Research and Applications

Predicting neutropenia risk in patients with cancer using electronic data

Pamala A Pawloski,^{1,2,3} Avis J Thomas,¹ Sheryl Kane,¹ Gabriela Vazquez-Benitez,¹ Gary R Shapiro,^{3,4} and Gary H Lyman^{5,6,7}

¹HealthPartners Institute, Minneapolis, Minnesota, USA, ²Health Care Systems Research Network/National Cancer Institute Cancer Research Network, USA, ³Regions Hospital Cancer Care Center, St. Paul, Minnesota, USA, ⁴Cancer Center of Western Wisconsin, New Richmond, Wisconsin, USA, ⁵Hutchinson Institute for Cancer Outcomes Research, Seattle, Washington, USA, ⁶Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, and ⁷University of Washington School of Medicine, Seattle, Washington, USA

Corresponding Author: Pamala A Pawloski, PharmD, HealthPartners Institute, 8170 33rd Ave S, MS23301A, Bloomington, MN 55425, USA. E-mail: pamala.a.pawloski@healthpartners.com; Tel. 952-967-5030; Fax: 952-967-5022

Received 19 May 2016; Revised 29 July 2016; Accepted 9 August 2016

ABSTRACT

Objectives: Clinical guidelines recommending the use of myeloid growth factors are largely based on the prescribed chemotherapy regimen. The guidelines suggest that oncologists consider patient-specific characteristics when prescribing granulocyte-colony stimulating factor (G-CSF) prophylaxis; however, a mechanism to quantify individual patient risk is lacking. Readily available electronic health record (EHR) data can provide patient-specific information needed for individualized neutropenia risk estimation. An evidence-based, individualized neutropenia risk estimation algorithm has been developed. This study evaluated the automated extraction of EHR chemotherapy treatment data and externally validated the neutropenia risk prediction model.

Materials and Methods: A retrospective cohort of adult patients with newly diagnosed breast, colorectal, lung, lymphoid, or ovarian cancer who received the first cycle of a cytotoxic chemotherapy regimen from 2008 to 2013 were recruited from a single cancer clinic. Electronically extracted EHR chemotherapy treatment data were validated by chart review. Neutropenia risk stratification was conducted and risk model performance was assessed using calibration and discrimination.

Results: Chemotherapy treatment data electronically extracted from the EHR were verified by chart review. The neutropenia risk prediction tool classified 126 patients (57%) as being low risk for febrile neutropenia, 44 (20%) as intermediate risk, and 51 (23%) as high risk. The model was well calibrated (Hosmer-Lemeshow goodness-of-fit test = 0.24). Discrimination was adequate and slightly less than in the original internal validation (c-statistic 0.75 vs 0.81).

Conclusion: Chemotherapy treatment data were electronically extracted from the EHR successfully. The individualized neutropenia risk prediction model performed well in our retrospective external cohort.

Key words: clinical decision support systems, computer-based decision support, febrile neutropenia, chemotherapy, risk model, granulocyte-colony stimulating factor

BACKGROUND AND SIGNIFICANCE

Febrile neutropenia (FN) and severe neutropenia (SN), characterized by a significant decrease in white blood cells with and without fever present, are serious sequelae resulting from myelosuppressive cancer chemotherapy treatment. FN is a medical emergency and is associated with treatment delays, dose reductions, hospitalizations, mortality, and a high cost burden that is estimated at \$1.2 billion annually.^{1,2} Granulocyte-colony stimulating factor (G-CSF) prophylaxis (filgrastim, pegfilgrastim, tbo-filgrastim, and filgrastim-sndz) decreases chemotherapy-induced neutropenia and infection,^{3–8} length and severity of neutropenia,^{9–20} and infection-related mortality and chemotherapy-associated early deaths,²¹ and reduces absolute and relative risk for all-cause mortality.²² However, G-CSF use often goes against recommended clinical practice guidelines.^{23–25} Current recommendations include use of G-CSF prophylaxis for high-risk FN regimens and avoiding it with low-risk regimens.^{3,7} However, in addition to its high cost, G-CSF is also associated with adverse events and is not warranted for all patients.^{26,27}

The incidence of FN and SN varies across diagnoses, treatment regimens, and lines of therapy. A retrospective analysis of adults receiving chemotherapy for solid tumors showed an FN rate of 13–21%;²⁸ however, a systematic review and meta-analysis of randomized controlled trials and observational studies showed significantly higher FN rates in the observational cohort (odds ratio [OR] = 1.74; 95% confidence interval [CI], 1.15–2.62; $P = .012$).²⁹ Individual patient characteristics, including comorbidities, are also associated with FN risk.³⁰ Current clinical guidelines recommend G-CSF prophylaxis for neutropenia risk >20% in adults with solid tumors and nonmyelogenous malignancies.^{3,7} The guidelines are primarily based on the chemotherapy regimen, with recommended consideration of individual patient characteristics when assessing neutropenia risk. Specific patient characteristics are provided within the guidelines; however, a mechanism to quantify patient-specific risk is lacking.³ Due to the widespread adoption of electronic health records (EHRs), extraction of patient data could inform individualized neutropenia risk prediction for patients at the point of chemotherapy prescribing.

