Data Quality of Chemotherapy-Induced Nausea and Vomiting Documentation

Melissa Beauchemin1,2, Chunhua Weng3, Lillian Sung4, Adrienne Pichon3, Maura Abbott1, Dawn L. Hershman2, Rebecca Schnall1

1 School of Nursing, Columbia University, New York, New York, United States
2 Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York, United States
3 Department of Biomedical Informatics, Columbia University, New York, New York, United States
4 The Hospital for Sick Children, Toronto, Canada

Address for correspondence Melissa Beauchemin, PhD, RN, CPNP, Columbia University Mailman School of Public Health, 722 West 168th Street, Room 719, New York, NY 10032, United States (e-mail: mmp2123@cumc.columbia.edu).


Abstract

Objective The objective of the study was to characterize the completeness and concordance of the electronic health record (EHR) documentation of cancer symptoms among multidisciplinary health care professionals.

Methods We examined the EHRs of children, adolescents, and young adults who received highly emetogenic chemotherapy and characterized the completeness and concordance of chemotherapy-induced nausea and vomiting (CINV) documentation by clinician type and by the International Classification of Diseases 10th Revision (ICD-10) coding choice.

Results The EHRs of 127 patients, comprising 870 patient notes, were abstracted and reviewed. A CINV assessment was documented by prescribers in 75% of patients, and by nurses in 58% of patients. Of the 60 encounters where both prescribers and nurses documented, 72% agreed on the presence/absence of CINV.

Conclusion Most patients receiving highly emetogenic chemotherapy had a documented assessment of CINV; however, many had incomplete or discordant documentation of CINV from different providers by role, implying the importance of incorporating pragmatic knowledge of EHR documentation patterns among multidisciplinary health professionals for EHR phenotyping and clinical decision support systems directed toward cancer-related symptom management.

Keywords data quality, electronic health records, clinical informatics, UMLS, cancer symptoms

Background and Significance

Clinical decision support (CDS) is the most important “meaningful use” of electronic health record (EHR) data to support workflow for providers and optimize high-quality treatment for patients.1 CDS systems (CDSS) provide an effective mode to equitably improve care delivery by systematically examining EHR or patient-generated data. However, data quality challenges must be considered when analyzing and interpreting EHR data, especially when delivering an EHR-guided intervention.2–6 Incomplete or discordant data hinder both identifying patients from a target cohort (phenotyping) and utilizing the data for delivery of care (decision support). Weiskopf et al defined four prototypes for completeness, from which we focus on the first two: documentation and breadth. Data are considered incomplete if they are not documented (“documentation”) or only partially documented (“breadth”) as occurs when qualifiers and modifiers

received October 29, 2020
accepted after revision March 2, 2021

© 2021. Thieme. All rights reserved.
Georg Thieme Verlag KG,
Rüdigerstraße 14,
70469 Stuttgart, Germany

ISSN 1869-0327.
are missing. Undocumented pertinent negative findings or incomplete description of findings not only can result in incomplete data but are also important factors when developing CDSS, as erroneous recommendations may be provided, potentially risking patient safety or rendering the CDSS futile.3,4

Concordance assesses the agreement between elements in the EHR or between the EHR data and another source.2 EHR data are discordant when there is disagreement between elements, observations, or values that are documented in the EHR and may occur when multiple providers document differing observations due to factors such as timing of assessment, change in patient status, or errors in documentation or clinical assessment.2,3 A baseline assessment of EHR data quality is necessary prior to harnessing it for clinical research purposes.7,8

Symptom management is an integral component of high-quality cancer care because appropriate identification and management of symptoms improves quality of life and reduces adverse effects of disease and cancer-related treatment.9 Chemotherapy-induced nausea and vomiting (CINV) affects up to 80% of cancer patients,10 and high-quality clinical practice guidelines are readily available to classify the emetogenicity of chemotherapy and provide recommendations to prevent and treat CINV.11–15 Nonetheless, these guidelines are not always followed in clinical practice, and provision of guideline-concordant care may occur even less consistently in pediatric oncology settings.16–19 Barriers to providing guideline-concordant care include difficulty identifying at-risk patients, lack of systematic symptom screening, and incomplete knowledge of the most up-to-date guidelines.

