Not all that glitters is gold

Platelet transfusions for preterm neonates

Fustolo-Gunnink, S.F.

06

Preterm neonates benefit from low prophylactic platelet transfusion threshold despite varying risk of bleeding or death.

Susanna F. Fustolo-Gunnink, Karin Fijnvandraat, David van Klaveren, Simon J. Stanworth, Anna Curley, Wes Onland, Ewout W. Steyerberg, Ellen de Kort, Esther J. d'Haens, Christian V. Hulzebos, Elise J. Huisman, Willem P. de Boode, Enrico Lopriore and Johanna G. van der Bom for the PlaNeT-2 MATISSE Collaborators

Revised version published in: Blood 2019 Oct 24. Epub ahead of print.

ABSTRACT

The Platelets for Neonatal Thrombocytopenia (PlaNeT-2) trial reported an overall unexpected benefit of a prophylactic platelet transfusion threshold of $25x10^9$/L compared to $50x10^9$/L for major bleeding and/or mortality in preterm neonates (7% absolute risk reduction). However, some neonates in the trial may have experienced little benefit or even harm from the $25x10^9$/L threshold. We aimed to assess this heterogeneity of treatment effect in the PlaNet-2 trial, in order to investigate whether all preterm neonates benefit from the low threshold.

We developed a multivariable logistic regression model in the PlaNet-2 data to predict baseline risk of major bleeding and/or mortality for all 653 neonates. We then ranked the neonates based on their predicted baseline risk and categorized them into four risk quartiles. Within these quartiles we assessed absolute risk difference between the $50x10^9$/L and $25x10^9$/L threshold group.

A total of 146 neonates died or developed major bleeding. The internally validated C-statistic was 0·63 (95% confidence interval 0·58 – 0·68). The $25x10^9$/L threshold was associated with absolute risk reduction in all risk groups, varying from 4·9% in the lowest to 12·3% in the highest risk group.

These results suggest that a $25x10^9$/L prophylactic platelet count threshold can be adopted in all preterm neonates, irrespective of predicted baseline outcome risk. Future studies are needed to improve the predictive accuracy of the baseline risk model.

INTRODUCTION

Preterm neonates with severe thrombocytopenia (platelet count $<50x10^9$/L) are often treated with prophylactic platelet transfusions, despite lack of evidence for their efficacy.[1,2] In the recently published Platelets for Neonatal Thrombocytopenia (PlaNeT-2) trial, a platelet count threshold of $50x10^9$/L for prophylactic platelet transfusion increased the risk of the composite outcome major bleeding and/or mortality when compared to a lower threshold of $25x10^9$/L (odds ratio (OR) 1·57, 95% confidence interval

(CI) 1·06 – 2·32).[3] It is likely, however, that there was heterogeneity of treatment effect, with some neonates benefitting more, some less, and some not at all, from using the lower threshold. And perhaps some neonates benefitted from the higher threshold.

This heterogeneity of treatment effect can be caused by variation in baseline risk, which can lead to an overall trial result that does not represent the treatment effect in the majority of patients in the trial, for two reasons.[4,5] Firstly, most, if not all, patients in clinical trials are non-average, and their individual baseline risks may differ substantially from the average baseline risk. Secondly, baseline risk in a trial population is usually not normally distributed, because most trial outcomes occur in a relatively small number of high-risk patients, while the majority of patients are at much lower than average baseline risk.[5,6] Previous studies have shown that large variations in baseline risk were present in the populations of a substantial percentage of trials, and often lead to clinically significant differences in absolute treatment effects.[7] If the overall trial result does not represent the treatment effect in the majority of patients in the trial, then there are limitations translating the results into clinical practice for individual patients. Thus, before translating the results of the PlaNeT-2 trial into clinical practice, the presence of heterogeneity of treatment effect needs to be evaluated.

> HETEROGENEITY OF TREATMENT EFFECT CAN BE CAUSED BY VARIATION IN BASELINE RISK, WHICH CAN LEAD TO AN OVERALL TRIAL RESULT THAT DOES NOT REPRESENT THE TREATMENT EFFECT IN THE MAJORITY OF PATIENTS IN THE TRIAL

Therefore, the objective of our study was to explore heterogeneity of treatment effect in the PlaNet-2 trial, in order to assess whether there are specific groups of neonates who do or do not benefit from a low platelet count transfusion threshold.

