource data to create composite symptom severity indices provided greater discriminatory power to assess common gynecologic indications for hysterectomy. These methods can improve the understanding in healthcare use in the setting of long-standing inequities and be applied across populations to account for previously unexplained variations across race, geography, and other social indicators.

Key words: electronic health record, health equity, hysterectomy, leiomyoma, quality of life

Introduction

Efforts to measure patient-reported symptoms are essential to better patient care and to identify the drivers of unexplained racial differences in symptomatology and treatment. Important medical indications for treatment may not be well measured, or such measurements can be biased by race or other social factors because of how they were originally constructed.[1,2]

Hysterectomy for benign disease is both common and marked with inequalities by race, ethnicity, insurance status, and geography.[3–11] For the common symptoms of uterine bulk, vaginal bleeding, and pelvic pain, there are other treatments available, and it is the severity of the symptom alongside the failure of previous medical treatments that drive the appropriateness of hysterectomy. However, within electronic health records (EHRs), there is a lack of standardization of assessing symptom severity and no commonly used patient-reported outcome measures for these symptoms.

In a context in which decision-making is informed by patient-reported symptoms, quality-of-life effects, and shared decision-making tools,[12–15] diagnostic codes and laboratory values alone are insufficient to account for differences in patient symptom severity and have failed to explain marked racial variations.[16] In addition, the cumulative effects of symptoms over time or treatments previously tried and failed to go unmeasured. Therefore, although several studies have documented variations in hysterectomy use, it has remained unclear to what extent such variations represent differences in clinical indication, symptom severity, patient preferences, or biases within the healthcare system.

Here, we demonstrated an approach to characterize gynecologic symptom severity in a racially and socioeconomically diverse sample of >1900 premenopausal women treated with hysterectomy. Our objective was to use data from administrative billing and the EHR from 10 hospitals and follow a rigorous, multistep process to construct symptom severity indices for the 3 most prominent gynecologic indications for hysterectomy: bulk (pressure) symptoms from uterine enlargement, vaginal bleeding, and pelvic pain.[9,17]

Cite this article as: Doll KM, Howard AG, Stürmer T, et al. Development of an algorithm to assess unmeasured symptom severity in gynecologic care. Am J Obstet Gynecol 2021;XX:x.ex–x.ex.

0002-9378/$36.00

https://doi.org/10.1016/j.ajog.2021.11.020

Click Supplemental Materials under article title in Contents at ajog.org

AJOG at a Glance

Why was this study conducted?

We created symptom severity indices from structured and free-text electronic health record data for the most common indications for benign hysterectomy to improve the understanding of drivers of disparity. Traditional comorbidity indices have limited use in younger populations and when outcomes of interest are in treatment choice and not morbidity, mortality, or readmission. Hysterectomy, a surgery with striking variation by race and geography, is driven by gynecologic symptom severity, which is currently unmeasured in most population research.

Key findings

For the 3 most common indications for hysterectomy—vaginal bleeding, bulk symptoms (pressure), and pelvic pain—groups can now be compared with a measure that incorporates diagnoses, laboratory values, imaging data, and patient-reported symptoms. These composite symptom severity indices were highly associated with clinically relevant objective measures, and we were able to identify severe cases not noted by diagnostic codes alone.

What does this add to what is known?

Comprehensive multisource symptom severity indices can now be used as control variables or proxies for quality of life in health system–derived data in gynecology research. Accounting for the degree of symptom severity was particularly important for investigating unexplained inequities in care.

Materials and Methods

Data sources

The Carolina Data Warehouse for Health (CDW-H) is a searchable federation of electronic health information and administrative data from the University of North Carolina (UNC) Health system, with information from the 10 hospitals and hundreds of affiliated practices. We queried the CDW-H for structured clinical data and supplemented the structured data with free text and images from the EHRs that were captured by a team of professional abstractors.

Cohort identification using Carolina Data Warehouse for Health

As part of a larger study to examine determinants of racial disparities in premenopausal hysterectomy, those eligible for the cohort included North Carolina residents aged 18 to 44 years who underwent hysterectomy for benign (noncancer-related) disease between April 4, 2014, and December 31, 2017. The upper age limit of 44 years was chosen as a conservative cutpoint to identify premenopausal women with a high degree of specificity, as menopausal status is a major determinant of surgical decision-making and <5% of women undergo natural menopause before the age of 45 years.[18] Administrative billing codes were used to identify all hysterectomies performed (Supplemental Digital Content 1) during the eligible date range, including an International Classification of Diseases, Ninth Revision (ICD-9) to International Classification of Diseases, Tenth Revision (ICD-10) crosswalk to ensure full capture. Included sites had to have implemented Epic at least 180 days before the patient's surgery. Women were excluded if they were pregnant at the time of surgery; were not a North Carolina resident; had previous or active breast, ovarian, uterine, or cervical cancer diagnoses; or had cancers with treatment plans that may involve hysterectomy (bladder, anal, or colorectal). The average follow-back time for the analysis sample was 691 days (standard deviation of 345 days).

Capture of structured data: Carolina Data Warehouse for Health

Sociodemographic, clinical, and laboratory data were collected from the CDW-H. For sociodemographic factors, we captured date of birth and a 6-level race variable (White, African American or Black, Asian, American Indian or Alaska Native, other, and refused or unknown), Hispanic or Latin ethnicity (yes or no) from Epic, age, height and weight, marital status, home address, and insurance at the time of hysterectomy. For clinical information, we captured the date of surgery; all physician-billed and hospital-billed diagnostic and procedure codes associated with the hysterectomy encounter; all hemoglobin (Hgb) values, blood transfusions, imaging procedures, gynecologic well-care visits, prescriptions for pain medication, emergency department visits, and hospital admissions up to 12 months before surgery; and all diagnostic codes at the time of hysterectomy for the primary symptoms of interest: vaginal bleeding, pelvic pain, and bulk symptoms. Several codes were mapped to each symptom to capture all potential symptom reports (Supplemental Digital Content 1).

