TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	Development and validation of a discriminatory model for hematotoxicity in r/r LBCL patients
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	**Page 1** Objectives: study patterns of hematopoietic reconstitution, evaluate potential biomarkers, develop discriminatory model for hematotoxicity Study Design: retrospective analysis, 1 training cohort (n=58), 2 external validation cohorts (Europe, n=91 and USA, n=109) Setting: multicenter retrospective analysis Sample size: 258 patients Predictors: markers of hematopoietic reserve (e.g. PLT, Hb, ANC) and baseline inflammation (e.g. CRP, ferritin) Outcome: duration of severe neutropenia (days 0-60) and binary outcome severe neutropenia ≥14/<14 days Statistical analysis: discriminative model development Results: Pooled validation (AUC 0.89, Sensitivity 89%, Specificity 68%), high risk patients observe higher incidence of cytopenia and longer median duration of neutropenia (12 vs. 5.5 days). Conclusions: outline of clinical applications.
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	**Page 2-3** - High incidence of hematotoxicity → predisposes for infectious complications. - Description of previously established risk factors - no existing models for CAR T-cell assoc. hematotox.
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	**Page 3** Aim 1: characterize the influence of clinical/lab features on neutropenia Aim 2: develop a clinical score that discriminates between high and low risk for hematotoxicity
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	**Page 3:** Retrospective observational study. Training cohort (n=58) and two independent validation cohorts (Europe, n=91 and USA, n=109).
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	See **Results section, Page 5**
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	**Page 3:** multicenter study, s. **Fig. 1A**
	5b	D;V	Describe eligibility criteria for participants.	See **Results section, Page 5** and **Fig. 1A**
	5c	D;V	Give details of treatments received, if relevant.	See **Table 1** (baseline patient characteristics
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	**Page 4:** binary outcome → severe neutropenia ≥14 vs. <14 days (AUC analysis)
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	D	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	Time of measurement: prior to lymphodepletion (**page 3**) Description of all predictors: **Fig. S2** and **Fig. S4**, **Table S5**
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	D;V	Explain how the study size was arrived at.	Sample size was based on maximal availability in exploratory model development. Minimal requirement for validation data was the size of the training cohort.
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	The intervening days between outpatient follow-ups were counted to the total days of neutropenia only if both values were < 500/µl. Patients were excluded for early death or when receiving myelotoxic chemotherapy prior to day 30 (**page 3**). Temporal analysis: see **Table S3**
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	See **page 3**: predictors were handled as continuous variables
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	**Page 4:** Discriminative model for the binary outcome. Factors were identified by AUC analysis of the ROC curve (**Table S5**). Optimal discriminatory threshold identified via Youden J statistic. Three models were tested in the training cohort (**Tables S6-8, Fig. S5**). Model performance assessed by AUC_{ROC} and concomitant odds ratio calculation (**Table S9**).
	10c	V	For validation, describe how the predictions were calculated.	Internal validation: not done
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	See **Table S9**
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	Not done
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	See **Table S8**. Risk groups were determined by optimizing the respective Youden J statistic for the binary outcome.

Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	See **Table 1** See **Fig. S7:** lower mean CAR-HEMATOTOX score in the validation cohorts compared to the training cohort.
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	See **Fig. 1A** and **Table S3**
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	See **Table 1**
	13c	D;V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	See **Table 1** Outcome data was not obtained for the validation cohorts. Concomitant immunotoxicities are reported in **Table S2**
Model development	14a	D;V	Specify the number of participants and outcome events in each analysis.	See **Fig. 1A**, **Fig. 3** (D) and **Fig. 4** (V) See **Table 2:** incidence rates
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	See **Table S2** and **Table S4**
Model specification	15a	D;V	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	The Spearman correlation coefficient (r) is depicted in the graph inset for all univariate analyses. The calculated slope is reported for comparison of the CAR-HEMATOTOX with the duration of neutropenia. See linear regression analysis function, **Table S6H**
	15b	D	Explain how to use the prediction model.	Discriminate long and short neutropenia
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	AUC (see **Table S9**).
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	Not done
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	**Page 10:** limited sample size of training cohort. Potential overfitting bias. Retrospective nature of the analysis. Incomplete data concerning BM infiltration.
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	**Page 11**
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	**Page 9-11**
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	**Page 10-11:** risk-stratification of hematotoxicity, antimicrobial strategies, potentially autologous stem cell backup, anticipation of patient needs.
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	See **Supplemental Figures** and **Tables**
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	See **Acknowledgements** section

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

Supplementary Material

Supplemental Figures

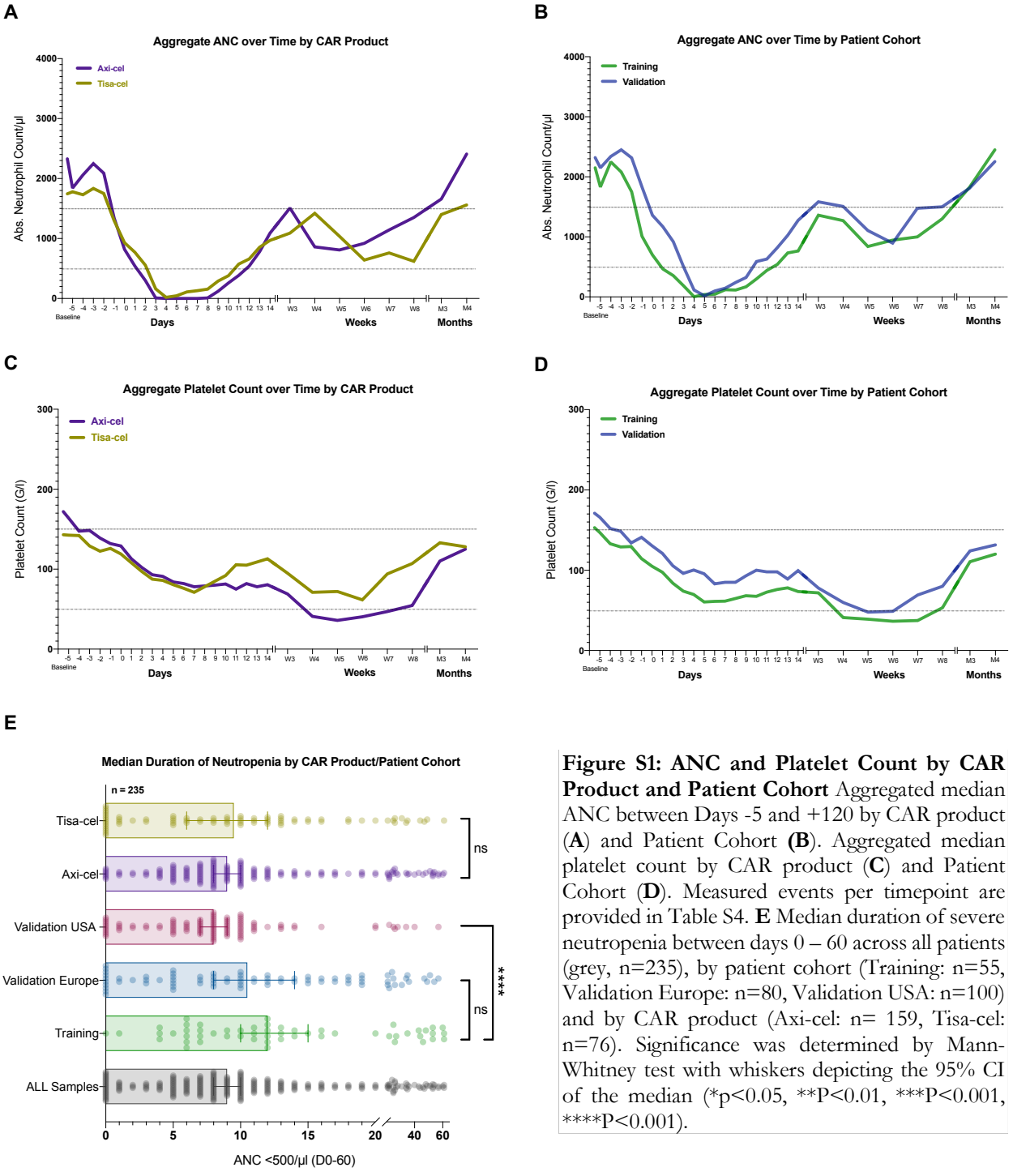


Figure S1: ANC and Platelet Count by CAR Product and Patient Cohort Aggregated median ANC between Days -5 and +120 by CAR product (**A**) and Patient Cohort **(B**). Aggregated median platelet count by CAR product (**C**) and Patient Cohort (**D**). Measured events per timepoint are provided in Table S4. **E** Median duration of severe neutropenia between days 0 – 60 across all patients (grey, n=235), by patient cohort (Training: n=55, Validation Europe: n=80, Validation USA: n=100) and by CAR product (Axi-cel: n= 159, Tisa-cel: n=76). Significance was determined by Mann-Whitney test with whiskers depicting the 95% CI of the median (*p<0.05, **P<0.01, ***P<0.001, ****P<0.001).