METHODS

We used data from the PlaNet-2 trial (n=660), a multicenter clinical trial with randomized treatment group assignment, open label treatment and open label end point evaluation. A prophylactic platelet transfusion threshold of $25x10^9$/L was compared to a threshold of $50x10^9$/L in neonates with severe thrombocytopenia, defined as a platelet count $<50x10^9$/L. The primary outcome was a composite of mortality and/or major bleeding

within 28 days of randomization. Neonates were randomized between June 2011 and August 2017 in 43 neonatal intensive care units in the United Kingdom, the Netherlands and Ireland. Definitions of major bleeding and other relevant details can be found in the protocol and report as published elsewhere.[8,9]

The PlaNeT-2 trial protocol was approved by independent ethics committees in the United Kingdom, the Netherlands, and Ireland. Parents or caretakers of all study participants gave written informed consent. The Current Controlled Trials number was ISRCTN87736839. The current study was conducted in accordance with the Declaration of Helsinki and reported according to The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines.[10]

VARIABLE SELECTION

A limited number of variables was selected prior to data analysis, a recommended strategy to avoid overfitting.[11,12] A set of 6 categorical subgroups was prespecified in the trial statistical analysis plan, as they were considered important predictors for outcome: gestational age at birth < 28 weeks, postnatal age at randomization <3 days, 3-7 days or >7 days, intra-uterine growth retardation (IUGR), necrotizing enterocolitis (NEC) at randomization, sepsis at randomization and administration of antenatal steroids. We defined IUGR as a birth weight in less than the 10th percentile in conjunction with an estimated fetal weight crossing percentiles downward during pregnancy, ultrasonographic evidence of uteroplacental insufficiency, or both. We defined NEC as stage 2A or higher, according to modified Bell's staging criteria[13], and sepsis as a culture positive sepsis or culture negative sepsis where a course of at least 5 days of antibiotics is to be administered for proven or clinically-suspected sepsis. Antenatal steroids were included as dichotomous variable: steroids were or were not administered (irrespective of whether a full or partial course was given). We have included gestational and postnatal age as continuous variables in our model. We also added sex to the model, as it is frequently included in existing prediction models for major bleeding or mortality. (Table S1) Treatment assignment was added to correct for any imbalance between the treatment groups despite randomization. Treatment assignment did not affect calculation of baseline risk, because this was calculated assuming the low threshold group for all neonates.

CODING OF VARIABLES

We dealt with missing data using single imputation, as the number of missing values was low (Table 1). We used median values for continuous and modal values (value that occurs most often) for categorical variables. We allowed for non-linearity of continuous variables by restricted cubic spline functions, with 2 degrees of freedom.[14]

MODEL SPECIFICATION AND ESTIMATION

We developed a logistic regression model to predict baseline risk of outcome. Baseline risk was calculated assuming a low threshold assignment for all neonates. No treatment interactions were added to the model as no prior evidence of strong interaction effects was available and our sample size was limited.[15] We used a full model approach: variables were selected *a priori* and were not removed from the model based on statistical significance. All analyses were performed in R, version 3.3.x.

MODEL PERFORMANCE AND VALIDITY

We expressed model performance as discrimination and calibration. Discrimination was quantified using the C-statistic (equal to the area under the receiver-operating characteristic curve for dichotomous outcomes). The C-statistic estimates the probability that out of two randomly chosen patients, the patient with the outcome has a higher predicted probability of major bleed and/or mortality than the patient without the outcome. We internally validated the C-statistic with a bootstrap procedure to correct for the optimism caused by using the same data for development and validation of the model.[12,14] Calibration refers to the agreement between predicted and observed risks and was assessed graphically with a validation plot.

ASSESSMENT OF HETEROGENEITY OF TREATMENT EFFECT

We predicted baseline risk of outcome for all neonates using the logistic regression model. We then ranked the neonates based on their predicted risk and categorized them into four risk quartiles (very low, low, moderate and high risk). Within these quartiles we assessed absolute risk differences between the high and low threshold group. We presented the absolute risk difference and confidence interval for each risk group.

