Citation: CPT Pharmacometrics Syst. Pharmacol. (2020) 9, 272–281; doi:10.1002/psp4.12486

ARTICLE

Clinical Trial Simulation To Optimize Trial Design for Fludarabine Dosing Strategies in Allogeneic Hematopoietic Cell Transplantation

Jurgen B. Langenhorst[1,2,3], Thomas P.C. Dorlo[4], Charlotte van Kesteren[1], Erik M. van Maarseveen[5], Stefan Nierkens[1,2], Moniek A. de Witte[6], Jaap Jan Boelens[1,7] and Alwin D.R. Huitema[4,5,*]

Optimal fludarabine exposure has been associated with improved treatment outcome in allogeneic hematopoietic cell transplantation, suggesting potential benefit of individualized dosing. A randomized controlled trial (RCT) comparing alternative fludarabine dosing strategies to current practice may be warranted, but should be sufficiently powered for a relevant end point, while still feasible to enroll. To find the optimal design, we simulated RCTs comparing current practice (160 mg/m^2) to either covariate-based or therapeutic drug monitoring (TDM)-guided dosing with potential outcomes being nonrelapse mortality (NRM), graft failure, or relapse, and ultimately overall survival (covering all three aforementioned outcomes). The inclusion in each treatment arm (*n*) required to achieve 80% power was calculated for all combinations of end points and dosing comparisons. The trial requiring the lowest *n* for sufficient power compared TDM-guided dosing to current practice with NRM as primary outcome (*n* = 70, NRM decreasing from 21% to 5.7%). We conclude that a superiority trial is feasible.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ The current practice of body-surface-area (BSA)-based fludarabine dosing provides highly variable plasma exposures, while actual body weight combined with renal function better predict individual drug clearance than BSA. Variable exposures lead to variable outcome: low fludarabine exposures associate with graft failure; high exposures with nonrelapse mortality (NRM).

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ What is the optimal design and expected outcome of a randomized controlled trial for individualized dosing of fludarabine?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ The overall mortality for therapeutic drug monitoring guided dosing is expected to be 26%, which is reduced compared with 39% for the current practice. A confirmatory trial to prove this effect prospectively should include 70 subjects per arm and have NRM as primary end point.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?
☑ A prospective trial can be designed rationally, including the minimal number of subjects necessary. In addition, individualized dosing is expected to have a relevant impact on survival after hematopoietic cell transplantation.

Allogeneic hematopoietic cell transplantation (HCT) is a potentially curative treatment for a variety of malignant and benign hematological disorders. Unfortunately, nonrelapse mortality (NRM; 10–40%) and disease relapse (20–50%) remain major causes of therapy failure,[1] thus further treatment optimization is potentially life-saving.

The conditioning regimen prior to HCT consists of a combination of cytotoxic agents (chemotherapy and serotherapy) administered to eradicate recipient's bone marrow and immune system. Minimizing the toxicity while maintaining the efficacy of such regimens is one of the key strategies to reduce NRM.[2] Fludarabine combined with busulfan and rabbit antithymocyte globulin (rATG) is a frequently used conditioning regimen for HCT. Variability in exposure has been shown to predict variable outcome for all these agents.[3–5]

Given that exposure-targeted dosing of busulfan through therapeutic drug monitoring (TDM) is widely applied and has been proven superior over fixed dosing in a randomized

[1]Pediatric Blood and Marrow Transplant Program, Princess Maxima Center for Pediatric Oncology, Utrecht, The Netherlands; [2]Laboratory of Translational Immunology, University Medical Centre Utrecht (UMCU), Utrecht University, Utrecht, The Netherlands; [3]Model-informed drug development consultant, Pharmetheus AB, Uppsala, Sweden; [4]Department of Pharmacy & Pharmacology, Antoni van Leeuwenhoek Hospital, Netherlands Cancer Institute, Amsterdam, The Netherlands; [5]Department of Clinical Pharmacy, University Medical Centre Utrecht (UMCU), Utrecht University, Utrecht, The Netherlands; [6]Department of Hematology, University Medical Centre Utrecht (UMCU), Utrecht University, Utrecht, The Netherlands; [7]Stem Cell Transplant and Cellular Therapies, Pediatrics, Memorial Sloan Kettering Cancer Centre, New York, New York, USA. *Correspondence: Alwin Huitema (a.d.r.huitema-2@umcutrecht.nl)
Received: September 26, 2019; accepted: December 3, 2019. doi:10.1002/psp4.12486

controlled trial (RCT),[6] the focus for dose optimization has now shifted to fludarabine. We previously described highly variable fludarabine plasma exposures following current practice of body-surface-area (BSA)-based dosing.[7] Ivaturi *et al.*[5] showed an association between low fludarabine exposure and graft failure with subsequent lower survival. Furthermore, other studies demonstrated that overexposure[8,9] (or lower predicted fludarabine clearance[9]) was associated with NRM. In follow-up to these findings, an optimal fludarabine exposure was identified at an area under the plasma concentration time curve (AUC) from the first dose until infinity ($AUC_{t_{0-\infty}}$) of 20 mg·hour/L.[10] This exposure target was associated with a minimal probability of having any of the negative events typically used to define HCT outcome (NRM, relapse, and graft failure). Our previous analysis suggested that individualized dosing, guided by TDM or based on the pharmacokinetic (PK) model of fludarabine,[7] is likely to improve PK target attainment compared with current practice, and possibly increase survival.

To implement and evaluate the superiority and feasibility of the suggested individualized fludarabine dosing algorithms in clinical practice, an RCT would be required to confirm the suggested advantage in terms of reduced NRM or increased overall survival (OS) of individualized dosing over current practice. However, the design of such a trial is complicated by various choices and practical circumstances/limitations, such as choice of end point, limited number of patients available (~ 9,000 HCTs/year throughout the United States[11]) and complicated calculation of the effect sizes (necessary for power calculation) associated with alternative dosing strategies.

To aid rational decision making regarding optimal trial design, we constructed a framework connecting dosing to (i) plasma exposure ($AUC_{t_{0-\infty}}$)[7] (ii) events (NRM, relapse, graft failure)[10] and (iii) survival by using the PK model, the previously developed cause-specific hazard (CSH) models, and newly developed Markovian transition elements. This framework was subsequently used to perform clinical trial simulations (CTSs) of RCTs comparing current practice to alternative dosing strategies. CTSs were used to assess the feasibility of an RCT evaluating the suggested individualized regimens that could be conducted in a reasonable number of patients, while maintaining clinical relevance.

METHODS

General framework

In **Figure 1**, the simulation framework is depicted. The framework consisted of a PK model prior to HCT and time-to-event models following HCT. How the various components and events were related is depicted by the arrows, while unconnected adjacent events were considered competing. For the PK components, the previously published PK model[7] was used. The time-to-event models can be further subdivided in the time until first event and, if applicable, a subsequent Markovian post-event survival transition.

