

RESEARCH ARTICLE

Improved risk prediction of chemotherapy-induced neutropenia—model development and validation with real-world data

Mikko S. Venäläinen¹  | Eetu Heervä^{2,3} | Outi Hirvonen^{2,4,5} | Sohrab Saraei¹ | Tomi Suomi¹ | Toni Mikkola⁶ | Maarit Bärlund^{7,8}  | Sirkku Jyrkkö² | Tarja Laitinen^{9,10} | Laura L. Elo^{1,11} 

¹Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland

²Department of Oncology, Turku University Hospital and FICAN West, Turku, Finland

³University of Turku, Turku, Finland

⁴Department of Clinical Oncology, University of Turku, Turku, Finland

⁵Palliative Center, Turku University Hospital, Turku, Finland

⁶Tays Research Services, Clinical Informatics Team, Tampere University Hospital and University of Tampere, Tampere, Finland

⁷Department of Oncology, Tays Cancer Centre, Tampere University Hospital, Tampere, Finland

⁸Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

⁹Department of Pulmonary Medicine, University of Turku and Turku University Hospital, Turku, Finland

¹⁰Administration Center, Tampere University Hospital and University of Tampere, Tampere, Finland

¹¹Institute of Biomedicine, University of Turku, Turku, Finland

Correspondence

Laura L. Elo, Turku Bioscience Centre, Tykistökatu 6 A, FI-20520 Turku, Finland.
Email: laura.elo@utu.fi

Funding information

LLE reports funding from the European Research Council (ERC) [grant number 677943], the Academy of Finland [grant numbers 296801, 310561, 314443, 329278, 335434, and 335611], and the Sigrid Juselius Foundation during the conduct of this study. MSV reports funding from the Academy of Finland [grant number 322123]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Background: The existing risk prediction models for chemotherapy-induced febrile neutropenia (FN) do not necessarily apply to real-life patients in different healthcare systems and the external validation of these models are often lacking. Our study evaluates whether a machine learning-based risk prediction model could outperform the previously introduced models, especially when validated against real-world patient data from another institution not used for model training.

Methods: Using Turku University Hospital electronic medical records, we identified all patients who received chemotherapy for non-hematological cancer between the years 2010 and 2017 ($N = 5879$). An experimental surrogate endpoint was first-cycle neutropenic infection (NI), defined as grade IV neutropenia with serum C-reactive protein >10 mg/l. For predicting the risk of NI, a penalized regression model (Lasso) was developed. The model was externally validated in an independent dataset ($N = 4594$) from Tampere University Hospital.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

Results: Lasso model accurately predicted NI risk with good accuracy (AUROC 0.84). In the validation cohort, the Lasso model outperformed two previously introduced, widely approved models, with AUROC 0.75. The variables selected by Lasso included granulocyte colony-stimulating factor (G-CSF) use, cancer type, pre-treatment neutrophil and thrombocyte count, intravenous treatment regimen, and the planned dose intensity. The same model predicted also FN, with AUROC 0.77, supporting the validity of NI as an endpoint.

Conclusions: Our study demonstrates that real-world NI risk prediction can be improved with machine learning and that every difference in patient or treatment characteristics can have a significant impact on model performance. Here we outline a novel, externally validated approach which may hold potential to facilitate more targeted use of G-CSFs in the future.

KEYWORDS

chemotherapy, clinical decision support, granulocyte colony-stimulating factor, machine learning, neutropenia

1 | INTRODUCTION

Prophylactic granulocyte colony-stimulating factors (G-CSFs) can be used to shorten neutropenia duration and prevent febrile neutropenia (FN), which are major dose-limiting and resource-intensive complications during cancer chemotherapy. Meta-analyses show that use of G-CSFs significantly reduces the risk of FN, with relative risks of 0.27–0.51 compared to patients with no G-CSF prophylaxis.^{1,2} However, risk assessment based on real-world settings are sparse.³

Significant underuse of G-CSFs in high- and intermediate-risk regimens, and overuse in low-risk regimens, have been reported both in the EU^{4,5} and in the US.^{6–8} G-CSFs contribute also significantly to increased healthcare costs.⁹ Therefore, selective use of G-CSFs based on predetermined FN risk is recommended to optimize the use of healthcare resources.^{10,11}

European and North American guidelines recommend the use of prophylactic G-CSF based on the predetermined risk of FN, classified as low, intermediate, or high, depending on the chemotherapy regimen.^{10,11} Prophylactic G-CSFs are recommended for high risk but not for low-risk regimens. For intermediate-risk regimens, a history of prior neutropenia or leukopenia, age >65 years, presence of comorbid conditions, advanced stage of disease (especially to bone marrow), poor performance status, female sex, and low hemoglobin support the use of prophylactic G-CSFs.^{10–14}

Risk prediction models for neutropenic complications based on retrospective data from the US^{14,15} or on prospective cohorts and trial settings in Europe^{12,16} or tailored for