Capture of unstructured data: electronic health record abstraction

Electronic health record abstraction: overview and rationale

We created an EHR data abstraction tool in RedCAP and accompanying protocol to capture candidate markers of symptom severity: the presence of gynecologic diagnoses, symptom descriptions, and surgeon-reported indication for hysterectomy for up to 12 months before surgery. Candidate markers of symptom severity were based on previous literature[19–23] and expert clinical input from the study team (K.M.D., E.C., E.M., and W.K.N.). In addition to the presence or absence of specific sequela of vaginal bleeding, pelvic pain, and abdominal or pelvic bulk symptoms, we captured healthcare utilization data (emergency department visits, blood transfusions, and opioid and other pain mediation use) and missed days of work or activity. Moreover, we planned an overlap with several data points captured by the CDW-H structured data to capture possible events and services

completed outside of the UNC Health system but documented in healthcare provider notes.

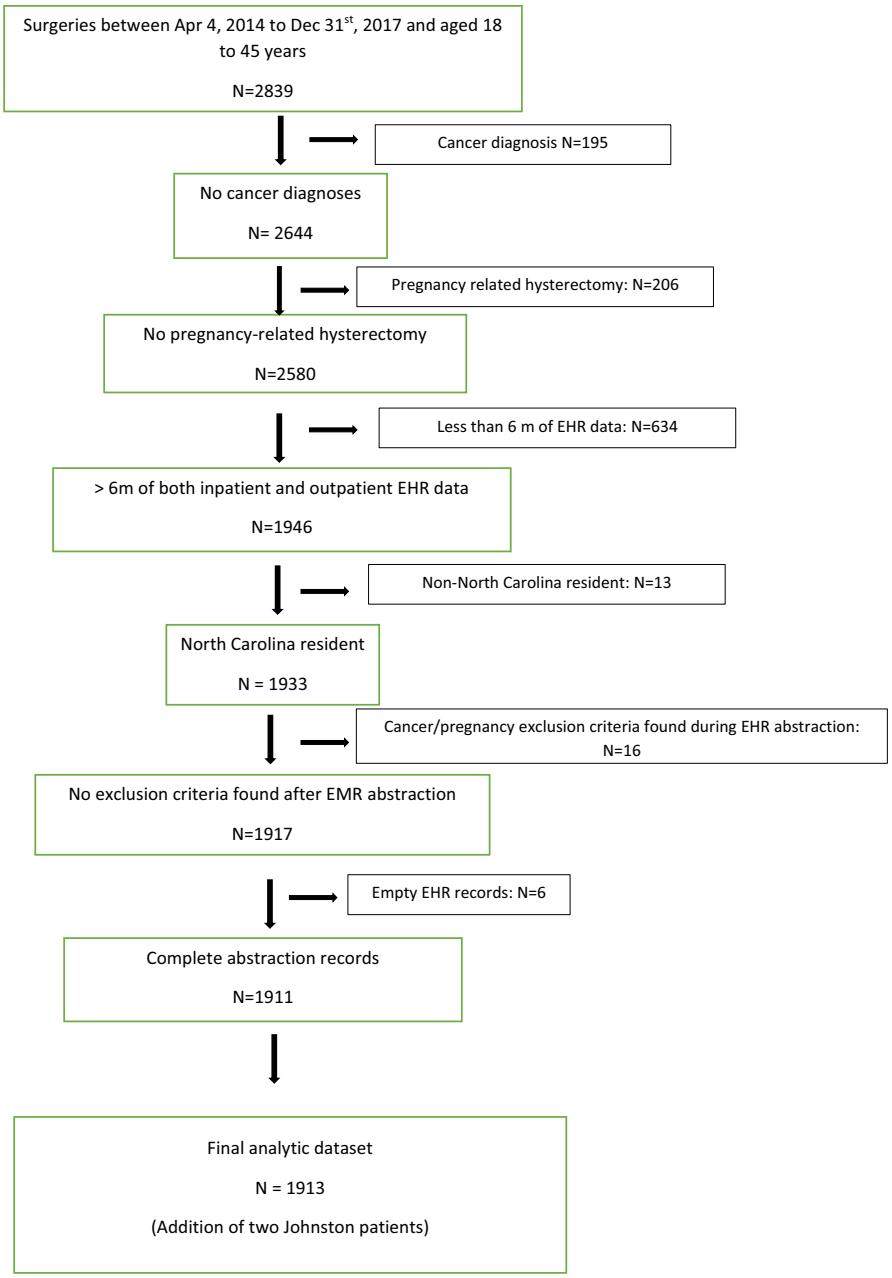
Electronic health record abstraction: pilot study to test and refine abstraction tool

Before finalizing the EHR data abstraction protocol, we conducted a pilot study to assess and refine it. We evenly sampled 52 cases among the hospital sites (a random sample of 5–6 records per site) and the 3 symptoms of interest, identified by relevant diagnostic codes and followed up by the abstraction protocol. On the basis of this experience, the protocol was updated to eliminate repetitive information with the structured data (eg, laboratory values for Hgb and opioid prescriptions). Moreover, we found planned data elements with missing rates too high to be of meaningful use (eg, tampon or pad count, documented in only 4 cases; inability to tolerate an examination, documented in 0 cases) and eliminated them for abstraction efficiency. We did not collect 0 to 10 pain scores as they were not consistently captured in notes or structured data, and they were not specific to any organ site or group. We had 3 possibilities for symptom reports—present, absent, or absent from the record. In the pilot study, only 1 symptom in 1 case was marked as "absent," and the rest were either present or not commented. Therefore, we adjusted the data entry to be "yes" or "absent from the record." Furthermore, the protocol was finalized, and full abstraction was completed.