For the first model in the time-to-event framework, the previously used data set and CSH models[10] were developed and refined further (Time-to-Event Model (Re-)estimation) to include only predictive covariates. Previously, covariate inclusion was on an *a priori* basis: including all prior known important covariates without testing for predictive power in this particular data set. The current analysis was focused on prediction of outcome based on exposure and, therefore, the covariate model was rebuilt.

The Markovian transition models were newly constructed (Simulating Events and Survival).

Simulation data set

The trial populations used for simulation were sampled from a population based on an in-house database of HCT recipients (henceforth referred to as "simulation data set," collected during 2005–2016: n = 194). These 194 subjects were selected from all transplantations recorded based on indication (myelodysplastic syndrome (MDS), leukemia, and lymphoma), age (> 20 years at HCT), and availability of creatinine values within 1 month of transplantation. These criteria were chosen to exclude benign (rare indication for HCT in adults and under-represented in the estimation data set) and plasma cell disorders (debatable eligibility for HCT due to poor outcome[12]), select the population with a high risk of overexposure with current practice (adults),[7,10] or omit imputation of creatinine values, respectively.

For the largest proportion of the in-house data set, covariates were well documented. To allow a complete set of input covariates necessary to predict PK and event probabilities, missing covariate values for remaining patients were imputed as follows. Body weights were sampled per sex from the known body weight distributions in each quintile of age. To impute heights that were considered plausible for the corresponding weight, body mass index (BMI) was sampled based on age and sex: for each subject with a missing value a random BMI was selected from all known values within his/her sex and 5% age quantile. The heights were then calculated from the known body weight and imputed BMI. The rATG exposures were not documented for all patients in the simulation data set and were, therefore, simulated as described in "Simulating PKs."

Simulating PKs

The simulation data set was constructed, assuming current practice of 160 mg/m^2 divided over 4 days for each patient. The previously published PK model,[7] a two-compartmental model with allometric scaling for weight and estimated glomerular filtration rate (eGFR) included as covariates, was used to simulate 25 sets of individual clearance values per input set of covariates, to take into account interpatient variability. Based on the simulated typical individual clearance, clearance values per day were simulated, taking into account the interoccasion variability previously estimated in the PK model.[7] Subsequently, five fludarabine plasma concentrations (also taking into account residual variability) were simulated on the first day of conditioning at times centered around 1, 5, 6, 7, and 8 hours after the start of infusion, where the distributions around these sample times were taken from the estimation data set.[7] These samples were simulated in order

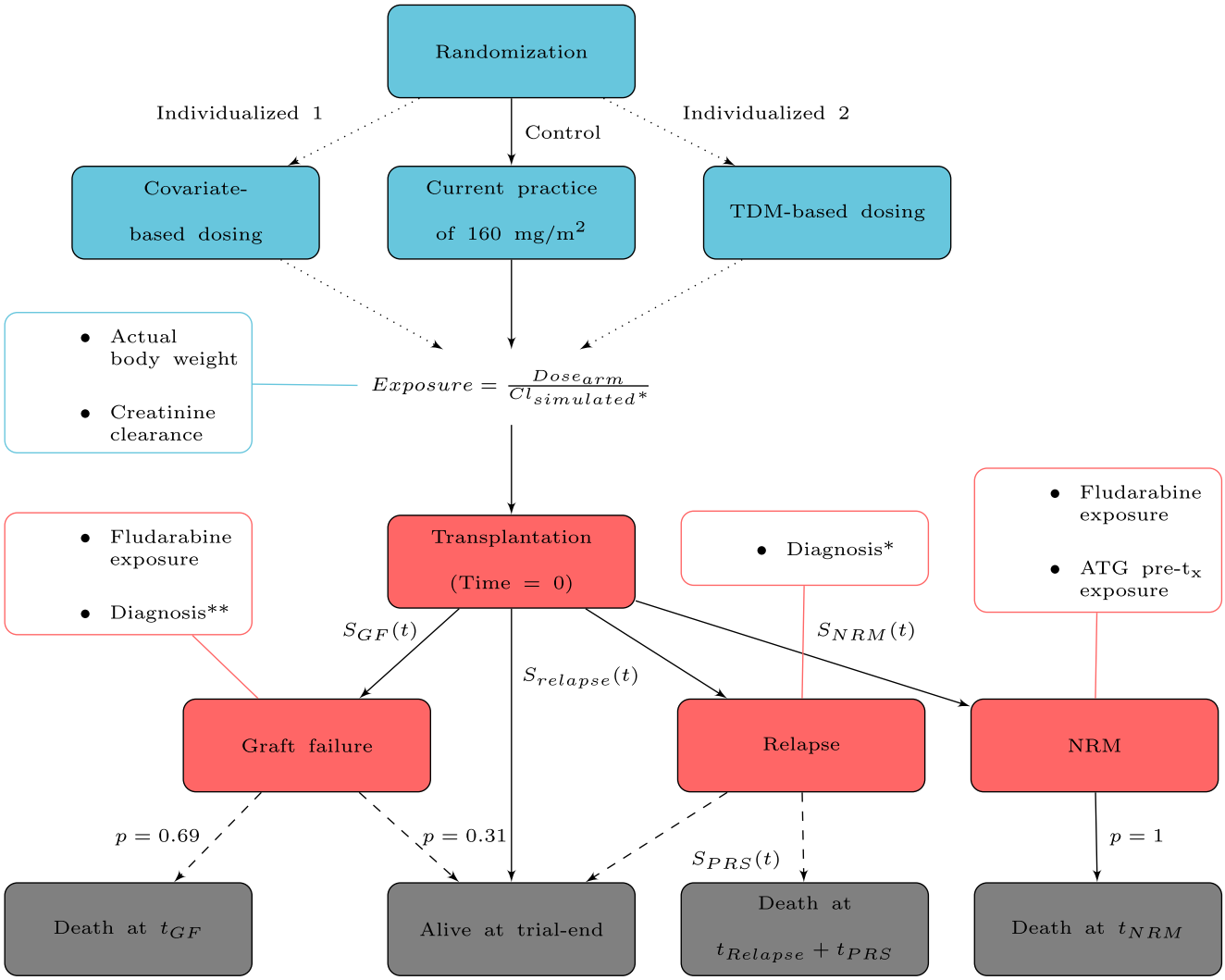

Figure 1 Simulation framework. Pharmacokinetic simulations are depicted by blue boxes, time-to-event simulations by the red boxes, and survival status by the gray boxes. Dotted lines are optional directions (i.e., individualized dosing is either therapeutic drug monitoring (TDM) or covariate-based, but the control is always current practice). Dashed lines depict the Markovian elements: transition from event to survival/death. Colored lines connect the covariates boxes to the corresponding model. Cl, clearance; GF, Graft failure; NRM, non-relapse mortality; PRS, post-relapse survival; rATG, rabbit anti-thymocyte globulin; S, survival; t_{GF}, time of graft failure; t_{NRM}, time of non-relapse mortality; $t_{relapse}$, time of relapse; t_x, time of transplantation. *Diagnosis here is subdivided in: leukemia/lymphoma, myelodysplastic syndrome, plasma cell disorders, and benign disorders. **Diagnosis here is subdivided in: benign and malignant disorders.

to perform TDM (Dosing Regimens). TDM samples were randomly excluded according to the previously observed incidence of missing samples (~ 8% randomly missing data).