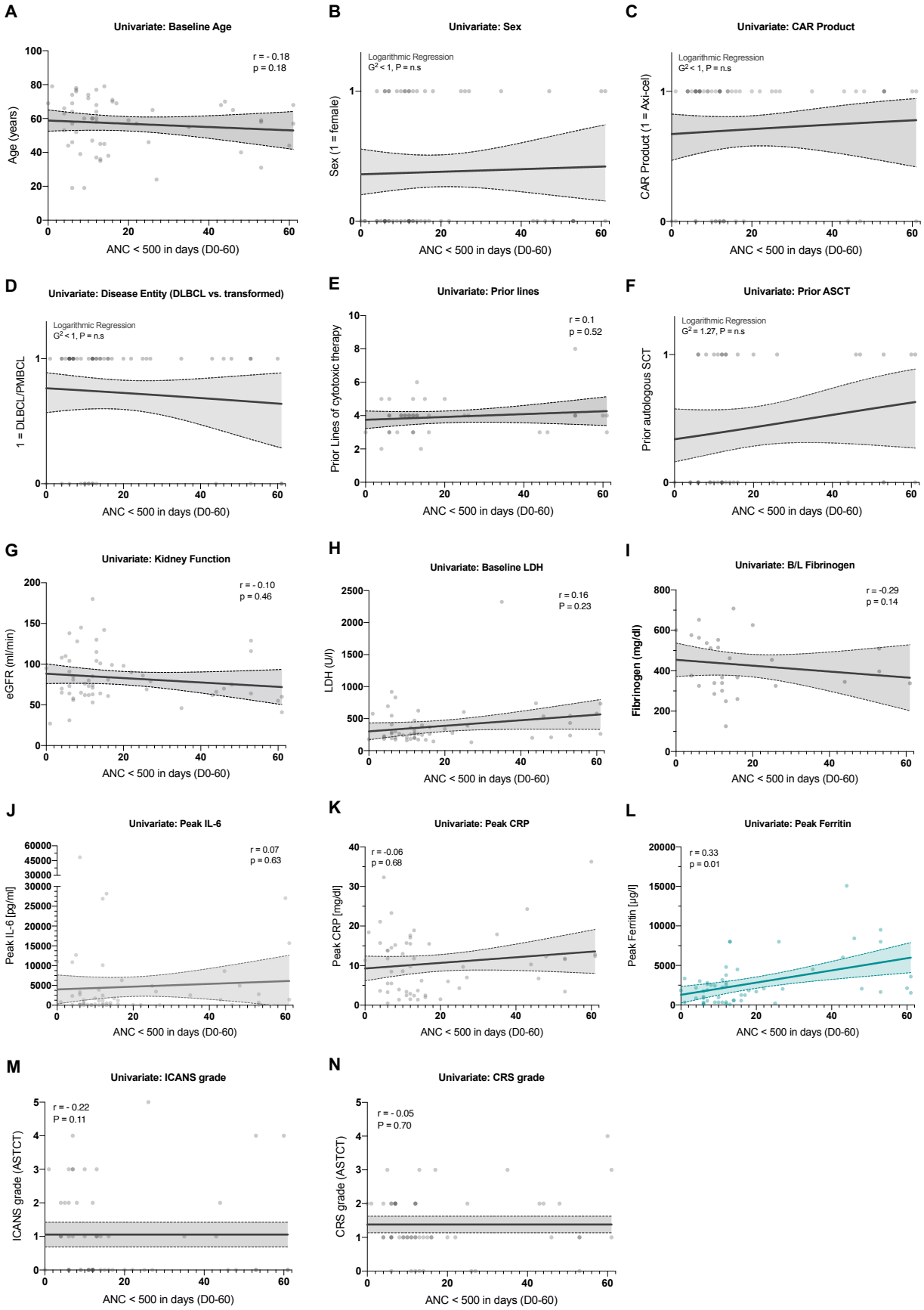

A
Univariate: Baseline Age
r = - 0.18
p = 0.18
Age (years)
ANC < 500 in days (D0-60)
B
Univariate: Sex
Logarithmic Regression
G² < 1, P = n.s
Sex (1 = female)
ANC < 500 in days (D0-60)
C
Univariate: CAR Product
Logarithmic Regression
G² < 1, P = n.s
CAR Product (1 = Axi-cel)
ANC < 500 in days (D0-60)
D
Univariate: Disease Entity (DLBCL vs. transformed)
Logarithmic Regression
G² < 1, P = n.s
1 = DLBCL/PMBCL
ANC < 500 in days (D0-60)
E
Univariate: Prior lines
r = 0.1
p = 0.52
Prior Lines of cytotoxic therapy
ANC < 500 in days (D0-60)
F
Univariate: Prior ASCT
Logarithmic Regression
G² = 1.27, P = n.s
Prior autologous SCT
ANC < 500 in days (D0-60)
G
Univariate: Kidney Function
r = - 0.10
p = 0.46
eGFR (ml/min)
ANC < 500 in days (D0-60)
H
Univariate: Baseline LDH
r = 0.16
P = 0.23
LDH (U/l)
ANC < 500 in days (D0-60)
I
Univariate: B/L Fibrinogen
r = -0.29
p = 0.14
Fibrinogen (mg/dl)
ANC < 500 in days (D0-60)
J
Univariate: Peak IL-6
r = 0.07
p = 0.63
Peak IL-6 [pg/ml]
ANC < 500 in days (D0-60)
K
Univariate: Peak CRP
r = -0.06
p = 0.68
Peak CRP [mg/dl]
ANC < 500 in days (D0-60)
L
Univariate: Peak Ferritin
r = 0.33
p = 0.01
Peak Ferritin [µg/l]
ANC < 500 in days (D0-60)
M
Univariate: ICANS grade
r = - 0.22
P = 0.11
ICANS grade (ASTCT)
ANC < 500 in days (D0-60)
N
Univariate: CRS grade
r = - 0.05
P = 0.70
CRS grade (ASTCT)
ANC < 500 in days (D0-60)

Figure S2: Univariate analysis of baseline patient features, peak cytokine levels, and CRS/ICANS grade
No significant correlation was observed between the duration of neutropenia (ANC < 500/µl, days 0-60= and the clinical parameters age, sex, CAR product, disease entity, number of prior therapy lines, prior autologous stem cell transplantation, kidney function, baseline LDH, baseline fibrinogen, peak IL-6, peak CRP, and maximal CRS or ICANS grade. While a significant positive correlation was observed between peak ferritin (**L**) and a prolonged duration of neutropenia, the overall correlation was decreased compared to ferritin prior to lymphodepletion (**Fig. 2E**). For simple linear regression, the graphical inset depicts: r = Spearman correlation coefficient; and p = two-tailed p-value. The 95% confidence bands of the best-fit line are shown in light shading. For logarithmic regression (non-continuous variables), the p-value is depicted for the likelihood test (G-squared) with light shading indicating the 95% asymptotic confidence bands.