Electronic health record abstraction: quality control

During the abstraction, we adhered to the protocol guide and kept a corollary log. Overall, these served as active documents with auditable updates based on abstractors' feedback and ongoing quality assurance review. A team of 4 abstractors with >20 years of cumulative experience completed all data abstraction. They could initiate a secondary review for any data ambiguity, and 5% of all records from each site were chosen at random for double abstraction by an abstractor with clinical experience or the abstractor team lead to ensure accuracy.[24] Discrepancies flagged by abstractors (3 of 100 unique records [3%] of data fields that were randomly sampled for quality check) were resolved through group discussion in consultation with clinical leads (K.D. and E.C) followed by appropriate protocol updates. All data were abstracted into REDCap and merged with administrative data for the final analytical data set.

FIGURE 1
Cohort identification



North Carolina residents treated with hysterectomy for benign, nonemergent indications using EHR data from a large healthcare system in the Southern United States, 2014–2017.

EHR, electronic health record; *EMR*, electronic medical record.

Doll et al. Symptom severity indices in gynecology. Am J Obstet Gynecol 2021.

TABLE 1
Descriptive characteristics of individuals between ages 18 and 44 years treated with hysterectomy in a large not-for-profit health system in the Southern United States, 2014–2017

Characteristic	Variable	n (%) or median (IRQ)
Race and ethnicity	Non-Hispanic White	1063 (56)
	Non-Hispanic African American or Black	580 (30)
	Non-Hispanic Asian	23 (1)
	Non-Hispanic American Indian or Alaska Native	17 (1)
	Hispanic	162 (8)
	Other	31 (2)
	Unknown or refused	37 (2)
Insurance status	Tricare	56 (3)
	Self-pay	149 (8)
	Private insurance	1375 (72)
	Medicare	70 (4)
	Medicaid	233 (12)
	Agency	30 (2)
Hospital type	Community	1047 (55)
	Rural	5 (0)
	Teaching	861 (45)
Year of surgery	2014	81 (4)
	2015	479 (25)
	2016	528 (28)
	2017	825 (43)
Age at hysterectomy		39 (19–45)
Uterine size		281 (20–7031)
Bulk diagnosis code[a] at surgery (DX)		366 (19)
Vaginal bleeding diagnosis code at surgery (DX)		1288 (67)
Pain diagnosis code at surgery (DX)		307 (16)

DX represents the administrative billing codes (ICD-9, ICD-10, and CPT).

CPT, Current Procedural Terminology; *ICD-9*, International Classification of Diseases, Ninth Revision; *ICD-10*, International Classification of Diseases, Tenth Revision.

[a] A complete list of all codes can be found in Supplemental Digital Content 1.

Doll et al. Symptom severity indices in gynecology. Am J Obstet Gynecol 2021.

Electronic health record abstraction: categorizing free text captured by abstractors

Although most fields in the REDCap abstraction tool forced data to be recorded in a structured format (ie, numerical, yes or absent), some fields allowed abstractors to record free text when they were unsure if the default options applied. All free-text entries were reviewed and either recoded into existing abstraction categories (eg, "pelvic floor tension myalgia" recoded into existing "pelvic pain" diagnosis) or used to create new categories not previously identified. Moreover, abstracted free-text data were used to reapply the exclusion criteria for cancer and pregnancy that administrative code definitions missed.

Construction of weighted indices for symptom severity

Our goal was to create a composite index for each symptom—bleeding, bulk, and pain—that was composed of appropriate symptom markers, weighted by their relative severity. First, we created histograms and descriptive tables for all symptom severity candidate markers—gynecologic diagnoses, symptom descriptions, and surgeon-reported indication for hysterectomy. The weights for each marker were first assigned by the investigators' clinical expertise with higher weights for markers more severe and more rare. For example, the presence of a report of "heavy bleeding" was given 1 point, iron supplementation was given 3 points, and history of blood transfusion for nonsurgery-related anemia was given 5 points.

Second, each marker was compared against an objective measure of severity. The objective measures were uterine weight from the pathology report, presence of anemia by laboratory criteria, and presence of opioid prescriptions for bulk, vaginal bleeding, and pelvic pain, respectively. We examined associations by visual inspection of histograms and sample distribution. Markers that were more strongly associated with objective measures of symptom severity were more highly weighted to improve the construct validity of the final composite symptom severity indices. In addition, a few candidate symptom markers (urinary symptoms, constipation, and weight gain) were not meaningfully associated with any of the 3 objective measures, were considered nonspecific, and were not included in the final 3 composite symptom severity indices.