For rATG, the dosing regimen (400 mg + 350·absolute number of lymphocytes starting at day –9) described by Admiraal *et al.*[3] was used and exposure was simulated using a PK model described in the same paper. Where absolute number of lymphocyte is the absolute number of lymphocytes a patient has preceding rATG administration in 10^9 cells per liter of blood.

Henceforth, rATG exposure was calculated by integration of the plasma-concentration-time curve as AUC from the first dose until time of transplantation ($AUC_{t_0-t_x}$) and AUC from time of transplantation until infinity ($AUC_{tx-\infty}$).

Time-to-event model (re-)estimation

To select predictive covariates for the CSH models and estimate associated effect sizes as well as developing the Markovian transition models, the data set (henceforth referred to as "estimation data set") from the previous publication was used.[10]

The same baseline hazards for the CSH models were used as previously presented in the PK-pharmacodynamic analysis.[10] To include predictive covariates to be used in the simulations, a selection based on forward addition ($P < 0.1$) and backward deletion ($P > 0.05$) was performed. Evaluated covariates were based on reports from literature, previous analyses, biological, or physiological plausibility: age at HCT,[3,13] rATG exposure pretransplantation and post-transplantation,[3,14] cumulative busulfan exposure,[4] graft source,[15,16] cytomegalovirus serostatus

(patient and donor),[17] diagnosis,[18] prior HCT,[19–21] and human leukocyte antigen disparity.[22]

Continuous covariates were tested as covariate on the baseline hazard using a (log-)linear, polynomial (first, second, third degree), and a polynomial spline (3, 4, 5 degrees of freedom) function. The relationship resulting in the lowest Akaike information criterion in the univariate analysis was selected as the most relevant form and was tested throughout the procedure.

For categorical covariates, the category with the highest number of patients was chosen as a reference. A proportional change in baseline hazard relative to the reference value was estimated for all other categories.

For the Markov elements, the probability of transition from relapse to death was modeled using a parametric survival model. Patients included were those who experienced a relapse and were followed from the time of relapse until death or last follow-up, whichever occurred first. Due to the small number of graft failures, the probability of death could not be estimated with reasonable precision and was taken from literature: 0.69.[23]

Framework evaluation

The different elements of the simulation framework were evaluated using visual predictive checks. For this, events, survival, and corresponding event-times were simulated 1,000 times for each subject. The cumulative incidence of events and OS were computed for each simulation. The mean and 95% prediction interval (PI) of 1,000 simulations were compared with the observed cumulative incidence and Kaplan-Meier curves for the events and OS, respectively. Results were stratified for indication (leukemia, lymphoma, and MDS) and/or fludarabine exposure (tertiles per indication), depending on event-type. The model should adequately reflect the observed survival differences as a result from exposure variation in all indications, to justify extrapolation of the exposure response relationship to the proposed clinical trial setting.

Dosing regimens

All doses were administered over 4 days at day –5 to –2 relative to HCT.

Current practice was defined as a cumulative dose of 160 mg/m^2 (40 mg/m^2/day).

Covariate-based dosing was based on the previously published PK model[7] and the herein included covariates eGFR and body weight. Doses were calculated using Eq. 1, with the total calculated dose being equally divided over the 4-day period.

$$\text{Cumulative dose} = 20\ \text{mg}\cdot\text{hour/L}\cdot(0.782\cdot\text{eGFR L/hour} + 3.24\,\text{L/hour})\cdot\frac{\text{BW}}{70\,\text{kg}}^{0.75} \quad (1)$$

where the cumulative dose is in mg of the fludarabine base (administered as phosphate), $20\,\text{mg}\cdot\text{hour/L}$ is the previously established optimal fludarabine exposure,[10] eGFR is estimated glomerular filtration rate in L/hour/70 kg and body weight (BW) is in kg.

TDM-based dosing was performed by first predicting the clearance from the simulated fludarabine concentration-time data (Simulating PKs) on the first of 4 days of conditioning. This was done with Bayesian forecasting, where the final estimates from the PK model (fixed and random effects) were used as priors and individual PK parameter estimates were generated using the post hoc step in NONMEM. Based on this, the individual exposure for current practice was derived (see Eq. 1). Henceforth, the dose adjustments for the subsequent days necessary to achieve the desired target exposure ($AUC_{t_{0-\infty}}$ of $20\,\text{mg}\cdot\text{hour/L}$) were calculated, taking into account the already administered exposure of the first dose (40 mg/m^2) and assuming linear PK behavior.

The fludarabine $AUC_{t_{0-\infty}}$ proved to be predictive for outcome[5,8–10] and was calculated for each dosing regimen (current practice, covariate-based dosing, and TDM-based dosing) using Eq. 2 and the simulated PK mentioned in "Simulating PKs."

$$AUC_{t_{0-\infty}} = \sum_{i=1}^{4}\frac{\text{Dose}_{\text{day}_i}}{\text{Clearance}_{\text{day}_i}} \quad (2)$$

Simulating events and survival

Using baseline characteristics and the simulated fludarabine $AUC_{t_{0-\infty}}$ per dosing regimen, daily event probabilities were simulated with the CSH models for each subject. A trial was simulated, by randomly assigning patients, stratified on diagnosis, to receive either current practice (160 mg/m^2) or individualized dosing (TDM or covariate-based). Events and OS were simulated using the survival probabilities corresponding to the assigned dosing regimen. Trials were simulated until 1 year after HCT.

To calculate the power of a trial design, the proposed trial was simulated 1,000 times and for each trial the *P* value was calculated using either Gray's test (separate events) or a two-sided log-rank test (OS). The power was defined as the fraction of trials reaching $P < 0.05$. The designs were evaluated by performing randomization with a varying number (50, 60, … , 220) of subjects included in each treatment arm (*n*). The trial design with the lowest *n* to achieve 80% power was considered optimal.

