Robust Methods for Assessment of Average and Scaled Average Bioequivalence

Robert Schall[1] & Divan Aristo Burger[2]

October 2019

[1] *Dept. of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa*

[2] *Department of Statistics, University of Pretoria, Pretoria, South Africa*

Objectives

1. Propose a flexible robust methodology for bioequivalence
2. Evaluate/motivate need for robust methods in bioequivalence
3. Compare robust and conventional methods empirically:
   - Apply methodology to a reasonably large data pool of conventional bioequivalence studies
   - Apply methodology to a number of replicate design bioequivalence studies

Do the above for:

- Conventional average bioequivalence assessment
  - Data from typical 2×2 crossover studies
- Reference-scaled average bioequivalence assessment
  - Data from replicate design crossover studies

Objectives

1. Propose a flexible robust methodology for bioequivalence
2. Evaluate/motivate need for robust methods in bioequivalence
3. Compare robust and conventional methods empirically:
   - Apply methodology to a reasonably large data pool of conventional bioequivalence studies
   - Apply methodology to a number of replicate design bioequivalence studies

Do the above for:

- Conventional average bioequivalence assessment
  - Data from typical 2×2 crossover studies
- Reference-scaled average bioequivalence assessment
  - Data from replicate design crossover studies

1. Propose a flexible robust methodology for bioequivalence
2. Evaluate/motivate need for robust methods in bioequivalence
3. Compare robust and conventional methods empirically:
   - Apply methodology to a reasonably large data pool of conventional bioequivalence studies
   - Apply methodology to a number of replicate design bioequivalence studies

Do the above for:

- Conventional average bioequivalence assessment
  - Data from typical 2×2 crossover studies
- Reference-scaled average bioequivalence assessment
  - Data from replicate design crossover studies

Violation of usual assumptions for "normal theory" analysis

- Outliers
- Heavy-tailed distribution (including outliers)
- Skewness of the distribution

Replace the normal distribution by the Student t distribution:

- Accommodates heavy tails/outliers
  - Small degrees of freedom
- Accommodates skewness
  - If skew Student t distribution is used
  - (Only for larger datasets)

Bayesian approach:

- Noninformative priors
- Good frequentist properties
- Robust approach successfully applied even for hierarchical nonlinear models

Burger & Schall (2018). Robust fit of Bayesian mixed effects regression models with application to colony forming unit count in tuberculosis research. *Stat. Med.*

Replace the normal distribution by the Student t distribution:

- Accommodates heavy tails/outliers
  - Small degrees of freedom
- Accommodates skewness
  - If skew Student t distribution is used
  - (Only for larger datasets)

Bayesian approach:

- Noninformative priors
- Good frequentist properties
- Robust approach successfully applied even for hierarchical nonlinear models

Burger & Schall (2018). Robust fit of Bayesian mixed effects regression models with application to colony forming unit count in tuberculosis research. *Stat. Med.*

Replace the normal distribution by the Student t distribution:

- Accommodates heavy tails/outliers
  - Small degrees of freedom
- Accommodates skewness
  - If skew Student t distribution is used
  - (Only for larger datasets)

Bayesian approach:

- Noninformative priors
- Good frequentist properties
- Robust approach successfully applied even for hierarchical nonlinear models

Burger & Schall (2018). Robust fit of Bayesian mixed effects regression models with application to colony forming unit count in tuberculosis research. *Stat. Med.*

Previous literature:

- De Souza et al. (2016):
  - Univariate and bivariate models $\Rightarrow 2 \times 2$ crossover designs
  - Extended generalized gamma distribution
  - Skew Student t distribution
- Ghosh & Ntzoufras (2005):
  - Population and individual bioequivalence
  - Student t

De Souza et al. (2016). The use of asymmetric distributions in average bioequivalence. *Stat. Med.*

Ghosh & Ntzoufras (2005). Testing population and individual bioequivalence: a hierarchical Bayesian approach

Previous literature:

- De Souza et al. (2016):
  - Univariate and bivariate models $\Rightarrow 2 \times 2$ crossover designs
  - Extended generalized gamma distribution
  - Skew Student t distribution
- Ghosh & Ntzoufras (2005):
  - Population and individual bioequivalence
  - Student t

De Souza et al. (2016). The use of asymmetric distributions in average bioequivalence. *Stat. Med.*

Ghosh & Ntzoufras (2005). Testing population and individual bioequivalence: a hierarchical Bayesian approach

Conventional model for standard 2×2 crossover study

$$y_{ij} = \mu + s_i + \zeta_h + \pi_m + \tau_j + e_{ij}$$

- y_{ij} log-bioavailability for subject i and formulation $j = T, R$
- μ overall mean
- s_i random effects (subject)
- ζ_h (sequence), π_m (period), τ_j (treatment): fixed effects
- e_{ij} residual
- $\text{var}(s_i) = \sigma_B^2$ and $\text{var}(e_{ij}) = \sigma_W^2$

Conventional model for standard 2×2 crossover study

$$y_{ij} = \mu + s_i + \zeta_h + \pi_m + \tau_j + e_{ij}$$

- y_{ij} log-bioavailability for subject i and formulation $j = T, R$
- μ overall mean
- s_i random effects (subject)
- ζ_h (sequence), π_m (period), τ_j (treatment): fixed effects
- e_{ij} residual
- $\text{var}(s_i) = \sigma_B^2$ and $\text{var}(e_{ij}) = \sigma_W^2$

Conventional model for standard 2×2 crossover study

$$y_{ij} = \mu + s_i + \zeta_h + \pi_m + \tau_j + e_{ij}$$

- y_{ij} log-bioavailability for subject i and formulation $j = T, R$
- μ overall mean
- s_i random effects (subject)
- ζ_h (sequence), π_m (period), τ_j (treatment): fixed effects
- e_{ij} residual
- $\text{var}(s_i) = \sigma_B^2$ and $\text{var}(e_{ij}) = \sigma_W^2$

Statistical decision rule

Calculate 90% two-sided confidence interval (CI) for $\exp(\tau_T - \tau_R)$

ABE concluded if the CI falls in bioequivalence range 0.80 and 1.25

1. Subject outlier (between-subject outlier):
   - For subject i, both observations, y_{iT} and y_{iR}, are extreme, in the same direction
   - Could be modeled as a mean shift in the subject effect s_i
   - No consequences since mean shift in s_i has no effect on point or interval estimates of relative bioavailability

2. Single data point outlier (within-subject outlier):
   - For subject i, either y_{iT} or y_{iR}, or both (but in opposite directions), is extreme
   - ... within-subject difference $y_{iT} - y_{iR}$ is extreme
   - Can severely affect results of the bioequivalence test

Schall, Ring, Endrenyi (2010). Residuals and outliers in replicate design crossover studies. *J. Biopharm. Stat.*

1. Subject outlier (between-subject outlier):
   - For subject i, both observations, y_{iT} and y_{iR}, are extreme, in the same direction
   - Could be modeled as a mean shift in the subject effect s_i
   - No consequences since mean shift in s_i has no effect on point or interval estimates of relative bioavailability

2. Single data point outlier (within-subject outlier):
   - For subject i, either y_{iT} or y_{iR}, or both (but in opposite directions), is extreme
   - ... within-subject difference $y_{iT} - y_{iR}$ is extreme
   - Can severely affect results of the bioequivalence test

Schall, Ring, Endrenyi (2010). Residuals and outliers in replicate design crossover studies. *J. Biopharm. Stat.*

1. Subject outlier (between-subject outlier):
   - For subject i, both observations, y_{iT} and y_{iR}, are extreme, in the same direction
   - Could be modeled as a mean shift in the subject effect s_i
   - No consequences since mean shift in s_i has no effect on point or interval estimates of relative bioavailability

2. Single data point outlier (within-subject outlier):
   - For subject i, either y_{iT} or y_{iR}, or both (but in opposite directions), is extreme
   - ... within-subject difference $y_{iT} - y_{iR}$ is extreme
   - Can severely affect results of the bioequivalence test

Schall, Ring, Endrenyi (2010). Residuals and outliers in replicate design crossover studies. *J. Biopharm. Stat.*

1. Subject outlier (between-subject outlier):
   - For subject i, both observations, y_{iT} and y_{iR}, are extreme, in the same direction
   - Could be modeled as a mean shift in the subject effect s_i
   - No consequences since mean shift in s_i has no effect on point or interval estimates of relative bioavailability

2. Single data point outlier (within-subject outlier):
   - For subject i, either y_{iT} or y_{iR}, or both (but in opposite directions), is extreme
   - ... within-subject difference $y_{iT} - y_{iR}$ is extreme
   - Can severely affect results of the bioequivalence test

Schall, Ring, Endrenyi (2010). Residuals and outliers in replicate design crossover studies. *J. Biopharm. Stat.*

Robust methodology

- Achieve robustness by:
  1. Use standard model for 2×2 crossover study (see above)
  2. But specify heavy-tailed distributions for residuals (and subject effects)
  3. Here, use Student t distributions with low degrees of freedom
  4. (Preferable to estimate the degrees of freedom, but can also be fixed)
- Fit the model using a Bayesian approach
- **Here:** Compare Student t with normal model (both Bayes)

Robust methodology

- Achieve robustness by:
  1. Use standard model for 2×2 crossover study (see above)
  2. But specify heavy-tailed distributions for residuals (and subject effects)
  3. Here, use Student t distributions with low degrees of freedom
  4. (Preferable to estimate the degrees of freedom, but can also be fixed)
- Fit the model using a Bayesian approach
- **Here:** Compare Student t with normal model (both Bayes)

Robust methodology

- Achieve robustness by:
  1. Use standard model for 2×2 crossover study (see above)
  2. But specify heavy-tailed distributions for residuals (and subject effects)
  3. Here, use Student t distributions with low degrees of freedom
  4. (Preferable to estimate the degrees of freedom, but can also be fixed)

- Fit the model using a Bayesian approach
- **Here:** Compare Student t with normal model (both Bayes)

Robust methodology

- Achieve robustness by:
  1. Use standard model for 2×2 crossover study (see above)
  2. But specify heavy-tailed distributions for residuals (and subject effects)
  3. Here, use Student t distributions with low degrees of freedom
  4. (Preferable to estimate the degrees of freedom, but can also be fixed)
- Fit the model using a Bayesian approach
- **Here:** Compare Student t with normal model (both Bayes)

Residuals and random effects

- Bayes$_T$
  - Student t distribution accommodates heavy tails/outliers
  - ... $e_{ij} \sim t\left(0, \sigma^{2*}_W, \nu_W\right)$ where $\sigma^2_W = \frac{\nu_W}{\nu_W - 2}\sigma^{2*}_W$
  - ... $s_i \sim t\left(0, \sigma^{2*}_B, \nu_B\right)$ where $\sigma^2_B = \frac{\nu_B}{\nu_B - 2}\sigma^{2*}_B$
  - ... ν_W, $\nu_B \Rightarrow$ degrees of freedom

- Bayes$_N$
  - Normal distribution – not robust to outliers
  - ... $e_{ij} \sim \text{Normal}\left(0, \sigma^{2*}_W\right)$ where $\sigma^2_W = \sigma^{2*}_W$
  - ... $s_i \sim \text{Normal}\left(0, \sigma^{2*}_B\right)$ where $\sigma^2_B = \sigma^{2*}_B$