CINV is a common and yet preventable cancer symptom, and we hypothesize that EHR data can be used to identify patients at risk of CINV, deliver a CDSS, and improve adherence to clinical practice guidelines.1,2,7,16 Further, given the known disparities in provision of guideline-concordant supportive care,20,21 we hypothesize that a CINV-focused CDSS is a key component to mitigate these disparities. This type of CDSS could feasibly be built from EHR-derived data, including chemotherapy regimen, age, and concomitant medications to deliver the CDSS to the prescribing clinician for prevention of CINV. Further, with integration of patient-reported data, this CDSS might also feasibly deliver guidance to modify a patient’s current CINV regimen based on symptom reports including the presence, severity, frequency, and temporality of symptoms. Finally, the use of billing data, including International Classification of Diseases 10th Revision (ICD-10) codes, might offer utility to conduct retrospective evaluation of patients, symptoms, and outcomes, although, historically, ICD codes have been limited in this capacity.22

Prior to developing and deploying a CDSS, we must identify and characterize EHR data quality. Some prior studies have assessed the completeness of problem lists in EHRs, but little is known about the data quality challenges for symptoms, especially for complex diseases such as cancer.23 We therefore conducted a comprehensive assessment of the completeness and concordance of CINV documentation in the EHR to assess the data quality.

### Methods

#### Data Source

We conducted a retrospective cohort study using the EHR data of pediatric and young adult oncology patients at a large, urban hospital that includes a stand-alone children’s hospital, and inpatient and outpatient cancer clinics all within an National Cancer Institute (NCI) designated Comprehensive Cancer Center. This study was approved by the Institutional Review Board at Columbia University; a waiver of informed consent was granted (IRB- AAAR9461).

#### Sample

All patients age 26 years or younger who received a highly emetogenic chemotherapy regimen (HEC), defined by pediatric and adult clinical practice guidelines available during the study period,11,24 for treatment of cancer from 2016 to 2018 inclusive were included in the analysis. Appendix A provides the list of HECs that qualified for inclusion. Age 26 was chosen as the upper age limit for two reasons: first, there are some data to suggest that provision of guideline-concordant supportive care is less common in pediatric oncology compared with medical oncology settings,16 and we therefore chose an overlapping age group, some of whom would be treated in medical oncology, to see if there were differences in documentation and data quality. Second, 26 years is the cutoff age for young adults to be on parental insurance, and disparities in high-quality care delivery have previously been documented by socioeconomic factors, such as insurance status.25,26 In fact, we recently reported that patients receiving emetogenic chemotherapy with commercial insurance were significantly more likely to receive guideline-concordant antiemetic prophylaxis compared with those with Medicaid (odds ratio [OR]: 2.4; 95% confidence interval [95% CI]: 1.0–4.8).19 All available EHR data from the first clinical encounter for HEC were included for each eligible patient.

#### Procedures

We queried the institutional data request system and identified all patients who received HEC between January 1, 2016 and December 31, 2018. Computerized provider order entry for all chemotherapy was the institutional mode of prescribing chemotherapy, and the Allscripts application was the EHR vendor for the duration of the study period. For each unique patient, we abstracted sociodemographic and clinical variables as well as clinical documentation directly from the EHR into a de-identified dataset. Data abstracted included the following: age, sex, diagnosis, chemotherapeutic regimen, race, ethnicity, location of chemotherapy administration (inpatient/outpatient), clinical setting (pediatric oncology/adult oncology), insurance (Medicaid/Medicare/Commercial), and all clinical documentation from the primary clinical team (from this institution, this includes physicians, nurse practitioners, and registered nurses [RN]) for the duration of the first follow-up encounter following administration of HEC. The EHR sections of interest are provided in Appendix B.

