Universiteit
Leiden
The Netherlands

Predicting time-to-event outcomes under different intervention strategies: methods and applications

Prosepe, I.

Citation

Prosepe, I. (2025, December 3). *Predicting time-to-event outcomes under different intervention strategies: methods and applications*. Retrieved from https://hdl.handle.net/1887/4284487

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4284487

Note: To cite this publication please use the final published version (if applicable).

Chapter 5

Decision-support algorithm for platelet transfusion in preterm infants with severe thrombocytopenia

Hilde van der Staaij, Ilaria Prosepe, Camila Caram-Deelder, Ruth H. Keogh, Emöke Deschmann, Christof Dame, Wes Onland, Sandra A. Prins, Florian Cassel, Esther J. d'Haens, Elke van Westering-Kroon, Peter Andriessen, Sabine L. Vrancken, Christian V. Hulzebos, Daniel C. Vijlbrief, Suzanne F. Fustolo-Gunnink, Karin Fijnvandraat, Enrico Lopriore, Johanna G. van der Bom, Nan van Geloven

Abstract

Importance: Preterm infants with severe thrombocytopenia (platelet count $< 50 \times 10^9/L$) frequently receive platelet transfusions. However, it is unclear in which cases prophylactic transfusion truly reduces bleeding risk or whether it does more harm than good.

Objective: To develop and validate a dynamic prediction model for major bleeding or mortality if prophylactic platelet transfusion were or were not to be given to infants with severe thrombocytopenia.

Design: Model development in an international multicenter cohort (2017-2021) and validation in a national multicenter cohort (2010-2014).

Setting: Fourteen neonatal intensive care units in the Netherlands, Sweden, and Germany participated in the development cohort, and seven Dutch centers in the validation cohort.

Participants: Infants of less than 34 weeks' gestation with severe thrombocytopenia.

Exposure: Two transfusion strategies were contrasted at each prediction time point: receiving a platelet transfusion within 6 hours (prophylaxis) versus no platelet transfusion for 3 days (no-prophylaxis).

Main outcome and measures: The primary outcome was the 3-day risk of major bleeding or mortality, re-estimated every 2 hours during the first week after severe thrombocytopenia onset. Predictors included gestational and postnatal age, small-for-gestational-age, necrotizing enterocolitis, sepsis, mechanical ventilation, vasoactive agents, platelet count, and prior platelet transfusion(s). Landmarking combined with the clone-censor-weight approach enabled dynamic prediction under the two transfusion strategies, accounting for time-varying confounding. Model performance was evaluated in the external validation cohort.

Results: Major bleeding or death occurred in 23% (235/1,042) of infants (median gestational age 28 weeks, median birth weight 900 gr) in the development cohort and 21% (135/637) in the validation cohort. The time-dependent area under the ROC curve was 0.69 (95%CI 0.60-0.76) for the prophylaxis and 0.85 (95%CI 0.76-0.92) for the no-prophylaxis strategy. Calibration plots showed good calibration. The model predicted substantial variation in benefit and harm of transfusion depending on the infant's current clinical condition.

Conclusion and relevance: Our model provides dynamic, individualized absolute risks of bleeding or death with and without subsequent platelet transfusion in infants with severe thrombocytopenia, showing good predictive performance. Its value as a decision-support algorithm for individualized prophylactic platelet transfusion and for reducing transfusion-related harm is ready for evaluation in future studies.

Based on an early version of the paper published in *JAMA*. DOI: 10.1001/jama.2025.14194.

5.1 Introduction

In neonatal intensive care units (NICUs), the majority of platelet transfusions are given to non-bleeding infants with severe thrombocytopenia with the goal to prevent bleeding. This prophylactic platelet transfusion strategy is based on the assumption that increasing the platelet count will reduce the risk of bleeding, given the crucial role of platelets in primary hemostasis. However, multiple studies have shown a weak or absent association between platelet count and bleeding risk in preterm infants[1–7]. In addition, infants with similar platelet counts may have different bleeding risks due to varying clinical conditions.

The PlaNeT-2/MATISSE randomized controlled trial on platelet transfusion thresholds found that using a platelet count threshold of $50 \times 10^9/L$ was associated with a higher risk of major bleeding or death compared with a $25 \times 10^9/L$ threshold [8]. However, this trial was not designed to investigate the risk of major bleeding or mortality under a no-prophylaxis strategy, where transfusions are given only in case of active bleeding. In addition, individual responses to transfusion may vary, and the risk of bleeding or mortality and the effect of prophylactic platelet transfusion may change over time as the infant's clinical condition evolves.

This highlights the need for an individualized approach, taking into account all time-varying information on platelet counts and other factors determining an infant's risk of bleeding and mortality, to identify infants who may benefit from transfusion versus those at risk of transfusion-related harm. Recently, methodology has been developed to train and evaluate prediction models for estimating individualized risks under different treatment strategies from observational data [9–13].

The aim of this study was to develop and validate a decision-support algorithm for prophylactic platelet transfusion to help clinicians evaluate the question: "What is the risk of major bleeding or death for a preterm infant with this risk profile if I were to give a prophylactic platelet transfusion, and what is the risk if I were not to give a transfusion?"

5.2 Methods

5.2.1 Data sources

For model development, we set up the 'PRedicting OutcomeS in Preterm nEonates with thromboCyTopenia' (PROSPECT) study, an international longitudinal cohort study in ten NICUs in the Netherlands, three in Sweden, and one in Germany between January 2017 and December 2021. The Institutional Review Board (IRB) of the Leiden University Medical Center (LUMC), Leiden, the Netherlands, approved the study (protocol 22-3028), followed by local IRBs of each center. Eight centers used an opt-out consent procedure, and the other centers waived the need for informed consent. The study protocol was published on `www.clinicaltrials.gov` (NCT06043050), and the statistical analysis plan was approved by the LUMC Department of Clinical Epidemiology (protocol A219). For model evaluation, we used data from the 'Monitoring Outcome in NEonatal Thrombocytopenia' (MONET) study, a longitudinal cohort study in seven Dutch NICUs between January 2010

and December 2014, as described elsewhere [14]. The studies complied with the Declaration of Helsinki and General Data Protection Regulation. We reported the study following the TRIPOD+AI guidelines, in addition to information on the causal aspects for development and evaluation of an interventional prediction model [9, 15, 16]. We collected data from medical records, laboratory systems, and cranial ultrasound reports, and we performed the analyses in Stata, version 16.1, and R, version 4.1.17 [17–20].

5.2.2 Population

We included all consecutive infants who met the following criteria: gestational age <34 weeks, at least one platelet count $< 50 \times 10^9/L$, and admission to the NICU. Exclusion criteria were severe congenital malformations, major bleeding before severe thrombocytopenia, fetal and neonatal alloimmune thrombocytopenia, and thrombocytopenia due to exchange transfusion. In the validation cohort, previous admission to another NICU was an additional exclusion criterion.

5.2.3 Interventions

We compared two transfusion strategies at each time point (referred to as prediction time points) at which a prophylactic platelet transfusion decision could be made: (i) prophylaxis strategy if an infant were to receive a prophylactic platelet transfusion within 6 hours, and (ii) no-prophylaxis strategy if an infant were not to receive a platelet transfusion for the next 3 days.

5.2.4 Outcome

The primary outcome was a composite of major bleeding or mortality within the next 3 days following each prediction time point. Major bleeding was defined as intraventricular hemorrhage (IVH) grade 3 or any grade IVH with parenchymal periventricular hemorrhagic infarction (PVHI) [21], cerebellar or solitary parenchymal hemorrhage on cranial ultrasound, subdural or epidural hemorrhage causing parenchymal compression, major pulmonary hemorrhage requiring mechanical ventilation or increased ventilatory settings, and gastrointestinal hemorrhage associated with hemodynamic instability. The secondary outcome, major bleeding or mortality within 14 days, was chosen based on the result of the PlaNeT-2/MATISSE trial, which indicated a potential delayed adverse effect of platelet transfusions [8, 22].

5.2.5 Predictors

Predictors were selected a priori based on literature review and expert advice. Time-fixed predictors included gestational age, small-for-gestational-age (SGA, birth weight <p10) [23], and postnatal age at severe thrombocytopenia onset. Time-dependent predictors were mechanical ventilation, most recent platelet count, cumulative number of platelet transfusions, necrotizing

enterocolitis (NEC) Bell's stage ≥IIA [24], clinically suspected/proven sepsis, and vasoactive agents. In addition, time since the first platelet count $< 50 \times 10^9/L$ was included, as well as an interaction between time and SGA, because the association between SGA and bleeding may vary immediately after severe thrombocytopenia onset compared with a few days later [14]. Interactions with platelet transfusion were taken into account by modeling the data separately for the prophylaxis and no-prophylaxis strategies. More details about the targets of estimation ('prediction estimands') are presented in Table 5.3 and 5.4 in the Appendix.

5.3 Statistical analyses

5.3.1 Model development

We combined landmarking with the clone-censor-weight approach [10, 25, 26]. For a detailed discussion of the method, see Sections A and B of the Appendix. In brief, our prediction time points (landmarks) represented 2-hour intervals at which a new prediction was made to accurately capture changes in the infant's condition, using the updated information about the infant up to that point. At every prediction time point, infants were cloned, with one clone assigned to the prophylaxis strategy and the other to the no-prophylaxis strategy. Next, we artificially censored infants when they no longer "adhered" to the designated transfusion strategy (Figure 5.4 of the Appendix) [27]. For the no-prophylaxis strategy, this meant that infants who received a platelet transfusion within 3 days from the prediction time point were censored at the time of transfusion. For the prophylaxis strategy, infants were artificially censored if they did not receive a platelet transfusion within 6 hours. Lastly, we accounted for potential bias due to the censoring of non-adherent infants, by applying inverse probability of censoring (IPC) weighting based on time-dependent confounders, as identified in a directed acyclic graph (Figure 5.5 of the Appendix) [10, 25, 28]. Infants were eligible for inclusion at prediction time points from the first platelet count $< 50 \times 10^9/L$ up to 7 days thereafter at each time their platelet count was $< 50 \times 10^9/L$. If a platelet transfusion was given, a new post-transfusion platelet count was required for inclusion in the next prediction time point. Once major bleeding or death occurred, infants were no longer included. We fitted two weighted Cox proportional hazards landmark models, one for each transfusion strategy, and used bootstrapping to obtain 95% confidence intervals for the hazard ratios [10].

5.3.2 Model evaluation

We assessed how the predictions from the model matched the observed outcomes in the validation dataset, which we reweighted with newly estimated IPC weights derived from the validation dataset [10]. Using the prediction formulas from the developed model, we estimated the 3-day risk of major bleeding or death under both transfusion strategies at each prediction time point for each infant in the validation cohort. In this dataset, some infants transferred to another NICU had censored outcomes, because follow-up ended after transfer to a non-participating NICU. We described the model's ability to correctly discriminate between patients who did and those who did not (yet) experience major bleeding or death using

the time-dependent area under the ROC curve (AUCt) and concordance index (c-index) [10]. To evaluate how well the estimated absolute risks corresponded to observed outcomes, we assessed calibration using the observed-to-expected ratio and calibration plots. Overall prediction accuracy was assessed using the (scaled) Brier score, which combines both calibration and discrimination [10, 29]. We used bootstrapping to construct 95% confidence intervals for the performance measures.

5.4 Results

5.4.1 Development and validation cohort characteristics

In the development cohort, 1,226 of 13,101 infants (9.4%) born <34 weeks had confirmed severe thrombocytopenia. After applying the exclusion criteria, 1,042 infants remained in the development cohort. In the validation cohort, severe thrombocytopenia was observed in 830 of 9,333 infants (8.9%), of whom 637 were included (Figure 5.1). In both cohorts, the median gestational age was 28 weeks, birth weight 900 grams, and severe thrombocytopenia onset at 4 days after birth. Other demographics were also similar, except that more infants received multiple platelet transfusions in the validation cohort (Table 5.1). Platelet transfusions in the development cohort were administered with a median volume of 14 ml/kg (interquartile range, IQR, 10-15), duration of 30 min (IQR 30-60), and infusion rate of 20 ml/kg/hour (IQR 15-30). In the validation cohort, the median volume was 10 ml/kg (IQR 10-15), with no information on duration (Section B of the Appendix). As completeness of data was carefully monitored during data collection, there were no missing data for measured predictors or confounders.

5.4.2 Major bleeding and mortality

Major bleeding or death occurred in 23% of infants (235 out of 1,042) within 10 days of the first platelet count $< 50 \times 10^9/L$ in the development cohort, including 108 major bleedings and 127 deaths. In the validation cohort, 21% of infants (135 out of 637) experienced major bleeding or death, with 63 major bleedings and 72 deaths. Grade 3 IVH was the most common type of major bleeding in both cohorts, followed by major pulmonary and cerebellar hemorrhage (Table 5.6 of the Appendix). Nearly all patients had at least one cranial ultrasound after the onset of severe thrombocytopenia, 98% (1,024/1,042) in the development cohort and 95% (604/637) in the validation cohort. Table 5.7 of the Appendix describes the number of infants with major bleeding or death and those discharged to a stepdown unit or transferred to a non-participating NICU by transfusion strategy, and the corresponding number of predictions.