To assess the impact of patient-specific characteristics on SN or FN risk, Lyman et al.³¹ developed and internally validated an algorithm for estimating neutropenia risk using a large prospective patient registry. Patient-specific variables included age, receipt of prior chemotherapy, cancer diagnosis, relative dose intensity (RDI) for cycle 1 chemotherapy, white blood cell count (WBC), glomerular filtration rate, aspartate aminotransferase level, alkaline phosphatase level, total bilirubin, concurrent immunosuppressive therapy, and receipt of G-CSF prophylaxis. Regression analysis demonstrated good model discrimination for SN or FN risk over multiple chemotherapy cycles. The objectives of the present study include verification of the quality of electronic chemotherapy treatment data extracted from the EHR and external validation of the existing neutropenia risk model.³¹

MATERIALS AND METHODS

Study setting and population

This retrospective data quality analysis and model validation was conducted at HealthPartners, an integrated health care system located in the Midwest serving 1.5 million members and more than 1 million patients annually. The study population included adults 18 years and older with incident breast, colorectal, lung, lymphoid, or

ovarian cancer diagnosed from January 1, 2007, to December 31, 2012. Subjects were required to have received cycle 1 of a new chemotherapy regimen from an owned HealthPartners cancer clinic and health plan enrollment for at least 10 of the 12 months preceding and following the relevant cancer diagnosis. Subjects with HIV, leukemia, or a history of stem cell transplant were excluded. Subjects were also excluded if they were treated with an oral chemotherapy regimen or received therapy at a contracted or external facility, because treatment records were not readily available for electronic data extraction. The HealthPartners Institutional Review Board granted approval, including a waiver of patient consent.

Data extraction and verification

Data were electronically extracted from 2 sources: the EHR (Epic Systems, Verona, WI, USA) and the site's virtual data warehouse (VDW). Variables extracted from the EHR reporting database (Clarity) included laboratory values (aspartate aminotransferase, alkaline phosphatase, serum creatinine, glomerular filtration rate, total bilirubin, WBC, and absolute neutrophil count [ANC]). The VDW is a federated, standardized data repository developed for single- or multisite research within the Health Care Systems Research Network.^{32–36} VDW variables extracted from the virtual tumor registry file included tumor type and characteristics, patient age at diagnosis, sex, race, first course of treatment, and receipt of prior chemotherapy. Virtual tumor registry data were populated from the institution's tumor registry, composed of medical chart data manually abstracted by trained abstractors according to North American Association of Central Cancer Registries standards.³⁷

Chemotherapy treatment data (height, weight, body surface area, chemotherapy regimen, G-CSF agent, doses, and administration dates) were electronically extracted from the EHR oncology treatment module (Epic Beacon). Chemotherapy treatment data were also manually abstracted by trained research staff. The electronically extracted data were compared with manually extracted data to verify electronic extraction of the variables. Discrepancies between the electronic and manual data were adjudicated by the principal investigator (PAP).

Variables derived from extracted data included the presence of neutropenia, SN, and FN using previously defined criteria.³ Neutropenia is defined as an ANC <1500/mcL ($<1.5 \times 10^9/L$). SN is defined as an ANC <500/mcL, and FN is defined by a single oral temperature $\geq 38.3^\circ C$ or $\geq 38.0^\circ C$ over 1 h; ANC <500 neutrophils/mcL or <1000 neutrophils/mcL with a predicted decline to $\leq 500/mcL$ over the next 48 h. All SN and FN events were chart abstracted and reviewed by the principal investigator. The confirmation rate was 91%. Renal function was assessed using the Cockcroft and Gault method.³⁸ Immunosuppressive therapy, defined as concurrent treatment with medication(s) indicated to decrease the immune response for noncancerous conditions, was identified from VDW prescription claims data from outpatient pharmacies using national drug codes.

Treatment regimen data were extracted and categorized according to standard regimens published in the treatment guidelines³ and primary literature. Where more than 1 standard dose was identified, the least aggressive dosing regimen was used. Patient data were excluded if the treatment regimen could not be attributed to a standard regimen. Chemotherapy treatment regimen data were used to calculate the RDI (the ratio of planned dose to the standard regimen dose) for each chemotherapy agent. The RDI was calculated following the method outlined by Lyman et al.³¹; however, the

administered dose was substituted for the planned dose due to limitations of the retrospective study design. RDIs for individual cytotoxic agents were averaged to obtain an overall RDI for each patient.

