Prospective Multi-Center Evaluation of the Duration of Therapy for Thrombosis in Children
(the "Kids-DOTT" Trial)

Document Date: 7/11/2019

NCT - NCT00687882

Statistical Analysis Plan

Protocol Title: Kids-DOTT

Sponsor: National Institutes of Health

Version Number: 2.00

Version Date: 11 July 2019

Prepared by: **John Kittelson, PhD, Lead Trial Statistician**

On Behalf of the Kids-DOTT Data Coordinating Center

CPC Clinical Research and University of Colorado

APPROVALS

Author

PhD

Date

CPC Approval

William Hiatt, MD

Date

Sponsor Approval

eilGoldenberg MD

Dat

Committee Approv

Sam Schulman, MD

Date

TABLE OF CONTENTS

Kids-DOTT Randomized Trial Cohort
Statistical Analysis Plan

The Kids-DOTT trial has two components: (1) a randomized controlled trial using a prospective randomized open-label blinded-endpoint (PROBE) design, wherein treatment is not blinded but outcome assessment is blinded, and (2) a parallel cohort comprised of children who are not eligible for randomization to short duration anticoagulation at the 6-week visit due to complete veno-occlusion or persistently positive antiphospholipid antibodies. The purpose of this statistical analysis plan is to describe the statistical analysis for the randomized controlled trial population. The analysis plan for the non-randomized parallel cohorts will be described in a separate document.

The original design of the Kids-DOTT trial was based on historical observational studies which report a 10% annual rate of recurrent venous thromboembolism (VTE) and a 10% annual rate of clinically relevant bleeding. In the first half of the trial the risk has been much lower than the original design assumptions; that is, the observed risk is approximately 3% for recurrent VTE and 1% for the bleeding endpoint. *This analysis plan updates initial plans to provide a revised bivariate decision curve. This analysis plan does not include an interim analysis because the reduced risk for the VTE and bleeding endpoints does not provide sufficient information to support an interim analysis prior to trial completion.*

1. OVERVIEW OF STATISTICAL DESIGN FOR THE Kids-DOTT RANDOMIZED TRIAL

1.1 Design overview:

(a) Original trial design: The primary hypothesis addresses non-inferiority (NI) of shortened-duration anticoagulation when compared with conventional-duration therapy. The difference between short and conventional-duration treatment is measured by the absolute risk reduction (ARR) for the symptomatic recurrent VTE (srVTE) (primary efficacy) and clinically-relevant bleeding (primary safety) endpoints. Based upon historical observational data in pediatric srVTE, and consistent with the proposed study population, the trial design assumed a 10% risk of srVTE within 1 year of randomization in patients receiving conventional anticoagulation, and a 10% risk of clinically relevant bleeding (including 2% risk of major bleeding and 20% risk of clinically relevant non-major bleeding). The trial was designed to enroll 815 participants in order to have 570 (85%) in the randomized comparison of primary efficacy and safety events.

(b) Updated trial design: This analysis plan updates the original design to accommodate the lower risk for the VTE and bleeding endpoints. *The updated design retains the original definition of the bivariate NI bound. The bivariate decision curve is updated to reflect the changes in the confidence interval widths for the ARR due to the lower VTE and bleeding risks that have been observed in the first half of the trial.*

As in the original design, the updated design is based on a bivariate decision curve, which formalizes the need to achieve a favorable balance between risk of srVTE and clinically relevant bleeding. This objective is even more critical in the context of the vulnerability of the pediatric population. Given this and given the logical hypothesis that bleeding will be reduced in children receiving shortened-duration anticoagulation, the updated statistical approach continues to employ a bivariate endpoint analysis (i.e., trade-off between srVTE and clinically relevant bleeding risks as described below). The bivariate decision curve is specified using 12-month risk difference or ARR for the srVTE and clinically relevant bleeding outcomes. The updated decision curve follows from the original decision curve as described in Section 1.3.