Residuals and random effects

- Bayes$_T$
  - Student t distribution accommodates heavy tails/outliers
  - ... $e_{ij} \sim t\left(0, \sigma_W^{2*}, \nu_W\right)$ where $\sigma_W^2 = \frac{\nu_W}{\nu_W - 2}\sigma_W^{2*}$
  - ... $s_i \sim t\left(0, \sigma_B^{2*}, \nu_B\right)$ where $\sigma_B^2 = \frac{\nu_B}{\nu_B - 2}\sigma_B^{2*}$
  - ... ν_W, $\nu_B \Rightarrow$ degrees of freedom
- Bayes$_N$
  - Normal distribution – not robust to outliers
  - ... $e_{ij} \sim \text{Normal}\left(0, \sigma_W^{2*}\right)$ where $\sigma_W^2 = \sigma_W^{2*}$
  - ... $s_i \sim \text{Normal}\left(0, \sigma_B^{2*}\right)$ where $\sigma_B^2 = \sigma_B^{2*}$

Prior specifications

- Vague priors
- ... μ, ζ_h, π_m, $\tau_j \sim N(0, 10000)$
- ... $\sigma_W^{-2*} \sim \text{Gamma}(0.0001, 0.0001)$
- ... v_W, $v_B \sim \text{Normal}(0, 10000)\, T(2, \infty)$ (half-normal)
- ... $\sigma_B^* \sim t(0, 10000, 2)\, T(0, \infty)$ (half-t)

Implementation

- Fit model using `JAGS` via `R` package `runjags`
- Student t: Mixture of normal & gamma distribution
  - Speeds up convergence
  - Most priors are conjugate $\Rightarrow$ fast convergence
- ABE $\Rightarrow$ calculate:
  - Posterior estimate of $\exp(\tau_T - \tau_R)$
  - 90% highest posterior density (HPD) interval: LCL and UCL
- Deviance information criterion (DIC): Discriminate between Bayes_N & Bayes_T

Implementation

- Fit model using `JAGS` via `R` package `runjags`
- Student t: Mixture of normal & gamma distribution
  - Speeds up convergence
  - Most priors are conjugate $\Rightarrow$ fast convergence
- ABE $\Rightarrow$ calculate:
  - Posterior estimate of $\exp(\tau_T - \tau_R)$
  - 90% highest posterior density (HPD) interval: LCL and UCL
- Deviance information criterion (DIC): Discriminate between Bayes_N & Bayes_T

Implementation

- Fit model using `JAGS` via `R` package `runjags`
- Student t: Mixture of normal & gamma distribution
  - Speeds up convergence
  - Most priors are conjugate $\Rightarrow$ fast convergence
- ABE $\Rightarrow$ calculate:
  - Posterior estimate of $\exp(\tau_T - \tau_R)$
  - 90% highest posterior density (HPD) interval: LCL and UCL
- Deviance information criterion (DIC): Discriminate between Bayes_N & Bayes_T

Implementation

- Fit model using `JAGS` via `R` package `runjags`
- Student t: Mixture of normal & gamma distribution
  - Speeds up convergence
  - Most priors are conjugate $\Rightarrow$ fast convergence
- ABE $\Rightarrow$ calculate:
  - Posterior estimate of $\exp(\tau_T - \tau_R)$
  - 90% highest posterior density (HPD) interval: LCL and UCL
- Deviance information criterion (DIC): Discriminate between Bayes_N & Bayes_T

Pool of datasets from conventional 2×2 bioequivalence studies:

- 65 datasets of T/R comparisons
- ... for both AUC and C_{max}

Schall (2012). The empirical coverage of confidence intervals: Point estimates and confidence intervals for confidence levels. *Biometrical J.*

Need for Robust Methodology: Empirical Study

Fit Bayes$_T$ model to each dataset in the data pool:

- Estimate degrees of freedom of Student t distribution
- Compare with conventional REML & HL methods:
  - Shifts in point and interval estimates
  - Confidence interval lengths

Need for Robust Methodology: Empirical Study

Fit Bayes$_T$ model to each dataset in the data pool:

- Estimate degrees of freedom of Student t distribution
- Compare with conventional REML & HL methods:
  - Shifts in point and interval estimates
  - Confidence interval lengths