All chemotherapy agents included in the original model were placed in rank order based on the model's estimated OR for marginal risk.³¹ Fludarabine, gemcitabine, and bevacizumab were not included in the original neutropenia risk model but were included in the current study. These 3 agents were inserted into the OR ranking by assessing the marginal risks associated with them, using the primary literature and the current guidelines for myeloid growth factor use.³ The ORs for the 3 agents were estimated as the average of the ORs immediately preceding and following them in the ranking.

DATA ANALYSIS

External validation of the preexisting neutropenia risk prediction algorithm

The existing neutropenia risk prediction model was applied to the current data with the 2 previously described modifications: administered chemotherapy treatment data were used for the RDI, and 3 chemotherapy agents not included in the original model were added.³¹ The data were partitioned into 3 neutropenia risk strata: low (<10%), intermediate (10–20%), and high (>20%). The average predicted risk and observed SN and FN rates were computed for each stratum for comparison within strata and overall. Model performance based on continuous risk estimates was formally evaluated using calibration and discrimination measures. A comprehensive description of the risk model can be found in the supplemental appendix (Appendix A).

Logistic regressions were performed using the model's risk score as a univariate predictor of neutropenia events. Calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test.³⁹ A regression was also used to estimate beta coefficients for the best fitting linear transformation of the risk score. Model discrimination was assessed using the receiver operating characteristic (ROC) curve and the area under the curve (c-statistic).^{40,41} Additional tabular and graphical analyses were done to evaluate the use of prophylaxis by hypothetical predicted risk level; hypothetical predicted risk was computed assuming no G-CSF prophylaxis. Subgroup analyses were performed for patients without G-CSF prophylaxis and those with estimated neutropenia risk <30%. Sensitivity tests were conducted to assess the impact of incorporating fludarabine, gemcitabine, and bevacizumab into the model. One test included all patients and estimated the additional risk from the new agents as zero. The other test excluded patients ($n=29$) who received these agents. G-CSF prophylaxis was compared to risk model recommendations and neutropenic events. All electronically extracted data were obtained using SAS v9.3 (SAS Institute Inc., Cary, NC, USA).

RESULTS

Of the 262 patients who met the inclusion criteria, 41 were excluded from analysis. Reasons for exclusion were as follows: standard chemotherapy treatment regimen was undetectable ($n=29$), patient did not complete cycle 1 therapy ($n=8$), incomplete data or no receipt of chemotherapy ($n=6$), receipt of oral chemotherapy ($n=1$), and G-CSF use could not be verified ($n=1$). Data extraction initially included records from 2007 to 2013; however, many records following implementation of the oncology treatment module were

incomplete (eg, inpatient chemotherapy treatment was not initially included in the oncology module, and some records were not available for electronic extraction due to incomplete workflow entries). Therefore, we chose to exclude 2007 chemotherapy treatment data from analysis. Primary findings from the comparison of electronically and manually extracted data demonstrated that the extracted data were largely correct. The following issues were identified: (1) data specifying treatment cycle and day were not readily extractable; (2) G-CSF administration was not consistently recorded in the treatment module secondary to alternate outpatient clinic or weekend/holiday administration, but was accessible from other EHR reporting files; and (3) investigational or approved chemotherapy agents used in clinical trials may not be readily identifiable from extracted data.

Table 1 summarizes the current and original study population used to derive the algorithm.³¹ Of 221 evaluable patients, 91 (41%) were 65 years or older, 155 (70%) were female, and breast cancer (37%) was the most common diagnosis. Stratification of patients by predicted neutropenia risk included low ($n=135$; 61%), intermediate ($n=35$; 16%), and high ($n=51$; 23%) (Table 2). The population average predicted risk was 14.6%, compared with an observed event rate of 13.1%. Within each stratum, the mean predicted risk from the model and observed event rates were similar; P values corresponding to differences in event rates ranged from 0.93 to 0.99. Of the 79 patients who received G-CSF prophylaxis, the mean predicted risk was 11% after adjustment for G-CSF prophylaxis; 8 patients (10%) experienced SN or FN.

The Hosmer-Lemeshow goodness-of-fit statistic was 0.57, indicating good model fit. Calibration parameters were obtained from the same regression; the best-fitting linear transformation would multiply the risk score (the logit of the predicted risk) by 0.74 (95% CI, 0.38–1.10) and add -0.53 (95% CI, -1.22 , 0.16). This translated to correcting a 20% predicted risk to 17.4% or a 10% predicted risk to 10.4%. Because the calibration coefficients were small and not significantly different from zero, they were not used. All reported predicted risks are from the uncalibrated model. The area under the ROC curve (c-statistic) was 0.75, indicating adequate discriminatory performance (Figure 1).^{40,41} As seen on the ROC curve, 80% percent sensitivity (the rate of true positives) was associated with a false positive rate of 40%.

A secondary analysis examined actual G-CSF prophylaxis by level of hypothetical risk. Hypothetical risk was estimated using the model and assuming no G-CSF prophylaxis was administered (Figure 2). Across all hypothetical predicted risk levels, 8 (4%) had an SN or FN event despite receiving prophylaxis.