We defined the follow-up encounter for each patient after receiving HEC as either (1) for patients who received HEC in
the inpatient setting, the acute phase of chemotherapy (from start of chemotherapy until 24 hours following completion of chemotherapy), or through discharge from the hospital, whichever came first or (2) for patients who received HEC in the outpatient setting, the subsequent clinical encounter where they were seen at the hospital, or outpatient clinic after receiving the first chemotherapy cycle with HEC. All documented assessments from the follow-up encounter were identified and abstracted from the EHR. Other variables abstracted included whether the documentation was entered by a prescriber (e.g., physician, nurse practitioner) or an RN and the number of documents assessed per patient.

From each follow-up encounter, the abstracted clinical documentation and structured data were assessed for the presence or absence of CINV. The symptom was first coded as “assessed” if there was a specific comment about the presence or absence of nausea, vomiting, or similar terms. This presence of documentation was the primary definition of completeness. ► Table 1 provides the definition of outcome measures. CINV was then coded as “present” if there was any mention in text or discrete structured data point, such as a symptom assessment, that acknowledged the presence of nausea, vomiting, and retching, or if there was a documented emesis event on the RN flow sheet. If CINV was present, we abstracted any text regarding the severity, temporality, or other relevant descriptors of the symptom; this information informed the “breadth” of completeness. We also abstracted three categories of ICD-10 codes from prescriber clinical notes: primary oncologic diagnosis, encounter for or encounter following chemotherapy, and CINV-related codes. The ICD-10 codes are structured data points pulled from the billing section of the EHR into the prescriber notes. One researcher (M.B.) familiar with the EHR system abstracted the data, and two researchers (M.B. and M.A.) independently abstracted and coded 20% of all cases to ensure reliability.

Any disagreements were resolved through discussion and, if necessary, through a third reviewer.

Data Analysis

Following data abstraction from the EHR, descriptive statistics were computed to assess frequency of documented assessment (completeness) and presence/absence of CINV. The proportion of patients for whom the assessment for CINV (present/absent) was concordant between prescriber and RN notes, as well between prescriber documentation and ICD-10 codes for oncologic diagnosis, visit for chemotherapy, and CINV, was calculated. Bivariate analysis was conducted using chi-squared and logistical regression to assess the association between the clinical and demographic variables and the outcomes of interest (i.e., CINV assessment, CINV present as reported by prescriber and RN, ICD-10 codes, and concordance by provider type and ICD-10 codes). Associations with p-value < 0.05 were considered significant. All analyses were conducted using SAS version 9.4 (SAS Institute, Cary, North Carolina, United States).

From the assessments that were coded as CINV present, qualifiers and modifiers were compiled to determine the breadth of the CINV documentation, the secondary definition of completeness. We explored the assessments to determine if they included four domains of symptoms (presence, temporality, frequency, and severity) that are used in both pediatric and adult validated tools to measure CINV. 27,28

Results

EHR Assessment of CINV

We identified 127 patients who received their first cycle of HEC over a 3-year period, defined as an episode. The characteristics of the sample are described in ► Table 2. In total, 127 episodes (one episode per patient) were reviewed, including 390 prescriber notes and 480 RN notes or flow sheets. Prescriber documentation was primarily abstracted from the oncology prescriber note(s), specifically the history of present illness (HPI) section, an unstructured data domain. Nursing documentation was abstracted from six unique locations including flow sheets (structured data), shift assessments (structured data), and nursing-specific notes (unstructured data).

Completeness of Documentation

We identified a documented CINV assessment, defined in our study as completeness, in the EHR for 112 patients (88%). Prescribers documented an assessment for 95 patients (75%), and CINV was present in 61 (64%) patients. Factors associated with an increased likelihood of documenting CINV assessment included chemotherapy regimen and sex. Receiving a cisplatin-based therapy as the HEC regimen was significantly associated with having CINV assessment documented in the EHR (OR: 4.3; 95% CI: 1.2–15.3). Male sex was significantly associated with lower odds of having CINV assessment documented compared with female sex (OR: 0.37; 95% CI: 0.15–0.88). All other factors, including clinical setting, were not significant in bivariate analysis.

