**Supplemental Data**

1. Methods
	1. Busulfan concentration assay

Plasma busulfan (BU) concentrations were determined on-site using high-performance liquid chromatography-tandem mass spectrometry (LC-MS/MS; Agilent 1260, Agilent Technologies, CA, USA) equipped with Mass spectroscopic analysis using an AB Sciex API 3200 triple quadrupole MS/MS (Applied Biosystems, Framingham, MA, USA). Twenty microliters of busulfan-d8 (internal standard) were added into the 200 µL plasma sample. Busulfan separation was carried out with a BDS Hypersil C8 column (150 × 4.6 mm; Thermofisher Scientific, Waltham, MA, USA) under isocratic chromatography conditions with a 0.5 mL/min flow rate with a mobile phase A:mobile phase B ratio of 60:40. Mobile phase A consisted of 0.1% formic acid in 0.02 mol/L ammonium acetate; mobile phase B, 0.1% formic acid in acetonitrile. The limit of quantification was 0.04 µg/mL. The coefficient of variation for intraday and interday precision was between 1.8%–9.7% with accuracy of 101.8%–104.5% of nominal values for, respectively, low (600 ng/mL), medium (2700 ng/mL), and high (4800 ng/mL) quality control concentration levels.

* 1. Genotyping

The *GST* genotyping was performed according to a method previously described by Nguyen et al. [1]. The *GST* genotypes investigated included *glutathione S-transferase alpha 1* (*GSTA1)* alleles (rs3957356, G-52A and rs3957357, C-69T, defining haplotypes -52G, -69C as *GSTA1\*A* and -52A, -69T as *GSTA1\*B*), *GSTM1* copy number variations and *GSTP1* A313G polymorphism (rs1695). The *GSTM1* and *GSTP1* genes were amplified by a TaqMan genotyping assay (Applied Biosystems, Waltham, MA, USA), and the *GSTA1* promoter region explored by Sanger Sequencing (Applied Biosystems, Foster City, CA, USA) as a genotyping service (U2Bio, Bangkok, Thailand).

* 1. Base population model development
1. Parameters were estimated with the first order conditional estimation and interaction (FOCE-I) algorithm using NONMEM 7.4 (ICON Development Solutions, Ellicott City, MD, USA). One- and two-compartment models were evaluated as structural models. Interindividual variability (IIV) and inter-occasion variability (IOV), occasionally defined as a treatment day, were explored using an exponential-error model. Since body size is known to have a strong influence on clearance (CL) and volume (V) parameters in pediatrics, several approaches were compared: body weight (BW) with estimated allometric exponent, fat-free mass with constant estimated allometric exponent [2], and BW with body weight-dependent estimated allometric exponents [3]. Systematic day-to-day changes in pharmacokinetic (PK) parameters were also explored, either as fixed effects or as a time-dependent phenomenon [2]. Additive, proportional, and combined error models were investigated to describe residual variability. Models were compared based on the log-likelihood ratio test and evaluated with standard goodness-of-fit plots and prediction-corrected visual predictive checks (pcVPC) [4]. Parameter estimation precision was assessed with the NONMEM covariance step.

1.4 Development of the covariate model

Categorical covariates were coded: $θ\_{i}=θ\_{TV}∙(θ\_{COV})^{COV\_{i}}$ where $θ\_{i}$ was the individual parameter value, $θ\_{TV}$ the typical parameter value, $θ\_{COV}$ the covariate effect to be estimated and $COV\_{i}$ the individual covariate value equal to 0 or 1. Continuous covariates were coded with an allometric function:$ θ\_{i}=θ\_{TV}∙(\frac{COV\_{i}}{median(COV)})^{θ\_{COV}}$ where $median(COV)$ is the median value of the covariate in the population and $COV\_{i}$ a continuous variable. The SCM procedure was conducted with Pearl-Speaks-NONMEM [5]. Statistical significance was defined, respectively, at p < 0.05 and p < 0.01 during forward and backward procedures.

1.5 Internal validation procedure (k-fold)

The capacity of the formulas to predict clearance at day 1 was evaluated by comparing the formula-predicted individual CL to the reference individual CL. The latter was defined as the MAP-Bayesian estimate from the base model in those patients with a day 1 exploration availableand with every available concentration considered. Precision and bias were assessed, respectively, by root-mean square error (RMSE) and mean percentage error (MPE).

$$RMSE=\sqrt{\frac{1}{n}∙\sum\_{i}^{n}(PRED\_{i}-OBS\_{i})^{2} }$$

$$MPE(\%)= 100∙\frac{1}{n} \sum\_{i}^{n}(PRED\_{i}-OBS\_{i})/OBS\_{i}$$

Formulas were internally validated with a k-fold cross-validation procedure. Data was split in k = 10 parts and the following procedure was applied ten times: coefficients were re-estimated in NONMEM from 90% of the initial data, applied to predict CL of the remaining 10%, and compared to the reference clearance. The mean of the 10 RMSE and MPE obtained was then compared to the RMSE and MPE obtained from the original data.

1. Results
	1. Internal validation of the formula

The predictive performances of these formulas were RMSE = 2.22 L/h and MPE = +3.5% for the first (n = 133 patients) and RMSE = 1.91 L/h and MPE = +2.4% for the second (n = 113 patients). The internal k-fold cross-validation procedure found consistent estimates of the formula coefficients, ranging within a +/- 4% interval around the value corresponding to the analysis of the whole dataset (Supplementary Figure 1). Consistently, mean RMSE and MPE were close to the original, with RMSE = 2.06 L/h and MPE = +3.3% for the full-cohort equation, and RMSE = 1.82 L/h and MPE = +2.5% for the genetics cohort equation.



**Supplementary Figure 1***. Parameter estimates of the typical value of clearance (TVCL), and of the effects of body weight (BW\_CL), day of treatment (DAY\_CLV), malignancy (MALIGN\_CL) and GSTA1 mutation (GSTA1\_CL) on clearance, obtained during the K-fold procedure. Values are relative to original estimates. Upper panel: full cohort covariate model. Lower panel: pharmacogenetics cohort model.*

Bibliography

1. Nguyen AH, Biswas M, Puangpetch A, Prommas S, Pakakasama S, Anurathapan U, et al. Effect of GSTA1 Variants on Busulfan-Based Conditioning Regimen Prior to Allogenic Hematopoietic Stem-Cell Transplantation in Pediatric Asians. *Pharmaceutics*. 2022;14:401.

2. Lawson R, Staatz CE, Fraser CJ, Ramachandran S, Teague L, Mitchell R, et al. Population pharmacokinetic model for once‐daily intravenous busulfan in pediatric subjects describing time‐associated clearance. *CPT Pharmacom & Syst Pharma*. 2022;11:1002–17.

3. Poinsignon V, Faivre L, Nguyen L, Neven B, Broutin S, Moshous D, et al. New dosing nomogram and population pharmacokinetic model for young and very young children receiving busulfan for hematopoietic stem cell transplantation conditioning. Pediatr Blood Cancer [Internet]. 2020 [cited 2023 Feb 27];67. Available from: https://onlinelibrary.wiley.com/doi/10.1002/pbc.28603

4. Bergstrand M, Hooker AC, Wallin JE, Karlsson MO. Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. *AAPS J*. 2011;13:143–51.

5. Lindbom L, Ribbing J, Jonsson EN. Perl-speaks-NONMEM (PsN)—a Perl module for NONMEM related programming. *Comput. Methods Programs Biomed*. 2004;75:85–94.