A multicentred study to validate a consensus bleeding assessment tool developed by the biomedical excellence for safer transfusion collaborative for use in patients with haematological malignancy

C. Dyer,[1,*] C. R. Alquist,[2] M. Cole-Sinclair,[3] E. Curnow,[5] N. M. Dunbar,[6] L. J. Estcourt,[7,8,9] R. Kaufman,[10] J. M. Kutner,[11] J. McCullough,[12] Z. McQuilten,[13,14] L. Potiphar,[15] B. Rioux-Masse,[16] S. Slichter,[4] A. Tinmouth,[17] K. Webert,[18,19] P. Yokoyama,[11] S. J. Stanworth[8,9,20,21] & On behalf of the BEST Collaborative

[1]*NHS Blood and Transplant Clinical Trials Unit, Headington, Oxford, UK*

[2]*Section of Transfusion Medicine and Histocompatibility, Department of Pathology & Laboratory Medicine, Ochsner Health System, New Orleans, LO, USA*

[3]*Laboratory Haematology, Pathology St Vincent's Hospital, Melbourne, Vic., Australia*

[4]*Bloodworks Northwest, Seattle, WA, USA*

[5]*Statistics and Clinical Studies, NHS Blood and Transplant, Bristol, UK*

[6]*Department of Pathology and Laboratory Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA*

[7]*NHS Blood and Transplant, Oxford, UK*

[8]*Radcliffe Department of Medicine, University of Oxford, Oxford, UK*

[9]*Oxford BRC Haematology Theme, Oxford, UK*

[10]*Pathology, Brigham and Women, Boston, MA, USA*

[11]*Hemotherapy and Cell Therapy Department, Hospital Israelita Albert Einstein, Sao Paulo, Brazil*

[12]*University of Minnesota, Minneapolis, MN, USA*

[13]*Department of Haematology, St Vincent's Hospital Melbourne, Melbourne, Vic., Australia* [14]*Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Vic., Australia* [15]*Birmingham City University, Birmingham, UK*

[16]*Department of Hematology and Transfusion Medicine, Centre Hospitalier de l'Universite de Montreal, Montreal, QC, Canada*

[17]*Ottawa Hospital Research Institute, Ottawa, ON, Canada*

[18]*Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada*

[19]*Medical Science and Innovation, Canadian Blood Services, Ancaster, ON, Canada*

[20]*Transfusion Medicine, NHS Blood and Transplant, Oxford, UK*

[21]*Department of Haematology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK*

Background There continues to be uncertainty about the optimal approach to documenting bleeding data in platelet transfusion trials, with a desire to apply a common assessment tool across all trials. With this in mind, a consensus bleeding assessment tool (BAT) has been developed by the Biomedical Excellence for Safer Transfusion (BEST) collaborative, based on review of data collection forms used in published randomized trials and following content validation with a range of healthcare professionals at seven haematology centres through BEST members. This study aimed to evaluate reliability and reproducibility of the consensus BAT.

Methods Replicated clinical assessments of bleeding were undertaken by participants with haematological malignancies recruited at four haematology centres in an international, multicentred, observational study. Concordance of repeat assessments was calculated for agreement in site and grade of bleeding observed.

Results Forty patients consented to participate, and 13 trained bleeding assessors collected these data. Bleeding assessments were carried out on 113 separate days. Of all 225 bleeding assessments, 204 were compared for grade concordance, and
160 were compared for site concordance. There was very good grade concordance (83%, 95% confidence interval 74–93%) and good bleeding site concordance (69%, 95% confidence interval 57–79%) in observations of bleeding. Discordance was primarily in relation to assessing skin bleeding.

Conclusions Alongside a structured training programme, levels of concordance for a consensus BAT were high. Researchers using assessment tools for bleeding need to balance comprehensive data collection against potential loss of accuracy for some types of bleeding, such as skin findings.

Key words: clinical trial, platelet transfusions, transfusion therapy.