5.4.3 Dynamic model and model performance

Table 5.2 reports the hazard ratios of the model and Figure 5.2 shows the calibration plot. The median predicted 3-day risk of major bleeding or mortality was 7.4% (IQR 4.2%-14.3%)

Table 5.1: Cohort characteristics at birth, at the onset of severe thrombocytopenia, and summary of covariates at the start of each new prediction time point.
Abbreviations: IQR, interquartile range; AC, antenatal corticosteroids; PC, plarelet count; hsPDA, hemodynamically significant patent ductus arteriosus treated with ibuprofen or indomethacin; IV, intravenous; LP, lumbar puncture; NA, not available.
[a] Gestational age ranged from 22 to 34 weeks in the development cohort and from 24 to 34 weeks in the validation cohort.
[b] At least 2 doses before delivery. These data were missing for n=48 (4.6%) in the development cohort and for n=21 (3.3%) in the validation cohort.
[c] Infants may contribute to multiple prediction time points (i.e., 2-hour landmarks from the onset of severe thrombocytopenia up to 7 days), using the infant's updated information at each new prediction time point.
[d] IV bolus received within the last hour of the prediction time point for the management of (suspected) hypotension. This variable was not collected in the validation cohort.
[e] Planned surgery in the next 6h and planned lumbar puncture in the next 2h of the prediction time point.

Cohort characteristics	**Development cohort**	**Validation cohort**
At birth	**n = 1,042 infants**	**n = 637 infants**
Gestational age in weeks, median (IQR)[a]	28 (26–30)	28 (26–30)
Male sex, no. (%)	613 (59)	370 (58)
Birth weight in grams, median (IQR)	900 (698–1230)	900 (710–1177)
Small-for-gestational age, no. (%)	601 (58)	361 (57)
Cesarean section, no. (%)	762 (73)	450 (71)
Multifetal pregnancy, no. (%)	217 (21)	162 (25)
Complete course of AC[b], no. (%)	591 (57)	409 (64)
Apgar score <5 at 5 min, no. (%)	83 (8)	64 (10)
Perinatal asphyxia, no. (%)	76 (7)	53 (8)
Outborn, no. (%)	115 (11)	69 (11)
At the onset of severe thrombocytopenia	**n=1,042 infants**	**n=637 infants**
Postnatal age in days, median (IQR)	4 (2–8)	4 (2–9)
PC ($\times 10^9$/L), median (IQR)	38 (28–45)	38 (29–45)
PT before PC <50×10^9/L, no. (%)	51 (5)	68 (11)
At the start of each prediction time point	**n = 19,910 predictions[c]**	**n = 13,846 predictions[c]**
Postnatal age in days, median (IQR)	4 (2–9)	7 (2–13)
PC ($\times 10^9$/L), median (IQR)	39 (30–45)	36 (27–44)
No. (%) of prior PT:		
0	13,141 (66)	6,930 (50)
1	3,378 (17)	2,525 (18)
2	1,572 (8)	1,806 (13)
$\geq$3	1,819 (9)	2,585 (19)
Mechanical ventilation, no. (%)	8,130 (41)	6,619 (48)
Necrotizing enterocolitis, no. (%)	3,358 (17)	1,913 (14)
Sepsis, no. (%)	10,798 (54)	8,528 (62)
Receiving vasoactive agents, no. (%)	2,518 (13)	1,713 (12)
IV bolus, no. (%)[d]	80 (0.4)	NA
Pharmacologic treatment for hsPDA, no. (%)	36 (0.2)	182 (1.3)
Planned surgery or LP, no. (%)[e]	446 (2.2)	331 (2.4)

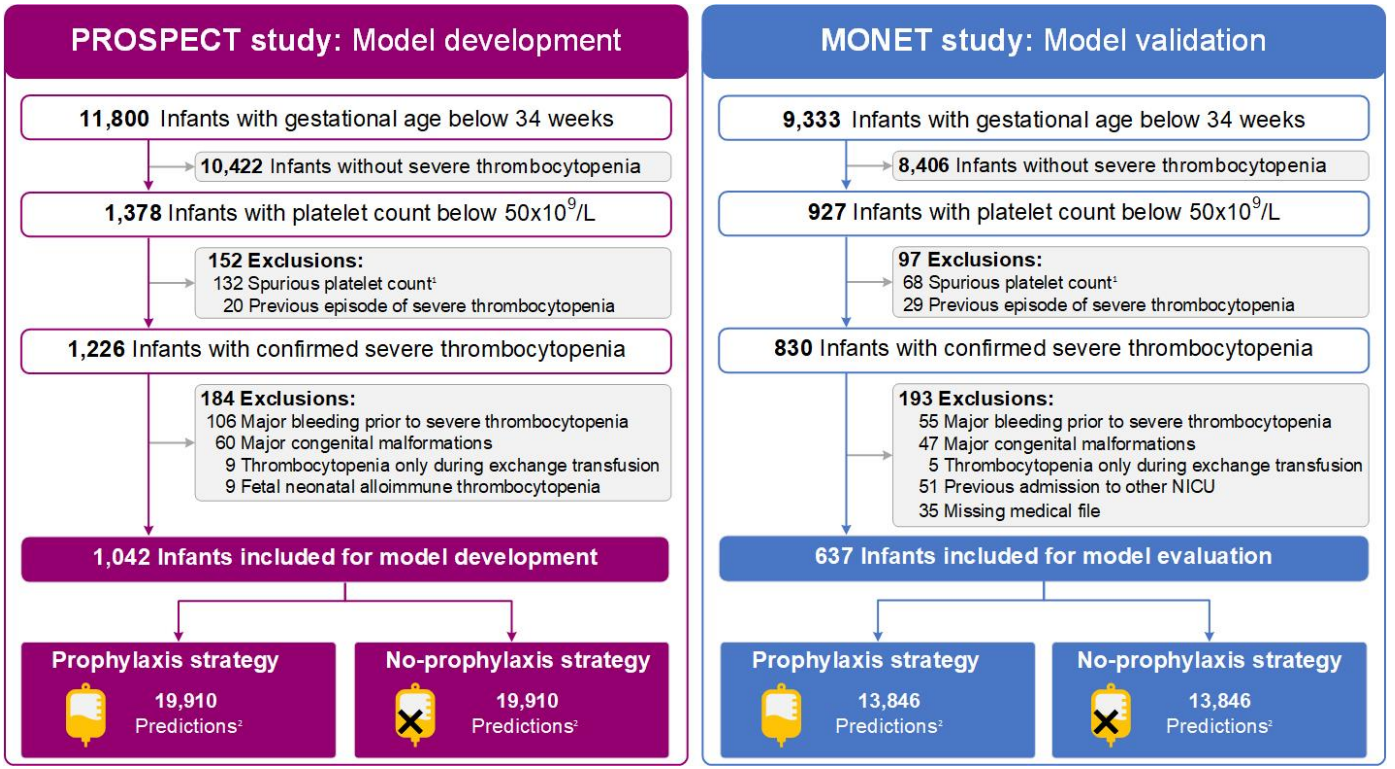


Figure 5.1: Flowchart of the model development cohort (PROSPECT study) and validation cohort (MONET study).
[a] Spurious platelet count due to clots, rapid spontaneous recovery, or labeled as laboratory error.
[b] Infants may contribute to multiple prediction time points (2-hour landmarks from severe thrombocytopenia onset up to 7 days), using the infant's updated information at each new prediction. Because of the 'cloning step' of the clone-censor-weight approach, the number of predictions is the same for both transfusion strategies.

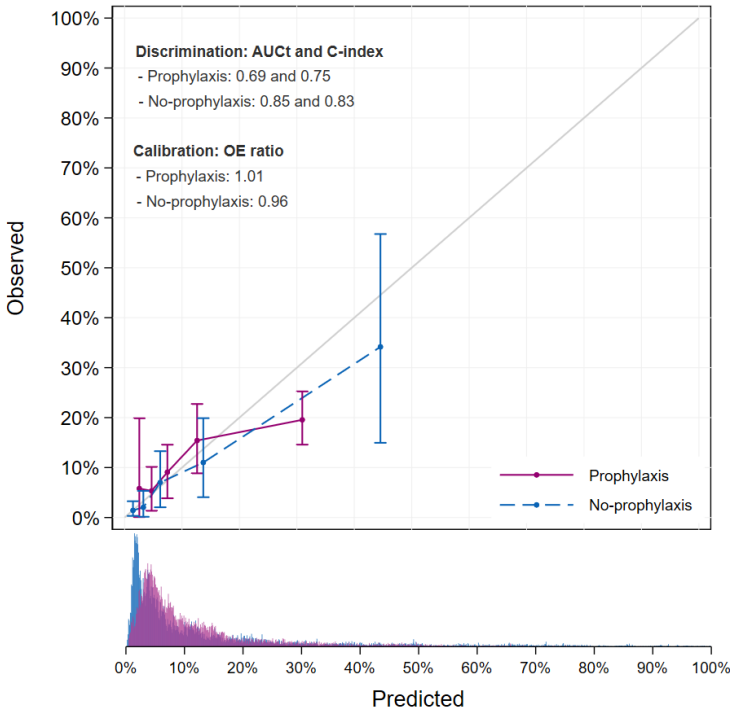


Figure 5.2: Calibration plot of the estimated 3-day risk of major bleeding or mortality under the prophylaxis and no-prophylaxis strategies in the validation cohort.
Abbreviations: AUCt, cumulative/dynamic area under the ROC curve; C-index, concordance index; OE ratio, ratio of observed and expected outcomes.
This calibration plot shows the agreement between the predicted and observed risks of major bleeding or mortality within the next 3 days for the prophylaxis (purple solid line) and no-prophylaxis (blue dashed line) transfusion strategies. The gray solid line at 45 degrees indicates perfect calibration, i.e., predicted and observed proportions are equal. The five dots represent quintiles of the predictions according to the predicted values, with on the x-axis the mean prediction of each quintile and on the y-axis the mean observed proportion of that group, including 95% confidence intervals obtained by bootstrapping using 500 bootstrap samples. The histograms along the x-axis show the distribution of risk estimates for both transfusion strategies.

Table 5.2: Abbreviations: CI, confidence intervals; PC, platelet count; SGA, small-for-gestational age (birth weight < 10^{th} percentile).
[a] Platelet transfusion strategy: prophylaxis (platelet transfusion within the next 6 hours) and no-prophylaxis (no platelet transfusion for the next 3 days). A hazard ratio >1.0 indicates that an increase in the value of the covariate is associated with a higher risk of major bleeding or death, and a hazard ratio <1.0 indicates a lower risk of the outcome. 95% CI were obtained via bootstrapping using 500 bootstrapped samples. It is important to note that the hazard ratios of the covariates in the model should not be interpreted as representing the causal effect on the outcome. The aim of the study is not to estimate the causal effect of each individual variable, but rather to accurately predict the risk of major bleeding or mortality under the two platelet transfusion strategies, using the combination of all predictors in the model. Details of the full prediction formulas are shown in Section D of the Appendix.
[b] Hazard ratios (95% CI) rounded to 3 decimal places: 0.999 (0.998 - 1.000) for the prophylaxis strategy and 0.999 (0.996 - 1.000) for the no-prophylaxis strategy.
[c] For time-varying covariates, the effect of the covariate may change over time, for example, the association between SGA and bleeding or death might differ immediately after the first platelet count $< 50 \times 10^9/L$ compared to days later, influenced by other risk factors and changes in the infant's clinical condition. Landmark time is the prediction time point, i.e., the time in hours (dived by 100) since the first platelet count $< 50 \times 10^9/L$.