The model was tested using subgroup analyses and sensitivity tests (Table 3). We examined only those patients who did not receive prophylaxis, patients with an estimated neutropenia risk <30%, and alternative methods for addressing cytotoxic agents not already included in the model. In each case, the model had reasonable fit (area under the ROC curve ranging 0.72–0.81).

DISCUSSION

This retrospective study demonstrates our ability to accurately extract electronic chemotherapy treatment data from the EHR oncology treatment module. This data had not previously been extracted electronically at our site, and manual verification confirmed accurate and reliable extraction. Our learnings from the data verification demonstrate that data variability occurred secondary to the module development, clinical workflow, and data documentation in the early phases of implementing the module. The data were

Table 1. Description of the current and the neutropenia risk model derivation cohort including rates of G-CSF primary prophylaxis and neutropenia events by patient characteristic among patients receiving cycle 1 chemotherapy treatment

Patient characteristic	Current population (n = 221) N (%)	Derivation population from Lyman et al. ²⁷ (n = 2425) (%)
Age ≥65	91 (41)	38.4
Female	155 (70)	70
Receipt of prior chemotherapy	2 (1) ^a	24.4
Race/ethnicity		
Non-Hispanic white	201 (91)	85
Non-Hispanic black	11 (5)	10
Hispanic	4 (2)	Unknown
Other	4 (2)	5
Baseline labs		
AST >35 U/L	37 (17)	13.5
ALP >120 U/L	26 (12)	19.5
Bilirubin >1 mg/dL	13 (6)	4.1
GFR <60 mL/min	31 (14)	19.5
WBC <5000 mm ³	29 (13)	12.8
Cancer type		
Colorectal	37 (17)	14
Lung (small cell)	14 (6)	5.8
Lung (non-small cell)	39 (18)	19
Ovary	11 (5)	8.4
Breast	81 (37)	38.7
Lymphoma	39 (18)	14.1
Concurrent immunosuppressive therapy	2 (1)	14.4
Planned ^b RDI ≥85%	207 (94)	69.9
Chemotherapy ^c		
Anthracyclines	92 (42)	37.8
Platinum(s)	89 (40)	32.5
Taxanes	45 (20)	29
Alkylating agents	111 (50)	42.1
Topoisomerase II inhibitors	1 (1)	6.2
Gemcitabine	2 (1)	6.9
Topoisomerase I inhibitors	26 (12)	1.1
Vinorelbine	2 (10)	3.4
Bevacizumab	22 (10)	—
Fludarabine	8 (4)	—
Pemetrexed	1 (0)	—
Primary G-CSF prophylaxis	82 (37)	19.6
Cycle 1 severe or febrile neutropenia event	29 (13)	18.9

^aDifference based on study inclusion criteria of newly diagnosed/recurrent disease.

^bActual dose administered rather than planned dose was used for the current RDI calculation.

^cChemotherapy use includes each chemotherapy agent or class received by every patient.

reliable and accurate following the first year of implementation; however, general themes outlining the utility of the data were identified. These findings will inform future proposed research at our site. However, the data variability identified at our site may not predict data quality following EHR implementation at other sites. It is possible that the observed challenges with workflow documentation are common among new EHR systems and their users.

This study also provided the first external validation of an existing patient-specific neutropenia risk prediction model.³¹ The current

Table 2. Severe and febrile neutropenia event rates stratified by predicted risk level among patients receiving cycle 1 chemotherapy treatment with any regimen

Predicted risk %		Observed events ^a				P value for difference in event rates
Range (%)	Mean	With an event	Without an event	Total patients	Event rate (%)	
<10	5.8	7	119	126	5.5	0.99
10–20	13.3	5	39	44	11.4	0.95
>20	37.6	17	34	51	33.3	0.93
Overall	14.6	29	192	221	13.1	

^aEvent = severe or febrile neutropenia.

study cohort was developed using similar inclusion criteria to those of the original cohort,³¹ except the current cohort included only patients diagnosed within the study period. The original cohort included 3638 patients with a diagnosis of breast, colorectal, lung, lymphoid, or ovarian cancer from 115 sites throughout the US.³¹ The final, internally validated model incorporated 11 patient characteristics: age, receipt of prior chemotherapy, abnormal liver and renal function, WBC, cancer type, concurrent immunosuppressive therapy, RDI for the planned chemotherapy regimen, and G-CSF prophylaxis.³¹

Key differences between the current and original cohorts include: (1) a higher proportion of patients with RDI ≥85%, (2) greater use of G-CSF prophylaxis, (3) lower SN or FN event rate, and (4) fewer patients receiving concurrent immunosuppressive therapy. These differences are likely due to myeloid growth factor guideline changes, including a decrease in the high neutropenia risk threshold to >20%, increases in treatment intensity and dose-dense therapy, and subsequent increased use of G-CSF prophylaxis.^{3,7,42–51} A lower reported rate of concurrent immunosuppressive therapy may be due to definitional differences (personal communication, G.H. Lyman).