Introduction

Bleeding is a clinically meaningful and relevant outcome in platelet transfusion trials, or alternatives to platelet transfusions, such as tranexamic acid [1]. If bleeding is used as a main outcome measure in clinical trials, it is important that it is defined and documented in a transparent, consistent and standardized way. There are several key considerations when bleeding is used as an outcome measure in patients with thrombocytopenia and haemato- logical malignancies: what signs and symptoms of bleeding are recorded; how is the assessment of bleeding documented; how is consistency of assessment assured across multiple participating centres; and how are these data translated into a score or grade of bleeding?

It is well recognized that there are variations in the methodology to record and grade bleeding between trials conducted in patients with thrombocytopenia as a complication of haematological malignancies and/or its treatment [2]. This creates challenges in comparing and interpreting data across studies. Because of this heterogeneity, it is not always possible to compare studies with any great confidence [3]. While there are existing bleeding scales, such as the Bleeding Severity Measurement Scale [4] and the more commonly used World Health Organization (WHO) system [5], which translate bleeding data into a grade, there is no validated tool to support standardized collection of data. In summary, there continues to be uncertainty about the optimal approach to document bleeding data in platelet transfusion trials, alongside a desire to apply a common assessment tool across all trials facilitating comparison of results between studies.

This study describes a prospective multicentred, obser vational cohort study. The consensus Bleeding Assessment Tool (BAT) used for this study was developed by review of all previous published tools [2], followed by hospital level content validation at seven Biomedical Excellence for Safer Transfusion (BEST) member centres. The overarching aim was to evaluate the consensus tool's ability to facilitate a standardized and reproducible approach to the collection of data required for assessment of bleeding. Secondary aims included assessments of ease of use; ability to capture all possible bleeding events; and translation of data into a clinically meaningful grade. For the purposes of this study, the WHO grading system was used, as it is the most commonly used and reported classification system of bleeding severity in platelet transfusion trials [2].

Methods

Study design

An international multicentre, observational cohort study with repeat bleeding assessments conducted at four haematology departments in Australia, Brazil, UK and USA.

Study centres and bleeding assessors

All four participating centres were selected by clinician members of the Biomedical Excellence for Safer Transfusion (BEST) Collaborative who acted as the centres' Principal Investigators. Bleeding assessments were performed either by the Principal Investigator or staff who had been trained on the use of the bleeding assessment tool.

Study participants

Adult in-patients diagnosed with a haematological malignancy, receiving myelosuppressive chemotherapy or a stem cell transplant and expected to develop a thrombo- cytopenia of 20 9 10^9/l or less for at least 3 days were included.

The consensus BEST Bleeding Assessment Tool (BAT)

The consensus BAT describes the data collection forms used to assess the severity and site(s) of bleeding in this study (Fig. 1). This was developed by review of all previous tools used to collect information on bleeding in the published literature [2], followed by content validation at seven (BEST member centres. Terms were refined using organ-specific definitions adapted from the different trial grading systems [2]. Up to six staff members, aiming for a mix of senior staff, junior staff in training and senior nurses, provided feedback as to whether the tool assessed the full breadth of potential bleeding complications seen in patients with haematological malignancies.

The consensus BAT was composed of four separate forms (Appendix S1). The forms collected information on:

- how bleeding was assessed,
- the severity of bleeding and follow-up and the site of bleeding.

Assigning a bleeding grade

Bleeding was defined using the WHO bleeding criteria. In the WHO system, bleeding episodes are categorized as grade 1 (mild), grade 2 (moderate), grade 3 (severe) or grade 4 (debilitating or life-threatening) [5]. Terms were further refined using organ-specific definitions adapted from grading systems [2, 5], and these definitions of the WHO grade were used to develop an algorithm to calculate bleeding grades (see Appendix S1). The algorithm used for this study was developed for the ongoing NHSBT sponsored UK/Australia TREATT trial (TRial to EvaluAte Tranexamic acid therapy in Thrombocytopenia; ISRCTN73545489); this TREATT trial is running alongside a similar study in USA (American Trial Using Tranexamic Acid in Thrombocytopenia A-TREAT; NCT02578901).