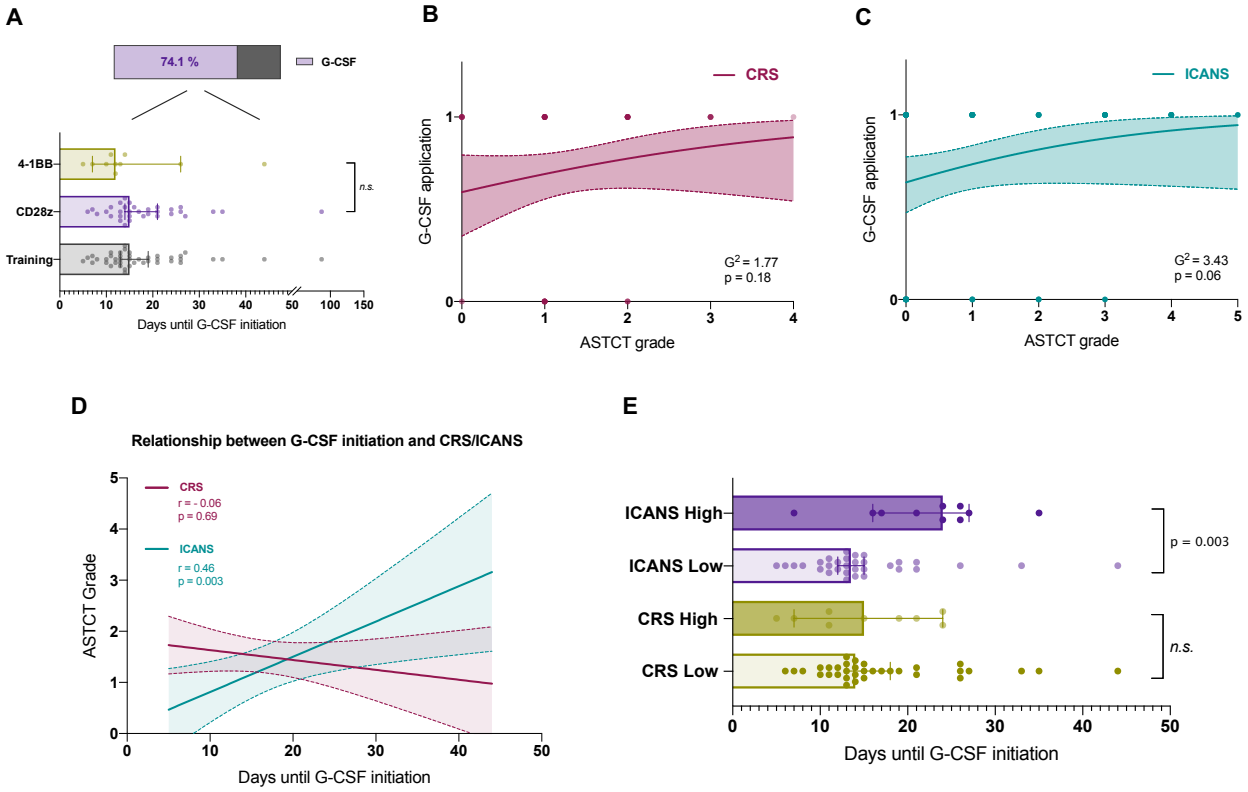


Figure S3: G-CSF was applied in approximately three out of four patients and initiated after a median of 15 days. Severe neurotoxicity was associated with a late G-CSF initiation.
A Top: Percentage of patients receiving G-CSF stimulation in the training cohort. Bottom: Graph displaying the median day on which G-CSF was initiated in relation to CAR T-cell transfusion. Error bars depict the 95% confidence interval. **B/C** Simple logistic regression analysis comparing G-CSF application to CRS (**B**) and ICANS (**C**) ASTCT grade; p-values according to Likelihood ratio test are depicted on the graph inset. **D** Simple linear regression analysis comparing the CRS/ICANS grade to the timepoint of G-CSF initiation in the training cohort. **E** Comparison of the mean time to G-CSF initiation in patients with severe CRS/ICANS (High = grade ≥ III) and an absent to moderate CRS/ICANS grade (Low = grade ≤ 2). Error bars depict the Mean +/- SEM with p-values being assessed via Mann-Whitney test.

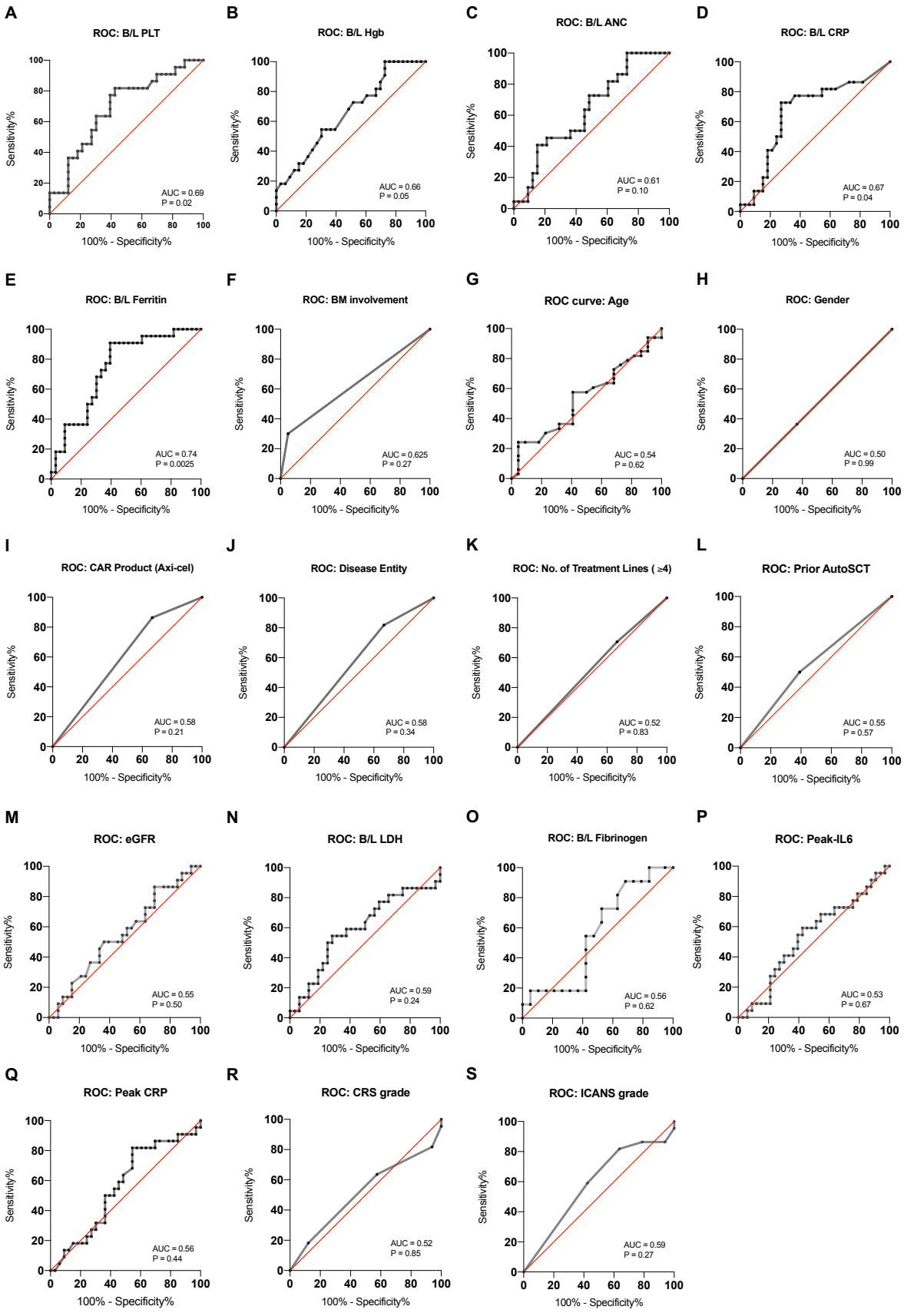
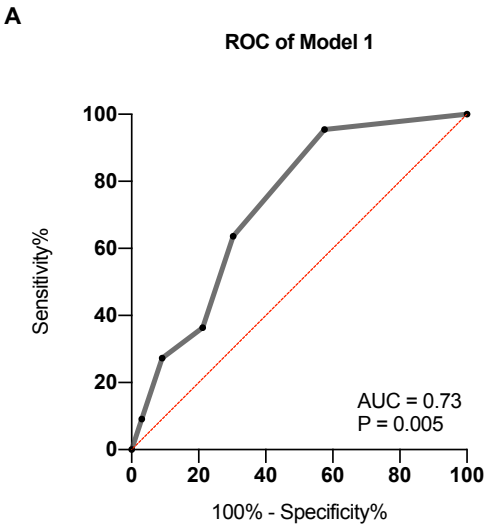
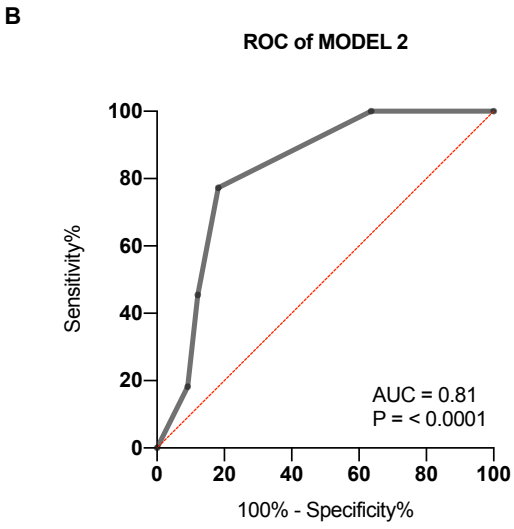