Table 1 Outcome measures

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness (prescriber and RN)</td>
<td>If CINV was assessed in the documentation by prescriber or RN (yes/no)</td>
</tr>
<tr>
<td>Completeness (secondary, “breadth”)</td>
<td>Details about qualifiers and modifiers of the CINV symptom (abstracted text)</td>
</tr>
<tr>
<td>Presence of CINV (prescriber and RN)</td>
<td>If completeness = yes, was CINV documented as present?</td>
</tr>
<tr>
<td>Concordance (prescriber)</td>
<td>If assessments by both the prescriber and RN were available, were the two assessments in agreement?</td>
</tr>
<tr>
<td>Concordance (ICD-10)</td>
<td>If the ICD-10 codes agreed with data abstracted from the prescriber’s documentation within the EHR for (a) primary oncologic diagnosis, (b) visit for chemotherapeutic encounter, and (c) presence of CINV symptoms</td>
</tr>
</tbody>
</table>

Abbreviations: CINV, chemotherapy-induced nausea and vomiting; ICD-10, International Classification of Diseases 10th Revision; EHR, electronic health record; RN, registered nurse.
Nursing assessment of CINV was documented in 72 patients (57%). Patient location during the follow-up encounter was significantly associated with RN documented assessment, with those seen in the inpatient setting less likely to have a documented CINV assessment (OR: 0.04; 95% CI: 0.01–0.32). All other factors were not significant in bivariate analysis. Of the 72 patients for whom assessment was documented, 40 (56%) reported the presence of CINV. Twenty-five (63%) of the cases in which CINV was documented as present were from inpatient structured flow sheets reporting the number of emesis episode(s); in these documented assessments, no further descriptors about the symptom were available in the documentation.

**Concordance of Documentation**

In 60 (47%) patient EHRs, both a prescriber and an RN documented a CINV assessment; we compared concordance of the documented assessment in these 60 cases. Of these, 43 (72%) reported concordant assessments. ► Table 3 provides examples of the 17 discordant assessments (28%). Of the 43 discordant assessments, 34 (79%) agreed that CINV was present, and the remaining 9 (21%) agreed that CINV was not present. ► Fig. 1 visually depicts the completeness and concordance of the 127 episodes.

We then examined the concordance of prescriber’s CINV assessment with the ICD-10 billing codes for CINV. Twenty (16%) patient/prescriber encounters included an ICD-10 code for CINV compared with the 61 (64%) cases in which prescribers documented CINV symptoms in unstructured notes. Of the 95 patients in whom CINV was assessed by the prescriber, 47 (50%) of the documented assessments agreed with the ICD-10 code for that encounter. Of the 20 patients in whom the ICD-10 code for CINV was present, 17 (85%) agreed with the provider assessment. ICD-10 codes for chemotherapy encounters were correctly documented in 81 patients (64%), whereas ICD-10 codes for primary oncologic diagnosis were correctly documented in 100% of the 127 patients in the cohort.

**Breadth of CINV Documentation**

Of the patients for whom a CINV assessment was documented by either prescriber or RN, few (n = 9) provided full information on the breadth of the symptom, including presence, temporality, frequency, and severity. ► Fig. 2 outlines the CINV terms identified from these EHR data and from validated CINV assessment tools to conceptualize the necessary components for complete (documented and breadth) CINV documentation.

**Discussion**

In this study, we found that the EHR documentation of CINV in children, adolescents, and young adults receiving HEC was neither uniformly complete nor concordant among health care professionals. These findings are important to understand the utility and limitations of CINV documentation in the EHR for future development of CDSS to improve adherence to CINV guidelines. Further, the findings highlight the need for team-based strategies for improving multidisciplinary collaborative documentation to improve data quality.