We have defined heterogeneity of treatment effect on an absolute scale, because this is generally considered to be more clinically relevant than a relative scale.[16,17] For example, an absolute risk difference of 5% may be clinically relevant, even though the relative risk difference (e.g. OR) is minimal. On the contrary, a twofold increase in relative risk is most likely clinically insignificant when baseline risk is extremely low (e.g. absolute risk increases from 0.01% to 0.02%).

ROLE OF THE FUNDING SOURCE

This research was supported by the National Health Service Blood and Transplant Research and Development Committee (Ref No: BS06/1); Sanquin Research, Amsterdam (grant PPOC-12-012027); Addenbrooke's

Charitable Trust; the Neonatal Breath of Life Fund 9145; and the National Institute for Health Research Clinical

Research Network. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

DATA SHARING STATEMENT

Any requests for de-identified trial data and supporting material (data dictionary, protocol, and statistical analysis plan) will be reviewed by the trial management group. Requests that have a methodologically sound proposal and whose proposed use of the data has been approved by the trial's independent trial steering committee will be considered. Proposals should be directed to the corresponding author in the first instance (acurley@nmh.ie) ; to gain access, data requestors will need to sign a data access agreement.

RESULTS

A total of 660 neonates were randomized in the PlaNeT-2 trial, with a median gestational age at birth of 26·7 weeks and median postnatal age at randomization of 7·5 days. Baseline characteristics of the study population are presented in Table 1. Seven neonates were either randomized in error or primary outcome data were missing and could not be inferred, leaving a total of 653 neonates for analysis. Neonatal death and/or major bleeding occurred in 19% of neonates in the 25-group and 26% of neonates in the 50-group, with an absolute risk difference of 7%. This corresponds to a number needed to treat (NNT) of 14, indicating that for on average every 14 neonates transfused according to the 25x10^9/L threshold, one major bleed or death may be prevented. More details on baseline characteristics and outcome descriptions can be found in the original report of the trial.[9]

In our baseline risk prediction model, presence of a previous major bleed, lower gestational age and treatment assignment to the 50x10^9/L group were independently associated with increased risk of outcome (Table 2). Postnatal age, antenatal corticosteroids, IUGR, female sex and sepsis were not independently associated with the outcome. The odds ratio for NEC was 1·62, with a 95% confidence interval of 0·95 - 2·78. The internally validated C-statistic was 0·63 (95% CI 0·58 - 0·68).

Table 1. Baseline characteristics (N=653)

	Total cohort (n=653)		Major bleed / death (n=146)		No major bleed / death (n=507)	
Gestational age in weeks: median (IQR)[a]	26·7	(24·9-28·7)	27·0	(25·0 - 29·0)	26·0	(24·6 - 28·0)
Postnatal age in days: median (IQR)[b]	7·5	(3·9 - 20·5)	8·7	(3·9 - 19·1)	7·0	(3·8 - 21·2)
Male, n (%)[c]	394	(60·3)	88	(60·3)	306	(60·4)
Intra-uterine growth retardation, n (%)	243	(37·2)	46	(31·5)	197	(38·9)
Antenatal corticosteroids, n (%)[d]	580	(88·8)	127	(87·0)	453	(89·3)
Sepsis, n (%)	412	(63·1)	102	(69·9)	310	(61·1)
Necrotizing enterocolitis, n (%)	107	(16·4)	31	(21·2)	76	(15·0)
Previous major bleeding, n (%)	120	(18·4)	38	(26·0)	82	(16·2)

IQR = interquartile range.
[a] In 5 cases the exact gestational age could not be determined due to uncontrolled pregnancies. It was estimated in full weeks.
[b] Missing data in 2 cases, single imputation using median.
[c] missing data in 1 cases, single imputation using mode.
[d] Missing data in 4 cases, single imputation using mode.

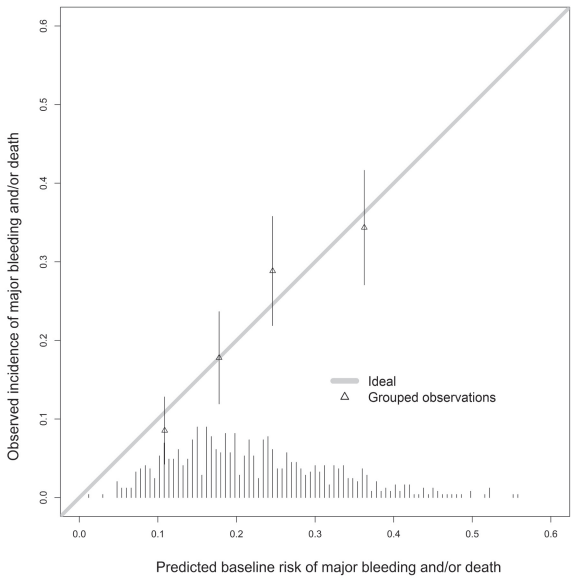
Table 2. Multivariable analysis of death or major bleeding within 28 days of randomization (N=653)