Data collection and Bleeding Assessment

A maximum of three consecutive days of repeated bleeding assessments was conducted once a participant's plate- let count was <30 9 10^9/l. Data collection stopped if there was a spontaneous increase in the participant's platelet counts from <30 9 10^9/l to >50 9 10^9/l; a participant was discharged home; or a participant withdrew their consent to continue in the study.

The bleeding assessment was performed at approximately the same time each day (to capture an approximate 24 h window of information), ideally in the morning, by two independent trained assessors. The assessment of patients included; comprehensive physical examinations, interviews and review of their medical and nursing notes. Participants were asked to report on any bleeding they had experienced since the previous assessment. If it was their first assessment they were asked to report on any bleeding they had experienced in the pre- ceding 24 h. After the first assessor had completed their bleeding assessment, the second assessor repeated the assessment within a maximum of a 2-h window, but usually immediately after the first assessor.

To enhance consistency in the assessment of bleeding between assessors, detailed training material and guide notes were provided; assessors received training from the lead researcher which took approximately 15 min to deliver; and body maps were provided for recording skin bleeding.

Bleeding assessors selected the relevant forms to complete once they had reviewed medical records, consulted with the study participant and conducted the physical examination.

Ease of use of the tool was assessed by asking bleeding assessors to complete a short postexamination qualitative survey on up to three separate occasions. The questions included in the survey were in relation to; the time taken to complete the assessment, any specific challenges regarding the assessment and general feedback regarding the tool and corresponding assessment.

Primary outcome and statistical considerations

The primary outcome measure used for the study was the concordance rates in WHO bleeding grade assigned and site(s) of bleeding identified between two independent assessors for participants recruited across all participating centres. For definition of assessment of site concordance, a strict interpretation was applied, with agreement on all questions and subquestions (and with no missing data).

Sample size

Assuming that the observed concordance proportion will be 80% and that assessments will be duplicated for each participant on at least two consecutive days, a sample size of 40 participants yielding 80 duplicated bleeding assessments will provide a 95% confidence interval for the true concordance proportion of (71, 89%), if there is no correlation between concordances for the same partici- pant on different days. If there is perfect correlation the 95% confidence interval would be (68, 92%). Based on a sample size of 40 participants yielding 80 duplicated assessments, the precision of the concordance estimate will be at worst plus or minus 16%.

Data analysis

Analysis for this study was largely descriptive. Summaries of participant characteristics, the number of bleeds reported and grade and site of bleeding have been presented as number, percentage. The percentage concordance and 95% confidence interval (CI) for grade were calculated using a

weighted kappa statistic for all duplicate assessments where grade was reported by two assessors. A weighted kappa was used to capture the extent of the discordance between assessors. Weights were calculated as per the Cicchetti-Allison method [6]. Site of bleeding concordance was defined as agreement on bleeding status for all body sites examined (refer to Table 1 for list of body sites examined), for all duplicate assessments where bleeding status for all body sites was reported by two assessors. Since it is not possible to calculate a single kappa statistic for measures combined in this way, the percentage concordance and 95% CI for site were calculated using the exact binomial distribution. It was assumed that there was no correlation between concordance on different days for the same participant (and this was reviewed in a sensitivity analysis). To check this assumption, concordance in grade/site for the first visit for each participant was compared to that for all visits and any differences were described. All statistical analyses were undertaken using computer software (SAS/STAT, version 9 of the SAS System for Windows; SAS Institute, Inc., Cary, NC, USA).