Stratification of patients by low, intermediate, and high neutropenia risk categories demonstrates that risk estimates were accurate within each stratum. Importantly, 44 (20%) of the 221 patients were classified as intermediate risk, where the use of G-CSF prophylaxis is highly dependent on clinical judgment and individualization of risk could potentially be most beneficial. The resultant area under the ROC curve of 0.75 indicates reasonable overall performance and is only modestly lower than in the original development cohort.^{31,40,41}

The secondary analysis assessed the extent to which adherence to model recommendations might affect the rate of prophylactic G-CSF use and neutropenic events. Few patients in the low and intermediate risk groups received G-CSF prophylaxis (Figure 1), suggesting that there may be stronger concordance with published guidelines than previously reported.^{23,25,52} This may be the result of published research, local and national initiatives (eg, ASCO Top 5⁵³), the use of standardized order sets in the EHR, and the study being conducted in an integrated health care system. However, among patients in the high-risk group, one-third did not receive G-CSF, despite guideline recommendations.³ Therefore, a degree of discordance with clinical guidelines remains among both low- and high-risk groups, albeit to a greater degree within the high-risk group. These cases likely represent prescribers' clinical judgment in determining G-CSF prophylaxis. The National Comprehensive Cancer Network guidelines specify patient characteristics to consider for neutropenic risk determination³; however, those character-

Figure 1. The area under the receiver operating characteristic (ROC) curve demonstrating model fit for the neutropenia risk prediction model.

Figure 2. Observed G-CSF prophylaxis and neutropenic events compared with predicted risk if no G-CSF prophylaxis were administered.

istics were tested and some were not found to be statistically significant in a multivariate analysis.³¹ This emphasizes the need for risk prediction methods when assessing individual neutropenic risk. Figure 1 demonstrates SN and FN events occurring across risk categories, suggesting the difficulty in ascertaining risk solely by judgment. This point was further illustrated in a study comparing physician estimation of neutropenia risk with the neutropenia risk model used in this study.⁵⁴ The median physician-assessed and model-predicted SN and FN risks across all chemotherapy cycles were 20.0% and

Table 3. Models to predict severe or febrile neutropenia using the area under the receiver operating characteristic (ROC) curve (c-statistic) among patients receiving cycle 1 chemotherapy treatment

Model	C-statistic	Description
1	0.75	Final model, based on 221 patients
2	0.81	Removing 82 who received prophylaxis; performed to assess possible inaccuracies in the risk algorithm related to G-CSF prophylaxis
3	0.72	Removing 31 patients with estimated risk >30%; performed to remove possible high-risk outliers with excessive leverage and to focus on the group for which decision-making is most difficult
4	0.76	Risk for 3 new agents (fludarabine, gemcitabine, and bevacizumab) estimated as zero
5	0.75	Removing 29 patients receiving fludarabine, gemcitabine, or bevacizumab

17.9%, respectively. However, physician-assessed FN risk and model-predicted risk correlated weakly: 0.249 (95% CI, 0.179–0.316). Further, the correlation between physician-assessed FN risk and orders for G-CSF prophylaxis was moderate (0.313; 95% CI, 0.135–0.472).

Among those in the low and intermediate neutropenia risk groups who did not receive prophylaxis, some experienced a neutropenic event. Despite the chemotherapy regimen being associated with a low risk for a neutropenic event, individual risk factors likely impacted the level of risk. In addition, among those at high risk who did not receive prophylaxis, not all experienced a neutropenic event; it is possible that not all patients classified as high risk require G-CSF prophylaxis, at least initially. Individualized neutropenia risk estimation based on planned chemotherapy and patient-specific characteristics could enable optimization of neutropenia risk prediction and individualize G-CSF use, especially among patients receiving chemotherapy treatments associated with low or intermediate risk.

Subgroup analyses also demonstrated good model performance (c-statistic 0.81 and 0.72, respectively), suggesting that the model is adequate for use in other patient populations and that performance is not markedly affected by G-CSF prophylaxis or patients with neutropenia risk >30%. Sensitivity analyses assessing the addition of fludarabine, gemcitabine, and bevacizumab did not affect model output in an appreciable manner, showing that it is possible to update the model with newer medications as they become available. The findings related to the addition of the 3 agents were expected because the estimated ORs were derived from neutropenia rates reported in the primary literature and were low relative to other chemotherapy agents in the model. Applying this risk prediction model to patient care will enable providers to identify individuals most likely to benefit from G-CSF prophylaxis and to avoid its use where neutropenia risk is minimal. The model relies on patient information (eg, age, sex, laboratory values, cancer diagnosis, planned treatment regimen and dose, and receipt of prior chemotherapy) available in the EHR. We expect that implementation of the same algorithm elsewhere, using either an automated system or manual input, would have similar validity. The ease of implementation would be determined by the ease of obtaining reliable patient information.