	Platelet transfusion strategy[a]	
	Prophylaxis Hazard ratio (95% CI)	**No-prophylaxis** Hazard ratio (95% CI)
Time-fixed variables		
Gestational age (days)	0.98 (0.97 - 0.99)	0.98 (0.95 - 1.00)
Postnatal age at first PC <50×10^9/L (hours)	1.00 (1.00 - 1.00)[b]	1.00 (1.00 - 1.00)[b]
Small-for-gestational age	0.62 (0.40 - 0.95)	0.70 (0.29 - 2.18)
Time-dependent variables		
Mechanical ventilation	1.64 (1.01 - 2.80)	2.54 (1.47 - 5.78)
Necrotizing enterocolitis	0.87 (0.55 - 1.35)	1.85 (0.60 - 4.72)
Sepsis	1.16 (0.83 - 1.76)	0.84 (0.47 - 1.69)
Platelet count ($\times10^9/L$)	0.98 (0.97 - 0.99)	0.94 (0.91 - 0.97)
No. of prior platelet transfusions	1.28 (1.12 - 1.55)	1.25 (0.82 - 1.79)
Vasoactive agents	2.99 (2.18 - 4.10)	2.81 (1.07 - 5.12)
Time-varying covariate effects[c]		
Landmark time (linear)	0.06 (0.02 - 0.19)	1.13 (0.17 - 11.4)
Landmark time (quadratic)	2.09 (0.99 - 4.62)	0.44 (0.09 - 1.11)
Landmark time × SGA	3.98 (1.52 - 10.45)	1.90 (0.36 - 13.6)

under the prophylaxis strategy, and 6.0% (IQR 2.6%-16.3%) under the no-prophylaxis strategy. For both strategies, estimated risks coincided with observed outcomes on average, with observed-to-expected ratios of 1.01 (95% confidence interval [CI], 0.73-1.32) for the prophylaxis strategy and 0.96 (95% CI, 0.51-1.58) for the no-prophylaxis strategy, though risks were somewhat overestimated for the highest 20% of predicted values under both strategies. The time-dependent AUCt was 0.69 (95% CI, 0.60-0.76) for the prophylaxis and 0.85 (95% CI 0.76-0.92) for the no-prophylaxis strategy. Corresponding c-indexes were 0.75 (95% CI, 0.69-0.80) and 0.83 (95% CI, 0.72-0.89). The Brier score was 0.02 (95% CI, 0.02-0.03) for the prophylaxis strategy and 0.07 (95% CI, 0.04-0.09) for the no-prophylaxis strategy. Scaled Brier scores were 15.3% (95% CI, 9.2%-21.9%) for the prophylaxis strategy and 20.9% (95% CI, -16.3%-52.4%) for the no-prophylaxis strategy, suggesting better performance than a null model without covariates fitted to the validation data, which estimates under each strategy the

same average risk for all infants. The wider confidence interval and negative lower bound for the no-prophylaxis strategy indicated more statistical uncertainty. The validation measures for the 14-day prediction window and mean estimated/observed survival curves are reported in Table 5.8 and Figure 5.7 of the Appendix.

5.4.4 Predictions of the dynamic model

Figure 5.8 of the Appendix illustrates the use of the model to repeatedly predict major bleeding or mortality in the next 3 days if an infant were to receive prophylactic platelet transfusion (prophylaxis strategy) and the risk if they were not (no-prophylaxis strategy). Figure 5.3 A/B shows the model's risk estimates for example patients. It demonstrates that an infant's risk of major bleeding or death can vary between the strategies and that infants with the same platelet count may have different risks depending on other clinical characteristics (see figure caption for details). The differences in estimated risks between the strategies help infer individualized platelet transfusion effects. Figure 5.3 C shows the risk under no-prophylaxis minus the risk under prophylaxis against platelet count at each prediction time point in the development dataset. These predicted risk differences vary across the whole platelet count range, indicating benefit of transfusion (lower risk under prophylaxis strategy) for most predictions with platelet counts $< 20 \times 10^9/L$, and more frequently a harmful effect at higher platelet counts. Figure 5.3 D shows the predicted risk difference against the estimated risk under the no-prophylaxis strategy. Overall, when the estimated "untreated risk" is low (<5%), most predictions suggest harm from transfusion, whereas as the "untreated risk" increases, more and more predictions suggest benefit from transfusion.

5.5 Discussion

We developed and validated a dynamic prediction model to estimate individualized absolute risks of major bleeding or death both with and without a subsequent prophylactic platelet transfusion in preterm infants <34 weeks' gestation with severe thrombocytopenia. The findings show substantial variation of potential beneficial and harmful effects of platelet transfusion depending on the infant's current clinical condition. This is, to the best of our knowledge, the first study to develop a decision-support algorithm for prophylactic platelet transfusion decisions in preterm infants.

Model performance measures indicated good performance for the 3-day prediction window [30, 31]. The calibration plot showed good overall agreement between predicted and observed outcomes for both the prophylaxis and no-prophylaxis strategies, with observed-to-expected ratios close to one, albeit the 95% confidence intervals were wide [32, 33]. For the highest 20% of predicted values, both strategies somewhat overestimated the observed risks. This overestimation was similar (up to approximately 10 percentage points) for both transfusion strategies, suggesting that the difference between the estimated risks may be less affected by this. The calibration plot for the 14-day prediction window showed poorer calibration, with observed-to-expected ratios greater than one. This indicates that, on average, the model

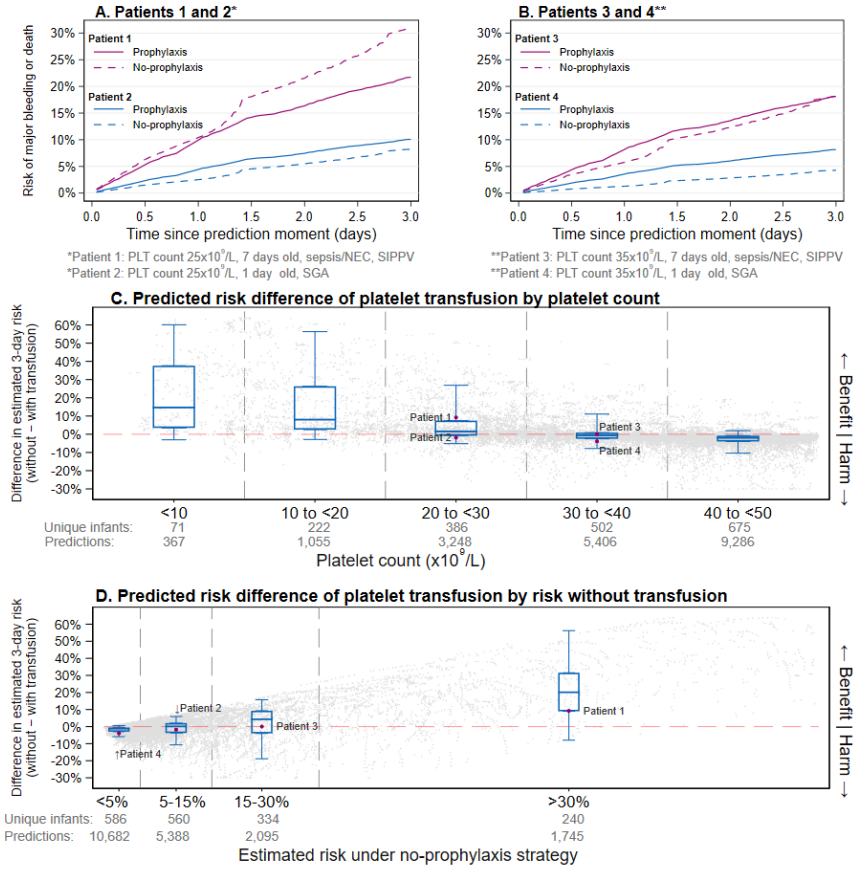


Figure 5.3: Parts A-D. Illustrations of predictions from the dynamic interventional prediction model.
Abbreviations: NEC, necrotizing enterocolitis; SIPPV, synchronized intermittent positive pressure ventilation; SGA, small-for-gestational age.
(A) Risk estimates under the prophylaxis (solid line) and no-prophylaxis (dashed line) strategies for two example patients of 28 weeks' gestation with a platelet count of $25 \times 10^9/L$. Patient 1 (purple lines), 7 days old, has necrotizing enterocolitis/sepsis and is on mechanical ventilation. Patient 2 (blue lines), 1 day old, is small-for-gestational-age and clinically stable.
(B) Risk estimates for patients 3 and 4, with similar characteristics as patients 1 and 2, respectively, but with a platelet count of $35 \times 10^9/L$.
(C) Estimated absolute-risk differences (risk under no-prophylaxis minus risk under prophylaxis strategy) at all prediction time points in the development dataset, plotted against platelet count on the x-axis. Blue dots (jitter plot) show the risk differences. Boxes indicate 25th–75th percentiles, central lines the median, and whiskers the 5th–95th percentiles. The red dashed line represents no risk difference; values >0 indicate benefit with transfusion, and values <0 indicate harm. Numbers below the x-axis show the number of predictions and unique infants per category. The estimated risk differences for example patients 1–4 are shown as purple dots.
(D) Similar to (C), but showing absolute-risk differences between strategies against the 'untreated' risk of major bleeding or death (i.e., under the no-prophylaxis strategy). For example, for patient 1 the estimated 3-day risk of major bleeding or mortality is 31% if receiving a transfusion and 22% if not receiving a transfusion, indicating a 9% lower risk with prophylactic platelet transfusion (potential transfusion benefit).

underestimated 14-day risks of major bleeding or mortality (Figure 5.7 of the Appendix).

Our findings corroborate those of the PlaNeT-2/MATISSE trial which demonstrated that prophylactic platelet transfusions given above a platelet count threshold of $25 \times 10^9/L$ were on average associated with a higher risk of major bleeding or death [8]. Additionally, we report three novelties. First, the presented model allows predicting major bleeding or death at any time point during the first week after the onset of severe thrombocytopenia for two potential (counterfactual) scenarios: i) if an infant were to receive a prophylactic platelet transfusion, and ii) if an infant were not to receive a prophylactic transfusion. The difference between these risks indicates the estimated individual causal effect of prophylactic platelet transfusion [13]. Second, these predicted effects showed considerable variation across different platelet counts, suggesting that there are other factors, besides platelet count, that influence the effect of prophylactic platelet transfusion. Thus, a more elaborate, individualized transfusion approach that incorporates multiple clinical variables, such as based on our decision-support algorithm, may be a better measure to guide prophylactic platelet transfusion decisions than a single platelet count threshold. Third, our findings show that prophylactic platelet transfusions were associated with a lower occurrence of major bleeding or death in some preterm infants with thrombocytopenia. This is biologically plausible, but has not previously been supported by quantitative evidence [34, 35].

Future studies should evaluate whether using the model during decision-making would improve clinical outcomes. This can best be assessed in a randomized impact trial comparing settings where the model is implemented as a prediction tool to inform prophylactic platelet transfusion decisions and settings where current (platelet-count-based) transfusion guidelines are followed [36, 37]. For some infants, the model's recommendations may align with current practice, adding no value, while for others, it may change transfusion decisions and affect outcomes. In preparation for an impact trial, observational data can be used to emulate such a trial to assess in which subgroups, e.g., based on gestational age, postnatal age, or intrauterine growth restriction, the model has the greatest potential to reduce exposure to unnecessary platelet transfusions and transfusion-related bleeding and mortality [38]. In different geographical settings or locations (e.g., non-tertiary care centers), additional validation studies may be needed [31, 39].

Platelet transfusion was modeled as a binary intervention reflecting 'real-world' practice, with a median volume of 15 mL/kg over 30 minutes, and infusion rate of 20 mL/kg/hour, representative for European NICUs [40]. However, the volumes administered in preterm infants are relatively high in relation to their body weight when compared with platelet transfusions in adults (5 ml/kg) [41]. One center in our study used a hyperconcentrated platelet product (3 ml/kg) [42], which may mitigate fluctuations in intravascular volume and blood pressure due to rapid volume expansion, potentially decreasing IVH risk [43]. Future studies should investigate whether smaller volumes and slower rates are safe and effective, and assess the combined volume effect of transfusing different blood components [34, 44, 45].

In addition, other parameters may improve bleeding prediction, such as measures of platelet production, reflected by immature platelet fraction, and measures of primary hemostasis, such as Platelet Function Analyzer-100 closure time and viscoelastic coagulation tests [5, 46–51]. These parameters were not included in the model because they are not (yet) commonly used

in clinical practice and their value in clinical decision-making should be further established.

Our study has several limitations. First, platelet transfusion practice was more liberal in the validation cohort (2010-2014) than in the development cohort (2017-2021) due to guideline changes following the PlaNeT-2/MATISSE publication, which may have affected calibration [31, 39]. Second, we could not predict major bleeding or mortality separately due to sample size limitations. Third, we cannot rule out residual confounding due to unknown or unmeasured factors. Fourth, like for any prediction model, predictions based on a limited sample carry uncertainty [52, 53]. Fifth, confirming that all platelet counts were measured before major bleeding is challenging, as intracranial bleeding is often asymptomatic and detected with time delay by routine cranial ultrasound screening.

Strengths of the study include a large development cohort of over 1,000 severely thrombocytopenic infants. Readily obtainable predictors and confounders were selected before analyses and data collection was meticulous. Evaluation of the model in a cohort distinct from the development dataset showed good predictive performance.

In conclusion, this dynamic prediction model can be used to repeatedly predict individualized absolute risks of major bleeding or death if an infant were and were not to receive a prophylactic platelet transfusion. This is a promising decision-support algorithm to individualize prophylactic platelet transfusion decisions and should be further evaluated in impact studies.