Novelty and Impact

There are several risk prediction models for chemotherapy-induced neutropenia, but the existing models may not always apply to real-life patients in different healthcare systems. A novel machine learning-based model was developed to predict neutropenic infection risk in cancer patients. The model performance was externally validated in an independent cohort, outperforming two previously introduced conventional models. In the future, our model may facilitate more targeted use of granulocyte colony-stimulating factors.

specific cancer type¹⁷ have been introduced. Model by Lyman et al. is commonly used, which is based on 3760 US patients with different types of cancer treated during 2002–2006¹⁴ and has subsequently been re-evaluated in a larger population.¹⁵ Despite the available models, they have not been typically externally validated against data from institutions other than the one used for model training.¹⁷ In the development of any risk prediction model, external validation is of great importance to verify the generalizability of the model to patients outside the training cohort and should always be done before its implementation and wider use in a clinical setting.¹⁸

In addition to the lack of external validation, the development of previous models has typically relied on conventional modeling strategies whose performances

might be improved with machine learning-based modeling approaches.¹⁹ For example, the widely used model by Lyman et al. is a multivariable logistic regression model developed by applying a stepwise variable selection procedure¹⁴ in which the variable selection process is known to be more unstable compared to penalized approaches and can magnify problems associated with model overfitting.^{19,20} The recently re-evaluated version of this model introduced by Li et al. also uses the same modeling approach and same variables as the original study but with few variable-related modifications made based on clinical and numerical rationale.¹⁵

The aim of the present study was to evaluate if a machine learning-based risk prediction model could outperform previously introduced models by Lyman et al. and Li et al.^{14,15} for neutropenic complications in real-world patient data. To avoid overfitting and to verify model generalizability, we carried out external validation against independent data from another university hospital.

2 | MATERIAL AND METHODS

2.1 | Patients

We gathered comprehensive clinical data on all patients who received intravenous chemotherapy between January 2010 and December 2017 at Turku University Hospital, Finland, covering the population of 480 000 inhabitants in Southwest Finland. This data included the patient's sex, age, time of death, weight, height, body temperatures, given diagnoses (ICD-10 codes), electronic prescriptions, intravenous chemotherapies and monoclonal antibodies, G-CSF therapies, laboratory values, and hospital in- and outpatient visits.

All patients aged at least 18 years with ICD-10 code C00-79 before receiving their first dose of intravenous chemotherapy, either in curative or palliative regimens, were included (Figure 1). Patients treated with investigational products or non-chemotherapy regimens such as immune checkpoint inhibitors (232 patients), patients with nonmelanoma skin cancer (9 patients), and patients having multiple primary malignancies (511 patients) were excluded. This left a total of 5879 patients in the analyses. For model training and internal testing, the data were randomized into separate training ($N = 3920$, two thirds of the data) and test ($N = 1959$, one third of the data) cohorts.

Using the same protocol, we collected an independent validation cohort from Tampere University Hospital with the catchment population of 515 100 inhabitants. In total, 4594 patients treated between January 2014 and June 2019 were identified.

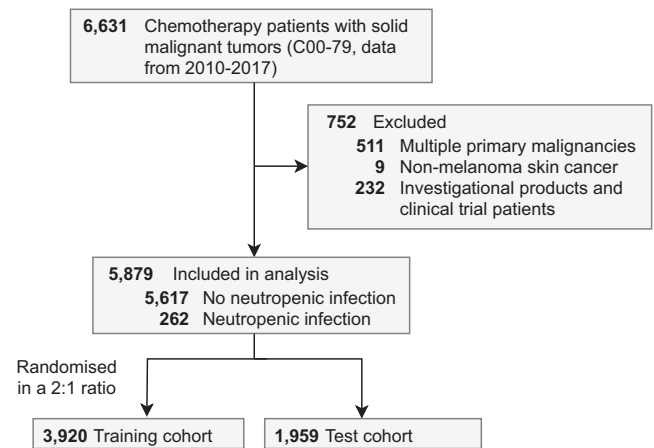


FIGURE 1 Selection of patients into the Turku University Hospital cohort

The data gathering and analysis were performed with research permissions granted by institutional review boards of Turku and Tampere University Hospitals. Additional information on data gathering and confidentiality can be found in the Supplementary Material.

2.2 | Study endpoint and candidate predictors

Febrile neutropenia is a classical endpoint in neutropenia studies, but the definition of FN varies depending on the source.^{14,21} We designed neutropenic infection (NI) as primary endpoint, which was defined as grade IV neutropenia (absolute neutrophil count (ANC) $<0.5 \times 10^9/l$ according to Common Terminology Criteria for Adverse Events (CTCAE) version 4.0) within 14 days of the first chemotherapy infusion combined with subsequent serum C-reactive protein (CRP) level $>10 \text{ mg/l}$ within 5 days (of neutropenia). Patients who did not undergo any laboratory measurements within 14 days of the first chemotherapy infusion or failed to fulfill either of the two criteria were considered patients without NI. To further verify the selected primary study endpoint, the occurrence of NI was compared with admissions to any tertiary care hospital ward during the NI episode. We also tested FN as an endpoint, defined as body temperature $\geq 38^\circ\text{C}$ and ANC $<1.0 \times 10^9/l$ according to CTCAE version 4.0 but ignoring the required 1-hour limit for fever. We focused on the first-cycle of chemotherapy when the patients are at highest risk of developing neutropenia.^{12,14}

The candidate predictors for estimating the risk of NI are provided in Table 1 and in the Supplementary Material with additional details.