Figure: Bayes$_T$: Estimates of Residual DF – AUC

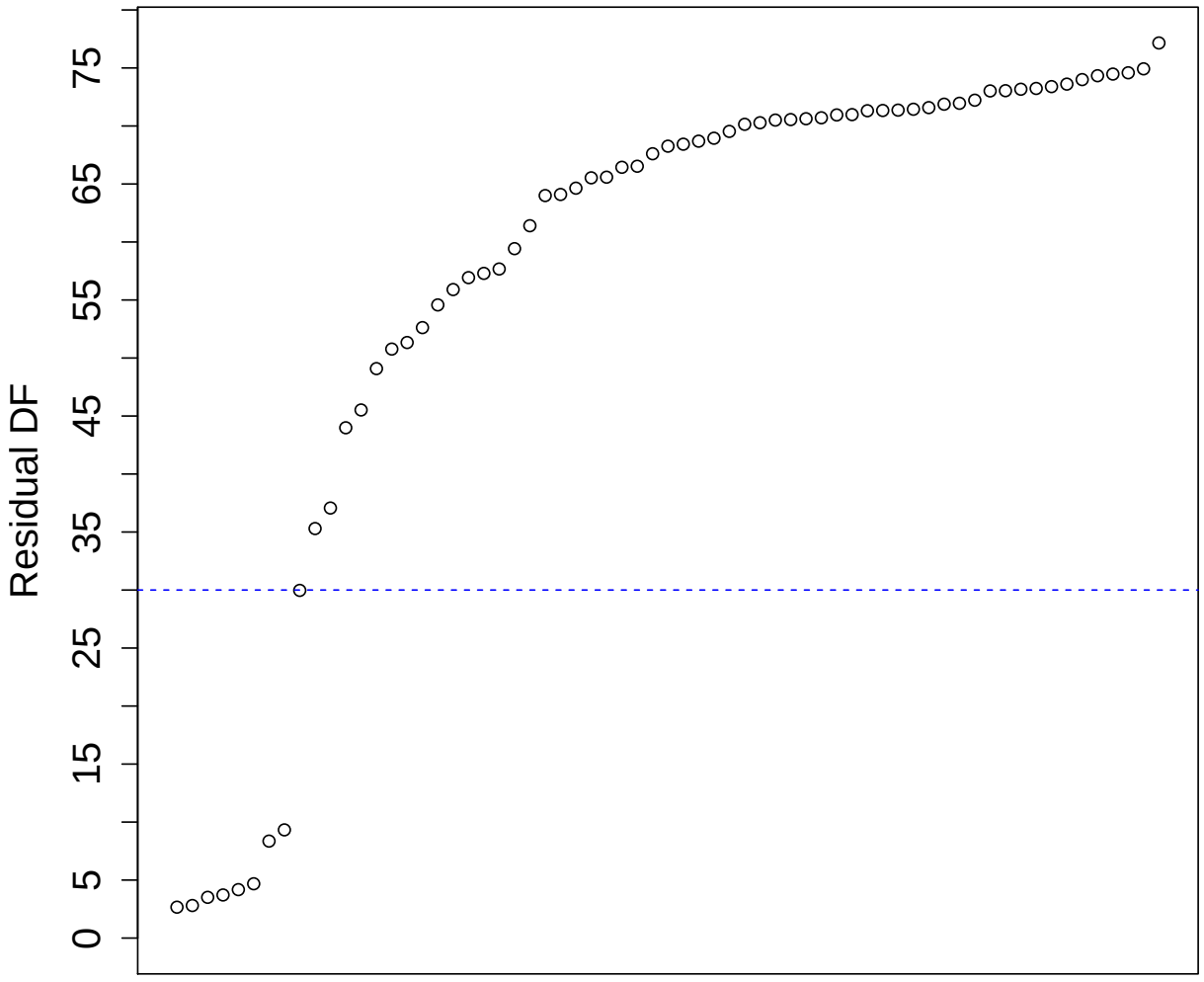


Figure: Bayes$_T$: Estimates of Random DF – AUC

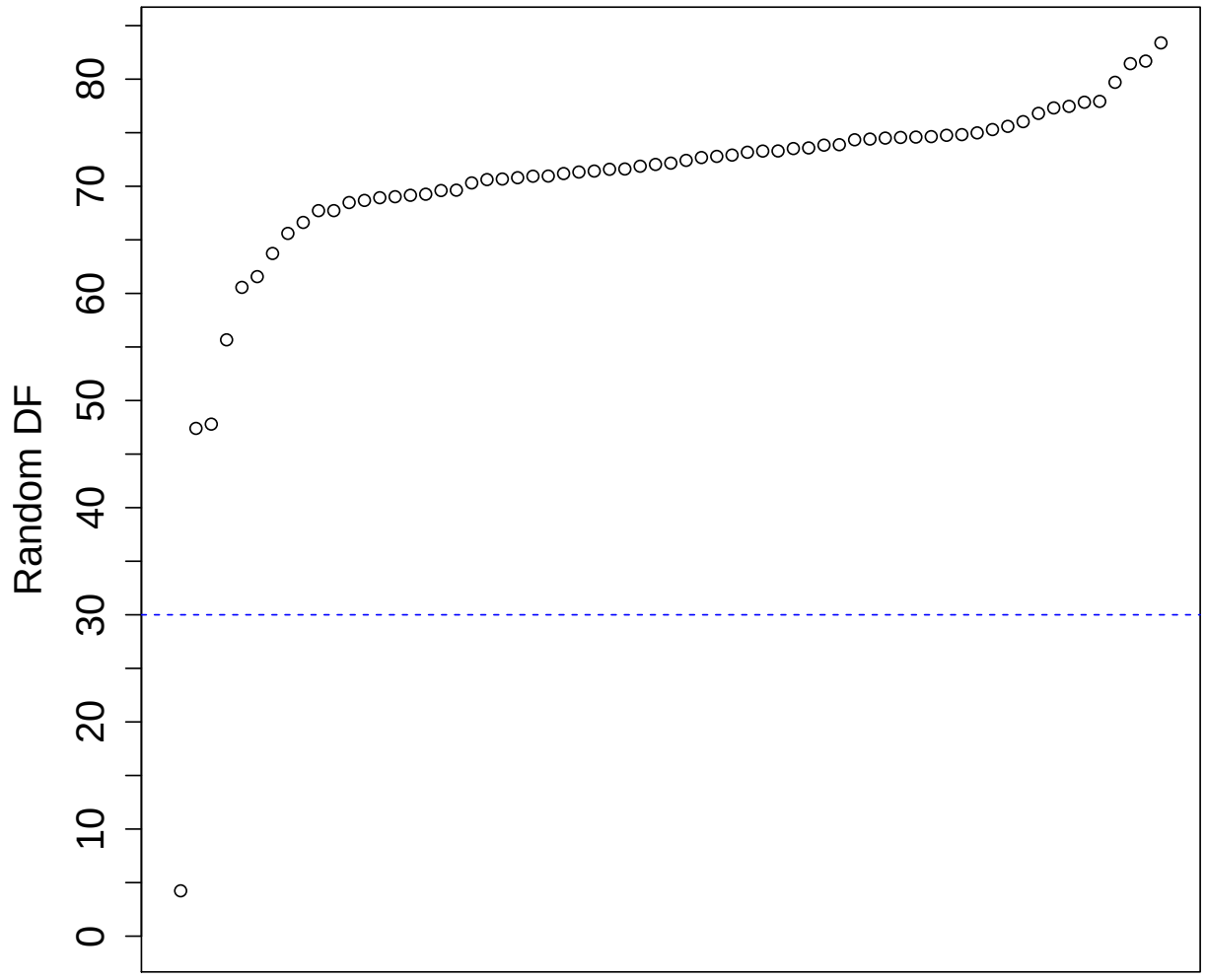


Figure: Bayes_T: Estimates of Residual DF – C_{max}

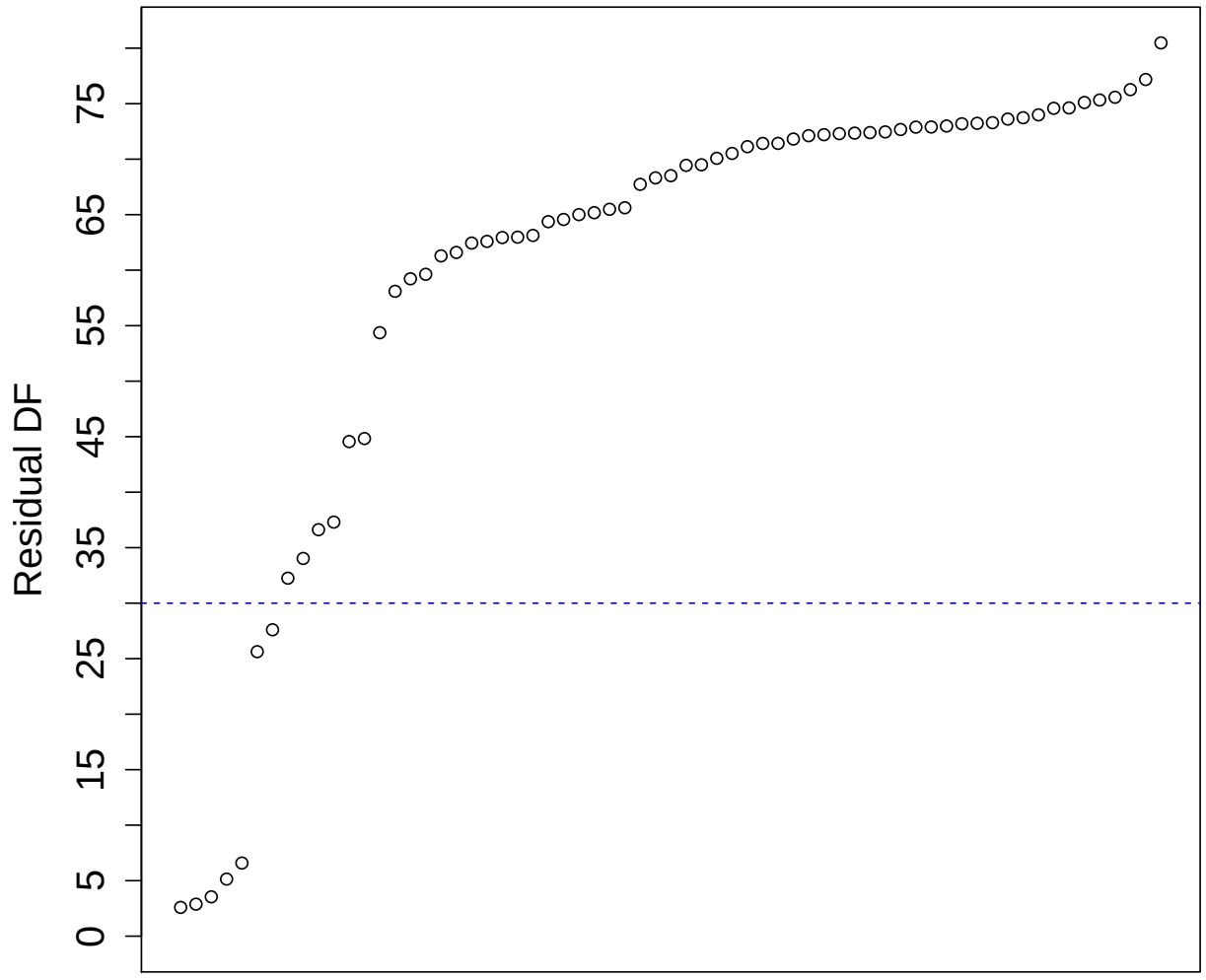
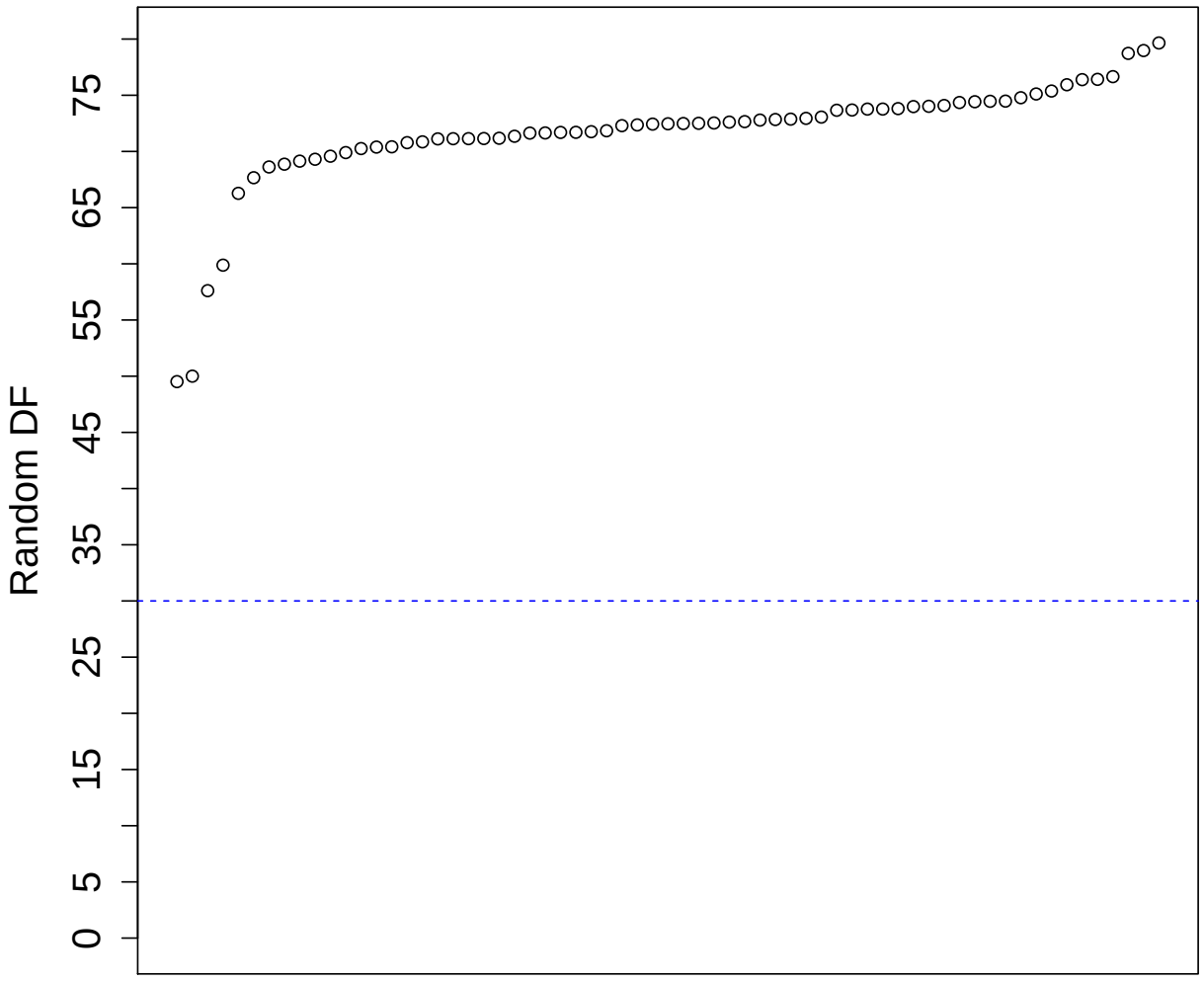


Figure: Bayes$_T$: Estimates of Random DF – C_{max}



Summary (degrees of freedom):

- 8/65 datasets for AUC and 5/65 datasets for C_{max} suggest a "heavy tailed" distribution of the residuals
- Suggests robust methodology might be needed in a small but non-negligible proportion of studies
- (Data pool possibly biased towards "neat" datasets / "successful" studies)
- Heavy tails in subject effect distribution are rare in this data pool (homogeneous subject populations?)
- In any case irrelevant for bioequivalence assessment

Summary (degrees of freedom):

- 8/65 datasets for AUC and 5/65 datasets for C_{max} suggest a "heavy tailed" distribution of the residuals
- Suggests robust methodology might be needed in a small but non-negligible proportion of studies
- (Data pool possibly biased towards "neat" datasets / "successful" studies)
- Heavy tails in subject effect distribution are rare in this data pool (homogeneous subject populations?)
- In any case irrelevant for bioequivalence assessment

Summary (degrees of freedom):

- 8/65 datasets for AUC and 5/65 datasets for C_{max} suggest a "heavy tailed" distribution of the residuals
- Suggests robust methodology might be needed in a small but non-negligible proportion of studies
- (Data pool possibly biased towards "neat" datasets / "successful" studies)

- Heavy tails in subject effect distribution are rare in this data pool (homogeneous subject populations?)
- In any case irrelevant for bioequivalence assessment

Summary (degrees of freedom):

- 8/65 datasets for AUC and 5/65 datasets for C_{max} suggest a "heavy tailed" distribution of the residuals
- Suggests robust methodology might be needed in a small but non-negligible proportion of studies
- (Data pool possibly biased towards "neat" datasets / "successful" studies)

- Heavy tails in subject effect distribution are rare in this data pool (homogeneous subject populations?)
- In any case irrelevant for bioequivalence assessment