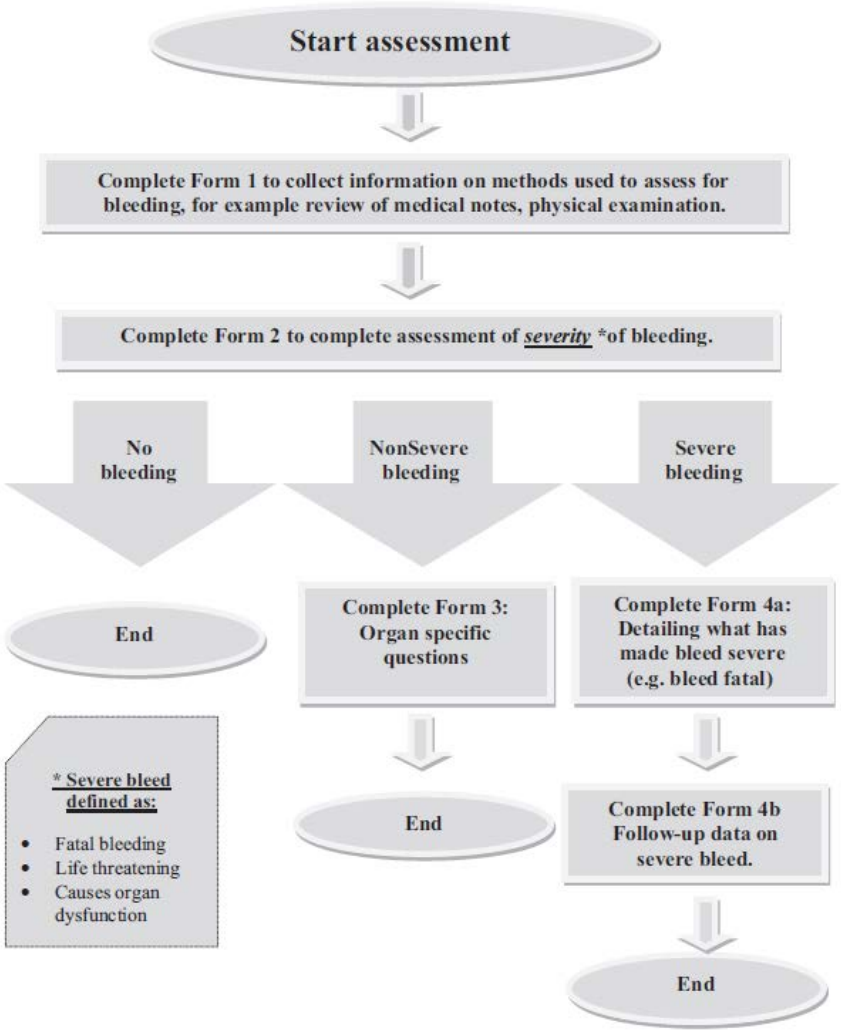

Fig. 1 Schematic diagram of the BAT process.

Ethical considerations

The study was undertaken according to the Declaration of Helsinki and Good Clinical Practice principles. The proto- col was approved for the UK participating centres by an external Research Ethics Committee (REC) – reference 16/ NW/0188. Ethical approval at the remaining international participating centres was sought and granted by the relevant ethics review bodies. All patients provided written informed consent. The study was conducted as specified by the protocol. As requested by the International Research Body (IRB) at the Sao Paulo Brazil centre, relevant study literature was submitted to them in both English and Portuguese.

Table 1 Sites and frequency of bleeding experienced by participants

Site of bleeding observed by assessors	Frequency of bleeding
Skin	102
Oropharyngeal	51
Epistaxis	49
Gastrointestinal	22
Blood in urine	18
Soft tissue and musculoskeletal	4
Haemoptysis	3
Invasive procedure	3
Abnormal vaginal bleeding	2
Retinal or vitreous	0
Body cavity/other	0
Subconjunctiva of the eye	0
Total	254

Note that participants may experience bleeding at more than one site at each assessment.

Table 2 Underlying diagnosis and treatment regimens of study participants

Diagnosis	Number of study participants $N = 40$ (%)
Underlying diagnosis	
Acute Myeloid Leukaemia	23 (58)
Myelodysplastic Syndrome	5 (13)
Acute Lymphoblastic Leukaemia	4 (10)
Multiple Myeloma	4 (10)
Myelofibrosis	2 (5)
Non-Hodgkin's Lymphoma	2 (5)
Treatment plan	
Haematopoietic stem cell transplant:	23 (58)
Allogeneic transplantation	16 (70)
Autologous transplantation	7 (30)
High dose chemotherapy:	13 (33)
Induction chemotherapy	4 (31)
Consolidation chemotherapy	5 (38)
Unknown	4 (31)
Other treatments	4 (10)
Lenalidomide	2 (50)
Decitabine	1 (25)
Carfilzomib	1 (25)

Results

Participant demographics and treatment regimens

Forty-five patients were invited to take part in the study. Five declined. Of the forty consented patients, 25 (62%) were men, and fifteen (38%) were women, their median age was 45. The majority of participants had acute leukaemia (27 of 40, 68%) and were undergoing a variety of treatments (Table 2).