TABLE 1 Characteristics of the Turku University Hospital cohort according to the occurrence of neutropenic infection during the first round of chemotherapy

	Neutropenic infection		<i>p</i> value ^b
	No	Yes	
	<i>N</i> = 5617	<i>N</i> = 262	
Demographics			
Sex, <i>N</i> (%)			<0.001
Male	2228 (40)	41 (16)	
Female	3389 (60)	221 (84)	
Age, <i>N</i> (%)			<0.001
<40	260 (5)	20 (8)	
40–65	2691 (48)	173 (66)	
>65	2658 (47)	69 (26)	
Mean BMI, kg/m ² (standard deviation)	26.6 (5.4)	26.3 (4.9)	0.7
Mean body surface area, m ² (standard deviation)	1.9 (0.2)	1.8 (0.2)	0.009
Use of prophylactic G-CSFs, <i>N</i> (%)	306 (5)	8 (3)	0.1
Comorbidities [ICD–10], <i>N</i> (%)			
COPD [J44, J96]	282 (5)	13 (5)	1.0
Coronary heart disease [I25]	369 (7)	10 (4)	0.1
Diabetes [E10-E14]	524 (9)	10 (4)	0.003
Heart failure [I50]	123 (2)	6 (2)	1.0
Renal impairment [N17-N19]	73 (1)	3 (1)	1.0
Liver failure [K70-K75]	64 (1)	3 (1)	1.0
Rheumatoid arthritis [M05-M07]	118 (2)	4 (2)	0.7
Ulcer disease [K25-K27]	89 (2)	2 (1)	0.4
Metastatic disease			
C77-C79 detected, <i>N</i> (%)	791 (14)	17 (6)	<0.001
Laboratory test results, mean (standard deviation)			
Absolute neutrophil count [$\times 10^9/l$]	4.4 (2.4)	3.6 (2.8)	<0.001
Alanine aminotransferase [U/l]	29.2 (31.1)	27.5 (21.4)	0.7
Alkaline phosphatase [U/l]	100.0 (126.8)	80.1 (64.0)	<0.001
Average red blood cell size [fl]	89.8 (5.3)	90.4 (4.7)	0.2
Blood hematocrit [%]	39.7 (4.1)	40.2 (3.4)	0.05
Blood hemoglobin [g/l]	131.4 (15.3)	133.8 (12.8)	0.03
Hemoglobin amount per red blood cell [pg]	29.8 (2.3)	30.2 (1.9)	0.01
Leukocyte count [$\times 10^9/l$]	7.4 (5.4)	6.6 (3.1)	<0.001
Plasma bilirubin [$\mu\text{mol/l}$]	9.5 (8.5)	9.4 (5.6)	0.2
Plasma potassium [mmol/l]	4.0 (0.4)	4.1 (0.3)	0.9
Plasma sodium [mmol/l]	140.3 (3.5)	140.8 (3.7)	0.002
Red blood cell count [$\times 10^{12}/l$]	4.4 (0.5)	4.5 (0.4)	0.3
Serum creatinine [$\mu\text{mol/l}$]	73.1 (19.8)	71.7 (16.8)	0.5
Thrombocyte count [$\times 10^9/l$]	300.0 (104.8)	267.9 (69.5)	<0.001
Planned relative dose intensity, <i>N</i> (%)			
< 85%	1103 (25)	24 (10)	<0.001
\geq 85%	3224 (75)	205 (90)	

TABLE 1 (Continued)

	Neutropenic infection		p value ^b
	No	Yes	
	N = 5617	N = 262	
Intravenous treatment regimens, N (%)			
Alkylating agents	880 (16)	15 (6)	<0.001
Anthracyclines	909 (16)	14 (5)	<0.001
Antimetabolites	2591 (46)	40 (15)	<0.001
Antitumor antibiotics	126 (2)	3 (1)	0.3
Monoclonal antibodies	439 (8)	81 (31)	<0.001
Platinum	2154 (38)	40 (15)	<0.001
Taxanes	1733 (31)	199 (76)	<0.001
Topoisomerase inhibitors	419 (7)	28 (11)	0.07
Vinca alkaloids	188 (3)	3 (1)	0.07
Cancer group [ICD-10], N (%)			
Breast [C50]	1845 (33)	204 (78)	<0.001
Central nervous system [C70-72]	80 (1)	0 (0)	
Colorectal [C18-20]	692 (12)	6 (2)	
Female reproductive [C51-57]	436 (8)	3 (1)	
Gastric [C15-16]	241 (4)	0 (0)	
Head and neck [C00-14, C30-32]	376 (7)	0 (0)	
Lung, non-small cell [C33-35]	530 (10)	20 (8)	
Lung, small cell [C33-35]	161 (3)	11 (4)	
Melanoma [C43]	49 (1)	0 (0)	
Other gastrointestinal [C17, C21, C22, C26]	78 (1)	1 (0)	
Pancreas and gallbladder [C23-25]	301 (5)	2 (1)	
Prostate [C61]	286 (5)	2 (1)	
Sarcoma [C40-41, C46-49]	69 (1)	5 (2)	
Testicular [C62]	93 (2)	3 (1)	
Urinary tract [C65-68]	261 (5)	3 (1)	
Other ^a	119 (2)	2 (1)	