References

1. Stanworth SJ, Clarke P, Watts T, *et al.* Prospective, Observational Study of Outcomes in Neonates With Severe Thrombocytopenia. *Pediatrics.* 2009;124(5):e826-e834.
2. Sparger KA, Assmann SF, Granger S, *et al.* Platelet Transfusion Practices Among Very-Low-Birth-Weight Infants. *JAMA Pediatrics.* 2016;170(7):687.
3. Muthukumar P, Venkatesh V, Curley A, *et al.* Severe Thrombocytopenia and Patterns of Bleeding in Neonates: Results from a Prospective Observational Study and Implications for Use of Platelet Transfusions. *Transfusion Medicine.* 2012;22(5):338–343.
4. Baer VL, Lambert DK, Henry E, Christensen RD. Severe Thrombocytopenia in the NICU. *Pediatrics.* 2009;124(6):e1095-e1100.
5. Deschmann E, Saxonhouse MA, Feldman HA, Norman M, Barbian M, Sola-Visner M. Association of Bleeding Scores and Platelet Transfusions With Platelet Counts and Closure Times in Response to Adenosine Diphosphate (CT-ADPs) Among Preterm Neonates With Thrombocytopenia. *JAMA Network Open.* 2020;3(4):e203394.
6. Resch E, Hinkas O, Urlesberger B, Resch B. Neonatal Thrombocytopenia—Causes and Outcomes Following Platelet Transfusions. *European Journal of Pediatrics.* 2018;177(7):1045–1052.
7. van der Staaij H, Hooiveld NMA, Caram-Deelder C, *et al.* Most Major Bleeds in Preterm Infants Occur in the Absence of Severe Thrombocytopenia: An Observational Cohort Study. *Archives of Disease in Childhood - Fetal and Neonatal Edition.* 2025;110(2):122–127.
8. Curley A, Stanworth SJ, Willoughby K, *et al.* Randomized Trial of Platelet-Transfusion Thresholds in Neonates. *New England Journal of Medicine.* 2019;380(3):242–251.
9. Luijken K, Morzywołek P, Van Amsterdam W, *et al.* Risk-Based Decision Making: Estimands for Sequential Prediction Under Interventions. *Biometrical Journal.* 2024;66(8):e70011.
10. Keogh RH, van Geloven N. Prediction Under Interventions: Evaluation of Counterfactual Performance Using Longitudinal Observational Data. *Epidemiology.* 2024;35(3):329–339.
11. van Geloven N, Swanson SA, Ramspek CL, *et al.* Prediction Meets Causal Inference: The Role of Treatment in Clinical Prediction Models. *European Journal of Epidemiology.* 2020;35(7):619–630.
12. Dickerman BA, Hernán MA. Counterfactual Prediction Is Not Only for Causal Inference. *European Journal of Epidemiology.* 2020;35(7):615–617.
13. Nguyen TL, Collins GS, Landais P, Le Manach Y. Counterfactual Clinical Prediction Models Could Help to Infer Individualized Treatment Effects in Randomized Controlled Trials—An Illustration with the International Stroke Trial. *Journal of Clinical Epidemiology.* 2020;125:47–56.
14. Fustolo-Gunnink SF, Fijnvandraat K, Putter H, *et al.* Dynamic Prediction of Bleeding Risk in Thrombocytopenic Preterm Neonates. *Haematologica.* 2019;104(11):2300–2306.
15. Collins GS, Moons KGM, Dhiman P, *et al.* TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods. *BMJ.* 2024:e078378.
16. van Amsterdam W. Appendix to: Prognostic Models for Decision Support Need to Report Their Targeted Treatments and the Expected Changes in Treatment Decisions. 2024.
17. Castor EDC CEDCS. 2016.
18. R Development Core Team, R: A Language and Environment for Statistical Computing. 2023.
19. StataCorp, Stata Statistical Software: Release 16. 2019.
20. Velde ET, Jansen T, Delgado V, Atsma D. Clair-DB: A Standardized Departmental Research Database. in *Computing in Cardiology*:503–506 2013.
21. Volpe JJ. Intraventricular Hemorrhage in the Premature Infant—Current Concepts. Part II. *Annals of Neurology.* 1989;25(2):109–116.
22. Moore CM, O'Reilly D, McCallion N, Curley AE. Changes in Inflammatory Proteins Following Platelet Transfusion in a Neonatal Population. *Pediatric Research.* 2023;94(6):1973–1977.
23. Hoftiezer L, Hof MH, Dijs-Elsinga J, Hogeveen M, Hukkelhoven CW, Van Lingen RA. From Population Reference to National Standard: New and Improved Birthweight Charts. *American Journal of Obstetrics and Gynecology.* 2019;220(4):383.e1-383.e17.
24. Bell MJ, Ternberg JL, Feigin RD, *et al.* Neonatal Necrotizing Enterocolitis: Therapeutic Decisions Based upon Clinical Staging. *Annals of Surgery.* 1978;187(1):1–7.
25. Hernán MA. How to Estimate the Effect of Treatment Duration on Survival Outcomes Using Observational Data. *BMJ (Clinical research ed.).* 2018;360:k182.

26. Van Houwelingen HC. Dynamic Prediction by Landmarking in Event History Analysis. *Scandinavian Journal of Statistics.* 2007;34(1):70–85.
27. Fu EL, Van Diepen M, Xu Y, *et al.* Pharmacoepidemiology for Nephrologists (Part 2): Potential Biases and How to Overcome Them. *Clinical Kidney Journal.* 2021;14(5):1317–1326.
28. Chesnaye NC, Stel VS, Tripepi G, *et al.* An Introduction to Inverse Probability of Treatment Weighting in Observational Research. *Clinical Kidney Journal.* 2022;15(1):14–20.
29. van Geloven N, Giardiello D, Bonneville EF, *et al.* Validation of Prediction Models in the Presence of Competing Risks: A Guide through Modern Methods. *BMJ.* 2022:e069249.
30. Collins GS, Dhiman P, Ma J, *et al.* Evaluation of Clinical Prediction Models (Part 1): From Development to External Validation. *BMJ.* 2024:e074819.
31. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: The Achilles Heel of Predictive Analytics. *BMC Medicine.* 2019;17(1):230.
32. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A Calibration Hierarchy for Risk Models Was Defined: From Utopia to Empirical Data. *Journal of Clinical Epidemiology.* 2016;74:167–176.
33. Riley RD, Archer L, Snell KIE, *et al.* Evaluation of Clinical Prediction Models (Part 2): How to Undertake an External Validation Study. *BMJ.* 2024:e074820.
34. Moore CM, Curley AE. Neonatal Platelet Transfusions: Starting Again. *Transfusion Medicine Reviews.* 2021;35(3):29–35.
35. Fustolo-Gunnink S, Huisman E, Van Der Bom J, *et al.* Are Thrombocytopenia and Platelet Transfusions Associated with Major Bleeding in Preterm Neonates? A Systematic Review. *Blood Reviews.* 2019;36:1–9.
36. Sachs MC, Sjölander A, Gabriel EE. Aim for Clinical Utility, Not Just Predictive Accuracy. *Epidemiology.* 2020;31(3):359–364.
37. De Hond AA, Vosslamber S, Lange S, *et al.* When, Where, Who, What, and Why? The Five Ws of Workflow Analysis for Implementing an AI Decision Support Tool at the Intensive Care. *Human Factors in Healthcare.* 2025;7:100095.
38. Janse RJ, Stel VS, Jager KJ, *et al.* When Impact Trials Are Not Feasible: Alternatives to Study the Impact of Prediction Models on Clinical Practice. *Nephrology Dialysis Transplantation.* 2024;40(1):27–33.
39. Van Calster B, Steyerberg EW, Wynants L, Van Smeden M. There Is No Such Thing as a Validated Prediction Model. *BMC Medicine.* 2023;21(1):70.
40. Houben NA, Lopriore E, Fijnvandraat K, *et al.* Platelet Transfusion in Neonatal Intensive Care Units of 22 European Countries: A Prospective Observational Study. *The Lancet Regional Health - Europe.* 2024;47:101086.
41. Scrivens A, Reibel NJ, Heeger L, *et al.* Survey of Transfusion Practices in Preterm Infants in Europe. *Archives of Disease in Childhood - Fetal and Neonatal Edition.* 2023;108(4):360–366.
42. Honohan Á, Van'T Ende E, Hulzebos C, *et al.* Posttransfusion Platelet Increments after Different Platelet Products in Neonates: A Retrospective Cohort Study. *Transfusion.* 2013;53(12):3100–3109.
43. Ferrer-Marín F, Sola-Visner M. Neonatal Platelet Physiology and Implications for Transfusion. *Platelets.* 2022;33(1):14–22.
44. Dannaway DC, Noori S. A Randomized Trial of Platelet Transfusions over 30 vs 120 Minutes: Is There an Effect on Post-Transfusion Platelet Counts?. *Journal of Perinatology.* 2013;33(9):703–706.
45. Davenport P, Sola-Visner M. Recent Advances in NICU Platelet Transfusions. *Seminars in Fetal and Neonatal Medicine.* 2025;30(1):101609.
46. Deschmann E. The Neonatal Hemorrhagic Risk Assessment in Thrombocytopenia Study-2. Clinical Trial Registration NCT04598750clinicaltrials.gov 2024.
47. Deschmann E, Saxonhouse MA, Feldman HA, Norman M, Barbian M, Sola-Visner M. Association Between In Vitro Bleeding Time and Bleeding in Preterm Infants With Thrombocytopenia. *JAMA Pediatrics.* 2019;173(4):393.
48. Cremer M, Paetzold J, Schmalisch G, *et al.* Immature Platelet Fraction as Novel Laboratory Parameter Predicting the Course of Neonatal Thrombocytopenia. *British Journal of Haematology.* 2009;144(4):619–621.
49. MacQueen BC, Christensen RD, Henry E, *et al.* The Immature Platelet Fraction: Creating Neonatal Reference Intervals and Using These to Categorize Neonatal Thrombocytopenias. *Journal of Perinatology.* 2017;37(7):834–838.
50. Raffaeli G, Tripodi A, Cavallaro G, *et al.* Thromboelastographic Profiles of Healthy Very Low Birthweight Infants Serially during Their First Month. *Archives of Disease in Childhood - Fetal and Neonatal Edition.* 2020;105(4):412–418.

51. Radicioni M, Bruni A, Bini V, Villa A, Ferri C. Thromboelastographic Profiles of the Premature Infants with and without Intracranial Hemorrhage at Birth: A Pilot Study. *The Journal of Maternal-Fetal & Neonatal Medicine.* 2015;28(15):1779–1783.
52. Thomassen D, Le Cessie S, Van Houwelingen HC, Steyerberg EW. Effective Sample Size: A Measure of Individual Uncertainty in Predictions. *Statistics in Medicine.* 2024;43(7):1384–1396.
53. Riley RD, Collins GS, Kirton L, *et al.* Uncertainty of Risk Estimates from Clinical Prediction Models: Rationale, Challenges, and Approaches. *BMJ.* 2025:e080749.

5.6 Appendix

Table 5.3: Estimand framework for a dynamic interventional prediction model [1].

Element	Estimand questions	Estimands in our study
Population	To which individuals will the prediction model be applied and in which health care setting?	The model is applicable to preterm infants <34 weeks' gestation and severe thrombocytopenia (platelet count $< 50 \times 10^9/L$), admitted to a tertiary care NICU in the Netherlands, Sweden, or Germany. Infants with a severe congenital malformation, recurrent major bleeding, or severe thrombocytopenia due to exchange transfusion were study exclusion criteria. In addition, the model is not designed to support platelet transfusion decisions prior to surgery or other invasive procedures.
Moment(s) of intended use	At which moment(s) is the prediction model (re)consulted to inform the intervention decision?	At any time point during the first week after the onset of severe thrombocytopenia, the model can be used to support prophylactic platelet transfusion decisions when platelet counts are $< 50 \times 10^9/L$.
Intervention options	Which intervention options are relevant at the moment(s) of making the intervention decision?	The intervention of interest is whether or not to give a prophylactic platelet transfusion during a period of severe thrombocytopenia, with the aim of reducing the risk of major bleeding or mortality. We are not focusing on therapeutic transfusions, which are given in case of active bleeding.
	For how long should the intervention strategy be fixed?	We compared two transfusion strategies at each time point at which a prophylactic platelet transfusion decision could be made: (i) Prophylaxis strategy if an infant were to receive a prophylactic platelet transfusion within a 'grace period' of 6 hours [2]. An infant complied with this strategy if receiving transfusion within 6 hours of the prediction time point. This timeframe was chosen because 75% of transfusions occurred within 6 hours of platelet count measurement. For model development and evaluation, infants who did not receive a transfusion within this period were artificially censored at 6 hours. (ii) No-prophylaxis strategy if an infant were not to receive a platelet transfusion for the next 3 days, so no transfusion during the entire 3-day prediction window. For model development and evaluation this meant that infants were artificially censored at the moment of transfusion. Of note, in the first 6 hours after each prediction time point, infants can contribute their data to both strategies. For example, if major bleeding occurred within 3 hours of the prediction time point, the infant's data is included up to the time of the event under both strategies. The strategies diverge at the time of transfusion if a transfusion is given within 6 hours of the prediction time point or after 6 hours if no transfusion is given (Figure 5.4).