A
ROC: B/L PLT
AUC = 0.69
P = 0.02
B
ROC: B/L Hgb
AUC = 0.66
P = 0.05
C
ROC: B/L ANC
AUC = 0.61
P = 0.10
D
ROC: B/L CRP
AUC = 0.67
P = 0.04
E
ROC: B/L Ferritin
AUC = 0.74
P = 0.0025
F
ROC: BM involvement
AUC = 0.625
P = 0.27
G
ROC curve: Age
AUC = 0.54
P = 0.62
H
ROC: Gender
AUC = 0.50
P = 0.99
I
ROC: CAR Product (Axi-cel)
AUC = 0.58
P = 0.21
J
ROC: Disease Entity
AUC = 0.58
P = 0.34
K
ROC: No. of Treatment Lines (≥4)
AUC = 0.52
P = 0.83
L
ROC: Prior AutoSCT
AUC = 0.55
P = 0.57
M
ROC: eGFR
AUC = 0.55
P = 0.50
N
ROC: B/L LDH
AUC = 0.59
P = 0.24
O
ROC: B/L Fibrinogen
AUC = 0.56
P = 0.62
P
ROC: Peak-IL6
AUC = 0.53
P = 0.67
Q
ROC: Peak CRP
AUC = 0.56
P = 0.44
R
ROC: CRS grade
AUC = 0.52
P = 0.85
S
ROC: ICANS grade
AUC = 0.59
P = 0.27
Sensitivity%
100% - Specificity%

Figure S4: Receiver Operating Characteristic (ROC) Curves analyzing the influence of baseline features on the duration of neutropenia

ROC was performed for the baseline features platelet count (**A**), Hemoglobin (**B**), Absolute Neutrophil Count (**C**), C-reactive protein (**D**), Ferritin (**E**), and BM infiltration (**F**), Age (**G**), Gender (**H**), CAR Product (**I**), Disease Entity (**J**), No. of treatment lines (**K**), prior autologous stem cell transplant (**L**), eGFR by CKD-EPI (**M**), LDH (**N**) and fibrinogen (**O**), comparing patients with a duration of neutropenia ≥ 14 days (n = 22) to patients with a duration of neutropenia < 14 days (n = 33). The dynamic factors peak IL-6 (**P**), peak CRP (**Q**), maximal CRS (**R**) and ICANS (**S**) grade using ASTCT grading were studied in the same manner. The calculated area under the ROC curve (AUC) and concurrent p-value are depicted on the graph inset.





Baseline Features	0 Point	1 Point
Platelet Count	< 100.000/µl	> 100.000/µl
Absolute Neutrophil Count (ANC)	> 1200/µl	< 1200/µl
Hemoglobin	> 9.0 g/dl	< 9.0 g/dl
C-reactive protein (CRP)	< 3.0 mg/dl	> 3.0 mg/dl
Ferritin	< 1200 ng/ml	> 1200 ng/ml
Low: 0-1	High: ≥ 2	

Baseline Features	0 Point	1 Point	2 Points
Platelet Count	< 175.000/µl	75.000 – 175.000/µl	< 75.000/µl
Ferritin	< 650 ng/ml	650 – 2000 ng/ml	> 2000 ng/ml
Low: 0-1	High: ≥ 2		

Figure S5: Model Comparison

A For model 1, discriminatory thresholds were determined on the basis of optimizing the respective Youden *J* statistic for each variable (see **Fig. S3**), requiring a specificity cutoff of at least 0.75. Simple weighting was applied (either 0 or 1 point for each factor). **B** Model 2 only includes the factors platelet count and ferritin, which were identified by multivariate analysis with a stepwise elimination (**Table S7**). Instead of dichotomization, a categorization into three groups was performed (0 vs. 1 vs. 2 points). The first point was given according to a sensitivity cutoff of 0.8 (ferritin >650 ng/ml, PLT <175 G/l), while the second point was applied according to a specificity cutoff of 0.85 (ferritin >2000 ng/ml, PLT<75 G/l).

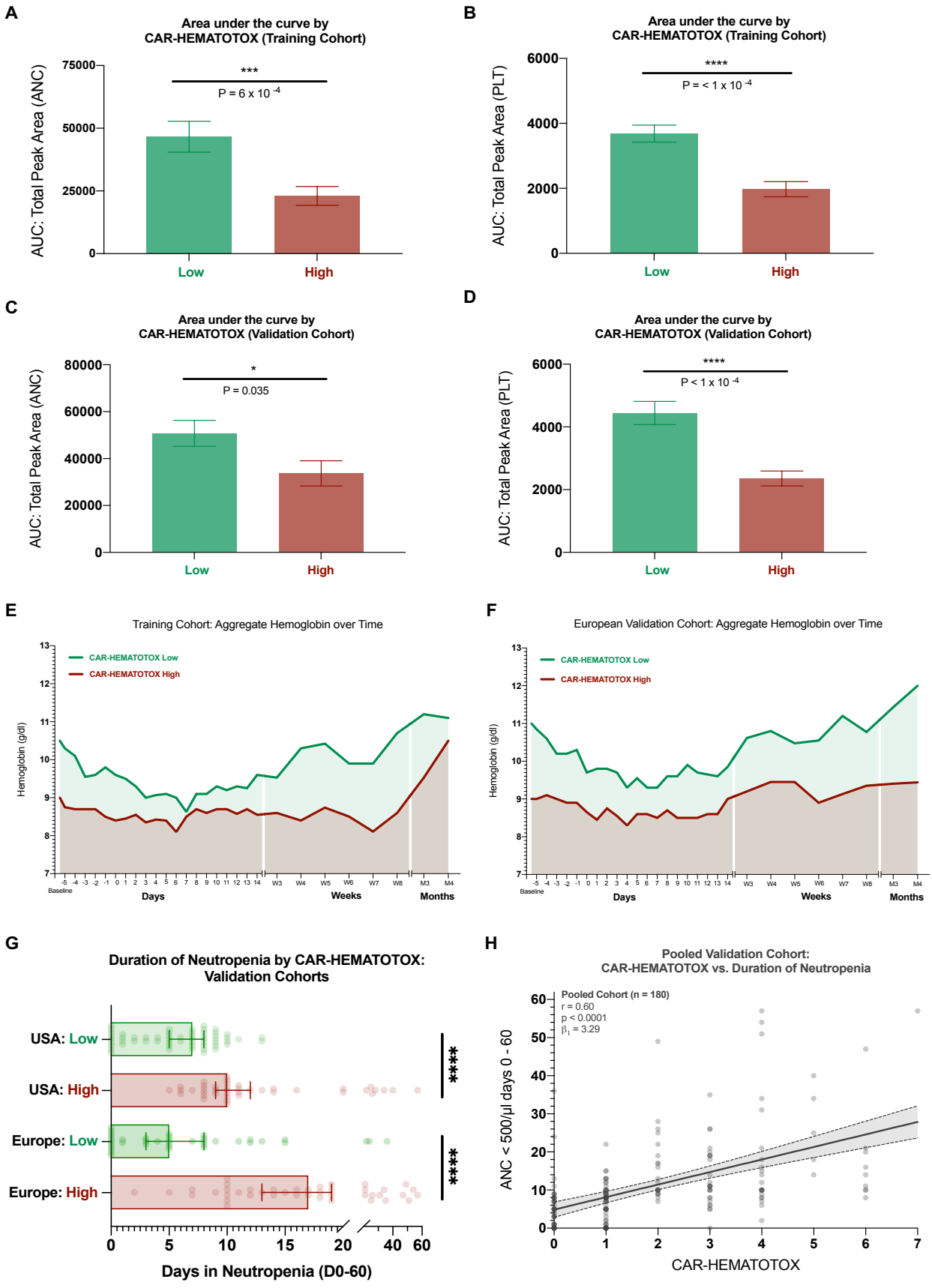
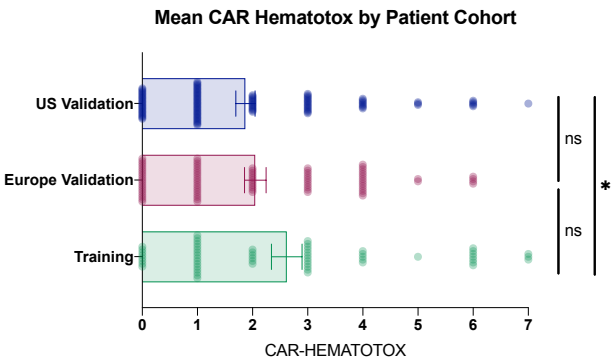
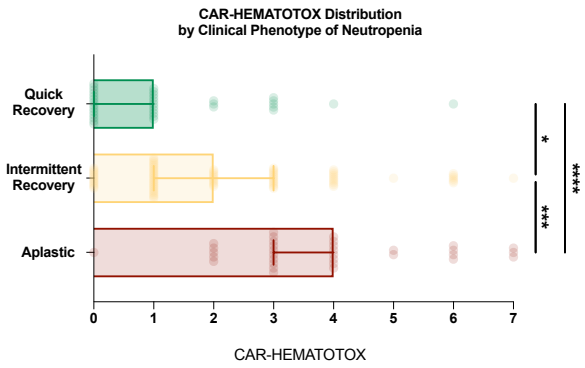