Figure: Bayes_T vs REML: eCDF of Shift in Estimates of ABE – AUC

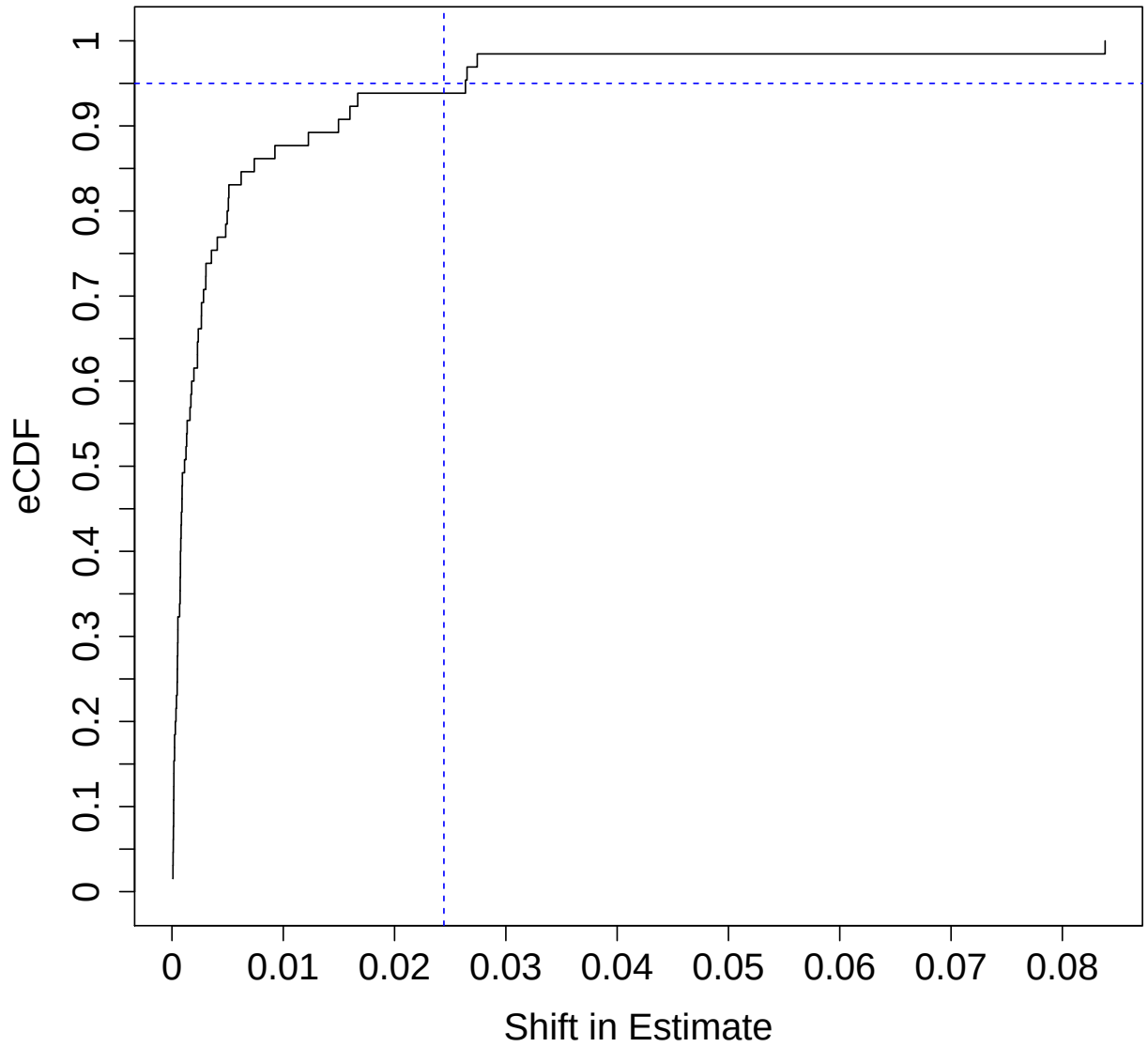


Figure: Bayes$_T$ vs REML: eCDF of Shift in LCLs of ABE – AUC

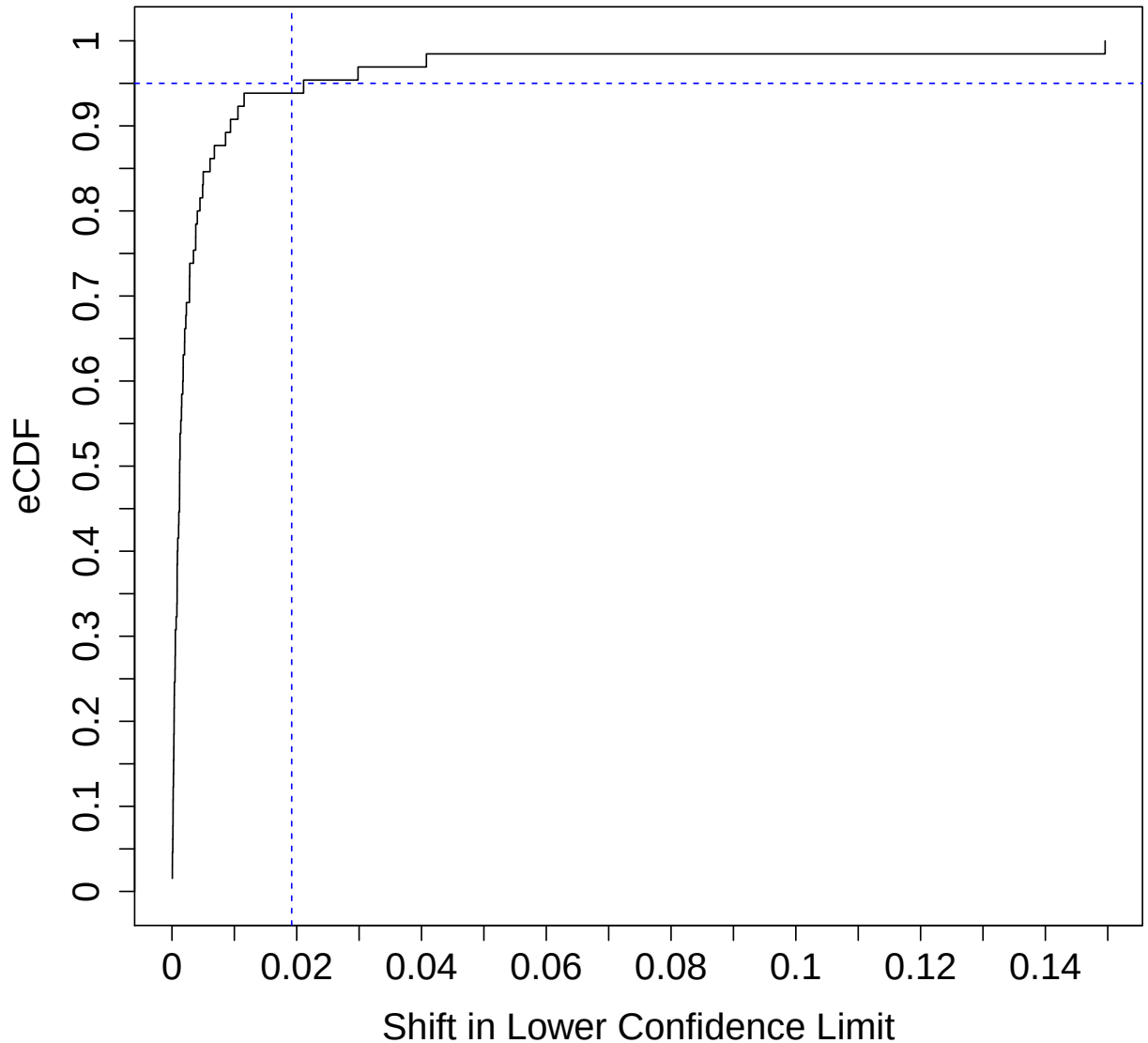


Figure: Bayes_T vs REML: eCDF of Shift in UCLs of ABE – AUC