Third, we noted the administrative coding patterns that were indicative of more severe values of the objective criteria and therefore gave them higher weights (points). For example, the presence of the ICD diagnostic code for uterine hypertrophy (which does not specify uterine size) in the year before surgery was uncommon and associated

TABLE 2
Presence of candidate markers of symptom severity in multisource data by presence or absence of symptom-specific diagnostic code at the time of hysterectomy

Candidate markers of symptom severity in multisource data	Diagnostic code present at the time of surgery	Diagnostic code absent at the time of surgery
Bulk symptom markers (data source)	**Bulk diagnostic codes present at surgery (n=366)**	**Bulk diagnostic codes absent at surgery (n=1547)**
Bloating (PT)	15	7
Pelvic pressure (PT)	12	6
Uterine size (50th–75th percentile)	21	26
Nonspecified bulk symptoms (PT)	5	1
Bulk as indication for surgery (MD)	5	1
Uterine size (≥75th percentile)	60	17
Vaginal bleeding symptom markers (data source)	**Vaginal bleeding diagnostic codes present at surgery (n=1288)**	**Vaginal bleeding diagnostic codes absent at surgery (n=625)**
Heavy bleeding (PT)	69	31
Irregular bleeding (PT)	50	17
Heavy bleeding as indication for surgery (MD)	49	3
Irregular bleeding as indication for surgery (MD)	35	5
Period lasts longer than 7 d (PT)	23	10
Lethargia or dizziness (PT)	20	13
Iron use (MD)	33	16
1 ED visit related to menorrhagia (DX)	5	3
Anemia diagnostic code at surgery (DX)	24	11
Anemia as indication for surgery (MD)	9	2
>1 ED visit related to bleeding (DX)	2	0
1 ED visit related to anemia (DX)	4	3
Anemia diagnostic code in the year before surgery (DX)	19	9
History of blood transfusion (MD)	7	3
>1 ED visit related to anemia (DX)	1	0
Anemia (Hgb<10) (LAB)	20	11
Pain symptom markers (data source)	**Pain diagnostic codes present at surgery (n=307)**	**Pain diagnostic codes absent at surgery (n=1606)**
Pelvic pain (PT)	72	42
Painful periods (PT)	51	30
Painful intercourse (PT)	21	9
Tylenol (PHARM)	17	11
NSAID (PHARM)	35	30
Pain as indication for surgery (MD)	44	18
Painful periods as indication for surgery (MD)	30	13
Other pain medication (PHARM)	10	3

Doll et al. Symptom severity indices in gynecology. Am J Obstet Gynecol 2021. *(continued)*

with larger uterine weight compared with such coding only present on the surgery encounter.

Fourth, we chose to include the objective measures in each respective symptom severity index, as these measures are important markers of symptom severity. By excluding them from the final scoring system, we would

TABLE 2
Presence of candidate markers of symptom severity in multisource data by presence or absence of symptom-specific diagnostic code at the time of hysterectomy *(continued)*

Pain symptom markers (data source)	Pain diagnostic codes present at surgery (n=307)	Pain diagnostic codes absent at surgery (n=1606)
Opioid (PHARM)	41	31
At least 1 pain-related ED visit (DX)	17	8
Muscle relaxant (PHARM)	13	5

PT indicates the symptom reported by the patient as recorded in the unstructured physician notes. MD indicates the physician-indicated reason for surgery in the preoperative or operative notes. DX indicates the administrative billing codes (ICD-9, ICD-10, and CPT). LAB indicates the results from laboratory tests. PHARM indicates the prescription information from the pharmacy billing data. Values are expressed as percentages.

CPT, Current Procedural Terminology; *ED*, emergency department; *Hgb*, hemoglobin; *ICD-9*, International Classification of Diseases, Ninth Revision; *ICD-10*, International Classification of Diseases, Tenth Revision; *NSAID*, nonsteroidal antiinflammatory drug.

Doll et al. Symptom severity indices in gynecology. Am J Obstet Gynecol 2021.

be omitting key data and undermeasuring severity in future use.

Refinement of weighted indices for symptom severity

We completed several index iterations to optimize variability while maintaining the logical progression of scoring from least to most severely symptomatic. Supplemental Digital Content 2 demonstrates the step-by-step process for the vaginal bleeding score, as an example. We tested whether the average follow-back time varied according to severity score to assess whether those with higher symptom severity scores were reflecting more follow-back time in which an individual's symptoms could be captured. Available look-back time did not vary by bleeding severity score (P=.61) or pain severity score (P=.86). It tended to be smaller for those with higher bulk scores (717 for ≥75th percentile) than for those with low bulk scores (860 for <25th percentile). This is the opposite of the pattern one would expect if a shorter follow-up time was biasing our estimates by missing symptom markers. At no point during this process were data stratified or analyzed by any demographic factor, including race.

Data analysis: evaluation of weighted indices for symptom severity compared with symptom-specific diagnostic codes

Descriptive statistics were used to assess agreement between each symptom severity index score with the diagnosis code of that symptom at the time of surgery—as evidenced by hospital billing for the procedure. Because of the skewed distribution of the index scores, we tested for a difference in severity score between the presence and absence of diagnostic codes using the Wilcoxon rank-sum test.

This study received approval from the UNC's Institutional Review Board on November 06, 2017 (study identification number: 17-2728).

Results

An initial 2830 individuals were identified through the CDW-H that were aged 18 to 44 years and had a hysterectomy within the study time frame. Of these individuals, 1933 met the inclusion criteria by query of the structured data and underwent EHR abstraction. With the removal of duplicates, empty records, and those whose abstracted information met the exclusion criteria, our final merged analytical cohort was 1913 (Figure 1). About a quarter (425) of the 1913 patients had a look-back period of <365 days. The first percentile was 186 days, whereas the 10% percentile was 252 days. The average follow-back time did not differ by age at the time of surgery (P=.30) or race and ethnicity (P=.15). Moreover, 10 individuals were missing uterine weight, and 153 individuals were missing Hgb laboratory results and were excluded from those respective symptom indices. Cohort characteristics and overall symptom prevalence are reported in Table 1.