A
Area under the curve by CAR-HEMATOTOX (Training Cohort)
AUC: Total Peak Area (ANC)
***
P = 6 x 10 -4
Low
High
B
Area under the curve by CAR-HEMATOTOX (Training Cohort)
AUC: Total Peak Area (PLT)
****
P = < 1 x 10 -4
C
Area under the curve by CAR-HEMATOTOX (Validation Cohort)
*
P = 0.035
D
Area under the curve by CAR-HEMATOTOX (Validation Cohort)
P < 1 x 10 -4
E
Training Cohort: Aggregate Hemoglobin over Time
CAR-HEMATOTOX Low
CAR-HEMATOTOX High
Hemoglobin (g/dl)
Baseline
Days
Weeks
Months
F
European Validation Cohort: Aggregate Hemoglobin over Time
G
Duration of Neutropenia by CAR-HEMATOTOX: Validation Cohorts
USA: Low
USA: High
Europe: Low
Europe: High
Days in Neutropenia (D0-60)
H
Pooled Validation Cohort: CAR-HEMATOTOX vs. Duration of Neutropenia
Pooled Cohort (n = 180)
r = 0.60
p < 0.0001
β1 = 3.29
ANC < 500/µl days 0 - 60
CAR-HEMATOTOX

Figure S6: Hematotoxicity over time by CAR-HEMATOTOX
Corresponding AUC analysis (total peak area) of the aggregated median ANC curves (**A, C**) and platelet count curves (**B, D**) for the training cohort (n = 55) and European validation cohort (n = 80) outlined in **Figures 3-4**. Statistical significance was determined by unpaired t test. **E-F** Aggregated median hemoglobin over time for the training cohort (**E**) and European validation cohort (**F**). Measured events per timepoint are provided in **Table S4**. **G** Median duration of severe neutropenia (Days 0-60) by CAR-HEMATOTOX (high: red, green: low risk) for both validation cohorts respectively (USA: top, Europe: bottom) **H** Univariate analysis comparing the CAR-HEMATOTOX to the duration of severe neutropenia in the pooled validation cohort (n=180). The calculated slope (β_1) of the simple linear regression is depicted, corresponding to an average increase in the duration of neutropenia of 3.29 days for every score increase of one. The calculated function of the linear regression analysis was determined as: $y_{pooled} = 4.865\ (\beta_0) + 3.29\ (\beta_1) * x$. β_0 = y-intercept, β_1 = slope.

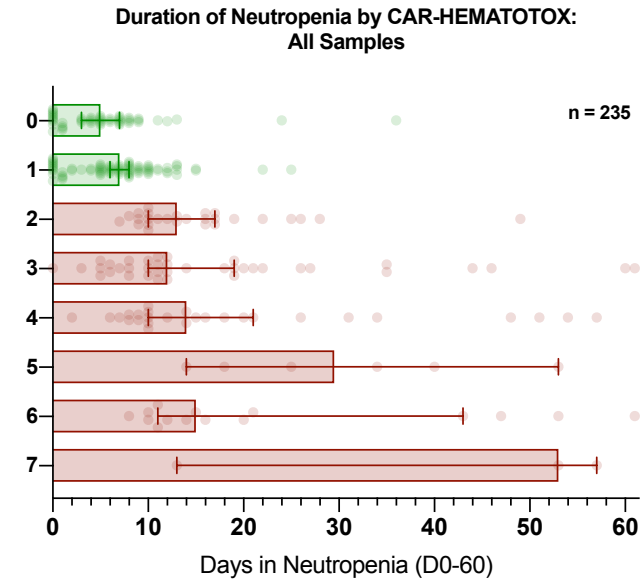
A



B



C



	n	Median	95% CI (Days)
0	48	5	3 – 7
1	72	7	6 – 8
2	26	13	10 – 17
3	38	12	10 – 19
4	27	14	10 – 21
5	6	29.5	14 – 53
6	15	15	11 – 43
7	3	53	13 – 57

D

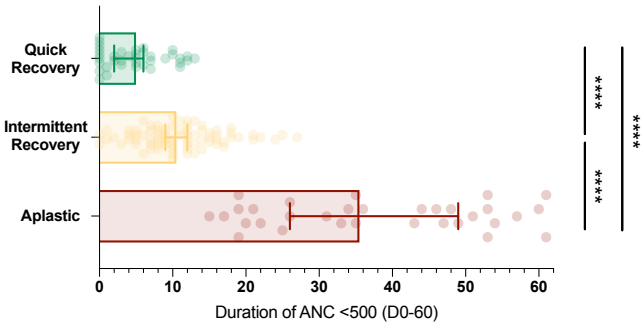


Figure S7: Characterization of neutrophil recovery and median duration of neutropenia by individual CAR-HEMATOTOX values

A Comparison of the mean CAR-HEMATOTOX by patient cohort. Bars indicate the mean CAR-HEMATOTOX with error bars depicting the SEM. Significance values were determined by Mann-Whitney test (*P<0.05, **P<0.01, ***P<0.001, ****P<0.0001). **B** Distribution of the CAR-HEMATOTOX by clinical phenotype of neutrophil recovery across all European patients (n = 149). Longitudinal CBC sampling was not obtained for the US validation cohort. Bars indicate the median CAR-HEMATOTOX within each phenotype group with whiskers depicting the 95% CI of the median. "Aplastic" defined as continuous severe neutropenia ≥ 14 days (n = 34). "Intermittent Recovery" defined as neutrophil recovery (ANC > 1500/μl) followed by a second dip with an ANC <1000/μl after day 21 (n = 78). "Quick Recovery" defined as sustained neutrophil recovery without a second dip below an ANC <1000/μl (n = 37). Significance values were determined by Mann-Whitney test (*P<0.05, **P<0.01, ***P <0.001, ****P<0.0001). **C** Analysis of the median duration of severe neutropenia (Days 0-60) for each individual CAR-HEMATOTOX score across all studied patients (n = 235). The table to the right of the graph shows the absolute number of patients (n), the median, and the 95% CI of the median for each individual CAR- HEMATOTOX score. **D** Median duration of severe neutropenia (Days 0-60) by clinical phenotype of neutrophil recovery. Significance values were determined by Mann-Whitney test (*P<0.05, **P<0.01, ***P <0.001, ****P<0.0001).