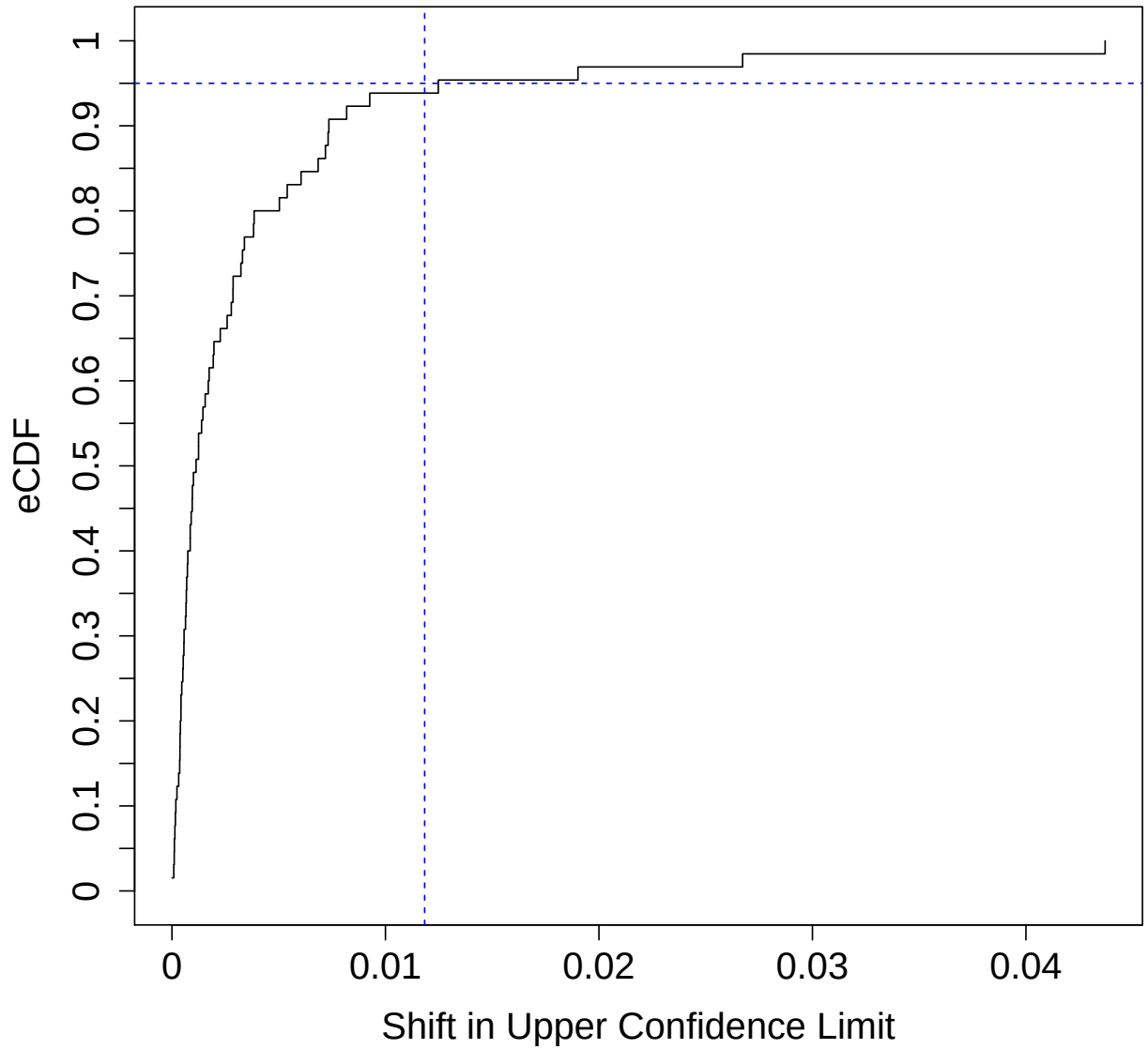


Figure: Bayes_T vs REML: eCDF of Shift in Point Estimate of ABE – C_{max}

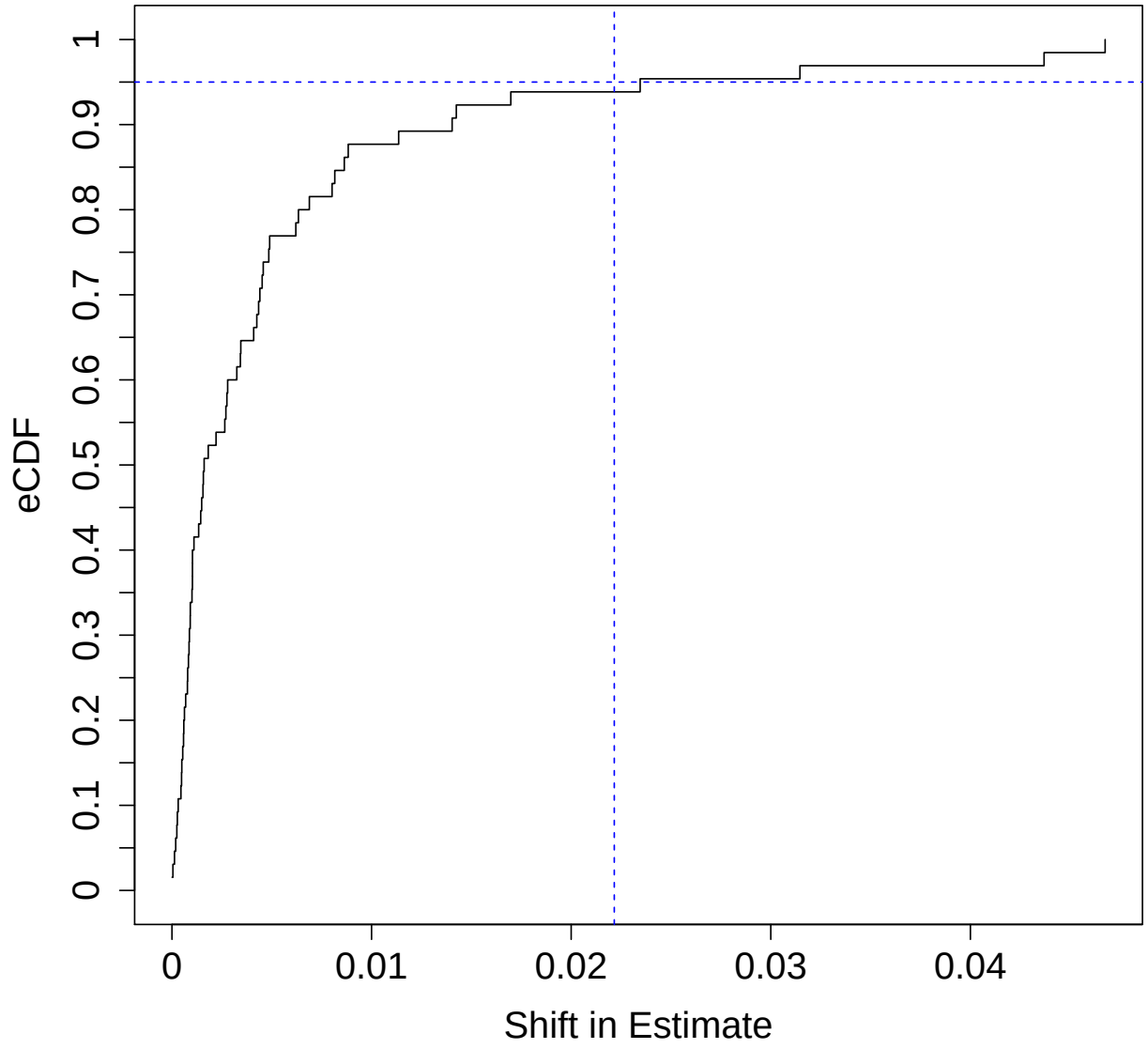


Figure: Bayes_T vs REML: eCDF of Shift in LCL of ABE – C_{max}

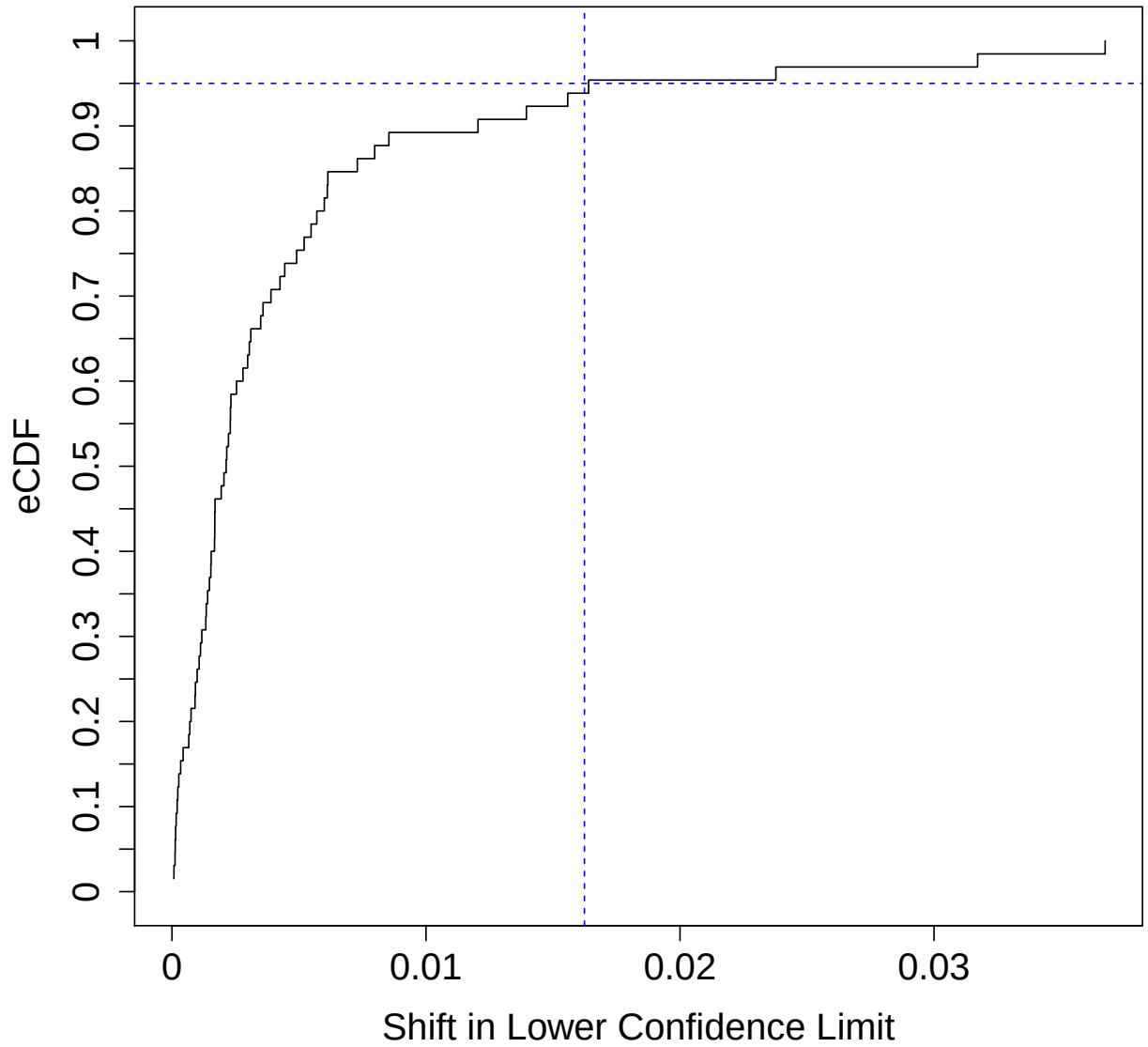
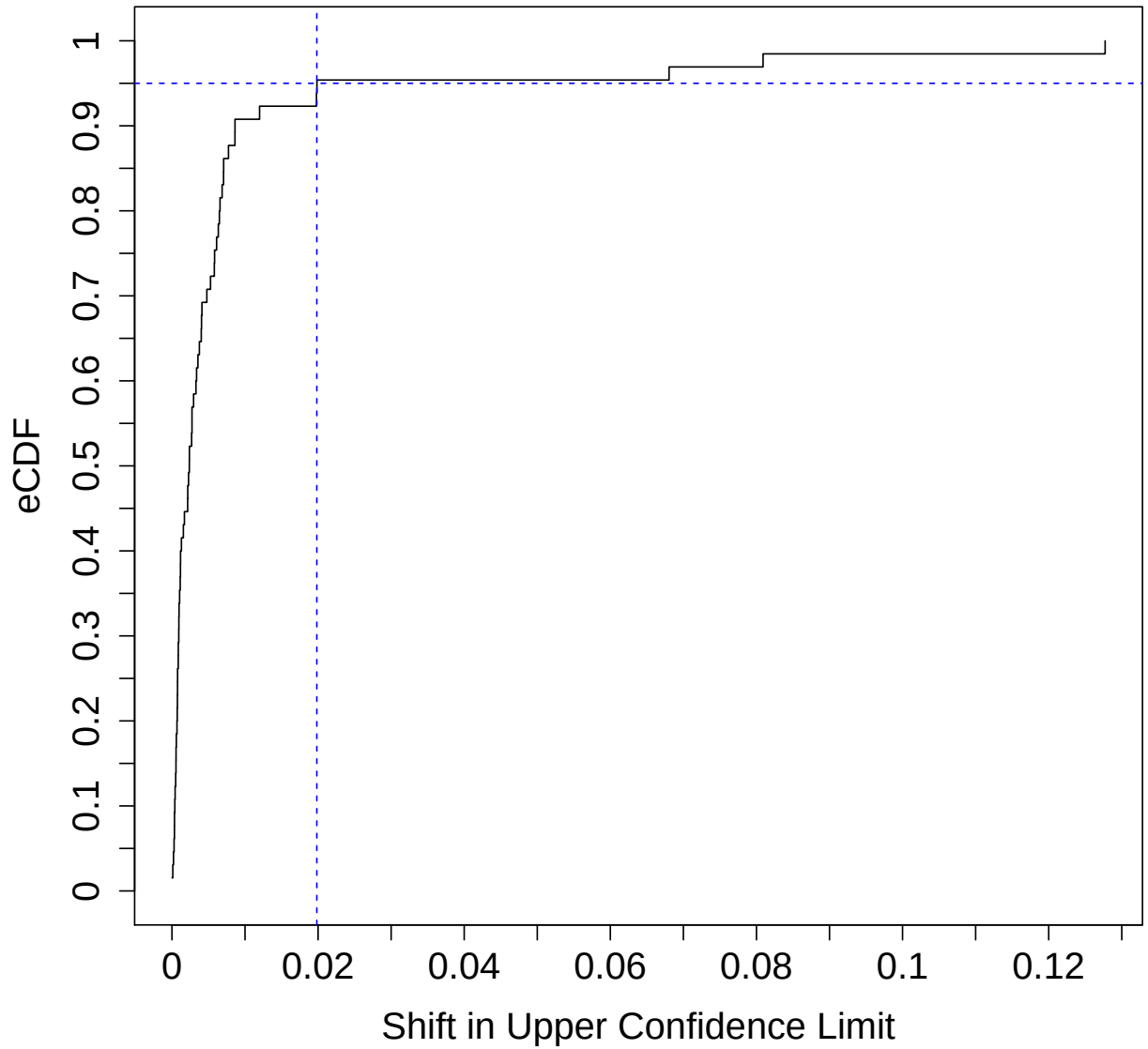