^a Category includes all remaining ICD-10 codes from C00-79.

^b Comparisons between the groups either having or not having NI were tested using the Mann-Whitney test for continuous variables and the chi-squared test or Fisher's exact test ($N < 5$) for categorical variables.

2.3 | Statistical analysis and model development

We applied penalized logistic regression with least absolute shrinkage and selection operator (Lasso) penalty to the training cohort to construct a multivariable model for predicting the individualized risk of NI. To identify the most influential predictors and to account for model variability due to random subsampling, the model construction was performed iteratively in multiple steps similarly as before.²²⁻²⁴ In previous studies, this approach has led to models with fewer variables but retaining the same prediction accuracy as the more complex models. During model

development and validation, only patients with complete data for the selected predictors were used.

The performance of the Lasso model was compared against previously introduced model for multiple cancer types by Lyman et al.¹⁴ as well as the revised version of the Lyman model introduced by Li et al.¹⁵ (Table S1). The comparisons were done using only those cancer types that were shared by all three studies (breast, ovarian, colorectal, small cell, and non-small cell lung cancer) and patients for whom all risk estimates could be determined. Furthermore, patients who received G-CSFs were excluded from this comparison since they were not included in the development of Li model. Finally, we compared the performance of the

Lasso model also with a model developed using the conventional stepwise variable selection procedure.

All statistical analyses and modeling were carried out using the R statistical computing environment version 3.4.3 (R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>). For penalized regression, implementation available in the R package *glmnet* (version 2.0–16)²⁵ was used. The discrimination performances of the risk assessment models were evaluated in terms of area under the receiver operating characteristic curve (AUROC) and compared using the DeLong test²⁶ implemented in the R package *pROC* (version 1.12.1). Comparisons between the groups either having or not having NI were tested using the Mann–Whitney test for continuous variables and the chi-squared test or Fisher's exact test ($N < 5$) for categorical variables. The level of significance was set at $p < 0.05$. Additional details on model development and statistical analyses can be found in the Supplementary Material.

3 | RESULTS

3.1 | Study populations

In total, the Turku University Hospital cohort consisted of 5879 patients (Table 1). Among these, NI occurred in 262 (4%) of the patients, of whom 225 (86%) were also subsequently admitted to a hospital ward. Out of 5879 patients, 314 (5%) received G-CSFs as primary prophylaxis, of whom eight (3%) developed NI. The patients who had NI were typically women ($p < 0.001$) with breast cancer ($p < 0.001$) treated with higher relative dose intensity (RDI) ($p < 0.001$) and had lower levels of blood ANC ($p < 0.001$), leukocytes ($p < 0.001$), and thrombocytes ($p < 0.001$) in the beginning of the treatment (Table 1). Despite the lower ANC and thrombocyte counts, for nearly all patients (98%) these were above the lower limit of the reference ($>1.5 \times 10^9/l$ for ANC and $>150 \times 10^9/l$ for thrombocyte count).

The validation cohort from Tampere University Hospital included 4594 patients who showed similar distributions in age, sex, comorbidities, cancer types, and RDI compared to the Turku University Hospital cohort (Table S2). The increased risk of NI associated with lower ANC and thrombocyte counts, however, were not observed in the validation cohort.

3.2 | Model development and validation

Of the more than 30 variables included (Table 1), a subset of 10 variables (Table 2) were selected by the Lasso

model for accurate NI risk predictions in the training cohort (AUROC 0.87, 95% confidence interval [CI] 0.84–0.89). The selected model showed similar performance also in the internal test cohort indicating good generalizability (AUROC 0.85, 95% CI 0.80–0.90). As expected, use of prophylactic G-CSFs and reduced RDI were among the most influential predictors decreasing NI risk (Table 2). Intravenous treatment regimens involving antimetabolites demonstrated also decreased NI risk compared to other treatment regimens. Of all cancer types, breast cancer and sarcoma patients showed increased NI risk. The other variables increasing NI risk were use of taxanes alone or in combination with monoclonal antibodies, use of topoisomerase inhibitors, and low pre-treatment ANC and thrombocyte counts. The majority of taxanes used were docetaxel 80 mg/m² Q3W to treat breast cancer (65%), docetaxel 50 mg/m² Q2W to treat metastatic castration-resistant prostate cancer (14%), and paclitaxel to treat gynecological cancers (11%). We observed that among patients treated with taxanes at full RDI, NI risk was 12%, compared to 3% in those treated with reduced RDI.