	Odds ratio	95% CI	P-value
Gestational age (days)[1]	0·58	0·40 – 0·84	0·004
Postnatal age (days)[1]	0·68	0·40 – 1·16	0·155
Antenatal corticosteroids	0·76	0·41 – 1·39	0·373
Intrauterine growth retardation	0·95	0·60 – 1·50	0·825
Female sex	0·96	0·65 – 1·43	0·848
Sepsis	1·32	0·86 – 2·02	0·204
Treatment ($50x10^9$/L threshold)	1·58	1·08 – 2·33	0·019
Necrotizing enterocolitis	1·62	0·95 – 2·78	0·079
Previous major bleed	1·63	1·00 – 2·65	0·049

[1] Interquartile range odds ratio.

Figure 1 shows that the model was well calibrated in all quartiles of predicted risk. The four triangles represent the predicted risk of outcome in the four risk groups. These triangles approximate the diagonal line, which represents the ideal calibration curve, where observed and predicted risks are identical.

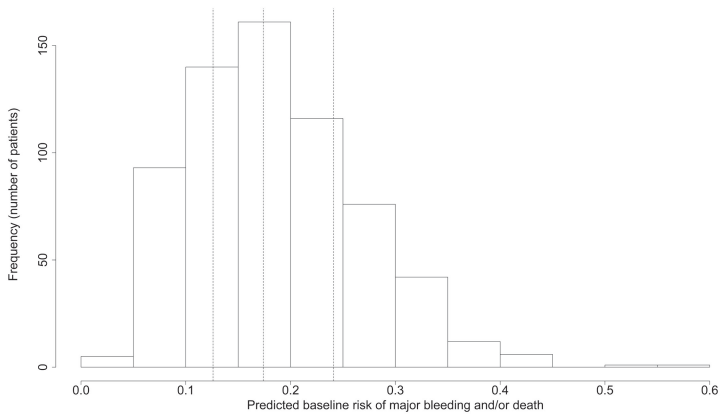
Figure 1. Validation plot.



Validity of predictions of major bleeding or mortality. The triangles indicate the predicted probabilities and observed frequencies for all four risk groups. The diagonal line represents perfect calibration: observed and predicted probabilities are identical. The distribution of predicted probabilities is shown at the bottom of the graphs.

The distribution of predicted baseline risk in the trial is presented in Figure 2. The median baseline risk was 21·0%. Cutoff points for the four risk quartiles in ascending order were 12·6%, 17·4% and 24·1% respectively.

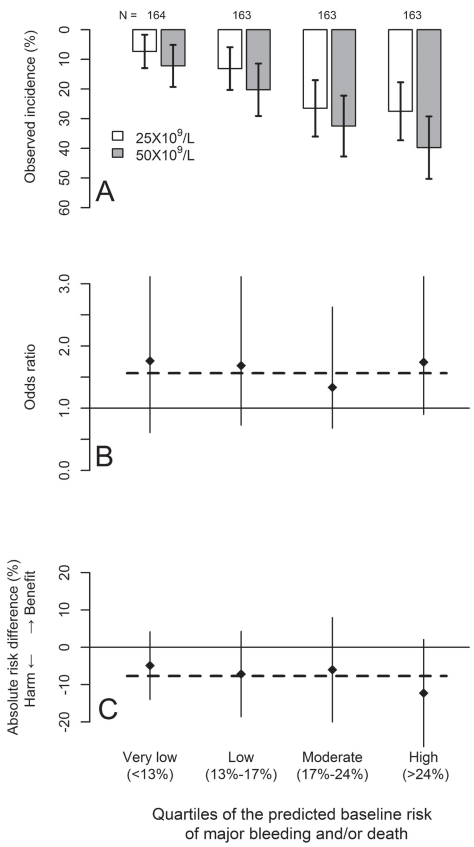
Figure 2 Distribution of predicted absolute risk of outcome (n=653).