Sensitivity analyses

Sensitivity analyses were performed only for the optimal trial (Power Calculations and Optimal Trial Results) to test uncertainties in certain assumptions. An important assumption is that the relationships estimated in the time-to-event model are true. In the models, a high increase of NRM probability was estimated already at exposures frequently observed for current practice, whereas the increase of graft failure probability was only found at exceptionally low exposures. The risk of relapse was not increased despite lower fludarabine (cytostatic) exposure. Consequently, a lower exposure (as is the case for both individualized dosing regimens: **Table 1**) would be predicted as beneficial. To relax the assumption of favorable low exposures and simulate worst-case model errors, adjustments in the models were made in favor of high exposures. For this, the probability of graft failure at exposures below the optimum was increased by a relative risk of 1.1 (10% higher). In a separate

simulation, the NRM probability at exposures above the optimum was adjusted by a relative risk of 0.9 (10% lower). A third analysis was conducted where a relationship between lower fludarabine exposures and relapse was introduced: the relapse probability was adjusted to increase exponentially below the optimum up to a relative risk of 2.0 for each 10 mg·hour/L decrease of fludarabine exposure (see Eq. 3; only used for exposures below 20 mg·hour/L).

$$P_{\text{relape,ID}=i,\text{Time}=j,\text{adjusted}_{\text{AUC}=x}} = P_{\text{relapse,ID}=i,\text{Time}=j,\text{unadjusted}_{\text{AUC}=x}} \cdot e^{(x-20\,\text{mg·hour/L})\cdot\frac{\log(2)}{10\,\text{mg·hour/L}}} \quad (3)$$

Where the unadjusted relapse is the point estimate from the CSH model for relapse at time equals *j* for subject *i* with fludarabine $AUC_{t_{0-\infty}}$ equals *x*.

Furthermore, in the simulations for TDM-based dosing, TDM was assumed to be successful in all subjects. In practice, samples may be lost or TDM may fail otherwise. Therefore, the possibility that TDM is unsuccessful in 10% of the patients, defined as all samples missing for these patients, was tested. For this, randomly selected subjects received covariate-based dosing instead of TDM-based dosing, which was regarded a feasible and good alternative in practice. For all sensitivity analyses, power was recalculated by simulating the trial 1,000 times, implementing aforementioned assumptions.

RESULTS

Population characteristics

Characteristics of the populations used for estimation[10] and sampling for clinical trial simulations are listed in **Table 1**. The characteristics used as input for the simulations were in correspondence with the range/distribution observed in the original population on which the models were based.

Time-to-event model (re-)estimation

Baseline hazard distributions for NRM, relapse, and graft failure were kept as previously described.[10] Predictive covariates for NRM were: rATG pre-transplant exposure (log-linear, $P = 0.04$), fludarabine $AUC_{t_{0-\infty}}$ (polynomial spline, $P < 0.001$); for relapse: diagnosis (benign, leukemia/lymphoma, plasma cell disorder, and MDS, $P < 0.001$); for graft failure: fludarabine $AUC_{t_{0-\infty}}$ (linear, $P = 0.04$), and diagnosis (benign, malignant, $P = 0.03$). For postrelapse survival, a Gompertz distribution was found optimal.

The visual predictive check (**Figure 2**) shows adequate description of OS with the current platform. The same holds true for cumulative incidence of separate events (data not shown).

Exposure simulation

In **Figure 3** the simulated exposures for the different dosing regimens are depicted. In general, the individualized dosing

Table 1 Patient characteristics

	Estimation data set	Simulation data set
Age at HCT, years	36 (0.23–74)	46 (20–74)
Weight at HCT, kg	65 (4.3–130)	74 (36–130)
Dose, mg/m^2	160	Current practice: 160
		Current practice: 160 TDM: 122 (76–186)[a]
		Covariate-based: 125 (86.7–142)[a]
Fludarabine $AUC_{t_{0-\infty}}$, mg·hour/L	24.7 (13.9–66)	Current practice: 26.8 (16.8–44)[a]
		Current practice: 26.8 (16.8-44) TDM: 20.3 (17.3–24.2)[a]
		Covariate-based[a]: 20.3 (13.4–30)[a]
Creatinine clearance, mL/minute/1.73 m^2	110 (25–140)	120 (44–140)
ATG pre-HCT exposure, AU·day/mL	54 (4.4–210)	77 (28–226)[a]
Diagnosis (*N*, %)		
Leukemia/lymphoma	71 (37%)	76%
MDS	30 (16%)	24%
Plasma cell disorder	23 (12%)	Excluded
Benign	68 (35%)	Excluded
Year of HCT		
2017	1 (0.52%)	
2016	60 (31%)	
2015	60 (31%)	
2014	32 (17%)	
2013	11 (5.7%)	
2012	13 (6.8%)	
2011	12 (6.2%)	
2010	3 (1.6%)	

ATG, anti-thymocyte globulin; AU, arbitrary units; $AUC_{t_{0-\infty}}$, area under the plasma-concentration-time-curve from the first dose until infinity; HCT, allogeneic hematopoietic cell transplantation; MDS, myelodysplastic syndrome; TDM, therapeutic drug monitoring.
Displayed values are: median (range) unless specified otherwise.
[a]2.5th and 97.5th percentiles.