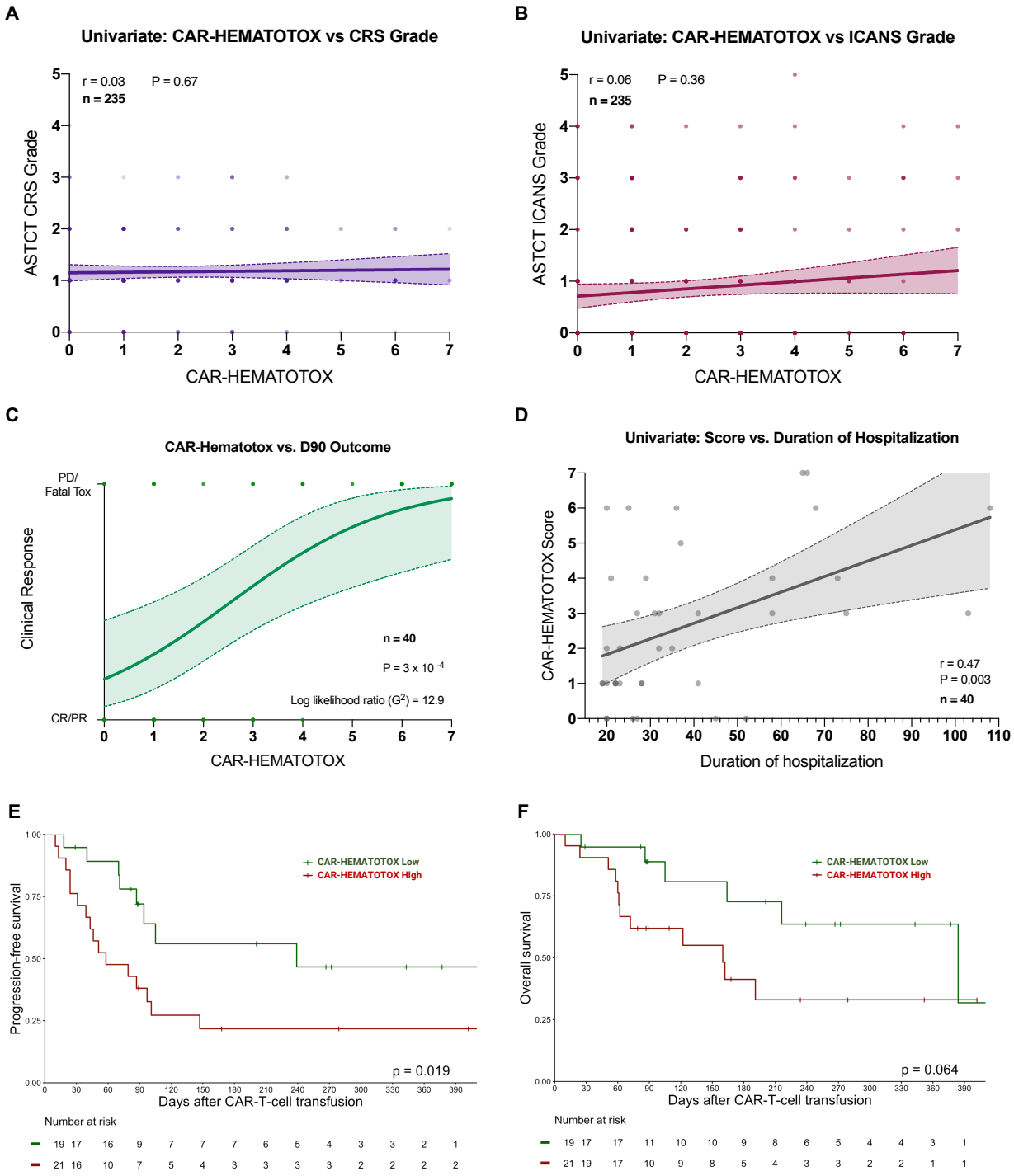


Figure S8: Univariate Analysis of CAR-HEMATOTOX vs. CRS/ICANS and patient-related outcomes
A CAR-HEMATOTOX vs. CRS Grade (ASTCT) **B** CAR-HEMATOTOX vs. ICANS Grade (ASTCT) **C** CAR-HEMATOTOX vs. clinical outcome after three months. Clinical outcome was defined according to Lugano criteria. Patients with progressive disease or who died secondary to toxicity are found in the "PD/Fatal Tox" category. **D** CAR-HEMATOTOX vs. duration of hospital stay. Hospital stay was defined from start of lymphodepletion (+2 days leniency) until first discharge. Patients who died prior to day 30 were censored from analysis. **E-F** Kaplan-Meier estimates of progression-free survival (**E**) and overall survival (**F**) according to CAR-HEMATOTOX when applying a CAR-HEMATOTOX threshold of ≥3 versus <3.

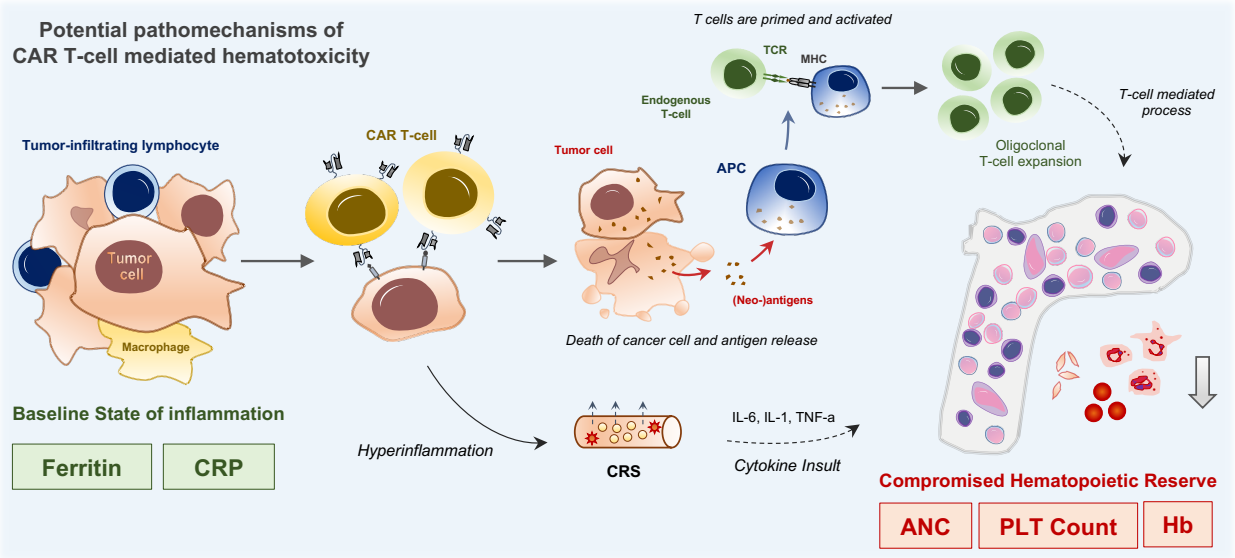


Figure S9: Potential pathomechanisms of CAR T-cell mediated hematotoxicity

Supplementary Tables

	Tisa-cel (n = 88)	Axi-cel (n = 170)	*p*
Age, years	64 (28 – 83)	64 (19 – 79)	
Median lines of prior therapy (range)	3 (2 – 11)	3 (2 – 9)	
Prior autologous stem cell transplantion	32 (36%)	37 (22%)	0.02
Disease Entity			
Diffuse Large B-Cell Lymphoma (DLBCL)	60 (68%)	116 (68%)	n.s.
Primary Mediastinal B-Cell Lymphoma (PMBCL)	2 (2%)	13 (8%)	n.s.
Transformed Lymphoma (trFL, trMCL, trMALT, trCLL, trHL)	26 (30%)	41 (24%)	n.s.
Complete Blood Count			
Median ANC, 95% CI (ANC/µl)	2265 (1770 – 2620)	2640 (2380 – 2950)	n.s.
Median Platelet Count, 95% CI (G/l)	153.5 (125 – 181)	169 (154 – 180)	n.s.
Median Hemoglobin, 95% CI (g/dl)	9.9 (9.5 – 10.3)	10.2 (9.6 – 11.0)	n.s.
Markers of Tumor Burden & Inflammation			
Median LDH, 95% CI (U/l)	291.5 (269 – 351)	261 (245 – 298)	n.s.
Median CRP, 95% CI (mg/dl)	1.14 (0.71 – 1.73)	0.98 (0.60 – 1.44)	n.s.
Median Ferritin, 95% CI (ng/ml)	763 (443 –1062)	449 (340 – 572)	n.s.

Table S1: Baseline demographic and clinical characteristics of treated CAR T-cell patients
Statistical significance ($P < 0.05$) between groups (Tisa-cel vs. Axi-cel) was determined by Fisher's exact two-sided t test. Tisa-cel = tisagenlecleucel (harbors 4-1BB co-stimulatory domain). Axi-cel = axicabtagene ciloleucel (harbors CD28z co-stimulatory domain).

Dynamic Clinical Variables	All patients (n = 258)	Training (n = 58)	Validation Europe (n = 91)	Validation USA (n = 109)	Tisa-cel (n = 88)	Axi-cel (n = 170)	*p*	Europe (n = 149)	USA (n = 109)	*p*
CRS (ASTCT)										
Grade I-II	**195 (75.6%)**	43 (74.1%)	69 (75.8%)	83 (76.1%)	60 (68.2%)	135 (79.4%)	*0.066*	112 (75.2%)	83 (76.1%)	*n.s.*
Grade III-V	**28 (10.9%)**	7 (12.1%)	11 (12.1%)	10 (9.2%)	11 (12.5%)	17 (10.0%)	*n.s.*	18 (12.1%)	10 (9.2%)	*n.s.*
ICANS (ASTCT)										
Grade I-II	**78 (30.2%)**	17 (29.3%)	14 (19.8%)	47 (43.1%)	17 (19.3%)	61 (35.9%)	*0.007*	31 (20.8%)	47 (43.1%)	*< 0.001*
Grade III-V	**45 (17.4%)**	11 (18.7%)	7 (8.8%)	27 (24.8%)	7 (8.0%)	38 (22.4%)	*0.003*	18 (12.1%)	27 (24.8%)	*0.01*
Tocilizumab applied	**154 (59.7%)**	42 (72.4%)	57 (62.6%)	55 (50.5%)	56 (63.6%)	100 (58.8%)	*n.s.*	99 (66.4%)	55 (50.5%)	*0.01*
Dexamethason applied	**100 (38.8%)**	17 (29.3%)	34 (37.4%)	49 (45.0%)	22 (25.0%)	78 (45.9%)	*0.001*	51 (34.2%)	49 (45.0%)	*0.09*

Table S2: Incidence of concurrent CAR T-cell mediated immunotoxicities and description of toxicity management
CRS/ICANS grading was performed according to the ASTCT consensus guidelines (Lee et al). The application of the anti-IL6 receptor antagonist Tocilizumab or the corticosteroid Dexamethason is reported for all patients. Statistical significance ($P < 0.05$) between groups (Tisa-cel vs. Axi-cel and Europe vs. USA) was determined by Fisher's exact two-sided t test.