Comparison of individual markers of symptom severity with symptom-specific diagnostic codes

Several markers of symptom severity did not fully overlap with the presence of the relevant symptom-specific diagnosis code (Table 2). In some cases, the women had reported symptoms in the EHR that were not captured in diagnostic codes. For example, 31% of women who did not have a diagnostic code for vaginal bleeding reported heavy bleeding as a problematic symptom from free-text information in the EHR. Among those with the diagnostic code for vaginal bleeding, the presence of severe symptom markers (eg, lightheadedness or dizziness [26%] or requiring blood transfusion [7%]) were present among a minority who would be indistinguishable from those with milder symptoms based on codes (Table 2).

Composite symptom severity indices

The final composite symptom severity indices for vaginal bleeding, bulk symptoms, and pelvic pain are shown in Table 3, with the distribution of scores across the study population depicted in Figure 2, along with an overlay of the categorization when using diagnostic codes alone. There were different ranges of scores across the 3 metrics with the composite bulk symptom severity score ranging from 0 to 14, composite vaginal bleeding symptom severity score ranging from 0 to 44, and composite pain symptom severity score ranging from 0 to 30.

Bulk index scores: comparison with bulk diagnostic codes

We found evidence ($P<.0001$) of differences in composite bulk symptom severity score between those with bulk diagnostic code present at surgery (median, 6; interquartile range [IQR], 3–7) and those without the code (median, 0; IQR, 0–1). However, 23 of 251 individuals (9%) with a composite bulk symptom severity index score of ≥6 did not have a bulk diagnosis code reported at surgery. In addition, as an example of the heterogeneity in symptoms, there were a total of 6 different combinations of symptoms that resulted in a composite bulk symptom severity index score of 7. One person with a composite bulk symptom severity score of 7 reported bloating, had a uterine size between the 50th and 75th percentiles, and had a bulk diagnosis code in the year before surgery. Another person had the same score but had bulk not otherwise specified and a uterine size ≥75th percentile. Supplemental Digital Content 3 shows the distributions of symptom presence or absence by composite bulk symptom severity score.

Bleeding index scores: comparison with anemia diagnostic codes

Similarly, composite bleeding symptom severity index scores differed based on diagnostic code of vaginal bleeding at surgery: median composite bleeding symptom severity score was 7 (IQR, 4–12) with the diagnostic code present and 1 (IQR, 0–4) ($P<.0001$) with the diagnostic code absent. As with bulk, there were individuals with high composite bleeding symptom severity index scores who did not have a diagnosis code of bleeding at the time of surgery. Specifically, 106 of 819 individuals (13%) with composite bleeding symptom severity scores of ≥7 did not have a bleeding diagnosis code recorded at surgery. Moreover, we observed variations in symptoms across composite bleeding symptom severity scores (Supplemental Digital Content 4).

TABLE 3
Composite symptom severity index scoring method for bulk, vaginal bleeding, and pelvic pain

Bulk severity index

Points	Symptoms
1	Bloating (PT)
	Pelvic pressure (PT)
	Uterine size (50th–75th percentile) (LAB)
2	Bulk diagnosis code at surgery (DX)
3	Nonspecified bulk symptoms (PT)
	Bulk as indication for surgery (MD)
	Bulk diagnosis code in the year before surgery (DX)
4	Uterine size (≥75th percentile) (LAB)

Vaginal bleeding severity index

Points	Symptoms
1	Vaginal bleeding diagnosis code at surgery (DX)
	Heavy bleeding (PT)
	Irregular bleeding (PT)
	Heavy bleeding as indication for surgery (MD)
2	Irregular bleeding as indication for surgery (MD)
	Vaginal bleeding diagnosis code in the year before surgery (DX)
	Period lasts longer than 7 d (PT)
	Lethargia or dizziness (PT)
3	Iron use (MD)
	1 ED visit related to bleeding (DX)
	Anemia diagnosis code at surgery (DX)
4	Anemia (Hgb<10) (LAB)
	Anemia as indication for surgery (MD)
	>1 ED visit related to bleeding (DX)
	1 ED visit related to anemia (DX)
	Anemia diagnosis code in the year before surgery (DX)
5	History of blood transfusion (MD)
	>1 ED visit related to anemia (DX)

Pain severity index

Points	Symptoms
1	Pelvic pain (PT)
	Painful periods (PT)
	Painful intercourse (PT)
	Tylenol (PHARM)

Doll et al. Symptom severity indices in gynecology. Am J Obstet Gynecol 2021. *(continued)*

TABLE 3
Composite symptom severity index scoring method for bulk, vaginal bleeding, and pelvic pain *(continued)*

Pain severity index	
Points	**Symptoms**
2	NSAID (PHARM)
	Pain as indication for surgery (MD)
	Painful periods as indication for surgery (MD)
3	Pain diagnosis code in the year before surgery (DX)
	Pain diagnosis code at surgery (DX)
	Other pain medication (PHARM)
Points	Symptoms
4	Opioid (PHARM)
	At least 1 pain-related ED visit (DX)
	Muscle relaxant (PHARM)

PT indicates the patient-reported symptom as recorded in the unstructured physician notes. DX indicates the administrative billing codes (ICD-9, ICD-10, and CPT). MD indicates the physician-indicated reason for surgery in the preoperative or operative notes. LAB indicates the results from laboratory tests or pathology reports. PHARM indicates the prescription information from the pharmacy billing data.