The results of our study demonstrate that CINV documentation in the EHR is not always complete and varies by provider type, as 75% of prescribers and 58% of RNs documented an assessment, the primary definition in our study for completeness. CINV should always be assessed, particularly when a patient is receiving HEC, and it cannot be assumed that missing data are the same as absence of symptoms. In a cohort of patients who recently received

---

**Table 2** Summary of patient characteristics (n = 127)

<table>
<thead>
<tr>
<th>Variable</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>68 (53.5%)</td>
</tr>
<tr>
<td>Female</td>
<td>59 (46.5%)</td>
</tr>
<tr>
<td><strong>Insurance (primary)</strong></td>
<td></td>
</tr>
<tr>
<td>Commercial</td>
<td>52 (40.9%)</td>
</tr>
<tr>
<td>Noncommercial</td>
<td>75 (59.1%)</td>
</tr>
<tr>
<td><strong>Age group</strong></td>
<td></td>
</tr>
<tr>
<td>0–5 mo</td>
<td>5 (3.9%)</td>
</tr>
<tr>
<td>6 mo–11 y</td>
<td>51 (40.2%)</td>
</tr>
<tr>
<td>12–17 y</td>
<td>27 (21.3%)</td>
</tr>
<tr>
<td>18 &lt; 26 y</td>
<td>44 (34.7%)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>80 (63%)</td>
</tr>
<tr>
<td>Not white</td>
<td>47 (37%)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>85 (66.9%)</td>
</tr>
<tr>
<td>Hispanic or other</td>
<td>42 (33.1%)</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td></td>
</tr>
<tr>
<td>Inpatient</td>
<td>92 (72.4%)</td>
</tr>
<tr>
<td>Outpatient</td>
<td>35 (27.6%)</td>
</tr>
<tr>
<td><strong>Provider location</strong></td>
<td></td>
</tr>
<tr>
<td>Pediatric</td>
<td>98 (77.2%)</td>
</tr>
<tr>
<td>Adult</td>
<td>29 (22.8%)</td>
</tr>
<tr>
<td><strong>Chemo type</strong></td>
<td></td>
</tr>
<tr>
<td>Cisplatin</td>
<td>34 (26.8%)</td>
</tr>
<tr>
<td>Noncisplatin</td>
<td>93 (73.2%)</td>
</tr>
<tr>
<td><strong>Cancer type</strong></td>
<td></td>
</tr>
<tr>
<td>Solid tumor</td>
<td>62 (48.8%)</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>38 (29.9%)</td>
</tr>
<tr>
<td>Central nervous system</td>
<td>15 (11.8%)</td>
</tr>
<tr>
<td>Leukemia</td>
<td>12 (9.5%)</td>
</tr>
<tr>
<td><strong>Cancer status</strong></td>
<td></td>
</tr>
<tr>
<td>First occurrence</td>
<td>115 (90.6%)</td>
</tr>
<tr>
<td>Relapse</td>
<td>12 (9.4%)</td>
</tr>
</tbody>
</table>

---

Applied Clinical Informatics  Vol. 12 No. 2/2021 © 2021. Thieme. All rights reserved.
HEC and with the highest risk of CINV with expected prevalence as high as 80%, it is unlikely that only 34% (RN documentation) and 48% (prescriber documentation) of the 127 included patients had any CINV symptoms, the proportion of those who had a CINV documented assessment and were reported to have symptoms. Indeed, the low proportion of CINV symptoms suggests the data quality is not plausible, compared with the known high prevalence of CINV.