Bleeding assessments

Thirteen independent bleeding assessors took part in the study. All assessors had medical or nurse training, and the majority had clinical haematology experience, which indicated that they were familiar with assessing bleeding in the patient group under review. A total of 225 bleeding assessments were carried out on 40 study participants over a period of 113 days.

Frequency and sites of bleeding observed

Of the 225 individual bleeding assessments conducted, 65 (29%) documented no bleeding; 151 (67%) reported nonsevere bleeding; and nine (4%) reported severe bleeding.

Bleeding (of any severity) occurred on 80 (71%) of the 113 days. On the other 33 days (29%), the bleeding assessors observed that the study participants had not experienced any bleeding

during that observational period. Severe bleeding occurred on 5 (4%) of the 113 days.

Bleeding grade

Assignment of WHO bleeding grade (0–4) was possible for 213 (95%) bleeding assessments. Across all centres, the rate of bleeding observed was as follows:

(1) Grade 0 was 50% (107 of 213)
(2) Grade 1 was 31% (65 of 213)
(3) Grade 2 was 15% (32 of 213)
(4) Grade 3 was 2% (five of 213)
(5) Grade 4 was 2% (four of 213)

Grading of 207 bleeding assessments was conducted using the algorithm. Eighteen assessments (8·7%) could not be graded due to missing data. Two independent investigators manually reviewed these 18 assessments to see if it was possible to derive a grade using information available from questions answered or details known about a participant. This manual review rendered a grade for six of these 18 assessments. Of all 213 graded assessments, 102 visits, or 204 assessments, for 40 patients were graded by two assessors and were included in the calculation of grade concordance; 80 visits, or 160 assessments, for 39 patients had bleeding status for all body sites reported by two assessors and were included in the calculation of site concordance (Our definition of site concordance relied on a complete bleeding assessment across all body sites, hence these numbers were smaller). Review of missing data helped to identify that certain questions related to presence of vaginal and subconjunctival bleeding as well as duration of mouth bleeding were more commonly left unanswered than others.

Primary outcome: concordance

Across all participating centres, the overall rate of concordance in bleeding grade was 83% (95% CI 74–93%), (Table 3). Manual review of the relevant severe or non- severe bleeding forms was conducted where there was discordance in the grade assigned. Reasons for discordance in grade assigned are detailed in Table 4. The most common area for disagreement between two assessors was the assessment of skin bleeding. If skin bleeding was omitted from the bleeding assessment tool, the overall rate of concordance in bleeding grade increased from 83 to 86% (95% CI 78–95%) with the greatest increase in concordance in grade 1 bleeds (Table 5).

Across all participating centres, the overall rate of concordance in site of bleeding identified by two independent assessors when reviewing both severe and nonsevere bleeding forms was 69% (95% CI 57–79%).

Concordance statistics for grade and site for the first visit were very similar to those for all visits (grade: 87%, 95% CI: 75–100% for first visit vs. 83%, 95% CI: 74–93% for all visits; site: 72%, 95% CI: 55–86% for first visit vs. 69%, 95% CI: 57–79% for all visits).

Ease of use of the BAT tool

In total, 46 surveys were completed by nine independent assessors; two from the UK, two from the USA and five from Brazil on different occasions. The mode time taken to complete a bleeding assessment was between 15 and 25 min. The following challenges were reported for 16 of the 46 (35%) survey responses:

(1) Difficulty in recording skin bleeding accurately.
(2) Uncertainty about specifying the duration of bleeding associated with oral blood blisters.
(3) Uncertainty about whether it was necessary to report bleeding associated with haemorrhoids.