CPT, Current Procedural Terminology; *ED*, emergency department; *Hgb*, hemoglobin; *ICD-9*, International Classification of Diseases, Ninth Revision; *ICD-10*, International Classification of Diseases, Tenth Revision; *NSAID*, nonsteroidal antiinflammatory drug.

Doll et al. Symptom severity indices in gynecology. Am J Obstet Gynecol 2021.

Pain index scores: comparison with pelvic pain diagnostic codes

In addition, our composite pain symptom severity score differed by pain diagnosis at or in the year before surgery ($P<.0001$) with a median composite pain symptom severity score of 10 (IQR, 6−16) in those with pain diagnosis present at surgery compared with that of 3 (IQR, 0−7) in those without the diagnosis. Of the 384 individuals, 227 (59%) with composite pain symptom severity scores of ≥10 did not have a pain diagnosis code at the time of surgery. We found evidence, as with the other severity scores, of variation in symptoms among individuals with the same composite pain symptom severity score (Supplemental Digital Content 5).

Comment

Principal findings

We developed 3 composite symptom severity indices for the 3 most common indications for hysterectomy in the United States. High scores on the indices were strongly associated with the presence of diagnostic codes. However, the indices detected a substantial proportion of cases (9%−59%) who were not coded for that symptom via diagnostic codes. Furthermore, the ranges of the indices allowed greater statistical discrimination among those with diagnosis codes present (range, 16%−67%) (Table 1). This is a needed step on the pathway to equity in hysterectomy in the United States.

Clinical implications

Without accounting for racial and ethnic differences in symptom severity, it is impossible to define a reasonable level of racial and ethnic differences in hysterectomy rates and design clinical interventions to achieve that target. We interpreted the discordance between diagnostic codes and symptom severity as arising from the fact that billing codes may just capture what is necessary for payment approval as distinct to the full experience of the patient, an important limitation when they are alone used to define gynecologic symptom status of an individual. Without a greater ability to account for currently undermeasured aspects of symptom severity in gynecology, our assessments of care quality of hysterectomy and other procedures did not take into account a major component of clinical decision-making. With the development of these measures, we have a path forward to evaluate, update, revise, and/or restructure current clinical guidance to ensure equitable distribution of hysterectomy among premenopausal women.