Predicted absolute risk of outcome is represented as the absolute risk of outcome on the x-axis and the frequency of each absolute risk category (0-0·05, 0·05-0·1, 0·1-0·15 etc) in the trial population on the y-axis.

Figure 3 shows the treatment effect, defined as absolute risk difference, in the four risk categories. Panel A shows increasing event rates in the four risk groups, presented separately for the 25x10^9/L and 50x10^9/L study groups. Panel B show a more or less constant relative risk reduction (OR), with the horizontal line indicating the overall trial result (OR 1·57). Panel C shows the variation in absolute risk difference between the risk groups, with the horizontal line indicating the overall trial result (7% risk difference). There is absolute risk reduction associated with the low threshold in all risk groups, varying from an absolute risk difference of 4·9% in the lowest risk group to 12·3% in the highest risk group. These values correspond with an NNT of 21 in the lowest and 8 in the highest risk group.

Figure 3. Absolute risk difference (ARD) between a high (50x10^9/L) and low (25x10^9/L) threshold for prophylactic platelet transfusion thresholds in preterm neonates with respect to major bleeding and/or mortality within 28 days after randomization.



Event rates (panel A), odds ratios (panel B) and absolute risk differences (panel C) are presented for all four risk categories, vertical lines represent 95% confidence intervals, horizontal lines represent overall trial results. A negative absolute risk reduction represents the risk decrease for a low prophylactic platelet transfusion threshold as compared to a high threshold.

DISCUSSION

We aimed to identify groups of neonates who experienced benefit from the low transfusion threshold in the PlaNet-2 trial. In order to investigate this, we assessed heterogeneity of treatment effect due to variations in baseline risk, using an internally validated baseline risk prediction model. Our results suggest that all neonates experienced benefit from the low threshold, as the low threshold was associated with absolute risk reduction in all risk groups. However, the absolute benefit varied considerably, from 4·9% in the lowest to 12·3% in the highest risk group.

These findings suggest that neonates with high predicted baseline risk are as vulnerable for harm associated with a higher transfusion threshold as neonates with low predicted baseline risk. These results appear contradictory to recommendations found in some guidelines that suggest platelet transfusion thresholds above 25x10^9/L for neonates with suspected higher baseline risk.[18–20] For example, these guidelines suggest using thresholds higher than 25x10^9/L for sick neonates with lower gestational age and/or birthweight. Clinicians who may have been reluctant to implement the results of PlaNet-2 in their smallest and sickest neonates, can now be more confident that even this population is likely to benefit from using a lower platelet transfusion threshold.

> THESE FINDINGS SUGGEST THAT NEONATES WITH HIGH PREDICTED BASELINE RISK ARE AS VULNERABLE FOR HARM ASSOCIATED WITH A HIGHER TRANSFUSION THRESHOLD AS NEONATES WITH LOW PREDICTED BASELINE RISK.

Gestational age, previous major bleeding and treatment assignment were independent predictors of outcome. Gestational age has been shown to predict bleeding and mortality in several prediction models (Table S1). Major bleeding prior to randomization (mainly IVH and pulmonary hemorrhage) occurred in 122 out of 660 neonates. Some of these neonates may have developed these bleeds during severe thrombocytopenia, as 39% of all neonates included in the PlaNeT-2 trial received platelet transfusions prior to randomization. This complicates accurate interpretation of these data, and further studies are needed to confirm previous major bleeding as a predictor for new major bleeding and/or death after onset of severe thrombocytopenia. Treatment assignment also predicted outcome, which was expected given the overall trial results. The remaining variables were not shown to be independent predictors in our model, but this may partially be due to lack of power (e.g. NEC). In addition, variables such as

postnatal age are thought to be important predictors for IVH, but as the incidence of IVH was low in the trial, their effect on the (composite) outcome may have been limited. Postnatal age was found to be an independent predictor of major bleeding in a recently published observational cohort study in which the incidence of IVH was higher.[21]