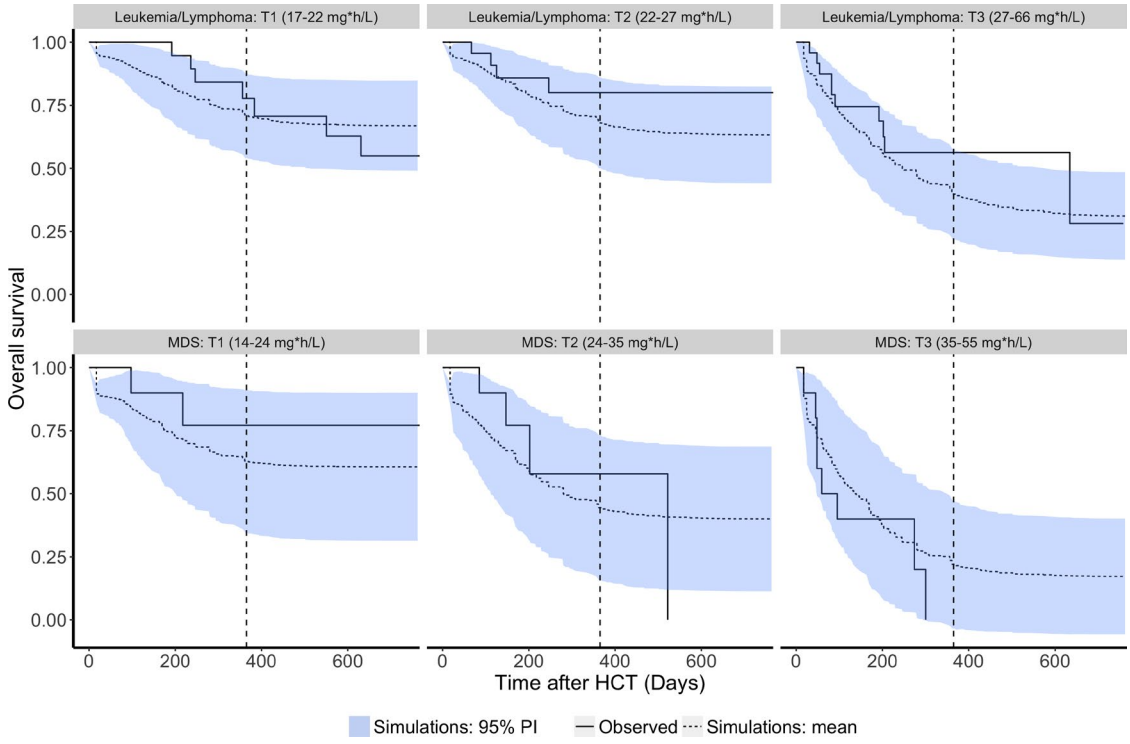

Figure 2 Visual predictive checks. Overall survival stratified for diagnosis and fludarabine exposure tertiles within each diagnosis. Solid lines indicate the observed Kaplan–Meier estimates and the dotted line shows the mean of simulations. The blue area represents the 95% confidence interval of simulations. The vertical dashed line depicts the end-time of the clinical trial simulation (1 year after transplantation). HCT, allogeneic hematopoietic cell transplantation; MDS, myelodysplastic syndrome; PI, prediction interval; T, tertile.

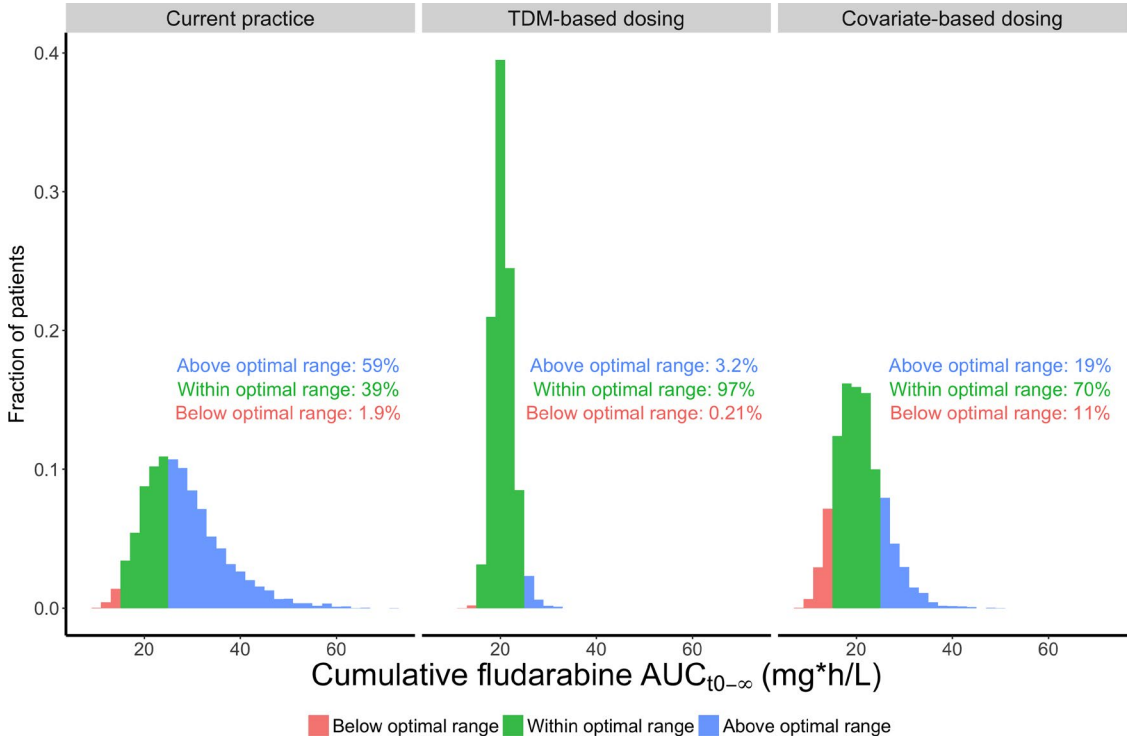

Figure 3 Fludarabine exposures for different dosing regimens. Histograms of the simulated fludarabine exposure following the different dosing regimens. The fraction on the *y*-axis is depicted within each trial arm and bins on the *x*-axis are every 2 mg·hour/L. The optimal range is defined as the optimum ± 25% (15–25 mg·hour/L). $AUC_{t_{0-\infty}}$, area under the plasma-concentration-time-curve from the first dose until infinity; TDM, therapeutic drug monitoring.

regimens resulted in lower exposures with more narrow distributions. This is illustrated by the high proportion (59%) of patients who exhibited an exposure > 25% higher than the defined optimum after current practice. In contrast, covariate-based dosing led to more patients (11% opposed to 1.9%) with exposures that were > 25% lower than the optimum. TDM performed best, with 97% of patients within ± 25% of the optimum.

Power calculations

In all simulated scenarios, the power to detect a significant difference in relapse or graft failure remained low (**Figure 4**: < 15%). When current practice was compared with covariate-based dosing, a power of 80% is reached with a trial size of 120 patients per arm for NRM, whereas this power threshold was not even achieved for OS with the maximal number of patients per arm of 220.

In scenarios where TDM-based dosing was the intervention, less patients were needed to have sufficient power, with 70 and 160 patients per arm necessary to reach 80% power for NRM and OS, respectively. The scenario where TDM-based dosing was compared with current practice with NRM as a primary end point was deemed the optimal trial format and evaluated in "Optimal Trial Results."

Optimal trial results

The overall mortality (1-OS) changed in favor of TDM-based dosing (**Figure 5**) from 39% (PI: 29–50%) to 26% (PI: 16–36%). This was due to a decrease in NRM from 21% (PI: 11–31) to 5.7% (PI: 0–11), comparable graft failure and an increase in relapse from 23% (PI: 14–34) to 26% (PI: 16–37).

Sensitivity analyses

The sensitivity analyses showed that increasing the event probability for relapse and graft failure at low exposures ($AUC_{t_{0-\infty}}$ < 20 mg·hour/L), as described in "Sensitivity Analyses," had a marginal effect on trial power (**Table 2**). However, reducing NRM probability at $AUC_{t_{0-\infty}}$ > 20 mg·hour/L by 10% did reduce power below the 80% threshold. In addition, unsuccessful TDM had an effect on power, but to a lesser extent (adjusted power: 78.8%).