Strengths of the present study include source data that support efficient, high-quality data extraction. In addition to the cohort similarities, identified differences demonstrate the model's validity in 2

somewhat dissimilar populations. This provides an opportunity for testing and possible implementation at other sites where high-quality EHR data are available. Study limitations include a small patient population from within a single site, which potentially limits generalizability. The small number of neutropenic events reduces the study power to adequately characterize model performance.⁵⁵ Despite receiving G-CSF prophylaxis, 8 patients across all predicted risk levels had an SN or FN event. Therefore, even with a more accurate individualized method to determine neutropenic risk, use of G-CSF prophylaxis will not prevent every event. The degree to which G-CSF may not work based on a given treatment regimen has been assessed as part of clinical trials; however, these assessments did not take into account individual patient risk. There are possibly additional patient-specific factors not yet identified that affect patient-specific neutropenia risk. Identifying additional risk factors may improve the model's discrimination.

CONCLUSION

Electronically extracted chemotherapy treatment data from the EHR were verified and provided information critical to understanding electronic chemotherapy treatment data structure and its limitations. We also provided the first external validation of an existing patient-specific neutropenia risk prediction model. The model demonstrated good model performance. This work will inform future research directed at the individualization of neutropenic risk.

CONFLICTS OF INTEREST

The authors report no conflicts of interest associated with this project.

RESEARCH SUPPORT

Funding for this project was provided through a HealthPartners internal research grant.

CONTRIBUTORS

Conception and design: PAP, GRS, AJT, SK, GVB. Provision of study materials or patients: SK. Collection and assembly of data: SK, AJT, PAP. Data analysis and interpretation, writing of manuscript, final approval of manuscript: all authors.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

The guidance of Terry Field and Diana Buist, members of the Health Care Systems (HCS) Cancer Research Network (CRN) and CRN Scholar Program mentors, contributed to the development of this research proposal through the CRN Scholars Program. The overall goal of the CRN is to conduct collaborative research to determine the effectiveness of preventive, curative, and supportive interventions for major cancers that span the natural history of those cancers among diverse populations and health systems. HealthPartners Institute is an affiliate member of the HCS Research Network. We would also like to acknowledge Laurie Van Arman and Diane Eggen for chart abstraction, and Cathy Iffert and Laurny Davin, who assisted with manuscript preparation and submission.

REFERENCES

1. Medicare Drug Spending Dashboard 2015-12-21. <https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-12-21.html>. Accessed March 17, 2016.
2. Lyman GH, Kuderer NM. Epidemiology of febrile neutropenia. *Supportive Cancer Therapy*. 2003;1:23–35.
3. Myeloid Growth Factors. *National Comprehensive Cancer Network*, V1. 2016. http://www.nccn.org/professionals/physician_gls/pdf/myeloid_growth.pdf. Accessed March 15, 2016.
4. Bohlius J, Reiser M, Schwarzer G, Engert A. Granulopoiesis-stimulating factors to prevent adverse effects in the treatment of malignant lymphoma. *Cochrane Database Syst Rev*. 2004;3:CD003189.
5. Sung L, Nathan PC, Alibhai SM, Tomlinson GA, Beyene J. Meta-analysis: effect of prophylactic hematopoietic colony-stimulating factors on mortality and outcomes of infection. *Ann Intern Med*. 2007;147:400–11.
6. Cooper KL, Madan J, Whyte S, Stevenson MD, Akehurst RL. Granulocyte colony-stimulating factors for febrile neutropenia prophylaxis following chemotherapy: systematic review and meta-analysis. *BMC Cancer*. 2011;11:404.
7. Smith TJ, Bohlke K, Lyman GH, et al. Recommendations for the use of WBC growth factors: American Society of Clinical Oncology Clinical Practice Guideline Update. *J Clin Oncol*. 2015;33:3199–212.
8. Wang L, Baser O, Kutikova L, Page JH, Barron R. The impact of primary prophylaxis with granulocyte colony-stimulating factors on febrile neutropenia during chemotherapy: a systematic review and meta-analysis of randomized controlled trials. *Supportive Care Cancer*. 2015;23:3131–40.
9. Gisselbrecht C, Haioun C, Lepage E, et al. Placebo-controlled phase III study of lenograstim (glycosylated recombinant human granulocyte colony-stimulating factor) in aggressive non-Hodgkin's lymphoma: factors influencing chemotherapy administration. Groupe d'Etude des Lymphomes de l'Adulte. *Leukemia Lymphoma*. 1997;25:289–300.
10. Timmer-Bonte JN, de Boo TM, Smit HJ, et al. Prevention of chemotherapy-induced febrile neutropenia by prophylactic antibiotics plus or minus granulocyte colony-stimulating factor in small-cell lung cancer: a Dutch randomized phase III study. *J Clin Oncol*. 2005;23:7974–84.
11. Trillet-Lenoir V, Green J, Manegold C, et al. Recombinant granulocyte colony stimulating factor reduces the infectious complications of cytotoxic chemotherapy. *Eur J Cancer*. 1993;29A:319–24.
12. Vogel CL, Wojtukiewicz MZ, Carroll RR, et al. First and subsequent cycle use of pegfilgrastim prevents febrile neutropenia in patients with breast cancer: a multicenter, double-blind, placebo-controlled phase III study. *J Clin Oncol*. 2005;23:1178–84.
13. Bui BN, Chevallier B, Chevreau C, et al. Efficacy of lenograstim on hematologic tolerance to MAID chemotherapy in patients with advanced soft tissue sarcoma and consequences on treatment dose-intensity. *J Clin Oncol*. 1995;13:2629–36.
14. Chevallier B, Chollet P, Merrouche Y, et al. Lenograstim prevents morbidity from intensive induction chemotherapy in the treatment of inflammatory breast cancer. *J Clin Oncol*. 1995;13:1564–71.
15. Crawford J, Ozer H, Stoller R, et al. Reduction by granulocyte colony-stimulating factor of fever and neutropenia induced by chemotherapy in patients with small-cell lung cancer. *N Engl J Med*. 1991;325:164–70.
16. Gatzemeier U, Kleisbauer JP, Drings P, et al. Lenograstim as support for ACE chemotherapy of small-cell lung cancer: a phase III, multicenter, randomized study. *Am J Clin Oncol*. 2000;23:393–400.
17. Muhonen T, Jantunen I, Pertovaara H, et al. Prophylactic filgrastim (G-CSF) during mitomycin-C, mitoxantrone, and methotrexate (MMM) treatment for metastatic breast cancer. A randomized study. *Am J Clin Oncol*. 1996;19:232–34.
18. Osby E, Hagberg H, Kvaloy S, et al. CHOP is superior to CNOP in elderly patients with aggressive lymphoma while outcome is unaffected by filgrastim treatment: results of a Nordic Lymphoma Group randomized trial. *Blood*. 2003;101:3840–48.
19. Pettengell R, Gurney H, Radford JA, et al. Granulocyte colony-stimulating factor to prevent dose-limiting neutropenia in non-Hodgkin's lymphoma: a randomized controlled trial. *Blood*. 1992;80:1430–36.