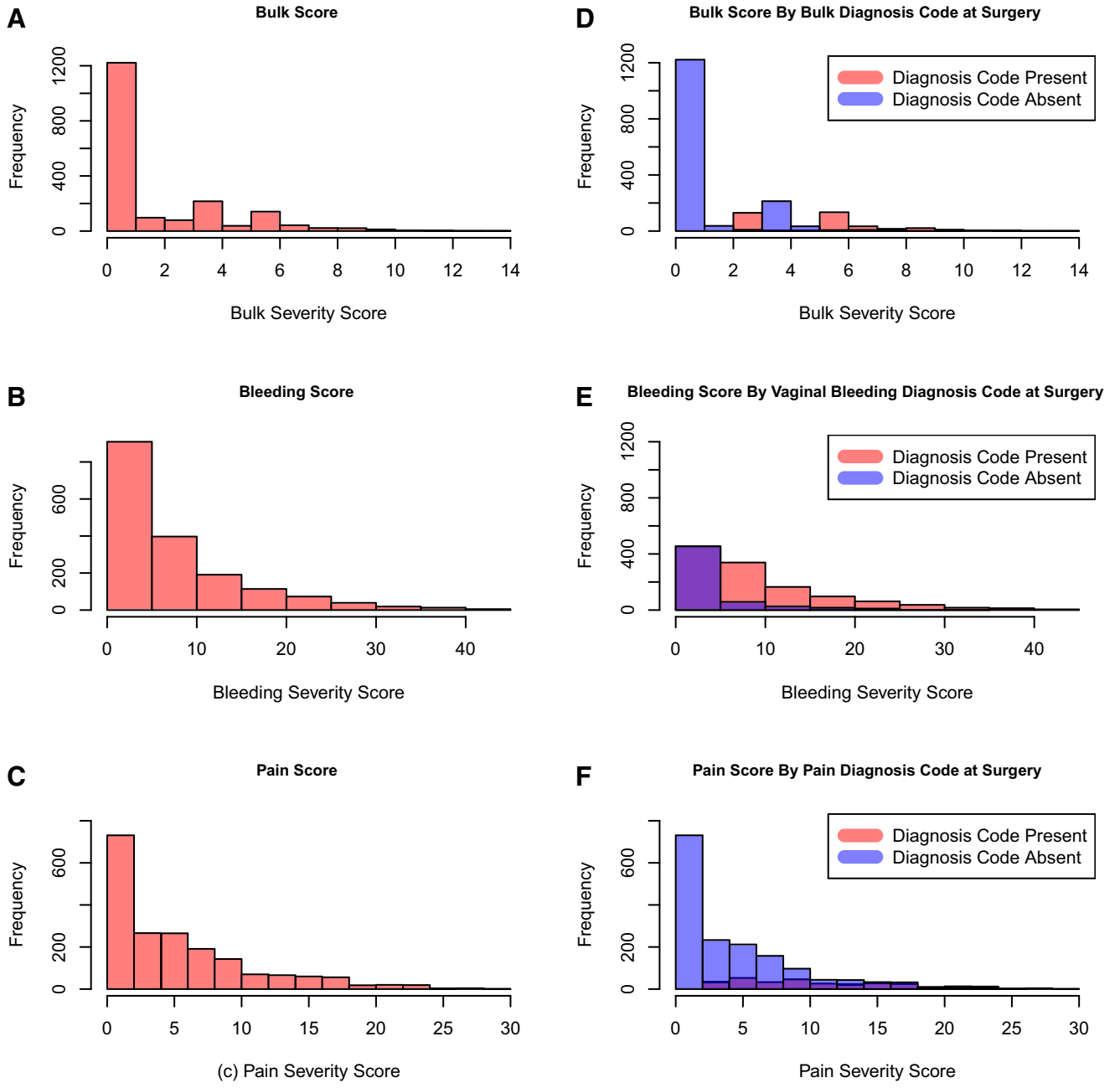
Research implications

This study was a part of a larger multiyear project to investigate the causes of hysterectomy rate disparity in the Southern United States. Given the extensive processes required to develop and refine these indices, this method was presented separately here. In addition, we created these measures blind to any racial or ethnic categorization to minimize bias for the larger study, whose analysis was underway. With these indices, we may move forward to uncover specific etiology and therefore appropriate intervention for longstanding hysterectomy disparities. We can examine how symptom severity may vary by groups of interest, indicating either a clinical need for increased access to and development of uterine-sparing treatments or systemic retraining on how symptom severity may be differentially assessed and acted on. Therefore, these indices were not limited to analyses of racial disparity but can be used in any research that seeks to account for gynecologic symptom burden as an important influential factor in the outcome of interest. This includes other racial and ethnic populations, groups defined by language or nativity, socioeconomic status, insurance status, sexual and gender minorities, and those with disabilities.

Results in the context of what is known

Our current ability to execute interventions to ensure equity in hysterectomy treatment was limited by the fact that existing research does not sufficiently account for patient symptom severity. Without this information, the causes of race- and ethnicity-based differences in treatment have remained contested throughout decades, with little productive movement toward a clear

FIGURE 2
Distribution of bulk, bleeding, and pain composite symptom severity indices



A, Bulk score. **B,** Bleeding score. **C,** Pain score. **D,** Bulk score by bulk diagnosis code at surgery. **E,** Bleeding score by vaginal bleeding diagnosis code at surgery. **F,** Pain score by pain diagnosis code at surgery. Severity scores stratified by diagnosis codes present (*red*) or absent at surgery (*blue*). Please note that the *purple color* results from overlapping data.

Doll et al. Symptom severity indices in gynecology. Am J Obstet Gynecol 2021.

consensus. As we have demonstrated in this analysis using multisource data, the severity of gynecologic symptoms can vary dramatically for the same diagnosis, from nonexistent to disabling. Our final algorithms demonstrated finer ability to differentiate among degrees of severity than the common use of administrative billing diagnostic codes, laboratory values, or pathology records alone.

Strengths and limitations

One limitation with this approach was that continued variation, in the form of differing symptom aspects, still existed within each level of symptom

severity score, which may indicate the limit of transformation of patient experience into quantitative data. Some women with the same symptom severity will have higher scores just because their providers document better. Moreover, we conceptualized the indices as measures that are specific but still have some limitations when it comes to sensitivity—better than codes alone, but not perfect. In addition, our composite score ranges differed (from 14 to 40) across the different composite symptom severity indices. In future use, scaling to a uniform 1 to 100 range would avoid differential weights in predictive analyses. The parent study cohort was defined by the performance of premenopausal hysterectomy, a salient clinical procedure with well-documented but unexplained differences in rates of treatment. These indices are useful for other studies of hysterectomy cohorts; however, they would need to be modified for non-hysterectomy cohorts given the inclusion of administrative data at the time of surgery and postoperative findings of uterine size. We acknowledge that EHR abstraction is expensive in both time and resources; therefore, our method may not be easily scalable. We hope that advancements in natural language processing can make this process more efficient and accessible. These data supported the value and feasibility of developing gynecologic-specific structured reporting of these common symptoms, especially about hysterectomy decision-making, which is increasingly possible with the customization of current major EHR vendors.

Conclusions

When put into use in healthcare disparities research, our proposed severity indices will provide critical tools to account for differences in patient health status. The ability to account for the variation in the severity of patient symptoms will improve work that seeks to identify what factors drive differences in healthcare utilization and outcomes. Patient-centered, health equity literature cannot progress until the field develops better methods for characterizing clinical indications of complex clinical phenomena. ■

Acknowledgments

We acknowledge the editorial assistance of the North Carolina Translational and Clinical Sciences Institute, which is supported by the National Center for Advancing Translational Sciences of the National Institutes of Health, through grant award number UL1TR002489.