### Data quality of CINV documentation

<table>
<thead>
<tr>
<th>Prescriber documentation</th>
<th>RN documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No acute events overnight. Afebrile, no cough or runny nose. No problems with constipation or diarrhea. No nausea/vomiting. Tolerating chemotherapy well so far. Appetite ok. No bleeding. No pain</td>
<td>Patient vomited immediately after first attempt of prednisone dosing at 1,700. Second dose of prednisone attempted at 1,800 with medication crushed in ice cream. Patient did not tolerate and vomited. Mother present at bedside</td>
</tr>
<tr>
<td>No significant events overnight. Afebrile. No vomiting or diarrhea. Constipation: no BM since Tuesday. Appetite has been good. No cough, runny nose, or other URI Sxs. No reports of hematuria. No other bleeding signs or Sxs reported. No problems with pain. No other problems or concerns reported</td>
<td>NO TEXT, CHECKED 1 EPISODE EMESIS</td>
</tr>
<tr>
<td>Started chemo 3/21, w/delayed vomiting Yesterday and today. Seemed to have jaw pain, but teething. Seems fussy changing position. Remains afebrile</td>
<td>NO TEXT, CHECKED NO SYMPTOMS</td>
</tr>
<tr>
<td>C/o nausea, but no vomiting. Is still eating and drinking</td>
<td>Nausea: none and Zofran given ATC at home. Vomiting: none</td>
</tr>
<tr>
<td>LP done on Friday 3/17. First dose of carboplatin given the same day. Since discharge has had frontal headache, Saturday slept a lot, no eating and no drinking. Was afebrile. Headache worse on Sunday despite Tylenol, caffeine (hot tea). Brought to ED. Given fluid bolus and morphine but still with headache. Headache worse when standing, much better lying down. No vision changes. No changes in nueri exam. No vomiting. Some nausea, on zofran ATC post carboplatin</td>
<td>NO NAUSEA OR VOMITING; ABDOMEN SOFT AND NONDISTENDED</td>
</tr>
</tbody>
</table>

Abbreviations: ATC, around the clock; BM, bowel movement; C/o, complains of; ED, emergency department; RN, registered nurse; Sxs, symptoms; URI, upper respiratory infection.

Fig. 1  Completeness and concordance of chemotherapy-induced nausea and vomiting (CINV) assessments.
components of poor data quality suggest that other factors, such as workflow challenges that inhibit complete documentation, may be partly responsible.\textsuperscript{34}

Further, the concordance of CINV assessments between prescribers and RNs, in patients whose data were available, was 72%. The discordance may be more suggestive of the fragmentation of documentation rather than truly discordant assessments when comparing RN documentation with prescriber documentation. This is further supported by our finding that there was no difference in completeness of CINV documentation by clinical setting, pediatric oncology compared with adult oncology. It is likely that CINV was correctly and even more completely assessed by many clinicians, and a potential solution may be consideration of EHR systems that support documentation as a synthesis rather than discreet task completion.\textsuperscript{35}

In EHR documentation where CINV was reported, incompleteness of documentation was notable with few documented assessments completely reporting the breadth of the symptom: the temporality, severity, and frequency. Validated CINV assessment tools include these characteristics, and guideline recommendations vary depending on them.\textsuperscript{12,13,27,28} Incomplete breadth of documentation of CINV may be related to lack of a validated assessment tool capturing the data and integrated into the EHR. Complete documentation of CINV symptoms by clinicians may not be feasible to implement and sustain in a busy clinical workflow, and a promising solution is through integration of patient-reported outcome measures into the EHR to improve the completeness of symptom assessments.\textsuperscript{36–38} A collaborative data entry between providers and patients may further mitigate other data quality challenges, such as inaccuracy and plausibility.\textsuperscript{2,7,38}

Although the observed documentation of CINV symptoms cannot currently inform an accurate CDSS to modify antiemetic regimens based on CINV guidelines for refractory or breakthrough nausea, our findings do provide preliminary guidance to develop a prophylactic CDSS based on known patient and treatment factors. As outlined in Fig. 2, with existing terminologies, specifically RxNorm, SNOMED-CT mapped through UMLS, CDSS can feasibly be developed to implement guideline-concordant antiemetic prophylaxis. Similar efforts are currently underway and offer additional guidance toward utilization of existing EHR data of children and adolescents with cancer to guide implementation of guideline-concordant CINV management.\textsuperscript{39}

Further, because the majority of patients had a documented assessment of the presence/absence of CINV, the utility of these data should be explored further. For example, a CDSS might utilize existing documentation, integrated with patient-reported CINV using standardized measures, to guide high-quality, guideline-based decision-making. This approach may also be scalable when applied to other commonly reported cancer symptoms or among other populations, such as older people with cancer.\textsuperscript{7,40} Importantly, this requires expansion of existing terminologies to focus on cancer-related symptoms and ensure they allow for documentation of complete breadth of symptoms.