20. Zinzani PL, Pavone E, Storti S, *et al.* Randomized trial with or without granulocyte colony-stimulating factor as adjunct to induction VNCOP-B treatment of elderly high-grade non-Hodgkin's lymphoma. *Blood*. 1997;89:3974-79.
21. Kuderer NM, Dale DC, Crawford J, Lyman GH. Impact of primary prophylaxis with granulocyte colony-stimulating factor on febrile neutropenia and mortality in adult cancer patients receiving chemotherapy: a systematic review. *J Clin Oncol*. 2007;25:3158-67.
22. Lyman GH, Dale DC, Wolff DA, *et al.* Acute myeloid leukemia or myelodysplastic syndrome in randomized controlled clinical trials of cancer chemotherapy with granulocyte colony-stimulating factor: a systematic review. *J Clin Oncol*. 2010;28:2914-24.
23. Potosky AL, Malin JL, Kim B, *et al.* Use of colony-stimulating factors with chemotherapy: opportunities for cost savings and improved outcomes. *J Natl Cancer Inst*. 2011;103:979-82.
24. Ramsey SD, Fedorenko C, Chauhan R, *et al.* Baseline Estimates of Adherence to American Society of Clinical Oncology/American Board of Internal Medicine Choosing Wisely Initiative Among Patients With Cancer Enrolled With a Large Regional Commercial Health Insurer. *J Oncol Pract*. 2015;11:338-43.
25. Waters GE, Corrigan P, Gatesman M, Smith TJ. Comparison of pegfilgrastim prescribing practice to national guidelines at a university hospital outpatient oncology clinic. *J Oncol Pract*. 2013;9:203-06.
26. Pawloski PA, Larsen M, Thoresen A, Giordana MD. Pegfilgrastim use and bone pain: a cohort study of community-based cancer patients. *J Oncol Pharm Pract*. 2016;22:423-29.
27. Dinan MA, Hirsch BR, Lyman GH. Management of chemotherapy-induced neutropenia: measuring quality, cost, and value. *J Natl Compr Canc Netw*. 2015;13:e1-7.
28. Weycker D, Li X, Edelsberg J, *et al.* Risk and consequences of chemotherapy-induced febrile neutropenia in patients with metastatic solid tumors. *J Oncol Pract*. 2015;11:47-56.
29. Truong J, Lee EK, Trudeau ME, Chan KK. Interpreting febrile neutropenia rates from randomized, controlled trials for consideration of primary prophylaxis in the real world: a systematic review and meta-analysis. *Ann Oncol*. 2016;27:608-18.
30. Li X, Luthra R, Morrow PK, *et al.* Comorbidities among patients with cancer who do and do not develop febrile neutropenia during the first chemotherapy cycle. *J Oncol Pharm Pract*. 2016;22:678-89.
31. Lyman GH, Kuderer NM, Crawford J, *et al.* Predicting individual risk of neutropenic complications in patients receiving cancer chemotherapy. *Cancer*. 2011;117:1917-27.
32. Hornbrook MC, Hart G, Ellis JL, *et al.* Building a virtual cancer research organization. *J Natl Cancer Inst Monogr*. 2005;35:12-25.
33. Aiello Bowles EJ, Tuzzio L, Ritzwoller DP, *et al.* Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. *Med Care*. 2009;47:1091-97.
34. Ritzwoller DP, Carroll NM, Delate T, *et al.* Patterns and predictors of first-line chemotherapy use among adults with advanced non-small cell lung cancer in the cancer research network. *Lung Cancer*. 2012;78:245-52.
35. Ritzwoller DP, Carroll N, Delate T, *et al.* Validation of electronic data on chemotherapy and hormone therapy use in HMOs. *Med Care*. 2013;51:e67-73.
36. Ross TR, Ng D, Brown JS, *et al.* The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. *EGEMS (Wash DC)*. 2014;2:1049.