Figure: Bayes_T vs REML: eCDF of Shift in UCL of ABE – C_{max}



Summary (shift in point and interval estimates of GMR):

- In 5% of datasets the point estimate and confidence limits shift by $\pm$ 0.02 units (AUC & C_{max})
- Suggests robust methodology might be needed in a small but non-negligible proportion of studies
- (Again note possible bias towards "neat" datasets / "successful" studies)

Summary (shift in point and interval estimates of GMR):

- In 5% of datasets the point estimate and confidence limits shift by $\pm$ 0.02 units (AUC & C_{max})
- Suggests robust methodology might be needed in a small but non-negligible proportion of studies
- (Again note possible bias towards "neat" datasets / "successful" studies)

Relative Confidence Interval Widths: Plot

- Relative CI widths against (estimated) residual degrees of freedom
- Relative CI widths against "need for robustness"

Relative Confidence Interval Widths: Plot

- Relative CI widths against (estimated) residual degrees of freedom
- Relative CI widths against "need for robustness"

Figure: Bayes$_T$ vs REML: Ratio of CI Widths – AUC

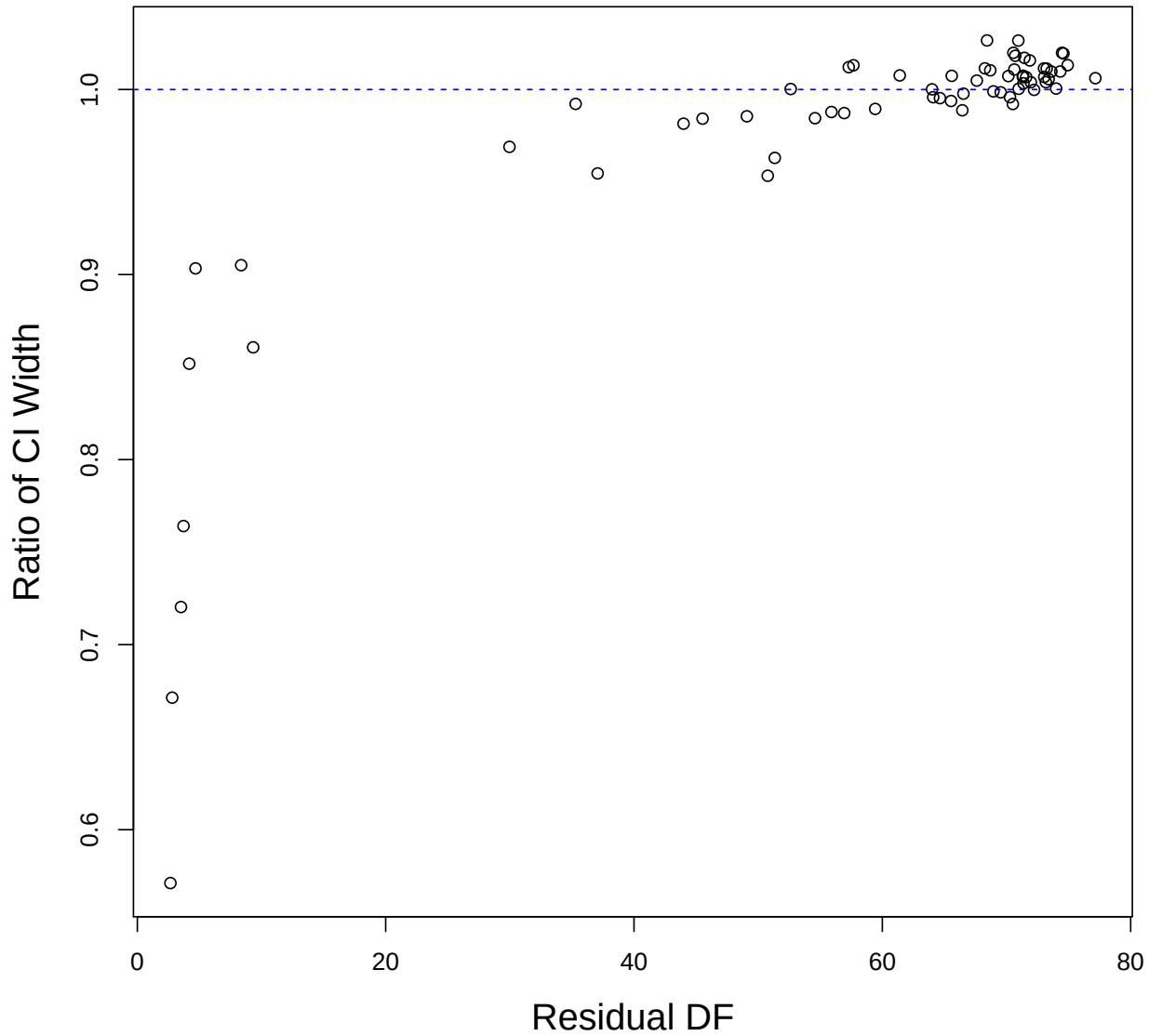
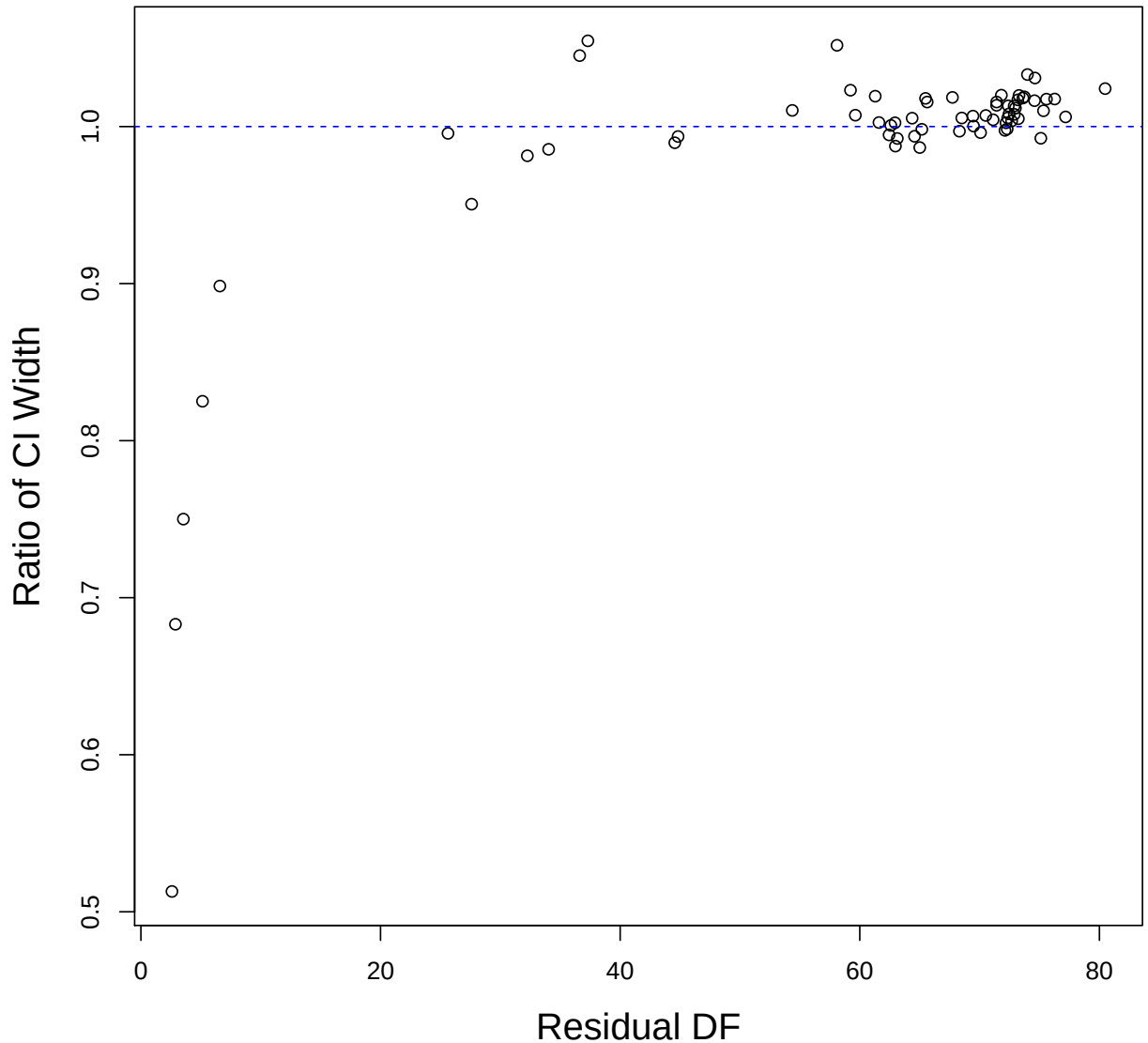


Figure: Bayes_T vs REML: Ratio of CI Widths – C_{max}



Summary (ratio of CI widths):

- Robust CIs are narrower than non-robust CIs when outliers are present (distribution is heavy-tailed)
- Robust CIs are similar to non-robust CIs when no outliers are present (distribution is not heavy-tailed)

Reason:

- Outliers/heavy tails inflate the residual variance

Summary (ratio of CI widths):

- Robust CIs are narrower than non-robust CIs when outliers are present (distribution is heavy-tailed)
- Robust CIs are similar to non-robust CIs when no outliers are present (distribution is not heavy-tailed)

Reason:

- Outliers/heavy tails inflate the residual variance

Summary (ratio of CI widths):

- Robust CIs are narrower than non-robust CIs when outliers are present (distribution is heavy-tailed)
- Robust CIs are similar to non-robust CIs when no outliers are present (distribution is not heavy-tailed)

Reason:

- Outliers/heavy tails inflate the residual variance

Method Comparison: Agreement Between Methods

Table: AUC and C_{max}: ABE Outcomes

		Method												
		REML				HL				$Bayes_N$				$Bayes_T$				
		Pass		Fail		Pass		Fail		Pass		Fail		Pass		Fail		Total
Method	Outcome	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
REML	Pass					117	(98.3)	2	(1.7)	118	(99.2)	1	(0.8)	118	(99.2)	1	(0.8)	119	(100.0)
	Fail					4	(36.4)	7	(63.6)	0	(0.0)	11	(100.0)	1	(9.1)	10	(90.9)	11	(100.0)
HL	Pass	117	(96.7)	4	(3.3)					116	(95.9)	5	(4.1)	118	(97.5)	3	(2.5)	121	(100.0)
	Fail	2	(22.2)	7	(77.8)					2	(22.2)	7	(77.8)	1	(11.1)	8	(88.9)	9	(100.0)
$Bayes_N$	Pass	118	(100.0)	0	(0.0)	116	(98.3)	2	(1.7)					117	(99.2)	1	(0.8)	118	(100.0)
	Fail	1	(8.3)	11	(91.7)	5	(41.7)	7	(58.3)					2	(16.7)	10	(83.3)	12	(100.0)
$Bayes_T$	Pass	118	(99.2)	1	(0.8)	118	(99.2)	1	(0.8)	117	(98.3)	2	(1.7)					119	(100.0)
	Fail	1	(9.1)	10	(90.9)	3	(27.3)	8	(72.7)	1	(9.1)	10	(90.9)					11	(100.0)
	Total	119	(91.5)	11	(8.5)	121	(93.1)	9	(6.9)	118	(90.8)	12	(9.2)	119	(91.5)	11	(8.5)	130	(100.0)

The DIC statistic preferred $Bayes_T$ over $Bayes_N$ in 44 out of 130 cases.