No problems were reported with participant acceptance of the assessment, although two responses indicated challenges examining the full skin of the back. General comments and suggestions for further refinements focused largely on difficulties associated with reporting skin bleeding and the format of the forms. All responders commented on the usefulness of the body map, although several responders requested that the body map be adapted to incorporate scope for recording mouth bleeding, which is currently lacking. All nine of the responders to the survey reported that they found guide notes and training very helpful.

Table 3 Concordance in bleed grade between assessors. 102 visits, or 204 assessments, for 40 patients could be graded by two assessors. Assessors shown in no particular order. Results shown as number and percentage, by grade

	Assessor 2				
Bleed grade	**0**	**1**	**2**	**3**	**4**
Assessor 1					
0	48 (91)	4 (8)	1 (2)	0 (0)	0 (0)
1	2 (6)	26 (84)	3 (10)	0 (0)	0 (0)
2	2 (14)	0 (0)	11 (79)	1 (7)	0 (0)
3	0 (0)	0 (0)	0 (0)	2 (100)	0 (0)
4	0 (0)	0 (0)	0 (0)	0 (0)	2 (100)

Discussion

Assessment of bleeding is important for both clinical practice and research. However, reported rates of bleeding vary considerably between published trials of platelet transfusion in patients with haematological malignancies, in major part reflecting different definitions and methods of data collection for bleeding. This study describes the findings from a study to assess reliability of a consensus bleeding assessment tool, developed and adapted from a synthesis of tools used in previous platelet transfusion trials [2]. Key findings of this study are a very good level of concordance of bleeding grade (83%) between independent assessors, and recognition that assessment of skin bleeding was a source of disagreement between assessors. Our study also highlighted the central role of education, documentation and training materials to support the application of a BAT across multiple centres. Information from the qualitative survey supported the initial work of the content validation exercise and no assessor reported being unable to capture a specific bleeding event, suggesting that the BAT is comprehensive in scope.

The very good level of agreement between independent bleeding assessors indicates that the consensus BAT pro- vides a standardized and robust method to assess bleeding. Whilst agreement between assessors on bleeding grade was high, there were discordant assessments,

which included disagreement in the assessment of skin bleeding. Several considerations need to be recognized by researchers when considering the need to collect data on all types of skin bleeding. Skin bleeding remains the most frequently observed type of bleeding in our study, and in other larger studies conducted in similar patient populations [7–9]. However, skin bleeding remains the most challenging type of bleeding to document. If skin bleeding was omitted from the bleeding assessment tool, then the overall rate of concordance in bleeding grade increased from 83 to 86%. There is debate about the significance of minor bleeding for patients, clinicians and researchers, though a systematic review has highlighted the paucity of direct primary research exploring patient vs. clinician perceptions of bleeding [9]. In addition, some types of minor bleeding may be anticipated irrespective of level of platelet count, for example, traumatic nose bleeding and bleeding with insertion of peripheral cannulas.

Table 4 Reasons for discordance in grade assigned across all centres

	Assessor 1		Assessor 2	
Site (s) of bleeding that led to discordance	**Grade**	**Comment**	**Grade**	**Comment**
Skin	1	Skin bleeding 1–25%	2	Skin bleeding >25%
Skin	0	No bleeding	1	New skin bleeding
Vaginal	0	No bleeding	1	Vaginal bleeding
Nose	0	Trauma-related nose bleed	1	Spontaneous nose bleed
Skin	1	Skin bleeding	0	No bleeding
n/a	2	No treatment for bleeding	3	Treatment for bleeding
Skin	0	No bleeding	1	New skin bleeding
Mouth skin	2	Mouth and skin bleeding	0	No bleeding
Mouth	2	Mouth bleeding	0	No bleeding
GI	1	Severity of bleeding	2	Assessor 1 reported GI bleeding Assessor 2 reported more extensive GI bleeding
GI	1	Mouth bleeding. No GI bleeding	2	Mouth & GI bleeding
GI	0	No bleeding	2	GI bleeding
Skin	1	New skin bleeding	0	No bleeding