Timepoint	B/L	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	W3	W4	W5	W6	W7	W8	M3	M4
No. of available measurements	149	138	144	146	148	149	149	149	149	148	148	148	147	149	149	149	147	140	141	134	132	140	131	118	100	98	93	115	80

Table S3: Measured events per timepoint for aggregated median analysis
The number of available measurements per timepoint is shown for all European patients (n = 149). Longitudinal CBC sampling was not obtained for the US validation cohort. The density of follow-up decreased after discharge, resulting in less available measurements. W = week, M = month. The measured CBC values are depicted in **Figures 1C-F**, **3C-D**, **4E-F**.

	Quick Recovery (n = 37)	Intermittent Recovery (n = 78)	Aplastic (n = 34)
Age, years	58 (28 – 83)	62 (19 – 79)	59 (31 – 79)
Median lines of prior therapy (range)	3 (2 – 9)	3 (2 – 7)	3 (2 – 9)
CAR Product			
4-1BB (Tisa-cel)	23 (34%)	32 (47%)	13 (19%)
CD28z (Axi-cel)	14 (17%)	46 (57%)	21 (26%)
Disease Entity			
Diffuse Large B-Cell Lymphoma (DLBCL)	20 (21%)	52 (55%)	23 (24%)
Primary Mediastinal B-Cell Lymphoma (PMBCL)	4 (36%)	4 (36%)	3 (28%)
Transformed Lymphoma (trFL, trMCL, trMALT, trCLL, trHL)	13 (30%)	22 (51%)	8 (19%)
Complete Blood Count			
Median ANC, 95% CI (ANC/µl)	2640 (2090 – 3030)	2440 (1750 – 2760)	1400 (1060 – 2080)
Median Platelet Count, 95% CI (G/l)	198 (165 – 226)	172 (139 – 186)	129.5 (66 – 156)
Median Hemoglobin, 95% CI (g/dl)	10.8 (9.8 – 11.4)	10.1 (9.4 – 10.5)	9.0 (8.2 – 9.6)
Markers of Tumor Burden & Inflammation			
Median LDH, 95% CI (U/l)	289 (218 – 326)	274 (250 – 327)	341 (276 – 421)
Median CRP, 95% CI (mg/dl)	0.88 (0.39 – 1.89)	2.15 (1.14 – 5.12)	2.62 (0.95 – 4.50)
Median Ferritin, 95% CI (ng/ml)	309 (182 – 572)	486 (359 – 806)	1203 (812 –2425)

Table S4: Baseline demographic and clinical characteristics of treated CAR T-cell patients by clinical phenotype of neutrophil recovery
"Aplastic" defined as continuous severe neutropenia ≥ 14 days (n = 34). "Intermittent Recovery" defined as neutrophil recovery (ANC > 1500/µl) followed by a second dip with an ANC <1000/µl after day 21 (n = 78). "Quick Recovery" defined as sustained neutrophil recovery without a second dip below an ANC <1000/µl (n = 37).

Variable *(optimal discriminatory value)*	**AUC (ROC)**	**P Value**	**OR (95% CI)**	**P Value**
Platelet Count *(<175 G/l)*	**0.69**	0.017	6.11 (1.70 – 18.91)	0.005
Hemoglobin *(<9.0 g/dl)*	**0.66**	0.05	2.67 (0.84 – 8.79)	0.10
ANC *(<1200/µl)*	**0.61**	0.10	3.88 (1.17 – 12.99)	0.05
CRP *(>3 mg/dl)*	**0.67**	0.04	3.13 (0.96 – 9.43)	0.08
Ferritin *(>650 µg/l)*	**0.74**	0.003	9.74 (2.37 – 34.53)	0.0007
Age *(>60 yo)*	0.54	0.62	0.51 (0.18 – 1.48)	0.27
Gender (female)	0.50	0.99	1.09 (0.38 – 3.52)	0.99
CAR-Product *(Axi-cel)*	0.58	0.21	2.25 (0.63 – 7.21)	0.35
Disease Entity *(DLBCL/PMBCL)*	0.58	0.34	2.25 (0.63 – 7.21)	0.35
No. of prior treatment lines *(≥4)*	0.52	0.83	0.88 (0.24 – 3.53)	0.99
Prior autoSCT *(yes)*	0.55	0.57	1.36 (0.33 – 5.52)	0.74
eGFR (<65 ml/min)	0.55	0.50	1.14 (0.40 – 3.66)	0.99
LDH *(>330 U/l)*	0.59	0.24	2.0 (0.66 – 6.43)	0.27
Fibrinogen *(>400 mg/dl)*	0.56	0.62	1.15 (0.24 – 5.05)	0.99
Peak-IL6 *(>2500 pg/ml)*	0.53	0.67	1.59 (0.48 – 5.27)	0.56
Peak CRP *(>10 mg/dl)*	0.56	0.44	0.69 (0.23 – 2.01)	0.59
CRS Grade *(ASTCT ≥ 2)*	0.52	0.85	0.78 (0.28 – 2.32)	0.78
ICANS Grade *(ASTCT ≥ 2)*	0.59	0.27	0.39 (0.12 – 1.48)	0.23

Table S5: Discriminatory capacity of individual clinical features for the outcome of severe neutropenia ≥ 14 days (test) vs. <14 days (control)

The area under the curve (AUC) and concomitant p-values are depicted for the ROC analyses from **Fig. S3**. ROC curves were utilized to identify the value defining the optimum discriminator (e.g. the value that maximized the Youden *J* statistic [sensitivity + specificity – 1]). To calculate the odds ratio (OR), continuous variables were dichotomized. Each variable was studied for the binary outcome of severe neutropenia ≥ 14 days (test) vs. <14 days (control). The Baptista-Pike method was used to calculate the confidence interval for the odds ratios with p-values determined by Fisher's exact test. To determine variables for further modelling, an AUC threshold of 0.6 was chosen (P-value ≤ 0.1). Moreover, an odds ratio greater than 2.5 (P-value ≤ 0.1) was intended. This left the following highlighted clinical characteristics: Platelet Count, Hemoglobin, ANC, CRP, and ferritin.

Variables in the Equation

Total (*n* = 55)	B	Standard Error	Significance	Exp(B)	95% C.I. for Exp(B) Lower	Upper
Step 1 [a]						
Platelet Count	- 0.006	0.004	0.186	0.994	0.986	1.003
Absolute Neutrophil Count (ANC)	0.000	0.000	0.342	1.000	0.999	1.000
Hemoglobin	- 0.178	0.226	0.430	0.837	0.538	1.302
C-reactive protein (CRP)	0.057	0.104	0.581	1.059	0.864	1.298
Ferritin	0.000	0.000	0.212	1.000	1.000	1.001
Step 2						
Platelet Count	- 0.005	0.004	0.219	0.995	0.987	1.003
Absolute Neutrophil Count (ANC)	0.000	0.000	0.405	1.000	0.999	1.000
Hemoglobin	- 0.239	0.200	0.233	0.788	0.532	1.166
Ferritin	0.000	0.000	0.164	1.000	1.000	1.001
Step 3						
Platelet Count	- 0.006	0.004	0.155	0.994	0.987	1.002
Hemoglobin	- 0.235	0.202	0.244	0.790	0.532	1.174
Ferritin	0.000	0.000	0.152	1.000	1.000	1.001
Step 4						
Platelet Count	- 0.006	0.004	0.112	0.994	0.986	1.001
Ferritin	0.001	0.000	0.087	1.001	1.000	1.001

[a] Variable(s) entered on step 1: Baseline Platelet Count, ANC, Hb, CRP, Ferritin.

Omnibus Tests of Model Coefficients

	Chi-square	df	Significance
Step 1			
Step	12.669	5	0.027
Block	12.669	5	0.027
Model	12.669	5	0.027
Step 2 [a]			
Step	- 0.304	1	0.581
Block	12.365	4	0.015
Model	12.365	4	0.015
Step 3 [a]			
Step	- 0.730	1	0.393
Block	11.635	3	0.09
Model	11.635	3	0.09
Step 4 [a]			
Step	- 1.416	1	0.234
Block	10.219	2	0.06
Model	10.219	2	0.06

[a] A negative Chi-squares value indicates that the Chi-Squares value has decreased from the previous step.