Summary (method comparison):

- Bayes$_T$ method agrees well with REML analysis
- Bayes$_T$ method agrees better with REML analysis than HL method
- DIC statistic prefers Bayes$_T$ to Bayes$_N$ in about a third of cases

Summary (method comparison):

- Bayes$_T$ method agrees well with REML analysis
- Bayes$_T$ method agrees better with REML analysis than HL method

- DIC statistic prefers Bayes$_T$ to Bayes$_N$ in about a third of cases

Summary (method comparison):

- Bayes$_T$ method agrees well with REML analysis
- Bayes$_T$ method agrees better with REML analysis than HL method

- DIC statistic prefers Bayes$_T$ to Bayes$_N$ in about a third of cases

Assess REML, Bayes$_N$ & Bayes$_T$:

- Model performance
  - Bias
  - RMSE
  - Interval coverage
- Statistical power
- Parameters chosen such that ABE ratio is:
  - 1.00
  - 1.10
  - 1.20
- Data simulated for Bayes$_N$ & Bayes$_T$
- ... for Bayes$_T$: $\nu_W = 2.5$, $\nu_B = 15$ (Heavy tailed)

Assess REML, Bayes$_N$ & Bayes$_T$:

- Model performance
  - Bias
  - RMSE
  - Interval coverage
- Statistical power
- Parameters chosen such that ABE ratio is:
  - 1.00
  - 1.10
  - 1.20
- Data simulated for Bayes$_N$ & Bayes$_T$
- ... for Bayes$_T$: $\nu_W = 2.5$, $\nu_B = 15$ (Heavy tailed)

Assess REML, Bayes$_N$ & Bayes$_T$:

- Model performance
  - Bias
  - RMSE
  - Interval coverage
- Statistical power
- Parameters chosen such that ABE ratio is:
  - 1.00
  - 1.10
  - 1.20
- Data simulated for Bayes$_N$ & Bayes$_T$
- ... for Bayes$_T$: $\nu_W = 2.5$, $\nu_B = 15$ (Heavy tailed)

Assess REML, Bayes$_N$ & Bayes$_T$:

- Model performance
  - Bias
  - RMSE
  - Interval coverage
- Statistical power
- Parameters chosen such that ABE ratio is:
  - 1.00
  - 1.10
  - 1.20
- Data simulated for Bayes$_N$ & Bayes$_T$
- ... for Bayes$_T$: $\nu_W = 2.5$, $\nu_B = 15$ (Heavy tailed)

Table: Simulation Study: Data from Normal Distribution

ABE	Method	Bias	RMSE	Coverage	Power
1.00	REML	0.0001	0.0199	89.9	100.0
	Bayes_N	0.0001	0.0199	89.5	100.0
	Bayes_T	0.0001	0.0199	89.6	100.0
1.10	REML	−0.0003	0.0221	89.4	100.0
	Bayes_N	−0.0003	0.0221	89.4	100.0
	Bayes_T	−0.0003	0.0221	89.3	100.0
1.20	REML	0.0002	0.0242	89.9	63.5
	Bayes_N	0.0002	0.0242	90.0	64.1
	Bayes_T	0.0001	0.0243	89.6	64.4

Table: Simulation Study: Data from Student t Distribution

ABE	Method	Bias	RMSE	Coverage	Power
1.00	REML	−0.0007	0.0404	91.0	98.6
	Bayes_N	−0.0007	0.0404	91.3	98.7
	Bayes_T	−0.0001	0.0290	92.3	100.0
1.10	REML	−0.0002	0.0524	89.5	93.1
	Bayes_N	−0.0002	0.0524	89.8	93.4
	Bayes_T	−0.0017	0.0339	90.0	99.4
1.20	REML	0.0016	0.0515	90.8	29.6
	Bayes_N	0.0016	0.0515	91.3	31.2
	Bayes_T	0.0006	0.0358	91.7	36.3

Summary:

- Data simulated from Bayes_N
- ... All models have good accuracy, precision & coverage
- ... All models yield similar statistical power

- Data simulated from Bayes_T
- ... All models have good accuracy & coverage
- ... Bayes_T has better precision
- ... Bayes_T yields higher statistical power

Summary:

- Data simulated from Bayes_N
- ... All models have good accuracy, precision & coverage
- ... All models yield similar statistical power

- Data simulated from Bayes_T
- ... All models have good accuracy & coverage
- ... Bayes_T has better precision
- ... Bayes_T yields higher statistical power

Conventional model for 2-treatment, 2-sequence, 4-period replicate crossover study

$$y_{ijk} = \mu + \zeta_h + s_{ij} + \pi_m + \tau_j + e_{ijk}$$

- y_{ijk} log-bioavailability for the k^{th} replicate of subject i and formulation $j = T, R$; $k = 1, 2$
- μ overall mean
- s_{ij} random effect of subject i and formulation j
- ζ_h (sequence), π_m (period), τ_j (treatment): fixed effects
- e_{ijk} residual

Conventional model for 2-treatment, 2-sequence, 4-period replicate crossover study

$$y_{ijk} = \mu + \zeta_h + s_{ij} + \pi_m + \tau_j + e_{ijk}$$

- y_{ijk} log-bioavailability for the k^{th} replicate of subject i and formulation $j = T, R$; $k = 1, 2$
- μ overall mean
- s_{ij} random effect of subject i and formulation j
- ζ_h (sequence), π_m (period), τ_j (treatment): fixed effects
- e_{ijk} residual

- Variance-covariance matrix of the $\boldsymbol{s}_i = (s_{iR}, s_{iT})'$:

$$\text{cov}(\boldsymbol{s}_i) = \boldsymbol{\Sigma}_B = \begin{pmatrix} \sigma^2_{BR} & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma^2_{BT} \end{pmatrix}$$

- Between-subject variances:
  - $\text{var}(s_{iR}) = \sigma^2_{BR}$
  - $\text{var}(s_{iT}) = \sigma^2_{BT}$
  - $\sigma^2_D = \text{var}(s_{iR} - s_{iT}) = (\sigma^2_{BR} + \sigma^2_{BT} - 2\rho\sigma_{BR}\sigma_{BT})$: subject-by-formulation interaction variance
- Within-subject variances:
  - $\text{var}(e_{iRk}) = \sigma^2_{WR}$
  - $\text{var}(e_{iTk}) = \sigma^2_{WT}$

- Variance-covariance matrix of the $\boldsymbol{s}_i = (s_{iR}, s_{iT})'$:

$$\text{cov}(\boldsymbol{s}_i) = \boldsymbol{\Sigma}_B = \begin{pmatrix} \sigma^2_{BR} & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma^2_{BT} \end{pmatrix}$$

- Between-subject variances:
  - $\text{var}(s_{iR}) = \sigma^2_{BR}$
  - $\text{var}(s_{iT}) = \sigma^2_{BT}$
  - $\sigma^2_D = \text{var}(s_{iR} - s_{iT}) = (\sigma^2_{BR} + \sigma^2_{BT} - 2\rho\sigma_{BR}\sigma_{BT})$: subject-by-formulation interaction variance
- Within-subject variances:
  - $\text{var}(e_{iRk}) = \sigma^2_{WR}$
  - $\text{var}(e_{iTk}) = \sigma^2_{WT}$

- Variance-covariance matrix of the $\boldsymbol{s}_i = (s_{iR}, s_{iT})'$:

$$\text{cov}(\boldsymbol{s}_i) = \boldsymbol{\Sigma}_B = \begin{pmatrix} \sigma^2_{BR} & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma^2_{BT} \end{pmatrix}$$

- Between-subject variances:
  - $\text{var}(s_{iR}) = \sigma^2_{BR}$
  - $\text{var}(s_{iT}) = \sigma^2_{BT}$
  - $\sigma^2_D = \text{var}(s_{iR} - s_{iT}) = (\sigma^2_{BR} + \sigma^2_{BT} - 2\rho\sigma_{BR}\sigma_{BT})$: subject-by-formulation interaction variance
- Within-subject variances:
  - $\text{var}(e_{iRk}) = \sigma^2_{WR}$
  - $\text{var}(e_{iTk}) = \sigma^2_{WT}$

Linearized reference-scaled average bioequivalence (RSABE) criterion:

$$\theta = \tau^2 - k^2\sigma_{WR}^2$$

where

- $k = \frac{\log(1.25)}{\sigma_0}$
- Choose: $\sigma_0 = 0.25$

Statistical decision rule

Two formulations are bioequivalent if the one-sided 95% upper confidence limit for θ is below zero

1. Subject outlier (between-subject outlier):
   - For subject i, all 4 observations y_{iT1}, y_{iT2}, y_{iR1}, y_{iR2} are extreme, in the same direction
   - Could be modeled as a mean shift in s_i
   - Again: has no consequences since a mean shift in s_i does not affect point or interval estimate of relative bioavailability

2. Subject-by-formulation outlier (within-subject but between-replicate outlier):
   - For subject i, the replicates y_{iT1} and y_{iT2}, jointly, are different from the replicates y_{iR1} and y_{iR2}
   - For subject i the within-subject, between replicate difference $(y_{iT1} + y_{iT2}) - (y_{iR1} + y_{iR2})$ is extreme
   - Could be modeled as a mean shift in either s_{iT} or s_{iR}
   - Can severely affect results of the bioequivalence test

1. Subject outlier (between-subject outlier):
   - For subject i, all 4 observations y_{iT1}, y_{iT2}, y_{iR1}, y_{iR2} are extreme, in the same direction
   - Could be modeled as a mean shift in $\boldsymbol{s}_i$
   - Again: has no consequences since a mean shift in $\boldsymbol{s}_i$ does not affect point or interval estimate of relative bioavailability

2. Subject-by-formulation outlier (within-subject but between-replicate outlier):
   - For subject i, the replicates y_{iT1} and y_{iT2}, jointly, are different from the replicates y_{iR1} and y_{iR2}
   - For subject i the within-subject, between replicate difference $(y_{iT1} + y_{iT2}) - (y_{iR1} + y_{iR2})$ is extreme
   - Could be modeled as a mean shift in either s_{iT} or s_{iR}
   - Can severely affect results of the bioequivalence test