**Strengths and Limitations**

This study provides foundational knowledge for EHR documentation to completely and accurately describe a common and important cancer-related symptom and identifies EHR design gaps in coordinating documentations among care provider team members of different roles. We acknowledge the limited generalizability of this study due to using data from...
a single institution, a single EHR system, and a sample limited
to pediatric, adolescent, and young adult patients. The find-
ings, specifically the completeness assessment, may vary by
hospital as well as by EHR system. However, the thorough
examination of at least two domains of data quality can inform
additional research on data quality of cancer symptoms and
also guide development of CDSS. Future studies should exam-
ine differences across institutions and/or EHR systems to test
the validity of this approach in multiple settings. Because CINV
is a universal cancer symptom and we leveraged standardized
terminologies, we anticipate that the preliminary figure out-
lining the necessary terminologies to fully capture CINV
symptoms (Fig. 2) will largely be generalizable across sites
that utilize EHR systems as well as across other demographic
groups (e.g., older adults with cancer).

We also focused on two domains of data quality: complete-
ness and concordance. Other domains—correctness, plausibil-
ity, and currency of data—are associated with challenges
utilizing EHR data for secondary use, and examination of these
domains would more fully inform future use of CINV docu-
mentation. We briefly comment on plausibility, comparing
the rates of positive CINV symptoms to published prevalence
data, but we did not thoroughly examine this domain. Both
plausibility and correctness, without another validation
source, such as paper charts or patient-reported measures
for comparison, could not be fully assessed. Finally, evaluation
of the data currency often requires review of data logs, and
access to EHR audit records was not attainable for this study.

Conclusion
This study characterizes the EHR data quality of CINV assess-
ment and provides a framework for a comprehensive data-
driven approach, needed for future CDSS, to capture a
common cancer symptom in the EHR. The findings highlight
the data quality limitations to completely capturing symp-
toms using clinical terminologies, a weakness that needs
further research to enable accurate phenotyping and predic-
tive modeling of cancer symptoms.

Clinical Relevance Statement
This research is important both to highlight the importance
of high-quality clinical documentation and to guide docu-
mentation improvements, specifically when reporting pa-
ient symptoms.

Multiple Choice Questions
1. How can incomplete data negatively impact future de-
velopment of CDS in guiding CINV symptom management?
Select all that apply.
   a. Incomplete data about the presence of symptoms might
trigger an inappropriate nudge for escalation of care.
   b. Incomplete data about the presence of symptoms might
trigger an inappropriate nudge for no escalation of care.
   c. Data quality would not affect a CDS algorithm.

Correct Answer: The correct answer is options a and b.

2. Does the discordance of CINV data quality identified in
some of the included patient documentation signify an
error by the prescriber and/or RN? Choose the best
answer.
a. Yes, either the prescriber or the nurse inaccurately
documented CINV symptoms.
b. No, the timings of the assessments are different and so
the discordance is expected.
c. Not necessarily—it is possible that the RN or the
prescriber did not document correctly, but it is also
plausible that there is a good reason for discordance,
such as different timing of assessments.

Correct Answer: The correct answer is option c.

Protection of Human and Animal Subjects
Protection of human and animal subjects was followed
according to the IRB. This study included data that had
already been collected for clinical care of included
patients, and all data were reported in aggregate to avoid
any patient identifiers.

Funding
M.B. was supported by the National Institute of Nursing
Research Training Grant (NR007969), the National Cancer
Institute Training Grant (CA094061), and the Doctoral
Degree Scholarship in Cancer Nursing (DSCN-18–068–01)
from the American Cancer Society. C.W. was supported by
the National Library of Medicine Grant (R01LM009886). R.
S. was supported by the National Institute of Nursing
Research of the National Institutes of Health under award
number K24NR 018621. The content is solely the respon-
sibility of the authors and does not represent the official
views of the National Institutes of Health or the American
Cancer Society.

Conflict of Interest
None declared.