Table 5 Concordance in bleeding grade between assessors if skin bleeding is omitted from the bleeding assessment tool. One additional visit could be graded by two assessors if skin bleeding omitted, therefore 103 visits, or 206 assessments, for 40 patients could be graded by two assessors. Assessors shown in no particular order. Results shown as number and percentage, by grade

	Assessor 2				
Bleed grade	**0**	**1**	**2**	**3**	**4**
Assessor 1					
0	47 (92)	3 (6)	1 (2)	0 (0)	0 (0)
1	0 (0)	30 (94)	2 (6)	0 (0)	0 (0)
2	2 (13)	1 (6)	12 (75)	1 (6)	0 (0)
3	0 (0)	0 (0)	0 (0)	2 (100)	0 (0)
4	0 (0)	0 (0)	0 (0)	0 (0)	2 (100)

A follow-up analysis of the TOPPS trial data (a randomized controlled trial that compared prophylactic platelet transfusion with a strategy of no prophylaxis in adults with haematological malignancies) reported no evidence that WHO grade 1 bleeding is a predictor for more serious

bleeding. This study included sensitivity analyses inclusive of skin bleeding [9]. However, an earlier study, based on a much older trial, [10] reported that WHO grade 1 bleeding (including all skin bleeding) on the previous day was associated with an increased risk of clinically significant bleeding (grades 2, 3 and 4). The differing results of these two studies might be due to the different methodologies and patient populations. Stan- worth *et al.* [10] included mainly autologous stem cell transplant patients, whereas Webert *et al.* [11] included patients with acute myeloid leukaemia receiving induction chemotherapy.

The concordance in WHO grade of bleeding was higher than concordance in site of bleeding (83 vs. 69%, respectively). It should be acknowledged that a strict interpretation was applied, requiring agreement on all questions and subquestions and with no missing data. It could be suggested that concordance in grade of bleeding is of greater importance than the site(s) of bleeding observed in clinical trials; given the former is used to interpret the clinical significance of bleeding data. It was noted that the majority of cases of discrepancies in site or grade occurred at participating centres (non-UK) with no previous experience of using the consensus BAT and who con- ducted fewer assessments for this study (data not shown). The full reasons for variability in bleeding site are not clear but may highlight the importance of experience and ongoing education. Our consensus BAT is able to produce data that can be translated into a clinically meaningful grade by WHO. We recognize the challenge of the small number of bleeding assessments (8·7%) for which it was not possible to assign a bleeding grade because of missing data. Further studies may help inform whether an electronic BAT could further reduce missing data queries.

This study has helped to identify a number of potentially important additional factors in terms of standardizing the assessment of bleeding in adults with haematological malignancies and specifically the role of assessor education. Training and documentation used in this study included a provision of detailed guide notes, a PowerPoint presentation to support the use of the BAT, and a body map for recording skin bleeding. The bleeding assessors in this study were all motivated and had received recent training, which is likely to have impacted the high quality of completed bleeding assessments. How- ever, our qualitative survey also confirmed an under- standable desire to simplify the bleeding assessment tool and explore alternative platforms such as an electronic BAT, alongside considerations of the desirability of comprehensiveness of all types of bleeding.

One strength of this study was that it involved multiple international participating centres, some with and without prior experience of bleeding assessment projects, and therefore, the findings are likely to be generalizable to centres enrolling into multicentred trials. Rates of bleeding observed were broadly similar to some studies using bleeding as a primary outcome measure, although higher rates have been recorded in other trials, which could reflect methodology and the sample size of our study and which we also started the onset of bleeding assessment at platelet counts below 30 9 109/l [7, 8]. The use of our tool does also support data collection and scoring for assessment through a previously published Bleeding Severity Measurement Scale (BSMS) [4], for example to include recording the assessment of need for an intervention in response to clinical bleeding, although there is less published experience with this BSMS tool. Through rigorous steps of development, the authors of this BSMS defined two grades of bleeding and also reported excel- lent inter-rater [intraclass correlation coefficient (ICC), 0·80] and intrarater (ICC, 1·0) reliability and good construct and criterion validity. In this study, we used our tool to assign a bleeding grade as defined by the WHO classification, but as the consensus BAT can be used to grade bleeding according to the BSMS, it could enable direct comparison between the two bleeding classifications in the future. Our study does have

limitations: it was relatively small and specifically in reference to the number of bleeding assessors that participated and the number of major bleeding events.