1.2 Statistical decision criteria using the bivariate (recurrent VTE:bleeding) outcome

1.2.1 Decision criteria for ARR in recurrent VTE and bleeding endpoints

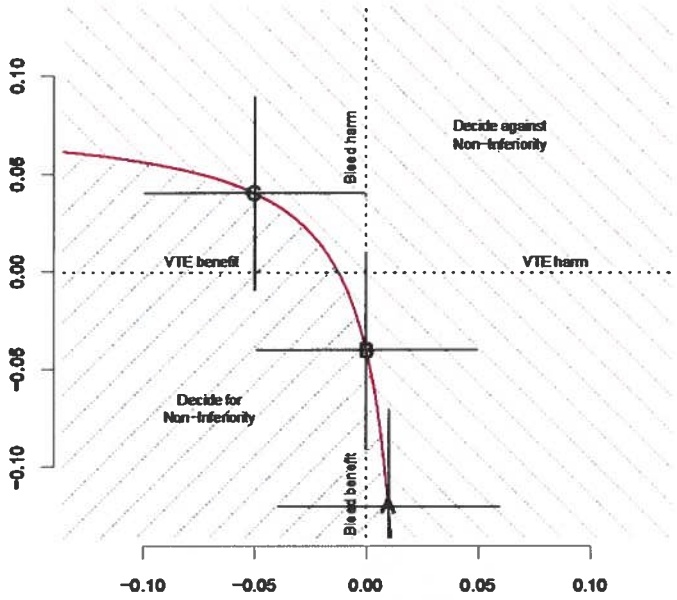
A bivariate recurrent VTE:bleeding tradeoff curve for treatment effects measured by ARR was selected to have appropriate bivariate inference at 3 reference design points (Table 1, below): point A represents an important reduction in bleeding risk with a slight increase in srVTE risk, point B represents a smaller reduction in clinically relevant bleeding risk with no effect on srVTE risk, and point C indicates a slight increase in clinically relevant bleeding risk that is offset by a significant reduction in srVTE risk. As in any trial design, such reference points are selected because they represent effects that should engender appropriate decisions and statistical inference. Specifically, if there is evidence of a highly significant reduction in clinically relevant bleeding risk for the study intervention (shortened-duration anticoagulation), then a non-inferiority conclusion can be defended even if there is a slight increase in srVTE risk (point A: -12% bleeding risk reduction [95% confidence interval: -0.169 to -0.071] and 1% recurrent VTE risk increase [95% confidence interval: -0.039 to 0.059). However, closer to the origin (point B), non-inferiority is decided when there is no difference in symptomatic recurrent VTE risk, but a slight reduction in clinically relevant bleeding risk. Although unlikely, point C represents a significant reduction in srVTE risk that offsets a 4% increase in clinically relevant bleeding risk.

Table 1. Statistical inference at the 3 reference design points on the Kids-DOTT decision curve assuming 285 subjects per group and 10% event risk.

Hypothetical	VTE risk difference		Bleed risk difference	
Result	Observed	95% CI	Observed	95% CI
A	0.01	(-0.039, 0.059)	-0.12	(-0.169, -0.071)
B	0.00	(-0.049, 0.049)	-0.04	(-0.089, 0.009)
C	-0.05	(-0.099, -0.001)	0.04	(-0.009, 0.089)

The methods of Kittelson et al. [1,2] are used to create a smooth curve through the three reference points to define bivariate decision criteria is given in Appendix equation (1) and is plotted in Figure 1.

Figure 1. Protocol-specified bivariate decision criteria for ARR for risk of srVTE and clinically relevant bleeding (Appendix equation 1).