37. Strategic Management Plan 2015 Update. <http://www.naaccr.org/About/NAACCR/StrategicManagementPlanSpring2014Update.aspx>. Accessed March 17, 2016.
38. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron*. 1976;16:31-41.
39. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stats Med*. 2004;23:2567-86.
40. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: John Wiley & Sons; 1998.
41. Hosmer DW, Lemeshow S. *Applied Logistic Regression*, 2nd ed. New York: John Wiley & Sons; 2000.
42. Lyman GH. Impact of chemotherapy dose intensity on cancer patient outcomes. *J Natl Compr Canc Netw*. 2009;7:99-108.
43. Bosly A, Bron D, Van Hoof A, *et al.* Achievement of optimal average relative dose intensity and correlation with survival in diffuse large B-cell lymphoma patients treated with CHOP. *Ann Hematol*. 2008;87:277-83.
44. Pettengell R, Schwenkglenks M, Leonard R, *et al.* Neutropenia occurrence and predictors of reduced chemotherapy delivery: results from the INC-EU prospective observational European neutropenia study. *Supportive Care Cancer*. 2008;16:1299-309.
45. Shayne M, Culakova E, Wolff D, *et al.* Dose intensity and hematologic toxicity in older breast cancer patients receiving systemic chemotherapy. *Cancer*. 2009;115:5319-28.
46. Chirivella I, Bermejo B, Insa A, *et al.* Optimal delivery of anthracycline-based chemotherapy in the adjuvant setting improves outcome of breast cancer patients. *Breast Cancer Res Treat*. 2009;114:479-84.
47. Loibl S, Skacel T, Nekljudova V, *et al.* Evaluating the impact of Relative Total Dose Intensity (RTDI) on patients' short- and long-term outcome in taxane- and anthracycline-based chemotherapy of metastatic breast cancer: a pooled analysis. *BMC Cancer*. 2011;11:131.
48. Lyman GH, Barron RL, Natoli JL, Miller RM. Systematic review of efficacy of dose-dense versus non-dose-dense chemotherapy in breast cancer, non-Hodgkin lymphoma, and non-small cell lung cancer. *Critical Rev Oncol/Hematol*. 2012;81:296-308.
49. Fauci JM, Whitworth JM, Schneider KE, *et al.* Prognostic significance of the relative dose intensity of chemotherapy in primary treatment of epithelial ovarian cancer. *Gynecol Oncol*. 2011;122:532-35.
50. Hanna RK, Poniewierski MS, Laskey RA, *et al.* Predictors of reduced relative dose intensity and its relationship to mortality in women receiving multi-agent chemotherapy for epithelial ovarian cancer. *Gynecol Oncol*. 2013;129:74-80.
51. Luciani A, Bertuzzi C, Ascione G, *et al.* Dose intensity correlate with survival in elderly patients treated with chemotherapy for advanced non-small cell lung cancer. *Lung Cancer*. 2009;66:94-96.
52. Ramsey SD, McCune JS, Blough DK, *et al.* Colony-stimulating factor prescribing patterns in patients receiving chemotherapy for cancer. *Am J Manag Care*. 2010;16:678-86.
53. Schnipper LE, Smith TJ, Raghavan D, *et al.* American Society of Clinical Oncology identifies five key opportunities to improve care and reduce costs: the top five list for oncology. *J Clin Oncol*. 2012;30:1715-24.
54. Lyman GH, Dale DC, Legg JC, *et al.* Assessing patients' risk of febrile neutropenia: is there a correlation between physician-assessed risk and model-predicted risk? *Cancer Med*. 2015;4:1153-60.
55. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stats Med*. 2016;35:214-26.