DISCUSSION

In this CTS, we showed that individualized fludarabine dosing through TDM might reduce overall mortality after HCT in adults with malignancies approximately from 39% to 26%. This is attributable to reduced NRM, without an increased risk of graft failure. The described simulations were based

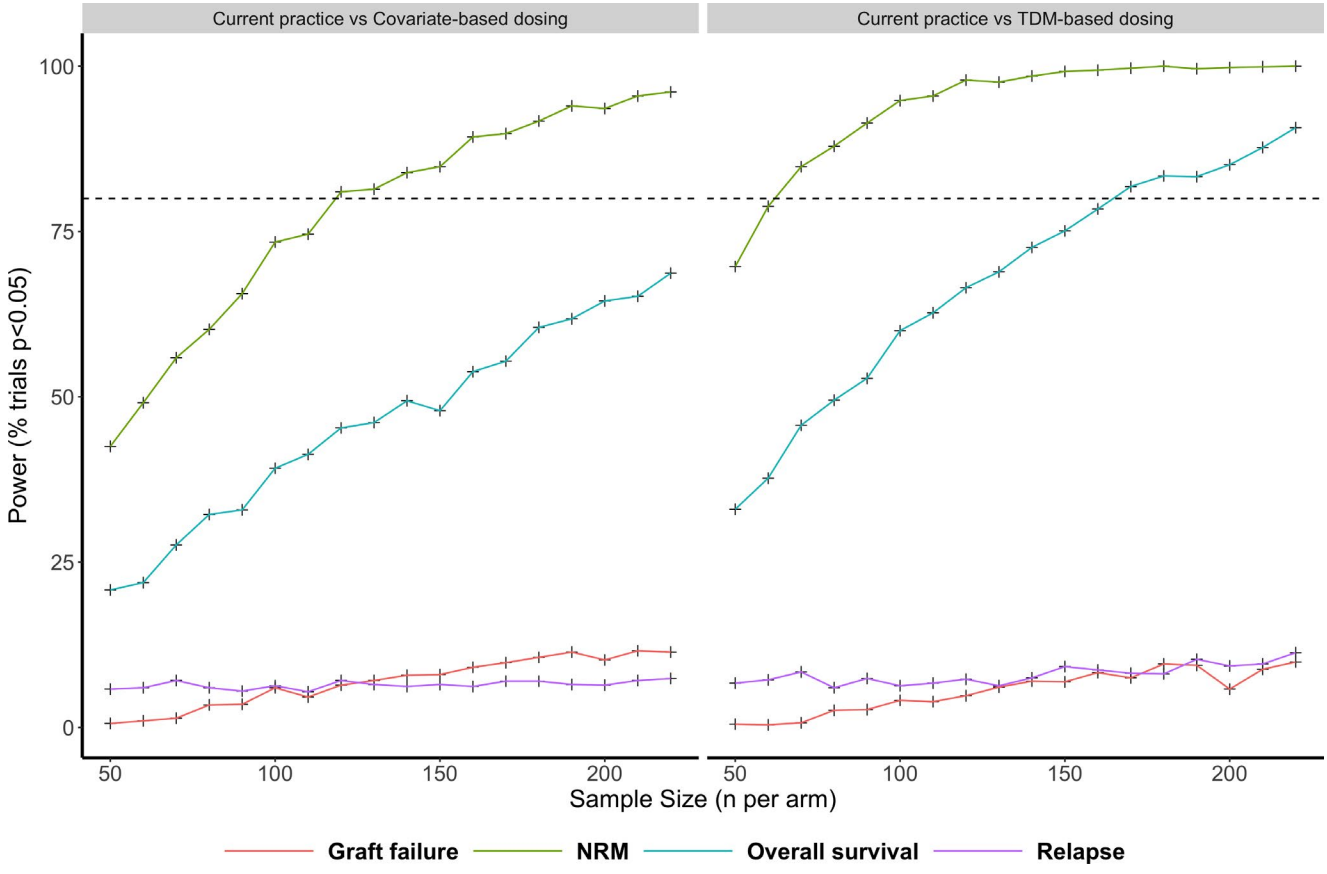

Figure 4 Power calculations for simulated trials. Each point depicts the percentage of 1,000 trials with $P < 0.05$ at the corresponding number of patients per arm. The horizontal dashed line is the threshold of 80% power. NRM, nonrelapse mortality; TDM, therapeutic drug monitoring.