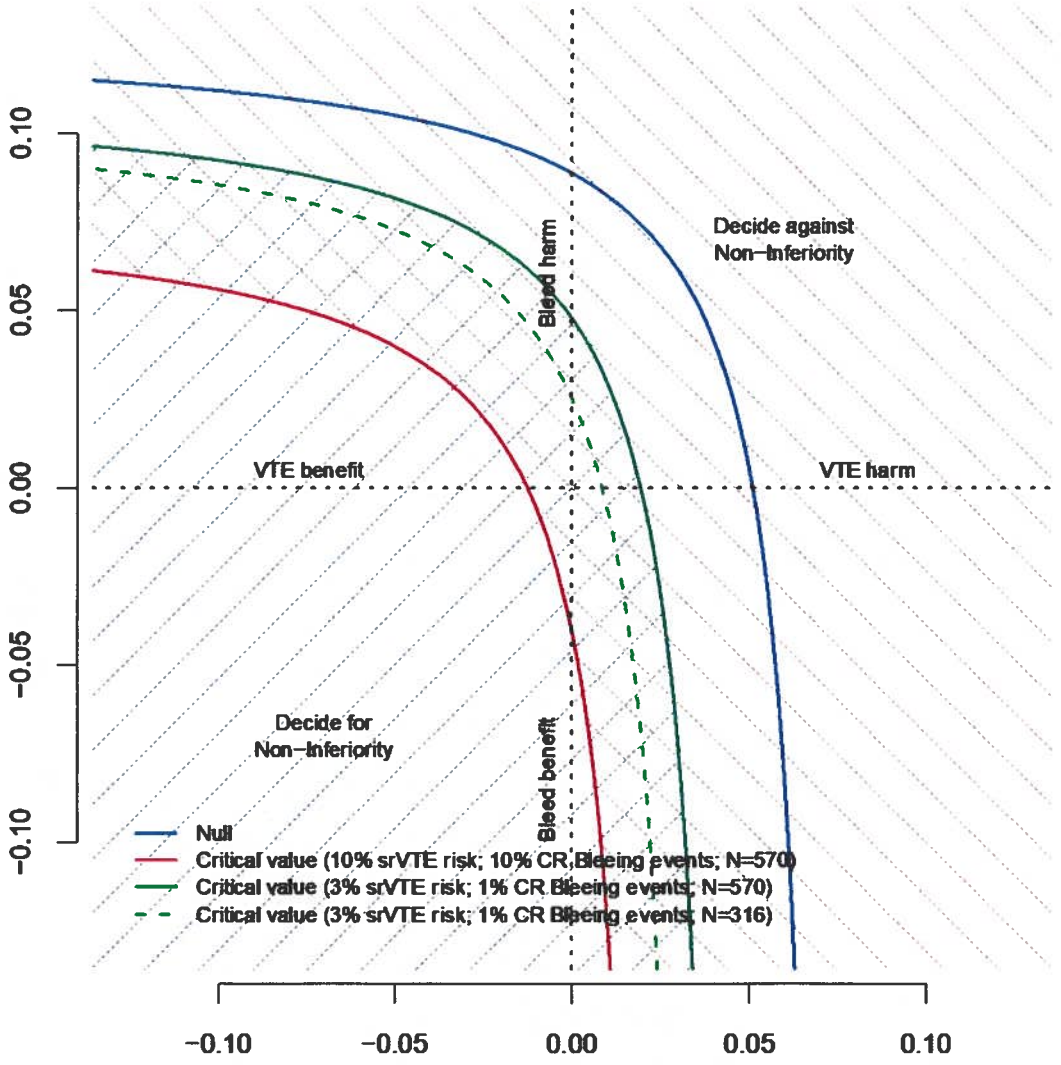


1.3 Bivariate non-inferiority bound and trial power

The decision curve of equation (1) and Figure 1 rejects the NI bound of equation (2) and is constructed under the assumption that srVTE and clinically relevant bleeding risks are both 10%. The curve must be shifted relative to the NI bound if the risks differ from 10%. Appendix equation (3) gives the equation for shifting the decision curve and is illustrated in Figure 2.

1.3.1 Bivariate null hypothesis

Figure 2: Bivariate decision curve with lower event risk



A bivariate NI bound is needed in order to determine how the bivariate decision curve should shift if the observed variance differs from the pre-trial design. Specifically, the original bivariate decision curve of Figure 1 is defined under the design assumption that the 12-month risks of srVTE and clinically relevant bleeding are both 10%. If the observed risks differ from these values, then the decision curve should shift relative to the NI bound. Equivalently, the standardized *Z*-score that is used to rule out the null curve should be the same regardless of the variance. The equations that define the NI bound and the final decision curve are given in the Appendix equations 1-3. Figure 2 shows the NI bound (blue line), the decision curve of Figure 1 (red line), and how that curve would shift if the 12-month srVTE event risk is 3% instead of 10%, and if the 12-month clinically relevant bleeding risk is 1% instead of 10%

(solid green line with 570 patients in the primary analysis population; dashed green line with 316 patients in the primary analysis population). The shift occurs because the variance of the ARR gets smaller as the risk gets closer to 0. The result is that with smaller variance the observed recurrent VTE-bleeding differences in the cross-hatched region would now lead to an NI decision. **Hence, with the lower variance associated with observed lower event risks than had been originally assumed based on historical data, there is a larger region of non-inferiority decision in the bivariate decision model, even if proposing a primary analysis population size that is smaller than initially proposed.** Power is therefore discussed further below.