Overall, the Lasso model predicted NI risk with high accuracy in the Turku University Hospital cohort (AUROC 0.86, 95% CI 0.84–0.89) as well as with fair accuracy in the Tampere University Hospital validation cohort (AUROC 0.73, 95% CI 0.69–0.77). For cancer types included in the development of previous models, the Lasso model significantly outperformed the previous models in both Turku University Hospital (Lyman model AUROC 0.47, $p < 0.001$; Li model AUROC 0.78, $p < 0.001$) and Tampere University Hospital cohorts (Lyman model AUROC 0.53, $p < 0.001$; Li model AUROC 0.70, $p = 0.01$), with AUROCs of 0.84 and 0.75, respectively (Table 3). The discrimination performance of the model developed using the conventional stepwise variable selection procedure was comparable to the Lasso model in the Turku University Hospital cohort but significantly worse in the Tampere University Hospital cohort (Supplementary Material).

In Turku University Hospital cohort, 56% of the patients with NI also met the criteria for FN. Overall, FN was observed in 221 (4%) of patients. When applied to predict the potential FN cases, the Lasso model significantly outperformed (AUROC 0.77) the previously introduced Lyman (AUROC 0.50, $p < 0.001$) and Li models (AUROC 0.73, $p = 0.007$) (Table 3).

3.3 | Assessing the effect of G-CSFs on predicted NI risk

We observed that, especially in the Turku University Hospital cohort, the higher the predicted risk of NI, the

TABLE 2 Coefficients and covariates in the Lasso risk assessment model for the occurrence of neutropenic infection during the first round of chemotherapy in the training cohort

Covariate	Coefficient ^a	Effect	OR (95% CI) ^b	<i>p</i> value
Intercept	0.477	Baseline risk	–	–
Cancer type				
Breast cancer	2.361	Increased risk	7.20 (5.39–9.77)	<0.001
Sarcoma	3.694	Increased risk	1.57 (0.55–3.55)	0.34
Laboratory test results				
Neutrophil count [$\times 10^9/l$] (per ln increase)	–0.282	Decreased risk	0.41 (0.31–0.56)	<0.001
Thrombocyte count [$\times 10^9/l$] (per ln increase)	–0.966	Decreased risk	0.44 (0.30–0.64)	<0.001
Treatment regimen				
Use of taxanes	1.262	Increased risk	7.10 (5.35–9.54)	<0.001
Combined use of taxanes and monoclonal antibodies	0.871	Increased risk	8.53 (6.37–11.35)	<0.001
Use of topoisomerase inhibitors	3.305	Increased risk	1.49 (0.97–2.19)	0.05
Use of antimetabolites	–0.787	Decreased risk	0.21 (0.15–0.29)	<0.001
Actions to reduce risk				
Use of G-CSF	–1.780	Decreased risk	0.55 (0.25–1.04)	0.10
Relative dose intensity <85%	–0.814	Decreased risk	0.34 (0.22–0.51)	<0.001

^aCoefficients indicate the impact of a 1-unit change in a predictor variable on the response variable when the other predictors are held constant.

^bThe odds ratios (OR) and corresponding *p* values were estimated separately using univariable logistic regression without penalization. CI denotes confidence interval.

TABLE 3 Comparison of the discrimination performances of the developed Lasso model, Lyman model,¹⁴ and Li model¹⁵ in Turku University Hospital and Tampere University Hospital validation cohorts for the occurrence of neutropenic infection and febrile neutropenia

Model	Turku University Hospital (<i>N</i> = 2101)				Tampere University Hospital Validation cohort (<i>N</i> = 1937)	
	Outcome: Neutropenic infection		Outcome: Febrile neutropenia		Outcome: Neutropenic infection	
	AUROC (95% CI)	<i>p</i> value ^a	AUROC (95% CI)	<i>p</i> value ^a	AUROC (95% CI)	<i>p</i> value ^a
Lasso	0.84 (0.81–0.86)	–	0.77 (0.73–0.81)	–	0.75 (0.69–0.77)	–
Lyman	0.47 (0.43–0.50)	< 0.001	0.50 (0.46–0.54)	< 0.001	0.53 (0.47–0.59)	< 0.001
Li	0.78 (0.75–0.80)	< 0.001	0.73 (0.70–0.76)	0.007	0.70 (0.66–0.74)	0.01

Note: Abbreviations: AUROC, area under the receiver operating characteristic curve; CI, confidence interval.