Implications and conclusions

The results of our study have several implications. It is unlikely that investigators will achieve 100% concordance in assessing all grades of bleeding in clinical studies, given our findings using trained and well-supported bleeding assessors. Resources in clinical trials should cover adequate training and educational materials for bleeding assessors to help standardize the assessment of bleeding, and these educational methods should be reported in trial publications. Greater uncertainties apply to the standardized recording of all forms of skin bleeding. The importance of assessing all types of skin bleeding remains unclear, and further research should establish whether it is a precursor to a more severe bleed, alongside studies to explore patients' perception of this type of bleeding and its effect on quality of life. Better assessment of bleeding remains key to many issues surrounding the use of platelets, including studies of the comparative efficacy of pathogen inactivated platelets compared to standard platelets [12]. Whilst it was not within the scope of this study to validate the system used to grade bleeding, including by BSMS, this validation exercise of the consensus BAT may provide clinicians and researchers with a tool to help reduce one source of bias in clinical studies that of outcome assessment. In addition, this study contributes to the wider debate regarding methods of assessment and documentation of all bleeding types, including all types of skin bleeding.

Acknowledgements

This work contributed to the Master's degree for C Dyer; We acknowledge helpful feedback on the draft from Samaher Sweity and Gillian Powter and Amber Degelia for additional assistance with the conduct of bleeding assessments in Oxford. We acknowledge the NHSBT CTU for supporting the study. We also acknowledge Nancy Heddle and Jill Corson who contributed to the content validation stage.

References

1 Desborough MJ, Smethurst PA, Estcourt LJ, et al.: Alternatives to allogeneic platelet transfusion. Br J Haematol 2016; 175: 381– 392

2 Estcourt LJ, Heddle NM, Kaufman RM, et al.: The challenges of measuring bleeding outcomes in clinical trials of platelet transfusions. Transfusion 2013; 53: 1531– 1543

3 Crighton GL, Estcourt LJ, Wood EM, et al.: A therapeutic-only versus prophylactic platelet transfusion strategy for preventing bleeding in patients with haematological disorders after myelosuppressive chemotherapy or stem cell transplantation. Cochrane Database of Systematic Reviews 2015; 30: CD010981

4 Webert KE, Arnold DM, Lui Y, et al.: A new tool to assess bleeding severity in patients with chemotherapy-induced thrombocytopenia. Transfusion 2012; 52: 2466– 2474

5 World Health Organisation (WHO). WHO Handbook for Reporting Results of Cancer Treatment. Geneva, WHO, 1979.

6 Fleiss JL: Statistical Methods for Rates and Proportions, 2nd edn. New York, Wiley, 1981

7 Stanworth S, Estcourt L, Powter G, et al.: A no-prophylaxis platelet-transfusion strategy for hematologic cancers. The New England Journal of Medicine 2013; 368: 1771– 1780

8 Slichter SJ, Kaufman RM, Assmann SF, et al.: Dose of prophylactic platelet transfusions and prevention of haemorrhage. The New England Journal of Medicine 2010; 362: 600– 613

9 Estcourt LJ, Pinchon D, Symmington E, et al.: Does bleeding affect patient reported outcome measures in patients with myelodysplasia or hematologic malignancies: a systematic review. Transfusion 2014; 54: 1166– 1179

10 Stanworth SJ, Hudson CL, Estcourt LJ, et al.: Risk of bleeding and use of platelet transfusions in patients with hematological malignancies: recurrent event analysis. Haematologica 2015; 100: 740– 747

11 Webert K, Cook RJ, Sigouin CS, et al.: The risk of bleeding in thrombocytopenic patients with acute myeloid leukaemia. Haematologica 2006; 91: 1530– 1537

12 Butler C, Doree C, Estcourt LJ, et al.: Pathogen-reduced platelets for the prevention of bleeding. Cochrane Database Syst Rev 2013; 28: CD009072