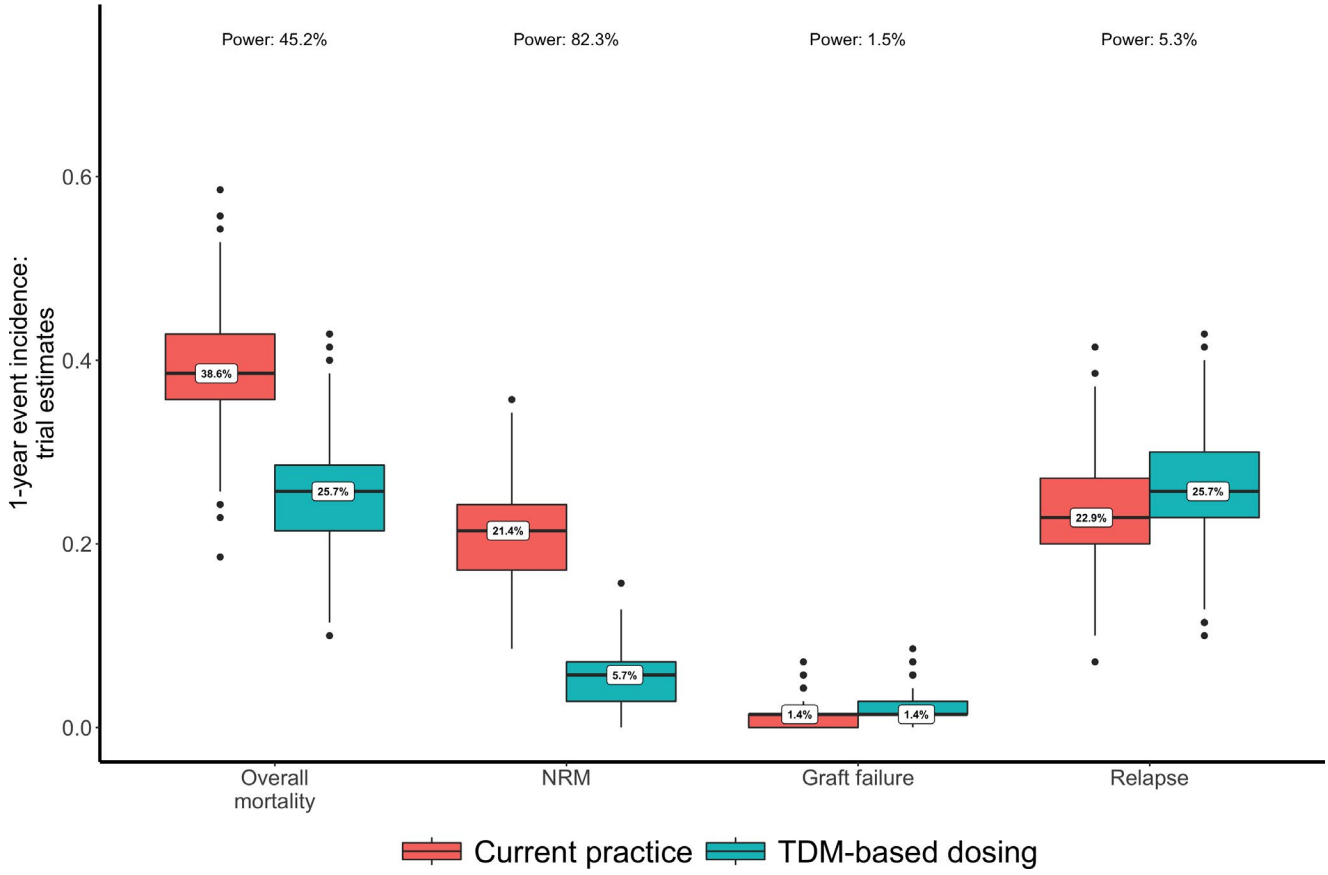

Figure 5 Simulated outcomes with optimal trial design. Boxplots of simulated trial results stratified by TDM-based dosing vs. current practice. Values of overall mortality or 1-OS represent 1 minus Kaplan–Meier estimate (overall mortality) or the cumulative incidence (graft failure, NRM, relapse). The horizontal center line and percentages correspond to the median estimates. Boxes go from the 25th to 75th percentile and whiskers are twice the interquartile range. NRM, nonrelapse mortality; OS, overall survival; TDM, therapeutic drug monitoring.

on the exposure-response relationship for fludarabine,[10] which was established on a data set of 192 HCT recipients of all ages. In order to confirm superiority of TDM to the current practice of 160 mg/m^2 dosing in an RCT with NRM as end point, 70 patients per arm would be required for a power of 80% under ideal circumstances.

To assess the feasibility of such a trial, the number of HCTs performed at the University Medical Centre in Utrecht (**Table 1**) can be taken as a reference. Here, ~ 60 patients/year were available in recent years, of whom half meet the diagnosis inclusion criteria (MDS/leukemia/lymphoma). To reach the total trial number of 140 patients (70 per arm), at least 5 years of inclusion would be needed for a single center. Therefore, such a trial would be most feasible in a multicenter setting, where all centers should implement the TDM procedure. However, TDM for busulfan is assumed to be clinical practice and, consequently, our previously developed method to quantify fludarabine and busulfan in one sample[24] could aid in implementation of TDM for fludarabine as well. The end point NRM is usually well-documented already, thus introducing no additional burden. In addition, a trial similar to the setting simulated here was performed for busulfan in HCT conditioning,[6] indicating the feasibility of the proposed fludarabine trial and putting the predicted survival gain and calculated power of our CTS in perspective. In the busulfan trial, TDM-based to BSA-based dosing was compared and an approximate survival gain of 20% was identified and statistical significance was achieved (P = 0.04) with 110 patients per arm, which is more than the 70 found in this study. However, the primary end point in the busulfan study was OS, which was also associated with lower power in our simulations for fludarabine. We simulated NRM to be the end point with most discriminative power for fludarabine, because an influence of fludarabine exposure on relapse could not be identified, probably because high fludarabine exposure leads not only to higher cytotoxicity but also to a reduced graft-vs.-disease effect, and graft failure is not of much

Table 2 Sensitivity analysis

Trial assumptions (n = 70, NRM primary end point)	Power (%)
Original assumptions	82.3
P_{NRM} reduced from $AUC_{t_{0-\infty}} > 20$ mg·hour/L: RR = 0.9	73.8
P_{GF} increased from $AUC_{t_{0-\infty}} < 20$ mg·hour/L: RR = 1.1	82.0
$P_{Relapse}$ increased from $AUC_{t_{0-\infty}} < 20$ mg·hour/L (see Eq. 3 for RR)	82.4
TDM fails in 10% of cases	78.8

GF, graft failure; NRM, nonrelapse mortality; P, event probability; RR, relative risk; TDM, therapeutic drug monitoring.

concern in malignancies.[25] For other HCT settings where graft failure plays a more prominent role, such as transplantation for benign disorders and/or a comparison with the also used lower BSA-based fludarabine dosing of 120–150 mg/m^2,[26–28] other end points could be more discriminative. The proposed CTS-framework could be applied to investigate this.

Generally, the results of a CTS are fully dependent on the underlying assumptions. Assumptions that were identified as uncertain were further investigated in the sensitivity analyses. Our CTS framework consisted of a PK model and survival models. Here, assumptions in the PK model are probably of less concern than in the survival models, given the difference in parameter uncertainty between these two models[7,10] and the adequate quantification of PK interpatient variability (i.e., low shrinkage[7]). The survival model parameters have a higher uncertainty. Importantly, in the estimation data set, the target population for simulations (adult leukemia, lymphoma, and MDS) was not the only indication. Other indications included in the estimation data set (plasma cell disorders and pediatric benign disorders) might have caused underestimation or overestimation of the true exposure-response relationship. Therefore, sensitivity analyses were conducted to evaluate the potential impact of unaccounted or incorrect relationships in the survival models, where a previously unobserved association between low exposure and relapse was introduced. In addition, the probability of graft failure at low exposures was increased. These unfavorable effects of low exposures (more relapses and graft failures) had relatively little effect on study power. A dampened relationship with NRM at high $AUC_{t_{0-\infty}}$ led to a loss of power, but uncertainty of the survival models was generally lower for these exposures, which makes it less likely that the assumed deviation holds true in the real situation. Reasonable deviations from the currently used exposure-response relationship are, therefore, not expected to have a large impact on study outcome.

Furthermore, we tested how unsuccessful TDM (10%) would affect power, which was negligible. Several other possible protocol deviations, such as the possibility of missing samples, were already implemented in the simulations and were based on our local TDM experience. Nevertheless, one could also expect physicians deviating from the TDM-based dosing recommendation, especially when the suggested adjustment is marginal or, by contrast, extremely large. Omitting extreme adjustments will vastly reduce power, and should, thus, be avoided at all cost. This can be done by clearly stating the dosing alteration procedure in the trial protocol.

In summary, CTS indicates the possibility of a substantial improvement in HCT survival by using TDM-based dosing of fludarabine. Furthermore, the proposed CTS framework can be used for prediction of efficacy for other treatment alterations in HCT. Based on the findings in this study, prospective evaluation studies can be designed to provide definite proof for the added value of individualized fludarabine dosing.

Funding. This work was supported by Foundation Children Cancerfree (KiKa) project number 190. T.D. was personally supported by a Dutch Research Council (NWO)/ZonMw Veni grant (project no. 91617140).

Conflict of Interest. J.L. works at Pharmetheus AB as a consultant for various pharmaceutical companies (www.pharmetheus.com). All other authors declare no competing interests for this work.

Author Responsibility. All authors reviewed and approved the final version of the manuscript. J.L., T.D., C.vK., J.J.B., and A.H. had full access to all data in the study and take full responsibility for the integrity of data and the accuracy of analysis.