^a*p* values are reported for comparisons with the Lasso model.

greater benefits from G-CSFs can be expected in terms of reducing the risk (Figure 2). Among patients with the highest predicted NI risk (>20%), the observed NI rate was 29% (79 out of 270 patients) without the use of G-CSF but roughly 3% when G-CSF was used (2 out of 59 patients). The proportion of patients receiving G-CSFs was also highest in this category (18%) compared to categories with predicted risk of 5–20% or <5% with 116 out of 850 patients (14%) or 52 out of 2502 patients (2%) receiving G-CSFs, respectively. A practical example on how to apply the fitted Lasso model to evaluate the effect of G-CSFs on NI risk is provided in the Supplementary Material.

4 | DISCUSSION

The decision whether to use prophylactic G-CSFs during chemotherapy is a daily question in oncological practice. Our study introduces a novel machine learning-based model to predict NI risk when initiating chemotherapy for cancer patients. The Lasso model uses a limited number of routinely used variables and showed excellent capabilities in predicting NI risk in both training and external validation real-world data cohorts, indicating good model generalizability.

The Lasso model outperformed both previously introduced models as well as a model based on stepwise

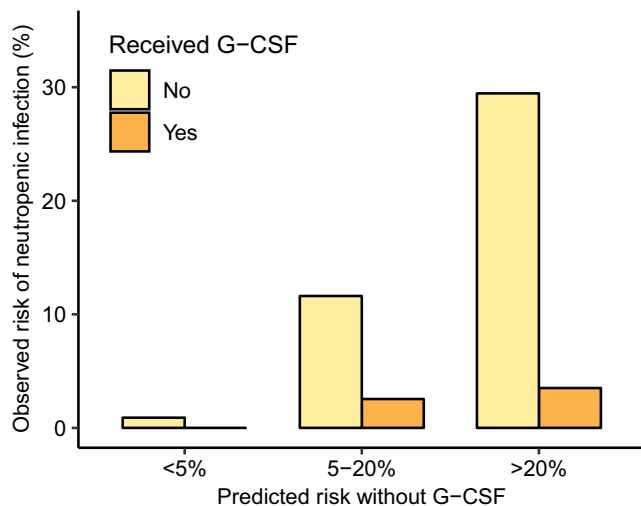


FIGURE 2 Observed rates of NI (y-axis) for patients with and without G-CSFs in different categories of predicted risk (x-axis) in the Turku University Hospital cohort. The risk of NI was determined only for patients with complete baseline information for variables included in the Lasso risk assessment model ($N = 3861$)

variable selection procedure in predicting NI risk with even fewer required input variables suggesting that machine learning-based modeling can improve both usability and predictive performance over more conventional approaches. This finding held true also in the validation cohort. Lack of external validation is a common problem in neutropenia prediction studies, limiting their implementation in clinical practice.

The improved performance is explained by the different input variables resulting from the applied iterative variable selection procedure used successfully also before.²²⁻²⁴ Alternatively, poorly adjusted regression coefficients due to inherent differences between patient populations and local healthcare practices or overfitting may explain the worse performance of previous models. For example, the Lyman model had drastically different coefficients for different treatment regimens compared to Lasso and Li models (Table 2, Table S1) which might explain its underperformance. This highlights also the importance of model validation in the target population before use.

The most common clinically monitored neutropenic complication is FN, for which several different diagnostic criteria can be found in the literature.^{10,14,21} However, FN was a challenging outcome measure in a real world, retrospective study setting, as comprehensive body temperature information is usually unavailable. Body temperature is not always documented or unreliable due to fever-lowering medication. Therefore, we chose CRP (>10 mg/l) accompanied with severe neutropenia ($ANC < 0.5 \times 10^9/l$) as the primary endpoint. In clinical practice, it is considered that these laboratory values are always

available when initiating antimicrobial treatment, even if the body temperature is below $38.0^{\circ}C$. Notably, the majority (86%) of the patients identified as having NI were also admitted to a hospital ward, supporting the validity of our primary endpoint. The occurrence of first-cycle NI (4%) was close to the occurrence of first-cycle FN reported in the literature.^{14,15} We also tested our risk assessment model for predicting FN, and observed a high level of accuracy, thus supporting the utility of our model for both FN and NI risk prediction.

Of the more than 30 risk factors studied, the Lasso model eventually required only a subset of 10 variables. As expected, the use of G-CSFs showed a similar risk-reducing effect against neutropenic complications as reported previously.^{1,2,14} Overall, patients receiving taxanes, with or without monoclonal antibodies to treat breast cancer formed the largest group of patients with increased NI risk. This is line with a recent meta-analysis that demonstrated a significantly increased risk of FN with the use of trastuzumab,²⁷ the most commonly used monoclonal antibody in our study cohort. The second risk group identified consisted of heterogeneous sarcoma patients, often treated with aggressive chemotherapy. Topoisomerase inhibitors, such as etoposide used in small-cell lung cancer and irinotecan used in gastrointestinal cancers, increased risk of NI. However, antimetabolites including fluorouracil commonly used in several types of cancer, were associated with reduced relative NI risk. These chemotherapy regimens remained as important predictors for NI regardless of the RDI used.