Strengths of our study are the randomized design of the main trial, overall high levels of completeness of data for primary outcome, pre-defined selection of variables to be included in the model, and agreeing on an analysis plan prior to starting the analyses. In addition, our risk-based analysis is superior to conventional subgroup analyses, as it focuses on absolute risk reduction, which has more clinical relevance than interactions on the relative scale which are assessed in conventional subgroup analysis. In addition, conventional subgroup analyses are often severely underpowered, because they require multiple testing, and many fail to meet best practices for subgroup testing. Although the method used in this paper is also underpowered (as the trial was powered only for the primary analysis), it has better power than a conventional subgroup analysis would have had, because only one variable (baseline risk) is compared between the study arms. Lastly, the current method allows for calculation of absolute risk reductions given a combination of multiple clinical characteristics, which is not possible with conventional subgroup analyses.[22–30]

Various limitations of our study need to be considered. Firstly, our sample size did not allow inclusion of interaction terms in our model, and no prior evidence of strong interaction effects was available. Clinically relevant interactions are another potential source of heterogeneity of treatment effect in addition to baseline risk variation, and therefore we may not have identified all heterogeneity of treatment effect that was present in the trial. A model with interaction terms can also be used to predict individualized treatment effect, which is the ultimate goal of personalized medicine.[31] Further studies are needed to assess this. Secondly, the C-statistic of our model indicates moderate discrimination, despite having selected variables that are generally considered to be important predictors for major bleeding and/or mortality. This could be due to the fact that there are important risk factors that we have not included in our model. For example, mechanical ventilation was shown to be a good predictor of major bleeding in a recently published dynamic prediction model.[21], but ventilation data were not collected in the PlaNet-2 study. Another explanation is that some risk factors applied mainly to IVH, and as its incidence was low, they did not perform well in our dataset.[9] It is also possible that baseline prediction models underperform because risk of outcome changes substantially as a result of clinical events that occur after baseline. This hypothesis is supported by the performance of the previously mentioned dynamic prediction model for major bleeding in preterm neonates, which had a median C-statistic of 0·74 (interquartile range 0·69-0·82). With this model, risk of major bleeding

within the subsequent three days could be predicted at any point in time during the first week after onset of severe thrombocytopenia.[21] Nevertheless, a C-statistic of 0·63 does allow for some degree of risk stratification, as it is higher than 0·5 (which equals chance). This is illustrated in Figure 3. Lastly, the model has to be externally validated before it can be implemented in future studies or clinical settings.

In conclusion, our study demonstrated that the $25x10^9$/L threshold was beneficial compared to the $50x10^9$/L threshold in all predicted baseline risk subgroups, although absolute benefit seemed to vary considerably. These findings suggest that a $25x10^9$/L transfusion threshold may be adopted in all preterm neonates, including those with high predicted risk of major bleeding and/or mortality. Future studies are needed, because our model had moderate discriminative capacity, did not include treatment interaction terms and needs to be externally validated. Ultimately, an improved and validated model will allow for a further refined prediction of individualized treatment effect for platelet transfusion in preterm neonates. This can be used to individualize our platelet transfusion guidelines, and potentially improve outcomes for preterm neonates with severe thrombocytopenia.

ACKNOWLEDGEMENTS

SFFG is a PhD candidate at the University of Amsterdam. This work is submitted in partial fulfillment of the requirement for the PhD.

AUTHORSHIP CONTRIBUTIONS

SFFG, KF, DK, SJS, AC, WO, EWS, EL and JGB designed the study. SFFG prepared the data for analysis, DK analyzed the data. SFFG and DK interpreted the data, SFFG, KF, DK, SJS, AC, WO, EWS, EL and JGB wrote the report. All authors revised and approved the final report. SFFG, KF, SJS, AC, WO, EL, EK, EJH, CVH, EJHu, WPB and the PlaNeT-2 MATISSE collaborators were involved in the PlaNeT-2 trial.

CONFLICTS OF INTEREST DISCLOSURES

SFFG, KF, DK, SJS, AC, WO, EWS, EL, EK, EJH, CVH, EJHu, WPB and JGB have nothing to disclose.

KEY POINTS

- A recent RCT showed increased risk of death or bleeding in neonates receiving platelet transfusions for platelet counts above $25x10^9$/L.
- The current analysis reveals that these harmful effects occur in neonates with high as well as low baseline risk of death or bleeding.

A revised version of this research was originally published in Blood. Fustolo-Gunnink SF et al. Preterm neonates benefit from low prophylactic platelet transfusion threshold despite varying risk of bleeding or death. Blood. 2019 Oct 24 (epub ahead of print).