Intervention options	Should the duration to fix the intervention option be aligned with the time until next moment of prediction?	No, the time until the next platelet transfusion decision moment is not aligned with the time until the next moment of prediction. Our landmarks (prediction time points) represented 2-hour intervals at which a new prediction was made to accurately capture changes in the infant's condition, using the updated information available about the infant up to that point.
Outcome and prediction window	Which outcome(s) are most informative for the intervention decision?	The primary outcome, most informative for transfusion decisions, is major bleeding or mortality (date and time as recorded in the medical record or cranial ultrasound report).
	What prediction window provides important information for the intervention decision?	We chose a 3-day prediction window, because transfused platelets are generally considered not to survive more than a few days in the neonatal recipient [3]. As secondary outcome, we also chose a 14-day prediction window, because platelet transfusions may have both short-term and longer-term effects.
	Should the outcome be defined differently because of the specified interventions?	No, the outcome major bleeding or mortality should be defined in the same way for both platelet transfusion strategies. NICU cranial ultrasound protocols recommended scans on days 1, 3, and 7 for infants <30 weeks' gestation, then biweekly until discharge; on days 1, 7, and before discharge for infants 30-32 weeks' gestation; and only when clinically indicated for infants >32 weeks' gestation. Additional scans were performed in cases of cerebral abnormalities or conditions known to increase bleeding risk, such as sepsis, necrotizing enterocolitis, or severe thrombocytopenia.
	Should the prediction window be aligned with the time until next moment of prediction?	No, the prediction window and prediction interval are not aligned: • Prediction window: short-term window of 3 days (primary outcome) and long-term window of 14 days (secondary outcome) relative to the prediction time point. • Time till next prediction time point: every 2 hours from first platelet count $< 50{\times}10^9/L$ until seven days later to adequately capture changes in the infants' clinical condition.
Predictors	Which predictors are predictive of the outcome of interest and based on which characteristics should the outcome risks be individualized?	We developed two models: one for the prophylaxis strategy and one for the no-prophylaxis strategy. Table 5.4 lists the predictors. Landmark time (i.e., time in hours since first platelet count $< 50 \times 10^9/L$) and the interaction between landmark time and SGA were included as covariates, as the association between SGA and bleeding may be stronger immediately after the onset of severe thrombocytopenia than several days later. We did not include interaction terms between platelet transfusion and other predictors, because we already accounted for these by modelling the data separately according to the prophylaxis and no-prophylaxis platelet transfusion strategies.

Table 5.4: Definitions of predictors in the outcome model. PC, platelet count; PT platelet transfusion; HFO, high-frequency oscillation; Landmark time, prediction time point at which a new prediction is made; NEC, necrotizing enterocolitis; SGA, small-for-gestational age.

	Definition	**Entry at each prediction time point**
Time-fixed variables		
Gestational age	Gestational age at birth.	Gestational age in days.
Postnatal age	Age in hours at first PC $< 50 \times 10^9/L$.	Age in hours.
SGA	Birth weight below the 10^{th} percentile according to Dutch birth weight curves.	SGA yes/no.
Time-dependent variables		
PC	Every PC during the study period ($< 50 \times 10^9/L$).	Most recent PC at each prediction moment.
Previous PTs	Cumulative number of PTs.	Cumulative number of PTs prior to each prediction moment.
Mechanical ventilation	Period of receiving mechanical ventilation, both conventional ventilation and HFO.	Ventilation episode yes/no, defined as uninterrupted period during which the patient is intubated (start moment, variable 'on') and mechanically ventilated until extubation (stop moment, variable 'off').
NEC	NEC defined as $\geq$ grade IIA according to Bell's criteria.	NEC episode yes/no, present when pneumatosis intestinalis was diagnosed on X-ray (variable 'on') until stop of antibiotics (variable 'off').
Sepsis	Culture-positive or clinically suspected sepsis for which antibiotics are given.	Sepsis episode yes/no, present from first dose (variable 'on') to last dose of antibiotics (variable 'off'). We included all sepsis cases from the start of antibiotics because the risk of bleeding is likely to be highest in the early stages of infection. In the validation cohort, sepsis was restricted to culture-positive or culture-negative cases with $\geq$5 days of antibiotics. We used a different definition because at the start of antibiotic treatment it is not known whether the blood culture will become positive or whether antibiotic treatment will be continued, and some infants may die of fulminant sepsis shortly after the start of antibiotics.
Vasoactive agents	Drugs that increase cardiac muscle contraction and/or lead to vasoconstriction, such as dobutamine, dopamine, epinephrine, isoprenaline, milrinone, and norepinephrine.	Receiving vasoactive agents yes/no. If different types of vasoactive agents were used during the same episode, the start moment was the time at which the first vasoactive agent type was initiated (variable 'on'), and the stop moment the time of the last vasoactive agent (variable 'off').
Landmark time (both linear and quadratic)	Time since first PC $< 50 \times 10^9/L$.	Time in hours (divided by 100) since the first PC $< 50 \times 10^9/L$.
Interactions		
Landmark time (linear) $\times$ SGA	Time since first PC $< 50 \times 10^9/L \times$ SGA.	Time in hours (divided by 100) since the first PC $< 50 \times 10^9/L \times$ SGA yes/no.

A. Data set-up for a dynamic interventional prediction model

To enable development of predictions under the two interventions (prophylaxis strategy of receiving a platelet transfusion within the next 6 hours versus no-prophylaxis strategy of not receiving a platelet transfusion for the next 3 days) at any time s_t during the first week after the onset of severe thrombocytopenia, we first modified our original dataset (structured longitudinally to include complete patient histories) to incorporate landmarking. A landmark represents a specific time point at which new predictions are made (in the main text, these are called 'prediction time points'). Individuals can contribute to multiple landmarks, referred to as 'person-landmarks'. We generated a separate copy of the dataset for each landmark time, defined at two-hour intervals up to seven days at time s_t = 0, 2, 4, ..., 164, 166, 168 hours. For each dataset copy, time was reset to zero at the corresponding landmark time, which served as a new baseline for estimating the risk of the outcome. Infants were included in person-landmarks if their last measured platelet count was $< 50 \times 10^9/L$. If the last platelet count was below $25 \times 10^9/L$, an additional requirement was that it was measured no longer than 24 hours before the landmark (in that case a new platelet count measurement required to be included again in a next person-landmark). If a platelet transfusion was given, inclusion in a next person-landmark required a new platelet count measurement post-transfusion (to avoid using pre-transfusion platelet counts in new predictions). Some extra choices had to be made for infants that were discharged to a stepdown unit, since these infants were unlikely to experience platelet transfusions, major bleeding, or death, follow-up time was set to the maximum study duration. In the landmarking process, individuals were excluded from the dataset copy corresponding to st if they had already been discharged by that time and therefore no longer belonged to the set of individuals for whom a transfusion decision needs to be made. In the validation dataset, some infants transferred to another NICU had censored outcomes, because follow-up ended after transfer to a non-participating NICU. Infants who had experienced the event or had been censored by st were excluded from the corresponding dataset copy, as their event or censoring occurred before the new baseline (landmark time). All copies were combined into one 'landmarked' dataset.

Starting from this 'landmarked' dataset, we created two new datasets using the clone-censor-weight approach [4]. This approach consists of three steps. The first step ('cloning') was to create two copies of the development dataset: dataset D^A for the prophylaxis strategy and dataset D^B for the no-prophylaxis strategy. In the second step ('censoring'), we modified each dataset such that each individual's follow-up is used for the duration for which their treatment status is consistent with the strategy under consideration (Figure 5.4). For the prophylaxis strategy, this meant that infants in dataset D^A were artificially censored at 6 hours if they did not receive a platelet transfusion within 6 hour. For the no-prophylaxis strategy this meant that infants in dataset D^B were artificially censored at the time of transfusion, for transfusions that took place within 3 days from the landmark time. During the first 6 hours after each landmark, an individual may meet the criteria for both the prophylaxis and no-prophylaxis strategies. In other words, the transfusion strategies began to differ at the moment a transfusion is given or after a maximum of 6 hours if no transfusion was given.

For the third step ('inverse probability of censoring weighting'), we divided the follow-up of each infant into periods of one hour, starting from the landmarking time, and used pooled

logistic regression to estimate the weights. The weight for a given individual at time s_t is the inverse of the conditional probability of remaining uncensored to time s_t for each transfusion strategy, given their past transfusion status and time-dependent covariate history. The following time-fixed covariates were included: gestational age, small-for-gestational age, postnatal age at the onset of severe thrombocytopenia, and perinatal asphyxia. The following time-dependent covariates were included: platelet count, the cumulative number of prior platelet transfusions, sepsis, necrotizing enterocolitis, pharmacological treatment for patent ductus arteriosus, mechanical ventilation, vasoactive agents, IV boluses for management of (suspected) hypotension, surgery or lumbar puncture, and landmark time (definitions in Table 5.5). To reduce the variance of the weights, we stabilized them by including all baseline (i.e., at landmarking time) values of the variables that were also used in the prediction model in the numerator of the stabilized weights [5]. As an additional measure to reduce variance, we truncated extreme weights at the 99.99$^{\text{th}}$ percentile in the development dataset and at the 99.95$^{\text{th}}$ percentile in the validation dataset (due to the smaller sample size) [6]. In the clone-censor-weight approach, individuals discharged within 3 days of the landmark time were assigned a weight of one at each subsequent time interval after discharge for the no-prophylaxis strategy, as it was sure they would adhere to this transfusion strategy. This means a constant cumulative weight from the moment of discharge onward. Individuals discharged within 6 hours of the landmark time were artificially censored at 6 hours for the prophylaxis strategy, as prophylactic platelet transfusions are typically administered in a NICU setting and discharge to a stepdown unit indicates that intensive monitoring is no longer required.

B. Causal assumptions

We assumed the following three key identifiability criteria for causal effects, which have been described in more detail elsewhere [7].

Sequential conditional exchangeability

This condition, sometimes called the 'no unmeasured confounding' assumption, requires that the observed transfusion decisions at each prediction time point are conditionally independent of the potential (counterfactual) outcome, given time-dependent confounders. We used the clone-censor-weight approach to emulate scenarios in which every infant followed the two transfusion strategies of interest: 'prophylaxis strategy' if receiving a platelet transfusion within the next 6 hours and 'no-prophylaxis strategy' if not receiving a transfusion within the next 3 days. This three-step approach is illustrated in Figure 5.4. In essence, it creates a pseudo-population that mimics random allocation of interventions, so that the risk of major bleeding or mortality under both transfusion strategies can be correctly estimated [4, 8, 9].

To evaluate the assumption of sequential conditional exchangeability, we made a directed acyclic graph (DAG) prior to data collection based on literature and expert advice to identify confounding variables that potentially impact both exposure to platelet transfusions and the risk of major bleeding (Figure 5.5). Based on this DAG, we selected a priori which variables