Author Contributions. J.B.L., T.P.C.D., C.vK., S.N., and A.D.R.H. wrote the manuscript. J.B.L., T.P.C.D., C.vK., E.M.vM., M.A.dW., J.J.B., and A.D.R.H. designed the research. J.B.L., T.P.C.D., C.vK., and A.D.R.H. analyzed the data.

1. Hahn, T. *et al*. Significant improvement in survival after allogeneic hematopoietic cell transplantation during a period of significantly increased use, older recipient age, and use of unrelated donors. *J. Clin. Oncol.* **31**, 2437–2449 (2013).
2. Gooley, T.A. *et al*. Reduced mortality after allogeneic hematopoietic-cell transplantation. *N. Engl. J. Med.* **363**, 2091–2101 (2010).
3. Admiraal, R. *et al*. Association between anti-thymocyte globulin exposure and survival outcomes in adult unrelated haemopoietic cell transplantation: a multicentre, retrospective, pharmacodynamic cohort analysis. *Lancet Haematol.* **4**, e183–e191 (2017).
4. Bartelink, I.H. *et al*. Association of busulfan exposure with survival and toxicity after haemopoietic cell transplantation in children and young adults: a multicentre, retrospective cohort analysis. *Lancet Haematol.* **3**, e526–e536 (2016).
5. Ivaturi, V. *et al*. Pharmacokinetics and model-based dosing to optimize fludarabine therapy in pediatric hematopoietic cell transplant recipients. *Biol. Blood Marrow Transplant.* **23**, 1701–1713 (2017).
6. Andersson, B.S. *et al*. Fludarabine with pharmacokinetically guided IV busulfan is superior to fixed-dose delivery in pretransplant conditioning of AML/MDS patients. *Bone Marrow Transplant.* **52**, 580–587 (2017).
7. Langenhorst, J.B. *et al*. Population pharmacokinetics of fludarabine in children and adults during conditioning prior to allogeneic hematopoietic cell transplantation. *Clin. Pharmacokinet.* **58**, 627–637 (2019).
8. Long-Boyle, J.R. *et al*. High fludarabine exposure and relationship with treatment-related mortality after nonmyeloablative hematopoietic cell transplantation. *Bone Marrow Transplant.* **46**, 20–26 (2011).
9. Sanghavi, K. *et al*. Personalized fludarabine dosing to reduce nonrelapse mortality in hematopoietic stem-cell transplant recipients receiving reduced intensity conditioning. *Transl. Res.* **175**, 103–115.e4 (2016).
10. Langenhorst, J.B. *et al*. Fludarabine exposure in the conditioning prior to allogeneic hematopoietic cell transplantation predicts outcomes. *Blood Adv.* **3**, 2179–2187 (2019).
11. Center for International Blood and Marrow Transplant, a contractor for the C.W. Bill Young Cell Transplantation Program operated through the U. S. Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau. Number of HCTs performed in the United States and reported to CIBMTR by year and donor type. <https://bloodcell.transplant.hrsa.gov/research/transplant_data/transplant_activity_report/year-donor_type.pdf>. Accessed February 26, 2018.
12. Dhakal, B., Vesole, D.H. & Hari, P.N. Allogeneic stem cell transplantation for multiple myeloma: is there a future? *Bone Marrow Transplant.* **51**, 492–500 (2016).
13. Yanada, M. *et al*. Allogeneic myeloablative transplantation for patients aged 50 years and over. *Bone Marrow Transplant.* **34**, 29–35 (2004).
14. Admiraal, R. *et al*. Association between anti-thymocyte globulin exposure and CD4+ immune reconstitution in paediatric haemopoietic cell transplantation: a multicentre, retrospective pharmacodynamic cohort analysis. *Lancet Haematol.* **2**, e194–e203 (2015).
15. Milano, F. *et al*. Cord-blood transplantation in patients with minimal residual disease. *N. Engl. J. Med.* **375**, 944–953 (2016).
16. Murata, M. *et al*. Allogeneic transplantation for primary myelofibrosis with BM, peripheral blood or umbilical cord blood: an analysis of the JSHCT. *Bone Marrow Transplant.* **49**, 355–360 (2014).
17. Ljungman, P. The role of cytomegalovirus serostatus on outcome of hematopoietic stem cell transplantation. *Curr. Opin. Hematol.* **21**, 466–469 (2014).
18. Wahid, S.F.A. Indications and outcomes of reduced-toxicity hematopoietic stem cell transplantation in adult patients with hematological malignancies. *Int. J. Hematol.* **97**, 581–598 (2013).
19. Fielding, A.K. *et al*. Outcome of 609 adults after relapse of acute lymphoblastic leukemia (ALL); an MRC UKALL12/ECOG 2993 study. *Blood* **109**, 944–950 (2007).
20. Poon, L.M. *et al*. Outcomes of second allogeneic hematopoietic stem cell transplantation for patients with acute lymphoblastic leukemia. *Bone Marrow Transplant.* **48**, 666–670 (2013).

21. Ruutu, T. *et al.* Second allogeneic transplantation for relapse of malignant disease: retrospective analysis of outcome and predictive factors by the EBMT. *Bone Marrow Transplant.* **50**, 1542–1550 (2015).
22. Huo, M.R. *et al.* The effect of HLA disparity on clinical outcome after HLA-haploidentical blood and marrow transplantation. *Clin. Transplant.* **26**, 284–291 (2012).
23. Rondon, G. *et al.* Long-term follow-up of patients who experienced graft failure postallogeneic progenitor cell transplantation. Results of a single institution analysis. *Biol. Blood Marrow Transplant.* **14**, 859–866 (2008).
24. Punt, A.M., Langenhorst, J.B., Egas, A.C., Boelens, J.J., van Kesteren, C. & van Maarseveen, E.M. Simultaneous quantification of busulfan, clofarabine and F-ARA-A using isotope labelled standards and standard addition in plasma by LC-MS/MS for exposure monitoring in hematopoietic cell transplantation conditioning. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **1055–1056**, 81–85 (2017).
25. Olsson, R. *et al.* Graft failure in the modern era of allogeneic hematopoietic SCT. *Bone Marrow Transplant.* **48**, 537–543 (2013).
26. Fedele, R. *et al.* Tolerability and efficacy of busulfan and fludarabine as allogeneic pretransplant conditioning therapy in acute myeloid leukemia: comparison with busulfan and cyclophosphamide regimen. *Clin. Lymphoma Myeloma Leuk.* **14**, 493–500 (2014).
27. Lee, J.H. *et al.* Randomized trial of myeloablative conditioning regimens: busulfan plus cyclophosphamide versus busulfan plus fludarabine. *J. Clin. Oncol.* **31**, 701–709 (2013).
28. Liu, H. *et al.* Busulfan plus fludarabine as a myeloablative conditioning regimen compared with busulfan plus cyclophosphamide for acute myeloid leukemia in first complete remission undergoing allogeneic hematopoietic stem cell transplantation: a prospective and multicenter study. *J. Hematol. Oncol.* **6**, 15 (2013).