REFERENCES

1. Fustolo-Gunnink SF, Huijssen-Huisman EJ, van der Bom JG, et al. Are thrombocytopenia and platelet transfusions associated with major bleeding in preterm neonates? A systematic review. Blood Rev. 2018;
2. Stanworth SJ, Clarke P, Watts T, et al. Prospective, observational study of outcomes in neonates with severe thrombocytopenia. Pediatrics. 2009;124(5):e826-34.
3. Curley A, Stanworth SJ, Willoughby K, et al. Randomized Trial of Platelet-Transfusion Thresholds in Neonates. N. Engl. J. Med. 2018;380(3):242–251.
4. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. Bmj. 2018;364:k4245.
5. Kent DM, Hayward RA. Limitations of Applying Summary Results of Clinical Trials to Invidivual Patients. J. Am. Med. Assoc. 2007;5–8.
6. Ioannidis JPA, Lau J. The impact of high-risk patients on the results of clinical trials. J. Clin. Epidemiol. 1997;50(10):1089–1098.
7. Nelson J, Kent DM, Dahabreh IJ, et al. Risk and treatment effect heterogeneity: Re-analysis of individual participant data from 32 large clinical trials. Int. J. Epidemiol. 2016;45(6):2075–2088.
8. Curley A, Venkatesh V, Stanworth S, et al. Platelets for neonatal transfusion - Study 2: A randomised controlled trial to compare two different platelet count thresholds for prophylactic platelet transfusion to preterm neonates. Neonatology. 2014;106(2):102–106.
9. Curley A, Stanworth SJ, Willoughby K, et al. Randomized Trial of Platelet-Transfusion Thresholds in Neonates. N. Engl. J. Med. 2018;380(3):242–251.
10. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. BMJ. 2014;(350):g7594.
11. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. Eur. Heart J. 2014;35(29):1925–1931.
12. Steyerberg EW. Clinical prediction models - a practical approach to development, validation and updating. 2009.

13. Bell MJ, Ternberg JL, Feigin RD, et al. Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. Ann. Surg. 1978;187(1):1–7.
14. Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer-Verlag New York, Inc. New York, USA; 2010.
15. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. Bmj. 2018;364:k4245.
16. Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: Understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. Int. J. Epidemiol. 2016;45(6):2184–2193.
17. Rothwell P, Mehta Z, Howard S, Gutnikov S, Warlow C. Treating Individuals 3: From subgroups to individuals: General principles and the example of carotid endarterectomy. Lancet. 2005;365(9455):256–265.
18. CBO. Blood Transfusion Guideline. Utrecht, the Netherlands: CBO; 2011.
19. Liumbruno G, Bennardello F, Lattanzio A, Piccoli P, Rossetti G. Recommendations for the transfusion of plasma and platelets. Blood Transfus. 2009;7(2):132–50.
20. Agence française de sécurité sanitaire des produits de santé. Transfusion de plaquettes: produits, indications. 2003.
21. Fustolo-gunnink SF, Fijnvandraat K, Putter H, et al. Dynamic prediction of bleeding risk in thrombocytopenic preterm neonates. 2019;
22. Hayward RA, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. BMC Med. Res. Methodol. 2006;6:1–11.
23. Sun X, Briel M, Busse JW, et al. Credibility of claims of subgroup effects in randomised controlled trials: Systematic review. BMJ. 2012;344(7852):1–9.
24. Wallac JD, Sullivan PG, Trepanowski JF, et al. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. JAMA Intern. Med. 2017;177(4):554–560.
25. Kent DM, Rothwell PM, Ioannidis JPA, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: A proposal. Trials. 2010;11:1–10.
26. Kent DM, Ruthazer R, Selker HP. Are some patients likely to benefit from recombinant tissue-type plasminogen activator for acute ischemic stroke even beyond 3 hours from symptom onset? Stroke. 2003;34(2):464–467.
27. Costa F, van Klaveren D, James S, et al. Derivation and validation of the predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet therapy (PRECISE-DAPT) score: a pooled analysis of individual-patient datasets from clinical trials. Lancet. 2017;389(10073):1025–1034.
28. Farooq V, Van Klaveren D, Steyerberg EW, et al. Anatomical and clinical characteristics to guide decision making between coronary artery bypass surgery and percutaneous coronary intervention for individual patients: Development and validation of SYNTAX score II. Lancet. 2013;381(9867):639–650.
29. Mulder MJHL, Venema E, Roozenbeek B, et al. Towards personalised intra-arterial treatment of patients with acute ischaemic stroke: A study protocol for development and validation of a clinical decision aid. BMJ Open. 2017;7(3):1–5.