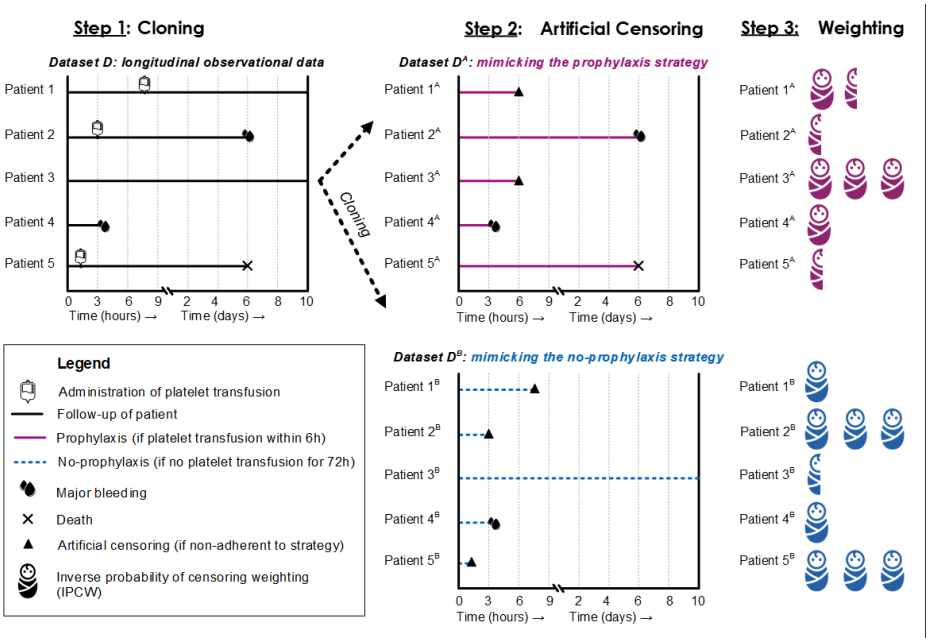


Figure 5.4: Graphical summary of the clone-censor-weight approach.
This figure shows the follow-up of five example patients, starting from one prediction time point (actually 'new' prediction time points every 2 hours), first in hours to show the time points where the two platelet transfusion strategies overlap, and then in days for a total follow-up of 10 days. To enable development of predictions under the two transfusion strategies, we used the clone-censor-weight approach.
Step 1 (cloning): each patient is duplicated ('cloned') and assigned to one of the two transfusion strategies. For example, patient 1, receives a platelet transfusion 8 hours after the prediction time point (Dataset D). After cloning, patient ('clone') 1A follows the prophylaxis strategy (platelet transfusion within the next 6h, purple solid line) in dataset D^A and patient 1B follows the no-prophylaxis strategy (no platelet transfusion for 3 days, blue dashed line) in dataset D^B.
Step 2 (artificial censoring): each clone is subsequently censored (Δ) when they become nonadherent to the transfusion strategy. Patient 1A is therefore censored at 6 hours for the prophylaxis strategy (no platelet transfusion received within 6h) and it's clone (patient 1B) is censored at 8 hours for the no-prophylaxis strategy at the time of transfusion. Note that during the first 6 hours after each prediction time point, an individual may meet the criteria for both strategies: they start to differ at the moment a transfusion is given or after a maximum of 6 hours if no transfusion is given.
Step 3 (weighting): transfusion-related censoring is informative because it depends on a physician's decision influenced by patient characteristics like severity of thrombocytopenia, illness level, and prematurity. Inverse probability of censoring weighting (IPCW) can be used to adjust for this. The idea behind IPCW is to calculate weights on each time point based on the (inverse) probability of remaining under observation (so uncensored) by upweighting those remaining in the study, who have similar characteristics to those who were censored. Those less likely to remain uncensored are up-weighted, and those more likely to remain uncensored are down-weighted. This creates a pseudopopulation in which the covariate balance between transfusion strategies is approximated over time, ensuring that censoring at transfusion is no longer influenced by previous exposure and patient characteristics related to censoring [8, 9].

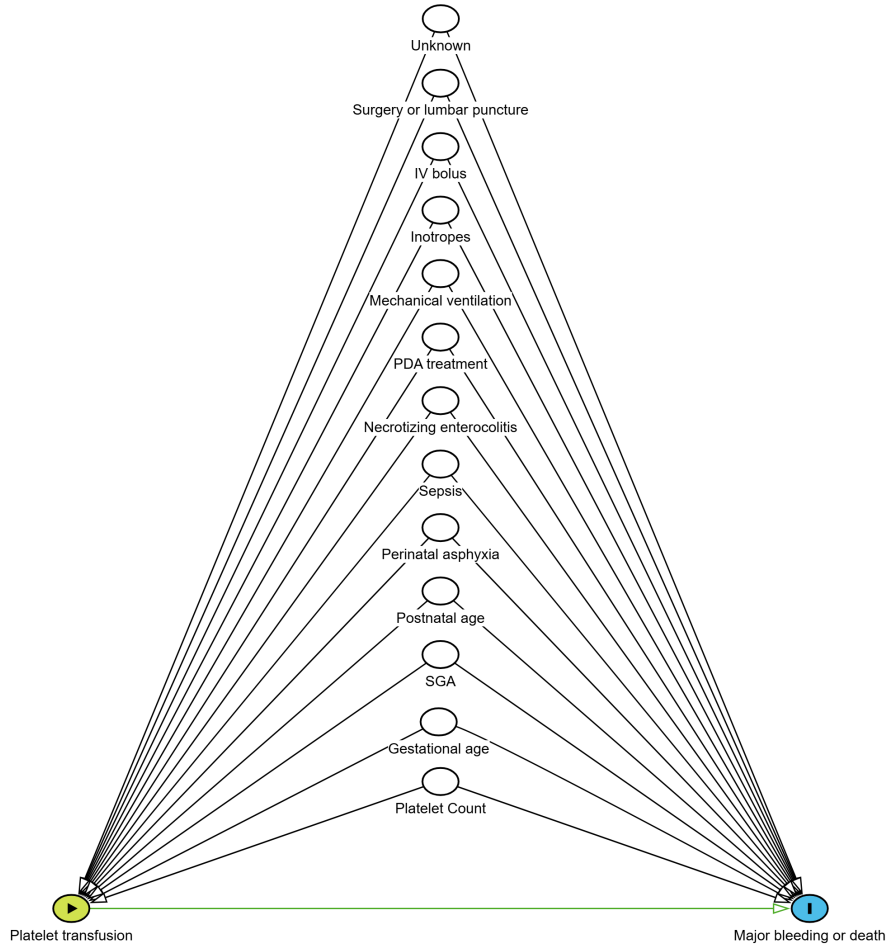


Figure 5.5: Directed acyclic graph for the association between prophylactic platelet transfusion and major bleeding or death.
Based on literature review and expert consensus (neonatologists, clinical epidemiologists, hematologist), we constructed a directed acyclic graph (DAG) to identify potential known and measured confounders for the association between the exposure 'prophylactic platelet transfusion' and the outcome 'major bleeding or death'. For simplicity, this DAG does not show time-dependent confounding, treatment-confounder feedback, and interactions between confounders.
Abbreviations: IV, intravenous; SGA, small-for-gestational age (birth weight <p10); PDA treatment, hemodynamically significant patent ductus arteriosus treated with ibuprofen or indomethacin.

Table 5.5: Definitions of (time-dependent) confounders included in the weights model of inverse probability of censoring weighting.
Abbreviations: IV, intravenous; PDA, patent ductus arteriosus.
[a] These variables were not included in the weights model for validation due to the smaller sample size of the validation cohort, resulting in a more parsimonious model.

Variable	Definition	Entry at each landmark point
Mechanical ventilation	Period of receiving mechanical ventilation, both conventional ventilation and high-frequency oscillation.	Ventilation episode (yes/no), defined as a continuous period of mechanical ventilation from intubation (variable 'on') to extubation (variable 'off').
Postnatal age	Age in hours at onset of severe thrombocytopenia.	Age in hours. A postnatal age of more than 4 weeks was categorized as $\geq$672 hours ($4\times7\times24$).
Gestational age	Gestational age at birth.	Gestational age in days.
Platelet count	Every platelet count ($\times10^9/L$).	Most recent platelet count at each landmark. Values below $10\times10^9/L$ and above $300\times10^9/L$ were grouped together.
Previous platelet transfusions	Cumulative number of previous platelet transfusions.	Cumulative number of platelet transfusions prior to each landmark. Values $\geq$ 5 were grouped together.
Small-for-gestational age	Birth weight <p10.	Small-for-gestational age yes/no.
Necrotizing enterocolitis	Necrotizing enterocolitis defined as $\geq$ grade IIA according to Bell's criteria.	Necrotizing enterocolitis episode yes/no, present when pneumatosis intestinalis was diagnosed on X-ray (variable 'on') until stop of antibiotics (variable 'off').
Sepsis	Sepsis defined as the start of antibiotic therapy for clinically suspected sepsis.	Sepsis episode yes/no, present from start (variable 'on') until stop of antibiotics (variable 'off').
Vasoactive agents[a]	Drugs that increase the force of cardiac muscle contraction and/or lead to vasoconstriction, such as dobutamine, dopamine, epinephrine, isoprenaline, milrinone, and norepinephrine.	Use of vasoactive agents yes/no, with timing defined from initiation of the first agent (variable 'on') to discontinuation of the last (variable 'off') if multiple agents were used.
Perinatal asphyxia	Perinatal asphyxia defined by at least 3 criteria: 5-min Apgar $\leq$5, arterial pH <7.0, base excess < -16 mmol/L, lactate >10 mmol/L within 1h of birth, $\geq$10 min of respiratory resuscitation, and/or cardiopulmonary resuscitation.	Perinatal asphyxia (meeting at least 3 of the criteria) yes/no.
Planned invasive procedures	Surgery planned in the next 6h and lumbar puncture planned in the next 2h.	Surgery planned in the next 6h (yes/no) and lumbar puncture planned in the next 2h (yes/no) after each landmark point (variable 'on') until 24h after surgery and 6h after lumbar puncture (variable 'off').
IV bolus therapy[a]	IV fluid bolus for the management of (suspected) hypotension.	Receiving an IV bolus within the previous hour from start of the landmark yes/no.
PDA treatment[a]	Hemodynamically significant PDA treated with ibuprofen or indomethacin.	Episode of PDA treatment yes/no, from the first dose (variable 'on') until the last dose of the course (variable 'off').
Interactions		
Platelet count $\times$ necrotizing enterocolitis	Last platelet count ($\times10^9/L$) $\times$ necrotizing enterocolitis.	Most recent platelet count x necrotizing enterocolitis yes/no.
Platelet count $\times$ planned invasive procedures	Platelet count ($\times10^9/L$) $\times$ planned invasive procedures.	Most recent platelet count $\times$ planned invasive procedures yes/no.

to include as predictors in the model, and which additional (time-varying) confounders to take into account during estimation in the 'weighting' step of the clone-censor-weight approach (Table 5.5).

Current transfusion protocols are based primarily on platelet count thresholds. Still, previous platelet transfusion protocols included other criteria for prophylactic platelet transfusion, which we had to consider as platelet transfusion protocols changed over time during the study periods. To clarify our selection of confounders, we describe the different transfusion protocols used during the study periods. At the time of the MONET study (validation cohort, 2010-2014) and during the first years of the PROSPECT study (development cohort, 2017-2021), national platelet transfusion protocols recommended prophylactic platelet transfusion with a platelet count threshold of $50 \times 10^9/L$ for sick or clinically unstable infants <32 weeks and birth weight <1500 grams, and a lower threshold of $20 \times 10^9/L$ for clinically stable infants. In some centers, prophylactic transfusions were also recommended at a higher platelet count threshold for the treatment of persistent ductus arteriosus (PDA). One center in the validation cohort followed a distinct transfusion protocol during the first year of the MONET study, according to which prophylactic platelet transfusions were not routinely given to non-bleeding preterm with (very) severe thrombocytopenia, but only in the need for surgery, invasive procedures, or indomethacin treatment for PDA [10]. We corrected for this by adding an additional 'center variable' to the weights model in the IPCW analysis during validation. In the last years of the PROSPECT study, transfusion protocols were gradually adapted following the PlaNeT-2/MATISSE trial results (published in January 2019) [11], with prophylactic transfusions recommended only for platelet counts below $25 \times 10^9/L$, regardless of severity of illness, gestational age, birth weight, and PDA treatment. In addition, the recommended dose according to the Dutch national prophylactic platelet transfusion protocol was changed from 10-15 ml/kg (i.e., $10 - 20 \times 10^9$/kg) to 10 ml/kg. Higher thresholds prior to exchange transfusion (exclusion criterion in our study), surgery and invasive procedures were recommended according to both old and new transfusion protocols.