References
1. Tcheng JE, Bakken S, Bates DW, eds. Optimizing Strategies for
Clinical Decision Support: Summary of a Meeting Series. Wash-
ington, DC: National Academy of Medicine; 2017
2. Weiskopf NG, Weng C. Methods and dimensions of electronic
health record data quality assessment: enabling reuse for clinical
3. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining
and measuring completeness of electronic health records for
4. Kennell TJ Jr, Willig JH, Cimino JJ. clinical informatics researcher’s
desiderata for the data content of the next generation electronic
5. Wei W-Q, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC.
Combining billing codes, clinical notes, and medications from
electronic health records provides superior phenotyping perfor-
mance. J Am Med Inform Assoc 2016;23(01):c20–c27
6. Zhao J, Papapetrou P, Askler L, Boström H. Learning from hetero-
Inform 2017;65:105–119
Multinational Association of Supportive Care in Cancer MASCC. Available at: https://www.mascc.org/about-mascc
Dusetzina SB. Investigating racial disparities in use of NK1 receptor antagonists for chemotherapy-induced nausea and vomiting among women with breast cancer. Cancer 2017;318(02):197–198
Appendix A  Chemotherapeutic agents classified as highly emetogenic

<table>
<thead>
<tr>
<th>Emetogenicity</th>
<th>Chemotherapeutic agent 1</th>
<th>AND agent 2 (if applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Cisplatin</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Carboplatin*</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Dacarbazine</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Dactinomycin*</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Procarbazine</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Cyclophosphamide</td>
<td>Doxorubicin</td>
</tr>
<tr>
<td></td>
<td>Cyclophosphamide</td>
<td>Etoposide*</td>
</tr>
<tr>
<td></td>
<td>Ifosfamide</td>
<td>Etoposide*</td>
</tr>
<tr>
<td></td>
<td>Thiotepa ≥ 300 mg/m²*</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Cytarabine 3 g/m²/dose*</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Cyclophosphamide ≥ 1 g/m²*</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Methotrexate ≥ 12 g/m²</td>
<td>n/a</td>
</tr>
</tbody>
</table>

\*Denotes regimens that are HEC for pediatric patients and MEC for adults.

Appendix B  Sections of electronic health record examined for data abstraction

<table>
<thead>
<tr>
<th>EHR system</th>
<th>Title of note</th>
<th>Section of interest</th>
<th>Variable within note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outpatient visits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescriber documentation</td>
<td>Follow-up visit (pediatric and adult oncology)</td>
<td>History of present illness (HPI)  Problem list (active)  GI medication prescriptions (home)</td>
<td>• CINV assessed(Y/N)  • CINV present (Y/N)  • Free text from HPI  • ICD-0 code for CINV (Y/N)  • ICD-10 code for primary disease (Y/N)  • ICD-10 code for antineoplastic visit (Y/N)</td>
</tr>
<tr>
<td></td>
<td>Home medication list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient visits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescriber documentation</td>
<td>Pediatric oncology note  Ob/Gyn encounter note  Medicine resident progress note  Hem/oncology attending follow-up note</td>
<td>History of present illness (active)  Clinical summary (ICD-0 codes)</td>
<td>• CINV assessed(Y/N)  • CINV present (Y/N)  • Free text from HPI  • ICD-10 code for CINV (Y/N)  • ICD-10 code for primary disease (Y/N)  • ICD-10 code for antineoplastic visit (Y/N)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nursing documentation</td>
<td>Ambulatory hem/oncology nursing assessment  Shift assessment  Flow sheets  Nursing chemotherapy/biotherapy record  Medication administration record  Nursing discharge note</td>
<td>GI symptoms  Emesis (volume)  Emesis (episode)  Chemotherapeutic and antiemetic agents administered in clinic</td>
<td>• Nausea/vomiting present (Y/N)  • Medication given (Y/N) (Drug)  • Chemotherapy type  • Confirm class of emetogenicity is HEC  • Appropriate regimen administered in clinic (Y/N)</td>
</tr>
</tbody>
</table>

Abbreviations: CINV, chemotherapy-induced nausea and vomiting; EHR, electronic health record; GI, gastrointestinal; HEC, highly emetogenic chemotherapy regimen; hem, hematology; Ob/Gyn, obstetrics and gynecology.