1.3.2 Trial power

With bivariate decision criteria, the power of the trial must be evaluated using a bivariate probability distribution. Table 2 shows the power for a non-inferiority decision of the bivariate design at the standardized mean treatment effects on srVTE and clinically relevant bleeding. Table 3 shows the transformation of the standardized mean treatment effects into the ARR scale using different standard errors. The power from Table 2 applies to the transformed effects in Table 3.

Table 2: Power under the bivariate decision rule for standardized true mean effects

VTE Risk	standardized effect on bleeding (δ_B)								
Difference (δ_V)	-1.96	-1.64	-1.28	-0.84	0.00	0.84	1.28	1.64	1.96
-1.96	0.979	0.955	0.909	0.813	0.512	0.203	0.103	0.050	0.024
-1.64	0.974	0.949	0.896	0.793	0.482	0.185	0.089	0.043	0.021
-1.28	0.966	0.938	0.875	0.765	0.450	0.160	0.076	0.036	0.016
-0.84	0.951	0.913	0.842	0.717	0.394	0.132	0.059	0.027	0.012
0.00	0.877	0.813	0.720	0.570	0.268	0.076	0.031	0.013	0.006
0.84	0.690	0.612	0.503	0.363	0.144	0.033	0.012	0.005	0.002
1.28	0.545	0.466	0.366	0.254	0.086	0.017	0.006	0.002	0.001
1.64	0.411	0.345	0.262	0.170	0.054	0.010	0.003	0.001	0.000
1.96	0.303	0.247	0.182	0.115	0.032	0.006	0.002	0.001	0.000

Table 3: True mean difference between short and conventional duration therapy in different scales.

Scale	True mean difference on recurrent VTE or bleeding risks								
Standardized	-1.960	-1.645	-1.282	-0.842	0.000	0.842	1.282	1.645	1.960
[1] S_{arr}	-0.049	-0.041	-0.032	-0.021	0.000	0.021	0.032	0.041	0.049
[2] S^V_{arr}	-0.028	-0.024	-0.018	-0.012	0.000	0.012	0.018	0.024	0.028
[3] S^B_{arr}	-0.016	-0.014	-0.011	-0.007	0.000	0.007	0.011	0.014	0.016

[1] Assuming 10% event risk (standardized scale times 0.0251).
[2] Assuming 3% event risk (standardized scale times 0.0143).
[3] Assuming 1% event risk (standardized scale times 0.0083).

As an example, using the decision curve of Appendix equation (1) the trial has 81.3% power under the hypothesis that there is no difference between short and standard duration therapy on the risk of symptomatic VTE recurrence ($\delta_V = 0$) and a 1.64 standardized reduction in the standardized mean effect on clinically relevant bleeding risk ($\delta_B = -1.64$). Assuming that the risk of srVTE is 3% and the true clinically relevant bleeding risk is 1%, then the trial has 81.3% power under the hypothesis that ARR for srVTE risk is 0 and ARR for clinically relevant bleeding risk is reduced by 1.4% (i.e., -1.64 times the standard error for bleeding, 0.0083). A reduction in the risk of srVTE and bleeding gives more power for the NI decision, as evidenced by the fact that the green lines in Figure 2 are shifted to the northeast relative to the red line.

In addition to trial power, it is useful to evaluate the likely width of the 95% confidence interval upon trial completion. Table 4 shows how the trial sample size affects the length for the srVTE and bleeding outcomes. As described above and illustrated Figure 2, the critical value curve is determined by the size of the confidence rectangle (formed from the 95% confidence intervals for bleeding and srVTE). For example, with a total of 570 patients in the primary analysis population (Figure 2 solid green line), the 95% confidence rectangle has total width of 5.6% horizontally and 3.2% vertically. If there are 316 patients in the primary analysis population, then the rectangle is 7.6% horizontally and 4.4% vertically (Figure 2 dashed green line).