Consistency

This assumption implies that observed outcomes are equal to counterfactual outcomes under corresponding level of treatment. However, this holds only if counterfactual treatments are single, well-defined values. The presence of *multiple versions of treatment* violates this assumption. In practice, treatment variations often occur, but the impact of variation in transfusion volumes and infusion rates on neonatal bleeding risk and mortality is unclear [12–14]. In our study, there was some variation in transfusion volumes and infusion rates, and for the validation dataset also in platelet storage media. These factors introduce multiple versions of treatment. For the prophylaxis strategy, variation arises from differences in transfusion volumes, infusion rates, and the frequency of multiple transfusions. For the no-prophylaxis strategy, variation is due to how often transfusions are given after 3 days under 'standard care'. In clinical practice, first a decision is made whether or not to provide a prophylactic platelet transfusion, and second on how the transfusion will be administered (e.g., transfusion volume and infusion rate). As the transfusion decision precedes the 'type of transfusion', causal estimates for each strategy can still be made; these estimates reflect the 'average pro-

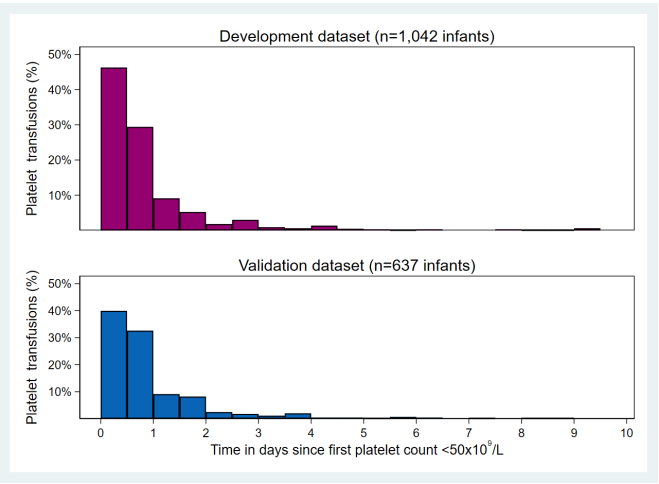


Figure 5.6: Timing of first platelet transfusion in development and validation cohorts. This figure shows histograms of the time to first platelet transfusion (in days) after the onset of severe thrombocytopenia (first platelet count $< 50 \times 10^9/L$) in the developmental cohort (top panel, n=1,042 infants) and the validation cohort (bottom panel, n=637 infants). Data are converted from hours to days and rounded to the nearest half-day. The y-axis represents the percentage of infants who received a first platelet transfusion at each time interval, while the x-axis represents the time in days since the onset of severe thrombocytopenia.

phylactic platelet transfusion' based on how these transfusions typically occurred in the development dataset [15]. Differences in platelet transfusion practices between the development and validation cohorts may impact the model's predictive performance during the evaluation of model performance. In the development dataset, data were available for 96% of transfusion volumes, 95% of transfusion durations and 94% of infusion rates, calculated by dividing transfusion volume (ml/kg) by duration (hours). The median volume of platelet transfusion was 14 ml/kg (IQR 10-15), the median duration was 30 minutes (IQR 30-60), and the median infusion rate was 20 ml/kg/hour (IQR 15-30). One center used hyper-concentrated platelet products with a median volume of 3 ml/kg (IQR 2-4), which accounted for 2% of all transfusions. Platelet products were obtained by apheresis from a single male donor, suspended in approximately 65% platelet additive solution type E (PAS-E) and 35% plasma, tested for ABO compatibility and Parvo B19, and leucocyte-depleted [16]. In the validation dataset, data on platelet transfusion volume were missing in only 1% of transfusions, with a median volume of 10 ml/kg (IQR 10-15). Information on transfusion duration and infusion rate was not available. All platelet transfusions were obtained from male donors by apheresis, 6% were hyper-concentrated platelet products (used in one center), 11% were suspended in PAS-E/plasma, and 83% were suspended in plasma. In addition, platelet products were tested for ABO compatibility and Parvo-B19, leucocyte-depleted and irradiated for infants weighing less than 1500 grams at birth and/or with a gestational age of less than 32 weeks. Other donor and product factors, like donor age and storage time, could also influence heterogeneity in platelet transfusions [17–19], but study data on these factors were unavailable. The time to

first platelet transfusion after the onset of severe thrombocytopenia was similar in both cohorts (Figure 5.6).

Conditional positivity

This assumption requires that both treatment strategies are observed for all covariate levels during the time period of interest for the prediction model. This requires that the probability of being assigned to both treatment strategies is always non-zero for all covariate levels, at least theoretically, as we want to compare two treatment strategies over time. For estimation purposes, this assumption must also hold to some extent in the data. We say "to some extent" because we can make some modelling assumptions that allow us to borrow information across covariate levels. We checked this assumption by evaluating the distribution of variables used in the weights estimation of inverse probability of censoring weights (IPCW), stratified by transfusion strategy over time, to assess common support between the prophylaxis and no-prophylaxis strategies. In addition, we made a visual inspection of the pair-wise space of the confounder pairs stratified by transfusion strategy and assessed that it seemed to be covered well enough, except for few combinations of covariates. Based on these evaluations, we concluded that the positivity assumptions held reasonably well for the vast majority of covariate combinations.

C. Results

Table 5.6: Total number of events and bleeding types in the development and validation cohorts. Abbreviations: IVH, intraventricular hemorrhage; PVHI, periventricular hemorrhagic infarction.
[a] The composite outcome of major bleeding or death within 21 days of the first platelet count $< 50 \times 10^9/L$ (final prediction time point on day 7 plus the 14-day prediction window for the secondary outcome), as well as the individual components (death and major bleeding separately), were estimated using the Kaplan-Meier method to account for censoring due to transfer of infants to a non-participating NICU in the validation cohort (see also Table 5.7). One infant had major bleeding followed by death in the same hour, contributing to both the individual component of death and major bleeding.
[b] Of the 143 deaths in the development cohort, 15 occurred after more than 10 days.
[c] Of the 90 deaths in the validation cohort, 18 deaths occurred after more than 10 days.
[d] Of the 116 first major bleedings in the development cohort, 8 occurred after more than 10 days (5 cerebellar hemorrhages, 1 IVH grade 3, and 2 major pulmonary hemorrhages). There were 61 cases of major bleeding followed by death, but they were recorded only as major bleeding to avoid overlap with mortality.
[e] Of the 66 first major bleedings in the validation cohort, 3 occurred after more than 10 days (1 IVH grade 3, 1 IVH grade 1-2 with parenchymal involvement, and 1 major pulmonary hemorrhage). There were 39 cases of major bleeding followed by death, but they were recorded only as major bleeding to avoid overlap with mortality.
[f] In the development cohort, there were 2 cases of cerebellar hemorrhage who were also diagnosed with IVH grade 3 based on the same cranial ultrasound.
[g] Severe gastrointestinal hemorrhage associated with hemodynamic instability, requiring volume boluses, inotropes or red blood cell transfusion in the same 24 hours.

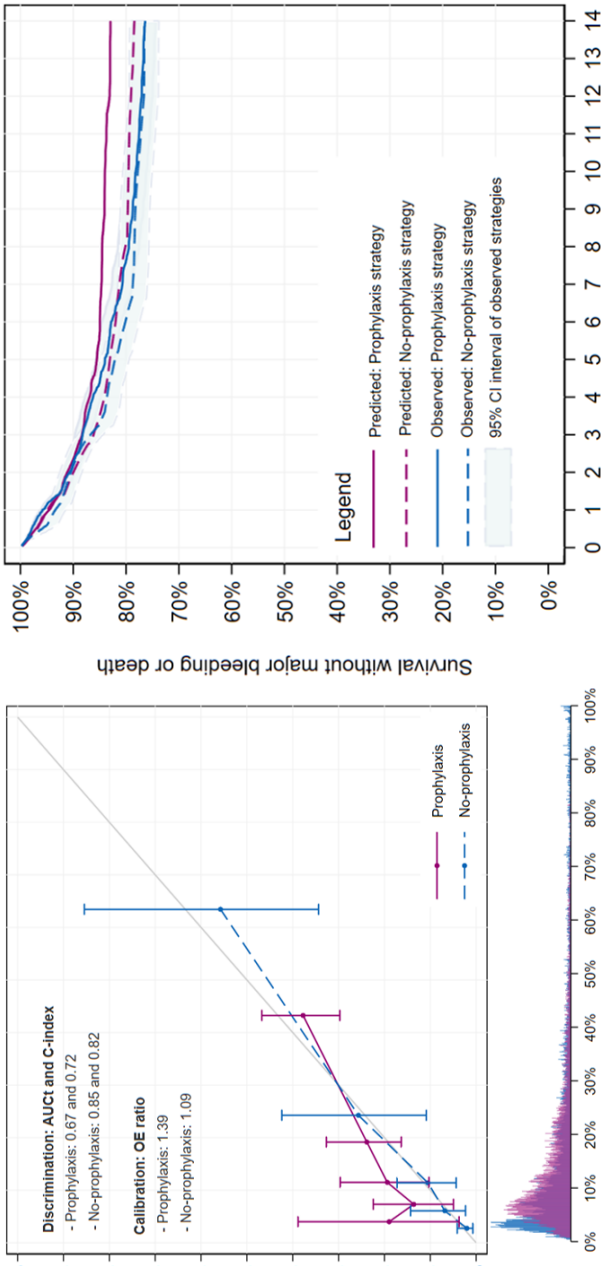
	Development cohort n = 1,042 infants	**Validation cohort n = 637 infants**
Major bleeding or death within 21 days[a], no. (%)	258 (24.8)	156 (25.0)
Death within 21 days[a], no. (%)	143 (13.6)[b]	90 (14.1)[c]
First major bleedings within 21 days[a], no. (%)	116 (11.9)[d]	66 (10.4)[e]
Type of first major bleedings, no. (% of first major bleedings)		
IVH grade 3 with or without PVHI	34 (29.3)	32 (48.5)
IVH grade 1 or 2 with PVHI	10 (8.6)	2 (3.0)
Solitary parenchymal hemorrhage	14 (12.1)	4 (6.1)
Cerebellar parenchymal hemorrhage[f]	21 (18.1)	10 (15.2)
Subdural or epidural hemorrhage	7 (6.0)	3 (4.6)
Major pulmonary hemorrhage	27 (23.3)	12 (18.2)
Severe gastrointestinal hemorrhage[g]	3 (2.6)	3 (4.6)

Table 5.7: Summary of the number of infants and their corresponding number of person-landmarks, stratified by transfusion strategy. Abbreviations: LM, landmark.
[a] Because infants can contribute to multiple landmarks (prediction time points) - each representing a new prediction based on updated information up to that point - we refer to these as 'person-landmarks'.

	Development cohort		**Validation cohort**	
	Prophylaxis	*No-prophylaxis*	*Prophylaxis*	*No-prophylaxis*
At patient-level, no. of:				
Unique infants	1,042	1,042	637	637
Infants with major bleeding or death	222	124	142	81
Discharged infants to stepdown unit	15	47	5	28
Transferred infants to non-participating NICU	0	0	22	18
At person-LM level[a], no. of:				
Total person-LMs	19,910	19,910	13,846	13,846
Person-LMs artificially censored	16,918	7,019	11,464	6,096
Major bleeding or death events per person-LM	908	1,395	776	1,012
Person-LMs with discharge to stepdown unit	33	740	10	341
Person-LMs transferred to non-participating NICU	0	0	123	244

Table 5.8: Validation measures of the predictive performance in the external validation dataset. Abbreviations: 95% CI, 95% confidence intervals, estimated via bootstrapping using 500 bootstrap samples; OE ratio, ratio of observed and expected outcomes, where a ratio <1 suggests an overestimation of risk and a ratio >1 suggests underestimation; AUCt, cumulative/dynamic area under the ROC curve, measuring discrimination at the prediction window; C-index, concordance index, summarizing discrimination over a range of follow-up times up to the prediction window.
[a] For the 3-day prediction window, the median predicted risk of major bleeding or mortality was 7.4% (IQR 4.2%-14.3%) under the prophylaxis strategy, and 6.0% (IQR 2.6%-16.3%) under the no-prophylaxis strategy.
[b] For the 14-day prediction window, the median predicted risk was 11.4% (IQR 6.6% - 21.6%) under the prophylaxis strategy, and 11.1% (IQR 4.9% - 28.5%) under the no-prophylaxis strategy.
[c] Brier score (lower is better): a measure that expresses the model's overall accuracy to predict whether a patient experiences the primary outcome at the prediction window, combining both discrimination and calibration. It summarizes how close estimated risks are to the observed primary event indicators at the end of the prediction window, with 0 indicating a perfect model and scores closer to 0 indicating better predictive performance [20].
[d] Scaled Brier score (higher is better): summarizes the percentage reduction in Brier score compared to a null model without covariates, with a score of 100% indicating a perfect model and a score of 0% indicating no predictive information beyond the null model [20].

	Platelet transfusion strategy	
	Prophylaxis	*No-prophylaxis*
	Estimated values (95% CI)	**Estimated values (95% CI)**
Prediction window of 3 days[a]		
Calibration:		
OE ratio	1.01 (0.73 – 1.32)	0.96 (0.51 – 1.58)
Discrimination:		
AUCt	0.69 (0.60 – 0.76)	0.85 (0.76 – 0.92)
C-index	0.75 (0.69 – 0.80)	0.83 (0.72 – 0.89)
Prediction accuracy:		
Brier score[c]	0.02 (0.02 – 0.03)	0.07 (0.04 – 0.09)
scaled Brier scored (%)	15.3 (9.2 – 21.9)	20.9 (-16.3 – 52.4)
Prediction window of 14 days[b]		
Calibration:		
OE ratio	1.39 (1.05 – 1.73)	1.09 (0.66 – 1.67)
Discrimination:		
AUCt	0.67 (0.59 – 0.74)	0.85 (0.77 – 0.94)
C-index	0.72 (0.67 – 0.77)	0.82 (0.74 – 0.89)
Prediction accuracy:		
Brier score[c]	0.03 (0.02 – 0.04)	0.09 (0.04 – 0.11)
scaled Brier scored (%)	24.4 (18.6 – 30.5)	30.4 (-2.51 – 64.58)



(a) Calibration plot of the estimated 14-day risk of major bleeding or mortality under the prophylaxis and no-prophylaxis platelet transfusion strategies in the validation cohort.