References

1. Ledford H. Millions of black people affected by racial bias in health-care algorithms. Nature 2019;574:608–9.
2. Madhusoodanan J. Is a racially-biased algorithm delaying health care for one million Black people? Nature 2020;588:546–7.
3. Callegari LS, Katon JG, Gray KE, et al. Associations between race/ethnicity, uterine fibroids, and minimally invasive hysterectomy in the VA Healthcare System. Womens Health Issues 2019;29:48–55.
4. LoGerfo JP. Variation in surgical rates: fact vs. fantasy. N Engl J Med 1977;297:387–9.
5. Vayda E, Morison M, Anderson GD. Surgical rates in the Canadian provinces, 1968 to 1972. Can J Surg 1976;19:235–42.
6. Jacoby VL, Fujimoto VY, Giudice LC, Kuppermann M, Washington AE. Racial and ethnic disparities in benign gynecologic conditions and associated surgeries. Am J Obstet Gynecol 2010;202:514–21.
7. Bower JK, Schreiner PJ, Sternfeld B, Lewis CE. Black-White differences in hysterectomy prevalence: the CARDIA study. Am J Public Health 2009;99:300–7.
8. Powell LH, Meyer P, Weiss G, et al. Ethnic differences in past hysterectomy for benign conditions. Womens Health Issues 2005;15:179–86.
9. Weiss G, Noorhasan D, Schott LL, Powell L, Randolph JF, Johnston JM. Racial differences in women who have a hysterectomy for benign conditions. Womens Health Issues 2009;19:202–10.
10. Wright JD, Herzog TJ, Tsui J, et al. Nationwide trends in the performance of inpatient hysterectomy in the United States. Obstet Gynecol 2013;122:233–41.
11. Katon JG, Bossick AS, Doll KM, et al. Contributors to racial disparities in minimally invasive hysterectomy in the US Department of Veterans Affairs. Med Care 2019;57:930–6.
12. Committee on Practice Bulletins—Gynecology. Practice bulletin no. 128: diagnosis of abnormal uterine bleeding in reproductive-aged women. Obstet Gynecol 2012;120:197–206.
13. Zandstra D, Busser JAS, Aarts JWM, Nieboer TE. Interventions to support shared decision-making for women with heavy menstrual bleeding: a systematic review. Eur J Obstet Gynecol Reprod Biol 2017;211:156–63.
14. ACOG Committee Opinion No. 557: management of acute abnormal uterine bleeding in nonpregnant reproductive-aged women. Obstet Gynecol 2013;121:891–6.
15. Committee Opinion No. 701: choosing the route of hysterectomy for benign disease. Obstet Gynecol 2017;129:e155–9.
16. Doll KM, Dusetzina SB, Robinson W. Trends in inpatient and outpatient hysterectomy and oophorectomy rates among commercially insured women in the United States, 2000–2014. JAMA Surg 2016;151:876–7.
17. Learman LA, Kuppermann M, Gates E, Gregorich SE, Lewis J, Washington AE. Predictors of hysterectomy in women with common pelvic problems: a uterine survival analysis. J Am Coll Surg 2007;204:633–41.
18. Rosner B, Colditz GA. Age at menopause: imputing age at menopause for women with a hysterectomy with application to risk of postmenopausal breast cancer. Ann Epidemiol 2011;21:450–60.
19. Magnay JL, O'Brien S, Gerlinger C, Seitz C. A systematic review of methods to measure menstrual blood loss. BMC Womens Health 2018;18:142.
20. Matteson KA, Scott DM, Raker CA, Clark MA. The menstrual bleeding questionnaire: development and validation of a comprehensive patient-reported outcome instrument for heavy menstrual bleeding. BJOG 2015;122:681–9.
21. Facchin F, Barbara G, Saita E, et al. Impact of endometriosis on quality of life and mental health: pelvic pain makes the difference. J Psychosom Obstet Gynaecol 2015;36:135–41.
22. Vilos GA, Allaire C, Laberge PY, Leyland N; SPECIAL CONTRIBUTORS. The management of uterine leiomyomas. J Obstet Gynaecol Can 2015;37:157–78.
23. Coyne KS, Margolis MK, Murphy J, Spies J. Validation of the UFS-QOL-hysterectomy questionnaire: modifying an existing measure for comparative effectiveness research. Value Health 2012;15:674–9.
24. Liddy C, Wiens M, Hogg W. Methods to achieve high interrater reliability in data collection from primary care medical records. Ann Fam Med 2011;9:57–62.

Author and article information

From the Department of Obstetrics and Gynecology, University of Washington School of Medicine, Seattle, WA (Dr Doll); Department of Health Services, University of Washington School of Public Health, Seattle, WA (Dr Doll); Department of Biostatistics, Gillings School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC (Dr Howard); Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC (Drs Howard and Robinson); Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC (Drs Stürmer and Robinson); Department of Medicine, University of North Carolina School of Medicine at Chapel Hill, Chapel Hill, NC (Dr Carey); The Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC (Dr Carey); Center for Women's Health

Research, Department of Obstetrics and Gynecology, University of North Carolina School of Medicine, Chapel Hill, NC (Dr Nicholson); Center for Health Promotion and Disease Prevention, University of North Carolina School of Public Health at Chapel Hill, Chapel Hill, NC (Dr Nicholson); Department of Obstetrics and Gynecology, University of North Carolina School of Medicine, Chapel Hill, NC (Dr Carey); Division of Reproductive Sciences, Department of Obstetrics and Gynecology, Duke University School of Medicine, Durham, NC (Dr Myers); Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, MI (Dr Nerenz); and Division of Women's Community and Population Health, Department of Obstetrics and Gynecology, Duke University School of Medicine (Dr Robinson).

Received May 14, 2021; revised Sept. 30, 2021; accepted Nov. 1, 2021.

The authors report no conflict of interest.

The research reported in this publication was supported by the National Institute of Minority Health and Health Disparities of the National Institutes of Health under award number 1R01MD011680. Moreover, this research is supported by the National Institute of Nursing Research of the National Institutes of Health under award number F31NR018786. The funding bodies had no role in the study design, collection, analysis, or interpretation of data; writing of the report; or decision to submit the article for publication.

Corresponding author: Kemi M. Doll. MD, MSCR. Kdoll@uw.edu