Model Summary

Step	-2 log likelihood	Cox & Snell R Square	Nagelkerke R Sqare
1	61.363	0.206	0.278
2	61.667	0.201	0.272
3	62.397	0.191	0.258
4	63.812	0.170	0.229

Table S6: Binary Logistic Regression Multivariate Analysis for the outcome of an ANC < 500/µl more than 14 days vs. <14 days (D0-60)

Stepwise analysis was performed using a backward elimination and likelihood ratio (probability for stepwise entry = 0.05, removal = 0.1)

Score Comparison	AUC (ROC)	P Value	OR (95% CI)	P Value
CAR-HEMATOTOX *Model 1*	0.73	0.005	4.0 (1.31 – 12.10)	0.03
CAR-HEMATOTOX *Model 2*	0.81	< 0.001	15.3 (3.86 – 61.56)	< 0.0001
CAR-HEMATOTOX *Model 3*	0.82	< 0.001	42.0 (6.00 – 452.1)	< 0.0001

Table S7: Discriminatory capacity of the three tested models for the outcome of severe neutropenia ≥ 14 days (test) vs. <14 days (control)
The AUC and concomitant p-values are depicted for the ROC analyses shown in **Fig. S5** (Models 1 and 2), as well as **Fig. 3B** (Model 3). For the calculation of the odds ratio, high vs. low risk patients by CAR-HEMATOTOX were studied for the binary outcome of severe neutropenia ≥ 14 days (test) vs. <14 days (control). A higher OR indicates increased a positive correlation between a high-risk CAR-HEMATOTOX and severe neutropenia ≥ 14 days. The Baptista-Pike method was used to calculate the confidence interval for the odds ratio with p-values determined by Fisher's exact test.

		Mann-Whitney Test		ROC Analysis				
Score Thresholds	n	Neutropenia, Median Duration D0-60 (95% CI)	Exact P-Value	Sensitivity (%)	Specificity (%)	Youden-Index	Positive LR	Negative LR
<1	8	6 (0 – 12)	6×10^{-4}	100	24.24	0.24	1.32	0.00
≥1	47	13 (11 – 17)						
<2	**23**	**7 (6 – 10)**	**$<1 \times 10^{-4}$**	**95.45**	**66.67**	**0.62**	**2.86**	**0.07**
≥2	**32**	**16.5 (13 – 43)**						
<3	29	7 (6 – 12)	$<1 \times 10^{-4}$	77.27	72.73	0.50	2.83	0.31
≥3	26	24 (12 – 46)						
<4	41	11 (7 – 13)	0.001	45.45	87.88	0.33	3.75	0.62
≥4	14	23 (12 – 53)						

Table S8: CAR-HEMATOTOX threshold determination
Thresholds were analyzed by ROC analysis for the end outcome of severe neutropenia ≥ 14 days. Sensitivity, specificity, Youden-Index, and positive and negative likelihood ratio tests are shown for each threshold. The median duration of neutropenia for each binary threshold was compared by Mann-Whitney Test with the 95% confidence interval and respective p-value depicted. A score threshold of 2 was chosen due to the optimal Youden J statistic.

Patient Cohort		Mann-Whitney Test		Binary Logistic Regression		Calibration Slope Estimate	ROC Analysis					
CAR-HEMATOTOX	n	Neutropenia, Median Duration D0-60 (95% CI)	*p*	Severe Neutropenia ≥14 days, no. (%)	*p*	(Validation: Training)	AUC	Sens. (%)	Spec. (%)	Youden-Index	Positive LR	Negative LR
Training - **Low**	23	7 (6 – 10)	<0.001	1 (4%)	<0.001		0.82	95	67	0.62	2.86	0.07
Training - **High**	32	16.5 (13 – 43)		21 (66%)								
Validation Europe - **Low**	38	5 (3 – 8)	<0.001	5 (13%)	<0.001	0.87	0.77	84	69	0.53	2.70	0.23
Validation Europe - **High**	42	17 (13 – 19)		28 (67%)								
Validation USA - **Low**	59	7 (5 – 8)	<0.001	0 (0%)	<0.001	0.48	0.91	100	67	0.67	3.03	0.00
Validation USA - **High**	41	10 (9 – 12)		12 (29%)								
Pooled Validation - **Low**	97	5.5 (5 – 8)	<0.001	5 (5%)	<0.001	0.70	0.89	89	68	0.56	2.74	0.17
Pooled Validation - **High**	83	12 (10 – 16)		40 (48%)								

Table S9: Performance metrics of CAR-HEMATOTOX by patient cohort
Applying a CAR-HEMATOTOX threshold of 2, the training, validation and pooled validation cohorts were studied by risk group (high vs. low). The respective median duration of neutropenia was compared by Mann-Whitney test with the 95% confidence interval and respective p-value depicted (middle, left). The difference in the incidence of the binary endpoint of severe neutropenia ≥ 14 vs. <14 days was compared by CAR-HEMATOTOX risk group using Fisher's exact two-sided t test (middle). The calibration slope estimates are depicted for the validation cohorts compared to the training cohort respectively (middle, right). ROC analysis for the binary outcome of severe neutropenia ≥ 14 vs. <14 days by CAR-HEMATOTOX (right). Area under the curve (AUC), sensitivity, specificity, Youden-Index, and positive and negative likelihood ratio tests are shown for each patient cohort.

Supplementary Methods

CAR T-cell timeline

Prior to CAR T-cell transfusion (= Day 0), patients received a lymphodepletion regimen on days -5 to -3. For the Tisa-cel product, fludarabine (25 mg/m^2) and cyclophosphamide (250 mg/m^2) were applied. For the Axi-cel product, the respective doses were 30 mg/m^2 for fludarabine and 500 mg/m^2 for cyclophosphamide. Response assessments were performed according to institutional guidelines based on clinical, laboratory, and imaging studies. When ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (PET-CT) was used, the response was graded according to Lugano criteria. Patients were monitored until death or loss of follow-up.

Confidence Interval Computation

For the analysis of the aggregated median cytopenia curves (**Fig. 1C-F),** the 95% confidence interval of the median was computed for each timepoint (**Table S3**). The implemented Prism v8.0 software calculated the confidence interval of the median is by a standard method explained in "J.H. Zar, Biostatistical Analysis, Fifth edition 2010, ISBN: 0131008463 (pages 548-549)", based on the binomial distribution. The respective curve was then graphed using Prism v8.0 software.

For linear regression analysis (**Fig. 3A, Fig. 4A-B**), the 95% confidence intervals of the best-fit line were plotted in light shading. The calculated slope is denoted with β_1 and the Spearmann correlation coefficient is denoted with *r* on the graph inset. Accepting the assumptions of linear regression, there is a 95% chance that the 95% confidence interval of the slope contains the true value of the slope, and that the 95% confidence interval for the intercept contains the true value of the intercept. The width of the confidence intervals is determined by the number of data points, their distances from the line, and the spacing of the X values.

As described in:
H. J. Motulsky, "Median and its CI", GraphPad Statistics Guide. 2021.
https://www.graphpad.com/guides/prism/latest/statistics/stat_median_and_its_confidence_inte.htm
H. J. Motulsky, "Confidence and prediction bands (linear regression)" and "Finding the best-fit slope and intercept", GraphPad Statistics Guide. 2021.
https://www.graphpad.com/guides/prism/latest/curve-fitting/slopeandintercept.htm

Area under the Curve (AUC) Computation

The AUC computation was performed with Graphpad Prism v8.0. For each graph, the software reports the total peak area of the AUC curve with the concomitant standard error and confidence interval for the AUC using the method described by Gagnon et al. The software's AUC calculations are equivalent to taking a

weighted average of all the Y values (*for example: measured ANC per patient per timepoint*), giving the Y values corresponding to the lowest and highest X values half the weight of the other points. To compare AUCs (e.g. CAR-HEMATOTOX high vs. low), the df for each group was assessed, which is dependent on the number of overall data points (and therefore also the number of missing data points) for each timepoint (see **Table S3**). Statistical significance was assessed by unpaired t test.

As described in:

H. J. Motulsky, "Area under the curve", GraphPad Statistics Guide. 2021.

https://www.graphpad.com/guides/prism/latest/statistics/stat_area_under_the_curve.htm

https://www.graphpad.com/support/faqid/2031/

References:

1. Robert C. Gagnon and John J. Peterson, Estimation of Confidence Intervals for Area Under the Curve from Destructively Obtained Pharmacokinetic Data, Journal of Pharmacokinetics and Pharmacodynamics, 26: 87-102, 1998.

2. Bailer A. J. (1988). Testing for the equality of area under the curves when using destructive measurement techniques. Journal of Pharmacokinetics and Biopharmaceutics, 16(3):303-309.

3. Jaki T. and Wolfsegger M. J. (2009). A theoretical framework for estimation of AUCs in complete and incomplete sampling designs. Statistics in Biopharmaceutical Research, 1(2):176-184.