(b) Mean estimated and observed survival curves

Figure 5.7: Abbreviations: AUCt, cumulative/dynamic area under the ROC curve; C-index, concordance index; OE ratio, ratio of observed and expected outcomes; 95% CI, 95% confidence intervals of the observed prophylaxis and no-prophylaxis strategies in the validation dataset.
Panel (a): This calibration plot shows the agreement between the predicted and observed risks of major bleeding or mortality within the next 14 days for the prophylaxis (purple solid line) and no-prophylaxis (blue dashed line) transfusion strategies. The gray solid line at 45 degrees indicates perfect calibration, i.e., predicted and observed proportions are equal. The five dots represent quintiles of the predictions according to the predicted values, with on the x-axis the mean prediction of each quintile and on the y-axis the mean observed proportion of that group, including 95% confidence intervals obtained by bootstrapping using 500 bootstrap samples. The histograms along the x-axis show the distribution of risk estimates for both transfusion strategies.
Panel (b): This graph shows the mean estimated survival curves (purple lines) without major bleeding or death for the two transfusion strategies up to 14 days after the first platelet count $< 50 \times 10^9/L$. The corresponding observed survival curves in the validation dataset are shown as blue lines, including the 95% confidence intervals (light blue). The prophylaxis strategy (platelet transfusion within 6 hours) is shown with the solid lines and the no-prophylaxis strategy (no platelet transfusion within 3 days) with the dashed lines.

D Prediction formulas of the dynamic interventional prediction model

The general formula for a Cox model is:

$$S(t \mid X) = S_0(t)^{\exp(\beta_1 X_1 + \cdots + \beta_n X_n)} = S_0(t)^{\exp(PI(X))}$$

where $S(t \mid X)$ is the function expressing the probability of major bleeding at time t after the onset of severe thrombocytopenia (i.e., the first platelet count $< 50 \times 10^9/L$) and $PI(X)$ is the prognostic index [21].

The predicted risk ($P(t \mid X)$) can be expressed as:

$$P(t \mid X) = 1 - S_0(t)^{\exp(PI(X))}.$$

Since reporting baseline hazards at the reference value (where all predictors are set to zero) is not clinically meaningful (e.g., a gestational age of zero is impossible), we instead present baseline risks for a hypothetical 'average' patient, who was assigned the median values of the time-fixed variables, as shown in Table 1, corresponding to a gestational age of 28 weeks (196 days), a postnatal age of 4 days (96 hours) at the onset of severe thrombocytopenia, and a platelet count of $39 \times 10^9/L$. Categorical time-dependent variables were set to zero. The estimated baseline risk over the 3-day prediction window (primary outcome) was 0.112 for the prophylaxis strategy and 0.043 for the no-prophylaxis strategy. Over the 14-day prediction window (secondary outcome), the baseline risk was 0.171 and 0.081 for the both strategies, respectively.

To estimate the absolute risk of major bleeding or mortality within the next 3 days under both strategies, we multiply the log-hazard ratios (taking the natural logarithms of the hazard ratios as reported in Table 2) and combine these with the baseline risks.

For the prophylaxis strategy, the $PI(X)$ can be calculated using the formula:

$$\begin{aligned} PI^{\text{prophylaxis}}(X) = &- 0.020 \cdot (\text{gestational age in days} - 196) - 0.001 \cdot (\text{postnatal age in hrs} - 96) + \\ &- 0.485 \cdot \text{SGA} + 0.495 \cdot \text{mechanical ventilation} - 0.139 \cdot \text{necrotizing enterocolitis} + \\ &+ 0.151 \cdot \text{sepsis} - 0.020 \cdot (\text{PC} - 39) + 0.246 \cdot \text{number of previous PT} + \\ &+ 1.095 \cdot \text{vasoactive agents} - 2.770 \cdot (\text{LM}_t/100) + 0.737 \cdot (\text{LM}_t^2/100^2) + \\ &+ 1.382 \cdot (\text{LM}_t/100) \times \text{SGA} \end{aligned}$$

with SGA = small-for-gestational age (birth weight <p10), PC = platelet count ($\times 10^9/L$), PT = platelet transfusion and LM_t = landmark time (hours since first platelet count $< 50 \times 10^9/L$).

For the no-prophylaxis strategy, the $PI(X)$ can be calculated using the formula:

$$\begin{aligned} PI^{\text{no prophylaxis}}(X) = &- 0.018 \cdot (\text{gestational age in days} - 196) - 0.001 \cdot (\text{postnatal age in hrs} - 96) + \\ &- 0.351 \cdot \text{SGA} + 0.934 \cdot \text{mechanical ventilation} + 0.614 \cdot \text{necrotizing enterocolitis} + \\ &- 0.180 \cdot \text{sepsis} - 0.061 \cdot (\text{PC} - 39) + 0.225 \cdot \text{number of previous PT} + \\ &+ 1.033 \cdot \text{vasoactive agents} - 0.121 \cdot (\text{LM}_t/100) - 0.839 \cdot (\text{LM}_t^2/100^2) \\ &+ 0.642 \cdot (\text{LM}_t/100) \times \text{SGA} \end{aligned}$$

with SGA = small-for-gestational age (birth weight <p10), PC = platelet count ($\times 10^9/L$), PT = platelet transfusion and LM_t = landmark time (hours since first platelet count $< 50 \times 10^9/L$).

We further illustrate this using Example Patient 1, whose predicted risks are also presented in Figure 5.3 of the manuscript. This infant was born at a gestational age of 28 weeks (196 days) with a normal birth weight (>10th percentile), a postnatal age of 7 days (168 hours) at severe thrombocytopenia onset, and has necrotizing enterocolitis/sepsis. The infant is receiving mechanical ventilation and hemodynamically stable, with no need for vasoactive agents. At the time of prediction, the platelet count is $25 \times 10^9/L$, and no prior platelet transfusions have been administered.

The absolute risk of major bleeding or mortality within the next 3 days under the prophylaxis strategy and the no-prophylaxis strategy can be calculated as follows:

$$\begin{aligned} PI^{\text{prophylaxis}}(X) = &-0.020 \cdot (196-196) - 0.001 \cdot (168-96) - 0.485 \cdot 0 + 0.495 \cdot 1 + \\ &- 0.139 \cdot 1 + 0.151 \cdot 1 - 0.020 \cdot (25-39) + 0.246 \cdot 0 + 1.095 \cdot 0 + \\ &- 2.770 \cdot (0/100) + 0.737 \cdot (0^2/100^2) + 1.382 \cdot (0/100) \cdot 0 = 0.722 \end{aligned}$$

$$P^{\text{prophylaxis}}(3 \mid X) = 1 - S_0(3)^{\exp(PI)} = 1 - (1-0.112)^{\exp(0.722)} = 0.217 = 21.7\%$$

$$\begin{aligned} PI^{\text{no prophylaxis}}(X) = &-0.018 \cdot (196-196) - 0.001 \cdot (168-96) - 0.351 \cdot 0 + 0.934 \cdot 1 + \\ &+ 0.614 \cdot 1 - 0.180 \cdot 1 - 0.061 \cdot (25-39) + 0.225 \cdot 0 + 1.033 \cdot 0 + \\ &+ 0.121 \cdot (0/100) - 0.839 \cdot (0^2/100^2) + 0.642 \cdot (0/100) \cdot 0 = 2.13 \end{aligned}$$

$$P^{\text{no prophylaxis}}(3 \mid X) = 1 - S_0(3)^{\exp(PI)} = 1 - (1-0.043)^{\exp(2.13)} = 0.309 = 30.9\%$$

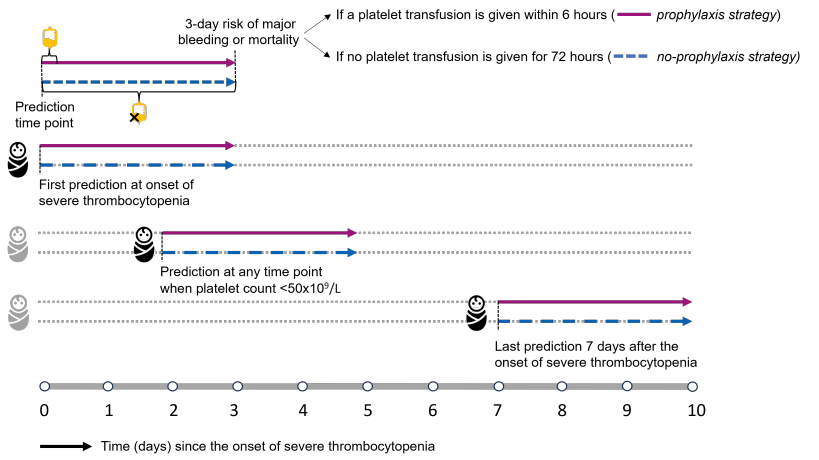


Figure 5.8: Illustration of the use of the model to estimate the 3-day risk of major bleeding or mortality with and without subsequent platelet transfusion. The decision-support algorithm can be used to repeatedly predict the risk of major bleeding or mortality in the next 3 days at any time the infant's clinical condition changes from the first platelet count $< 50 \times 10^9/L$ up to 7 days for two scenarios: if an infant were to receive prophylactic platelet transfusion (prophylaxis strategy, solid purple line) and if an infant were not to receive transfusion (no-prophylaxis strategy, dashed blue line).

References

1. Luijken K, Morzywołek P, Van Amsterdam W, *et al.* Risk-Based Decision Making: Estimands for Sequential Prediction Under Interventions. *Biometrical Journal.* 2024;66(8):e70011.
2. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American Journal of Epidemiology.* 2016;183(8):758–764.
3. Lambert MP, Sullivan SK, Fuentes R, French DL, Poncz M. Challenges and Promises for the Development of Donor-Independent Platelet Transfusions. *Blood.* 2013;121(17):3319–3324.
4. Keogh RH, van Geloven N. Prediction Under Interventions: Evaluation of Counterfactual Performance Using Longitudinal Observational Data. *Epidemiology.* 2024;35(3):329–339.
5. Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology.* 2000;11(5):550–560.
6. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology.* 2008;168(6):656–664.
7. Hernán M, Robins JM. *Causal Inference*. Chapman & Hall/CRC Monographs on Statistics & Applied ProbabilityBoca Raton: Chapman & Hall/CRC 2023.
8. Hernán MA. How to Estimate the Effect of Treatment Duration on Survival Outcomes Using Observational Data. *BMJ (Clinical research ed.).* 2018;360:k182.
9. Chesnaye NC, Stel VS, Tripepi G, *et al.* An Introduction to Inverse Probability of Treatment Weighting in Observational Research. *Clinical Kidney Journal.* 2022;15(1):14–20.
10. von Lindern JS, Hulzebos CV, Bos AF, Brand A, Walther FJ, Lopriore E. Thrombocytopaenia and Intraventricular Haemorrhage in Very Premature Infants: A Tale of Two Cities. *Archives of Disease in Childhood - Fetal and Neonatal Edition.* 2012;97(5):F348-F352.
11. Curley A, Stanworth SJ, Willoughby K, *et al.* Randomized Trial of Platelet-Transfusion Thresholds in Neonates. *New England Journal of Medicine.* 2019;380(3):242–251.
12. Honohan Á, Van'T Ende E, Hulzebos C, *et al.* Posttransfusion Platelet Increments after Different Platelet Products in Neonates: A Retrospective Cohort Study. *Transfusion.* 2013;53(12):3100–3109.
13. Tobian AA, Fuller AK, Uglik K, *et al.* The Impact of Platelet Additive Solution Apheresis Platelets on Allergic Transfusion Reactions and Corrected Count Increment (CME). *Transfusion.* 2014;54(6):1523–1529.
14. Dannaway DC, Noori S. A Randomized Trial of Platelet Transfusions over 30 vs 120 Minutes: Is There an Effect on Post-Transfusion Platelet Counts?. *Journal of Perinatology.* 2013;33(9):703–706.
15. VanderWeele TJ, Hernan MA. Causal Inference under Multiple Versions of Treatment. *Journal of Causal Inference.* 2013;1(1):1–20.
16. Sanquin . Bloedwijzer: Informatie Voor Gebruikers van Korthoudbare Bloedproducten En Omniplasma. 2023.
17. Thomas S, Kelliher S, Krishnan A. Heterogeneity of Platelets and Their Responses. *Research and Practice in Thrombosis and Haemostasis.* 2024;8(2):102356.
18. Stanworth SJ, Shah A. How I Use Platelet Transfusions. *Blood.* 2022;140(18):1925–1936.
19. Rinder HM, Murphy M, Mitchell JG, Stocks J, Ault KA, Hillman RS. Progressive Platelet Activation with Storage: Evidence for Shortened Survival of Activated Platelets after Transfusion. *Transfusion.* 1991;31(5):409–414.
20. van Geloven N, Giardiello D, Bonneville EF, *et al.* Validation of Prediction Models in the Presence of Competing Risks: A Guide through Modern Methods. *BMJ.* 2022:e069249.
21. Hunault C, Habbema J, Eijkemans M, Collins J, Evers J, te Velde E. Two New Prediction Rules for Spontaneous Pregnancy Leading to Live Birth among Subfertile Couples, Based on the Synthesis of Three Previous Models. *Human Reproduction.* 2004;19(9):2019–2026.