Reduced RDI, another mechanism to reduce neutropenia risk along with G-CSFs, was also identified as an important variable and could explain why age or comorbidities did not influence the Lasso model. This further reflects the oncology practice in Finland, where older and comorbid patients seem to be treated with reduced doses of chemotherapy. The risk-reducing effect of reduced RDI was consistent with previous reports.^{14,15} Finally, among all the pre-treatment laboratory test results, only lower ANC and thrombocyte counts were associated with elevated NI risk in the final model. Similar effects of ANC and thrombocyte counts on the risk of neutropenic complications have been reported also before.^{14,28} Lower thrombocyte count may reflect bone marrow dysfunction and could therefore be linked to simultaneous leukopenia, but this was not observed in the validation cohort and should be interpreted with caution.

In conclusion, morbidity due to neutropenic complications affects the patient's quality of life, creates substantial costs, and may even threaten the outcome of cancer treatment if the treatment schedule is postponed because of infection. To improve the targeted use of prophylactic G-CSFs, well-calibrated risk models applicable to real-world

data are needed. Here, we demonstrate that risk prediction of neutropenic complications can be improved with machine learning and that even previously validated models do not necessarily lead to correct predictions in all patient populations. Our novel machine learning-based model outperformed both previously introduced models and can be easily applied to identify individuals at high risk of neutropenic complications especially in countries with similar clinical practices. These findings were confirmed in an external validation cohort thus supporting the generalizability and clinical applicability of our model. Overall, the presented method holds potential for avoiding resource-intensive and life-threatening neutropenic complications and could facilitate the proper use of G-CSFs in the future.

ACKNOWLEDGMENTS

The authors thank both Auria and Tays Clinical Informatics Teams for gathering the patient data used in the analysis.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Sirkku Jyrkkiö, Tarja Laitinen and Laura L. Elo contributed to conceptualization and reviewing, and editing the manuscript. Mikko S. Venäläinen and Toni Mikkola were responsible for formal analysis and visualization. Mikko S. Venäläinen, Sohrab Saraei, and Tomi Suomi contributed to methodology. Toni Mikkola performed external validations. Mikko S. Venäläinen and Eetu Heervä wrote the original draft of the manuscript. Eetu Heervä, Outi Hirvonen, Maarit Bärlund, Sirkku Jyrkkiö, and Tarja Laitinen interpreted the data and reviewed the manuscript critically for important intellectual content. Tarja Laitinen and Laura L. Elo provided resources for the project. Laura L. Elo was responsible for funding acquisition and supervised the project. All authors read and approved the final manuscript.

ETHICAL STATEMENT

The requirements for ethical review in Finland are stipulated primarily in the Medical Research Act (488/1999, as amended) and the Act of the Medical Use of Human Organs, Tissues, and Cells (101/2001 as amended). Ethical review is statutorily required for interventional medical research and in some circumstances for studies on human organs, tissues, or cells. According to Finnish legislation, no ethical assessment or approval by an independent review board is mandatory for register studies, meaning research on patient records, or other database data. Therefore, ethical review was not necessary nor was required beforehand this study. However, the data

gathering and analysis were performed with research permissions granted by institutional review boards of Turku and Tampere University Hospitals.

DATA AVAILABILITY STATEMENT

Patient health records are confidential information and cannot be made publicly available. However, access to the datasets can be applied from Auria Clinical Informatics and Tays Research Services, with data permits granted by the Hospital District of Southwest Finland and the Pirkanmaa Hospital District, respectively.