Table 4: Length (half-width) of 95% confidence interval as a function of sample size.

Number randomized	**570**	**416**	**354**	**332**	**316**	**292**	**270**
CI half-width for srVTE (3% risk)	**0.028**	**0.033**	**0.036**	**0.037**	**0.038**	**0.039**	**0.041**
CI half-width for bleeding (1% risk)	**0.016**	**0.019**	**0.021**	**0.021**	**0.022**	**0.023**	**0.024**

2. ANALYSIS PLAN

2.1. Primary analysis

2.1.1 Definition of analysis populations

The primary analysis will be based on the per-protocol (PP) analysis for the primary endpoint. The companion intent-to-treat (ITT) analysis will also be completed and reasons for disagreement will be evaluated. A "positive" trial is one in which the primary hypothesis is true for both the PP and ITT analyses.

a) ITT: All randomized patients (i.e., excluding those in the parallel cohorts), analyzed according to randomization assignment.

b) PP: All randomized patients (i.e., excluding those in the parallel cohorts) who satisfy all of the following conditions:

- Completed follow-up through at least the 6 month visit (note that per-protocol, the 6 month visit and subsequent long-term follow-up visits may be conducted via phone).
- Received 80-120% of doses of anticoagulant medication based on protocol-prescribed randomized duration of anticoagulation.
- Had no protocol violations in eligibility criteria.

c) Safety: All randomized patients who have received at least 1 dose of anticoagulant post-enrollment.

2.1.2 Definition of endpoints

Primary efficacy and safety endpoints are defined as:

Efficacy: Occurrence of symptomatic recurrent VTE within 1 year.

Safety: Occurrence of clinically-relevant (major plus clinically-relevant non-major) bleeding within 1 year.

Treatment effects will be measured by the ARR for the efficacy and safety events over the 12-months interval using the centrally adjudicated events.

2.1.3 Analysis methods for primary endpoints

The primary efficacy analysis will estimate/test the ARR for clinically relevant bleeding and symptomatic VTE recurrence. The ARR and 95% confidence intervals will be estimated using the exact methods of Fay et al.[3] using the R package exact2x2. Non-inferiority will be decided if the 95% confidence rectangle rules out the non-inferiority curve defined by Appendix equation (2). Note that the exact confidence intervals of Fay et al. will result in a discrete decision curve instead of the smooth curve illustrated by the lines in Figure 2. The primary analysis will use data from the PP population, but the ITT analysis will also be conducted and reported as confirmatory.

2.2. Secondary analyses

2.2.1 Secondary endpoints

The secondary efficacy endpoints include:

1) Occurrence of symptomatic recurrent VTE or development of post-thrombotic syndrome (PTS) (composite endpoint) within 1 year;

2) Occurrence of symptomatic recurrent VTE within 2 years;

3) Development of PTS within 1 year;

4) Development of PTS within 2 years.

Post-thrombotic syndrome can only occur in subjects in which the index VTE occurred in a lower or upper extremity or inferior/superior vena cava; the secondary efficacy analyses that involve PTS (including the composite secondary efficacy endpoint) will therefore be restricted to those subjects. The definition of PTS will be based upon the presence of objective signs of chronic venous insufficiency, using the CEAP component of the Manco-Johnson Instrument, as described in the recommendations for endpoint definitions for pediatric VTE trials from the International Society on Thrombosis and Haemostasis [4].

2.2.2 Analysis methods for secondary endpoints

The main secondary efficacy analysis will compare the risk of the composite of symptomatic recurrent VTE or PTS within 12-months of the index VTE event in subjects with a lower or upper extremity (including vena cava) index deep vein thrombosis event (referred to as the "PTS/VTE" composite). As in the primary analysis the risks within each treatment group will estimated using the exact methods of Fay et al.[3] as implemented in the R package exact 2x2.

Risk difference (ARR) for the secondary efficacy endpoint will be calculated in both the ITT and PP analysis populations (defined in Section 2.1.1). The risk difference for the secondary safety endpoint will also be calculated in the safety population. The risk difference for the secondary efficacy endpoints will be calculated across all stratification variables (i.e., age group, thrombus location group), and separately within each level of the stratification variables. The differences between groups and 95% confidence intervals will be reported separately for each comparison. Non-inferiority will be judged by the estimated ARR and 95% confidence interval; there will be no formal bivariate decision curve for the secondary endpoint analysis.