30. Venema E, Mulder MJHL, Roozenbeek B, et al. Selection of patients for intra-arterial treatment for acute ischaemic stroke: Development and validation of a clinical decision tool in two randomised trials. BMJ. 2017;357:.
31. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. Bmj. 2018;2:k4245.

SUPPLEMENTARY MATERIALS

Table S1. Published prediction models assessing major bleeding or mortality in preterm neonates.

Author, year	Model description	Reference
Luque, 2014	A prediction model for severe intraventricular hemorrhage (IVH) was developed in 4747 very low birthweight infants in 6 South-American countries. Model variables were selected using stepwise selection procedures, and no external validation was performed. It is unclear to what extent the variables in the model are predictors or consequences of the bleeds, as no attempt was made to correct for timing of the exposure relative to the outcome. Significant predictors for major IVH were lower gestational age (GA), lower birth weight (BW), male, antenatal steroids, cesarean section, lower 1 min Apgar, mechanical ventilation, respiratory distress syndrome (RDS). Mechanical ventilation is not available in the PlaNet-2 database, and RDS is diagnosed later in life (not at baseline or moment of randomisation).	[1]
Vogtman, 2012	A prediction model for severe IVH was developed in 1782 neonates with GA <32 weeks or BW <1500 grams in Germany. Thirty predictors were tested using stepwise selection while only 149 events occurred during the study. Significant predictors for major IVH were lower GA, 1 min Apgar <6, early sepsis (<3 days), no pathological Doppler results, intrapartum tocolytics. Doppler results and presence or absence of intrapartum tocolytics are not available in the PlaNet-2 database.	[2]
Singh, 2013	A prediction model for severe IVH was developed in 2917 neonates with GA <34 weeks and who had not received indomethacin. Neonates with missing outcome records (n=253) were excluded. It is unclear how selection of predictors was performed. Significant predictors for major IVH were low GA, mode of delivery, low BW, gender, antenatal corticosteroids, caesarean section, outborn delivery and low 5 minute Apgar score.	[3]

Table S1. Continued.

Author, year	Model description	Reference
Fustolo-Gunnink, 2019	A dynamic prediction model for severe bleeding was developed in 640 neonates with GA <34 weeks and at least one platelet count <50x10^9/L. In this model, though not a baseline model, postnatal age and mechanical ventilation were predictors with a time-constant effect. Intra-uterine growth retardation (IUGR) and platelet count were predictors with time-varying effect, their effect may not have been present at baseline. In a sensitivity analysis, the model also accurately predicted mortality risk.	[4]
NICHD, 2008	The NICHD outcomes model predicts a set of outcomes including mortality in infants with GA 22-25 weeks. This model contains GA, BW, sex, singleton birth and antenatal corticosteroids.	[5]

REFERENCES

1. Luque MJ, Tapia JL, Villarroel L, et al. A risk prediction model for severe intraventricular hemorrhage in very low birth weight infants and the effect of prophylactic indomethacin. J. Perinatol. 2014;34(1):43–8.
2. Vogtmann C, Koch R, Gmyrek D, Kaiser A, Friedrich A. Risk-adjusted intraventricular hemorrhage rates in very premature infants. Dtsch. Arztebl. Int. 2012;109(31–32):527–533.
3. Singh R, Gorstein S V, Bednarek F, et al. A predictive model for SIVH risk in preterm infants and targeted indomethacin therapy for prevention. Sci. Rep. 2013;3:2539.
4. Fustolo-gunnink SF, Fijnvandraat K, Putter H, et al. Dynamic prediction of bleeding risk in thrombocytopenic preterm neonates. 2019;
5. Tyson JE, Parikh NA, Langer J, Green C, Higgins RD. Intensive Care for Extreme Prematurity — Moving beyond Gestational Age. N. Engl. J. Med. 2008;358(16):1672–1681.