ORCID

Mikko S. Venäläinen  <https://orcid.org/0000-0003-1777-4259>

Maarit Bärlund  <https://orcid.org/0000-0002-1469-5654>

Laura L. Elo  <https://orcid.org/0000-0001-5648-4532>

REFERENCES

- 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.
- Renner P, Milazzo S, Liu JP, Zwahlen M, Birkmann J, Horneber M. Primary prophylactic colony-stimulating factors for the prevention of chemotherapy-induced febrile neutropenia in breast cancer patients. *Cochrane Database of Systematic Reviews*. 2012;10:CD007913.
- Rapoport BL, Aapro M, Paesmans M, et al. Febrile neutropenia (FN) occurrence outside of clinical trials: occurrence and predictive factors in adult patients treated with chemotherapy and an expected moderate FN risk. Rationale and design of a real-world prospective, observational, multinational. *BMC Cancer*. 2018;18:917.
- Van Ryckeghem F, Haverbeke C, Wynendaele W, et al. Real-world use of granulocyte colony-stimulating factor in ambulatory breast cancer patients: a cross-sectional study. *Support Care Cancer*. 2019;27:1099-1108.
- Bokemeyer C, Gascón P, Aapro M, et al. Over- and under-prophylaxis for chemotherapy-induced (febrile) neutropenia relative to evidence-based guidelines is associated with differences in outcomes: findings from the MONITOR-GCSF study. *Support Care Cancer*. 2017;25:1819-1828.
- Barnes G, Pathak A, Schwartzberg L. G-CSF utilization rate and prescribing patterns in United States: associations between physician and patient factors and GCSF use. *Cancer Med*. 2014;3:1477-1484.
- Baig H, Somlo B, Eisen M, Stryker S, Bensink M, Morrow PK. Appropriateness of granulocyte colony-stimulating factor use in patients receiving chemotherapy by febrile neutropenia risk level. *J Oncol Pharm Pract*. 2019;25:1576-1585.
- Zullo AR, Lou U, Cabral SE, Huynh J, Berard-Collins CM. Overuse and underuse of pegfilgrastim for primary prophylaxis of febrile neutropenia. *J Oncol Pharm Pract*. 2019;25:1357-1365.
- Dinan MA, Hirsch BR, Lyman GH. Management of chemotherapy-induced neutropenia: measuring quality, cost, and value. *J Natl Compr Cancer Netw*. 2015;13:e1-7.

10. Aapro MS, Bohlius J, Cameron DA, et al. 2010 update of EORTC guidelines for the use of granulocyte-colony stimulating factor to reduce the incidence of chemotherapy-induced febrile neutropenia in adult patients with lymphoproliferative disorders and solid tumours. *Eur J Cancer*. 2011;47:8-32.
11. Crawford J, Becker PS, Armitage JO, et al. Myeloid Growth Factors, Version 2.2017, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*. 2017;15:1520-1541.
12. López-Pousa A, Rifà J, Casas de Tejerina A, et al. DELFOS Study Group. Risk assessment model for first-cycle chemotherapy-induced neutropenia in patients with solid tumours. *Eur J Cancer Care*. 2010;19:648-655.
13. Schwenkglens M, Pettengell R, Jackisch C, et al. Risk factors for chemotherapy-induced neutropenia occurrence in breast cancer patients: data from the INC-EU Prospective Observational European Neutropenia Study. *Support Care Cancer*. 2011;19:483-490.
14. Lyman GH, Kuderer NM, Crawford J, et al. Predicting individual risk of neutropenic complications in patients receiving cancer chemotherapy. *Cancer*. 2011;117:1917-1927.
15. Li Y, Family L, Chen LH, et al. Value of incorporating newly identified risk factors into risk prediction for chemotherapy-induced febrile neutropenia. *Cancer Med*. 2018;7:4121-4131.
16. Aapro M, Ludwig H, Bokemeyer C, et al. Predictive modeling of the outcomes of chemotherapy-induced (febrile) neutropenia prophylaxis with biosimilar filgrastim (MONITOR-GCSF study). *Ann Oncol*. 2016;27:2039-2045.
17. Lyman GH, Abella E, Pettengell R. Risk factors for febrile neutropenia among patients with cancer receiving chemotherapy: a systematic review. *Crit Rev Oncol/Hematol*. 2014;90:190-199.
18. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
19. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ*. 2015;351:h3868.
20. McNeish DM. Using lasso for predictor selection and to assuage overfitting: a method long overlooked in behavioral sciences. *Multivar Behav Res*. 2015;50:471-484.
21. Klastersky J, de Naurois J, Rolston K, et al. Management of febrile neutropenia: ESMO Clinical Practice Guidelines. *Ann Oncol*. 2016;27:v111-118.
22. Venäläinen MS, Klén R, Mahmoudian M, Raitakari OT, Elo LL. Easy-to-use tool for evaluating the elevated acute kidney injury risk against reduced cardiovascular disease risk during intensive blood pressure control. *J Hypertens*. 2020;38:511-518.
23. Venäläinen MS, Panula VJ, Klén R, et al. Preoperative risk prediction models for short-term revision and death after total hip arthroplasty: data from the finnish arthroplasty register. *JBJS Open Access*. 2021;6:e20.00091.
24. Mahmoudian M, Venäläinen MS, Klén R, Elo LL. Stable iterative variable selection. *Bioinformatics*. 2021. 10.1093/bioinformatics/btab501
25. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1-22.
26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837-845.
27. Funakoshi T, Suzuki M, Muss HB. Infection risk in breast cancer patients treated with trastuzumab: a systematic review and meta-analysis. *Breast Cancer Res Treat*. 2015;149:321-330.
28. Pfeil AM, Vulsteke C, Paridaens R, et al. Multivariable regression analysis of febrile neutropenia occurrence in early breast cancer patients receiving chemotherapy assessing patient-related, chemotherapy-related and genetic risk factors. *BMC Cancer*. 2014;14:201.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Venäläinen MS, Heervä E, Hirvonen O, et al. Improved risk prediction of chemotherapy-induced neutropenia—model development and validation with real-world data. *Cancer Med*. 2021;00:1–10. doi:[10.1002/cam4.4465](https://doi.org/10.1002/cam4.4465)