2.2.3 Missing data

The PTS component of the PTS/recurrent VTE composite endpoint is evaluated at the 12-month visit and participants who do not complete the 12-month visit will be missing the PTS component of the composite endpoint. The analysis of section 2.2.2 will be performed using data only from subjects who have completed the 12-month visit and have both PTS and recurrent VTE assessment without imputation for missing data.

2.3 Analysis of tertiary and exploratory endpoints

2.3.1 Analysis of subgroups

If there are a sufficient number of events, the ARR for the primary and secondary efficacy and safety endpoints will be calculated separately within each level of the stratification variables.

3. APPENDIX: Bivariate decision criteria

The bivariate decision curve for the design of Kids-DOTT was derived from the 3 reference points in Table 1 that were selected to capture a clinically meaningful balance between the potential benefits of symptomatic recurrent VTE risk reduction and the potential for increased clinically relevant bleeding risk. The bivariate decision curve of Figure 1 is given by equation (1) below, which is based on the standard error of the ARR (S_{arr}) assuming a 10% risk for recurrent VTE and a 10% risk for clinically relevant bleeding. The bivariate NI bound is given by equation (2), which is the curve that is shifted by 1.96 standard errors to the northeast of the decision curve. If the bleeding and VTE risks differ from the original assumptions, then the revised decision curve shifted by 1.96 standard errors to the southwest of the NI bound (equation 3). The shifted decision curve (Figure 2, green line) is given by equation (3) using a 3% risk of symptomatic recurrent VTE (π_V= 0.03) and a 1% risk of clinically relevant bleeding (π_B= 0.01).

$$\text{ARR decision curve:} \quad (ARR_V - 0.025)(ARR_B - 0.08) = 0.003 \tag{1}$$

$$\text{ARR null curve:} \quad (ARR_V - 0.025 - 1.96S_{arr})(ARR_B - 0.08 - 1.96S_{arr}) = 0.003 \tag{2}$$

$$\text{Shifted decision curve:} \quad (ARR_V - 0.025 - 1.96S_{arr} + 1.96S^V_{arr})(ARR_B - 0.08 - 1.96S_{arr} + 1.96S^B_{arr}) = 0.003 \tag{3}$$

where $S_{arr} = \sqrt{0.1 \times 0.9 \times 2/285} = 0.0251$, $S^V_{arr} = \sqrt{\pi_V \times (1 - \pi_V) \times 2/285}$, and $S^B_{arr} = \sqrt{\pi_B \times (1 - \pi_B) \times 2/285}$, where π_V and π_B denote the 1-year risk of recurrent VTE and clinically relevant bleeding, respectively.

REFERENCES

1. Kittelson JM, Spyropoulos AC, Halperin JL, Kessler CM, Schulman S, Steg G, Turpie AG, Cutler NR, Hiatt WR, Goldenberg NA; Antithrombotic Trials Leadership and Steering (ATLAS) Group. Balancing risk and benefit in venous thromboembolism trials: concept for a bivariate endpoint trial design and analytic approach. *J Thromb Haemost*. 2013; 11(8): 1443-8.

2. Kittelson JM, Steg PG, Halperin JL, Goldenberg NA, Schulman S, Spyropoulos AC, Kessler CM, Turpie AG, Cutler NR, Hiatt WR; Antithrombotic Trials Leadership and Steering (ATLAS) Group. Bivariate evaluation of thromboembolism and bleeding in clinical trials of anticoagulants in patients with atrial fibrillation. *Thromb Haemost*. 2016; 116(3): 544-53.

3. Fay MR, Proschan MA, and Brittain E. Combining one-sample confidence procedures for inference in the two-sample case. *Biometrics* 2015; 71:146-156.

4. Mitchell LG, Goldenberg NA, Male C, Kenet G, Nowak-Gottl U, Monagle P; Perinatal and Paediatric Haemostasis Subcommittee of the SSC of the ISTH. Definition of clinical efficacy and safety outcomes for clinical trials in deep venous thrombosis and pulmonary embolism in children. *J Thromb Haemost* 2011; 9(9):1856-8.