1. Subject outlier (between-subject outlier):
   - For subject i, all 4 observations y_{iT1}, y_{iT2}, y_{iR1}, y_{iR2} are extreme, in the same direction
   - Could be modeled as a mean shift in $\boldsymbol{s}_i$
   - Again: has no consequences since a mean shift in $\boldsymbol{s}_i$ does not affect point or interval estimate of relative bioavailability

2. Subject-by-formulation outlier (within-subject but between-replicate outlier):
   - For subject i, the replicates y_{iT1} and y_{iT2}, jointly, are different from the replicates y_{iR1} and y_{iR2}
   - For subject i the within-subject, between replicate difference $(y_{iT1} + y_{iT2}) - (y_{iR1} + y_{iR2})$ is extreme
   - Could be modeled as a mean shift in either s_{iT} or s_{iR}
   - Can severely affect results of the bioequivalence test

1. Subject outlier (between-subject outlier):
   - For subject i, all 4 observations y_{iT1}, y_{iT2}, y_{iR1}, y_{iR2} are extreme, in the same direction
   - Could be modeled as a mean shift in $\boldsymbol{s}_i$
   - Again: has no consequences since a mean shift in $\boldsymbol{s}_i$ does not affect point or interval estimate of relative bioavailability

2. Subject-by-formulation outlier (within-subject but between-replicate outlier):
   - For subject i, the replicates y_{iT1} and y_{iT2}, jointly, are different from the replicates y_{iR1} and y_{iR2}
   - For subject i the within-subject, between replicate difference $(y_{iT1} + y_{iT2}) - (y_{iR1} + y_{iR2})$ is extreme
   - Could be modeled as a mean shift in either s_{iT} or s_{iR}
   - Can severely affect results of the bioequivalence test

1. Subject outlier (between-subject outlier):
   - For subject i, all 4 observations y_{iT1}, y_{iT2}, y_{iR1}, y_{iR2} are extreme, in the same direction
   - Could be modeled as a mean shift in $\boldsymbol{s}_i$
   - Again: has no consequences since a mean shift in $\boldsymbol{s}_i$ does not affect point or interval estimate of relative bioavailability

2. Subject-by-formulation outlier (within-subject but between-replicate outlier):
   - For subject i, the replicates y_{iT1} and y_{iT2}, jointly, are different from the replicates y_{iR1} and y_{iR2}
   - For subject i the within-subject, between replicate difference $(y_{iT1} + y_{iT2}) - (y_{iR1} + y_{iR2})$ is extreme
   - Could be modeled as a mean shift in either s_{iT} or s_{iR}
   - Can severely affect results of the bioequivalence test

3. Single data point outlier (within-subject, within-replicate outlier):
   - For subject i, one of the four observations y_{iT1}, y_{iT2}, y_{iR1}, or y_{iR2} is extreme
   - At least one of the within-subject, within-replicate differences $(y_{iT1} - y_{iT2})$ or $(y_{iR1} - y_{iR2})$ is extreme
   - Can severely affect the results of the bioequivalence test

Schall, Ring, Endrenyi (2010). Residuals and outliers in replicate design crossover studies. *J. Biopharm. Stat.*

3. Single data point outlier (within-subject, within-replicate outlier):
   - For subject i, one of the four observations y_{iT1}, y_{iT2}, y_{iR1}, or y_{iR2} is extreme
   - At least one of the within-subject, within-replicate differences $(y_{iT1} - y_{iT2})$ or $(y_{iR1} - y_{iR2})$ is extreme
   - Can severely affect the results of the bioequivalence test

Schall, Ring, Endrenyi (2010). Residuals and outliers in replicate design crossover studies. *J. Biopharm. Stat.*

Figure: Study 1a – Conditional Residuals: AUC (Slide 58)

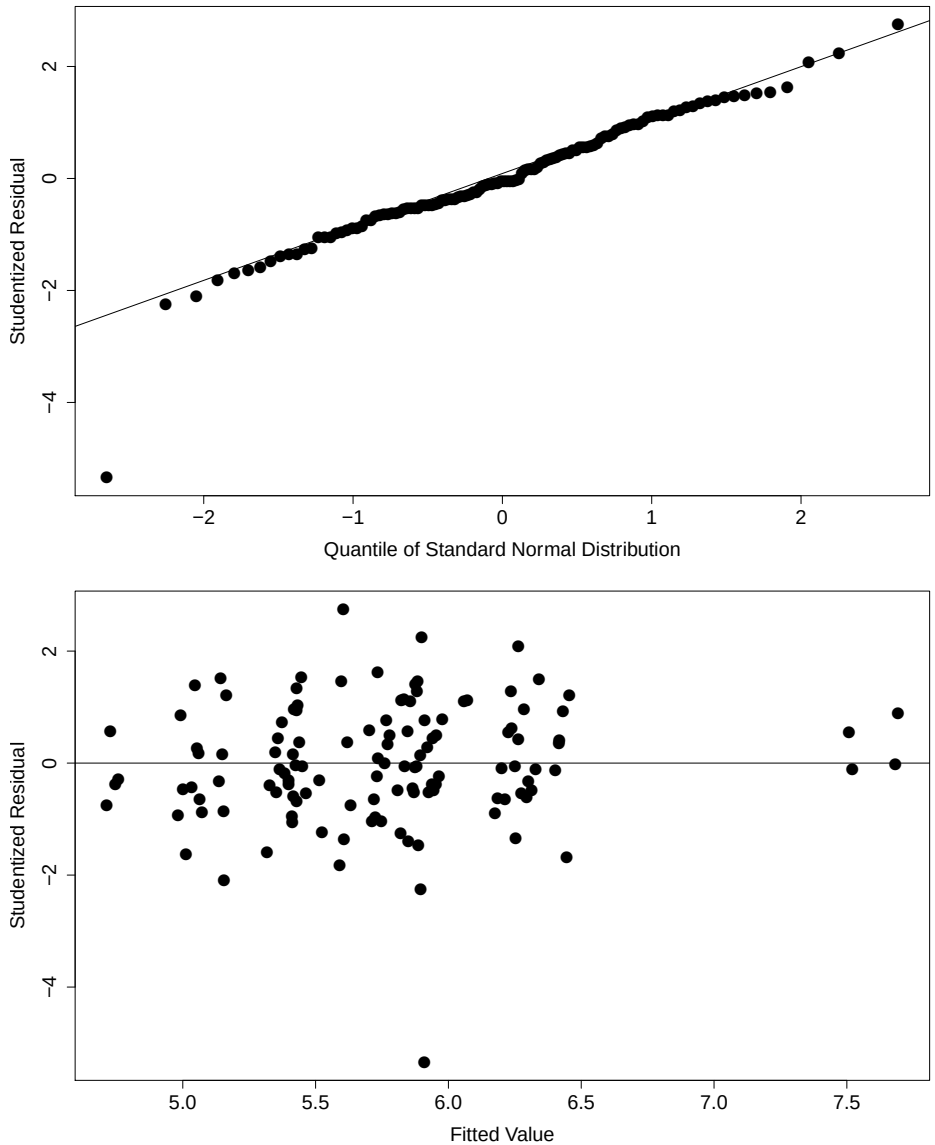


Figure: Study 1a: s_{DT} – Residuals: AUC

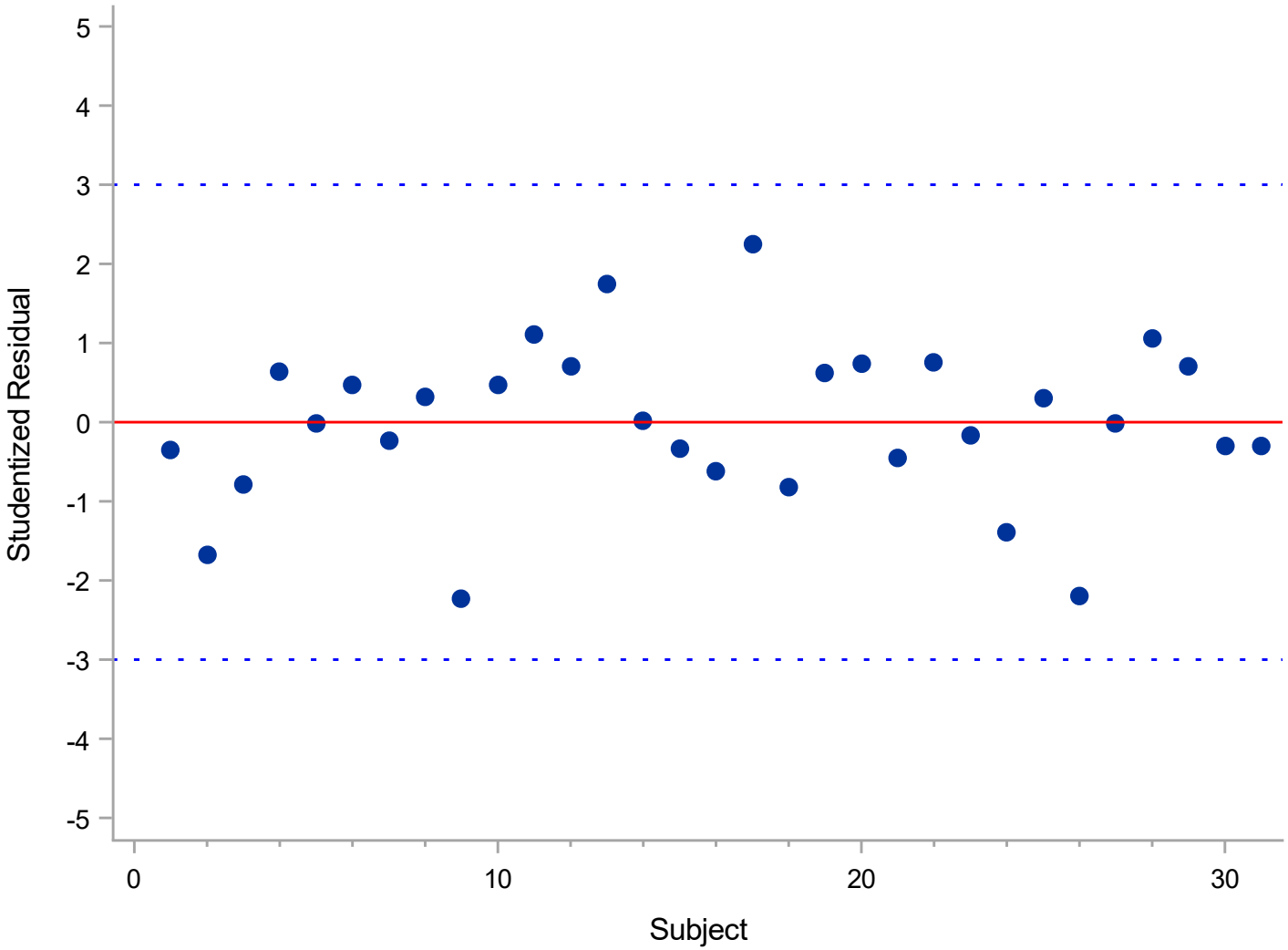


Figure: Study 1a: s_{DR} – Residuals: AUC

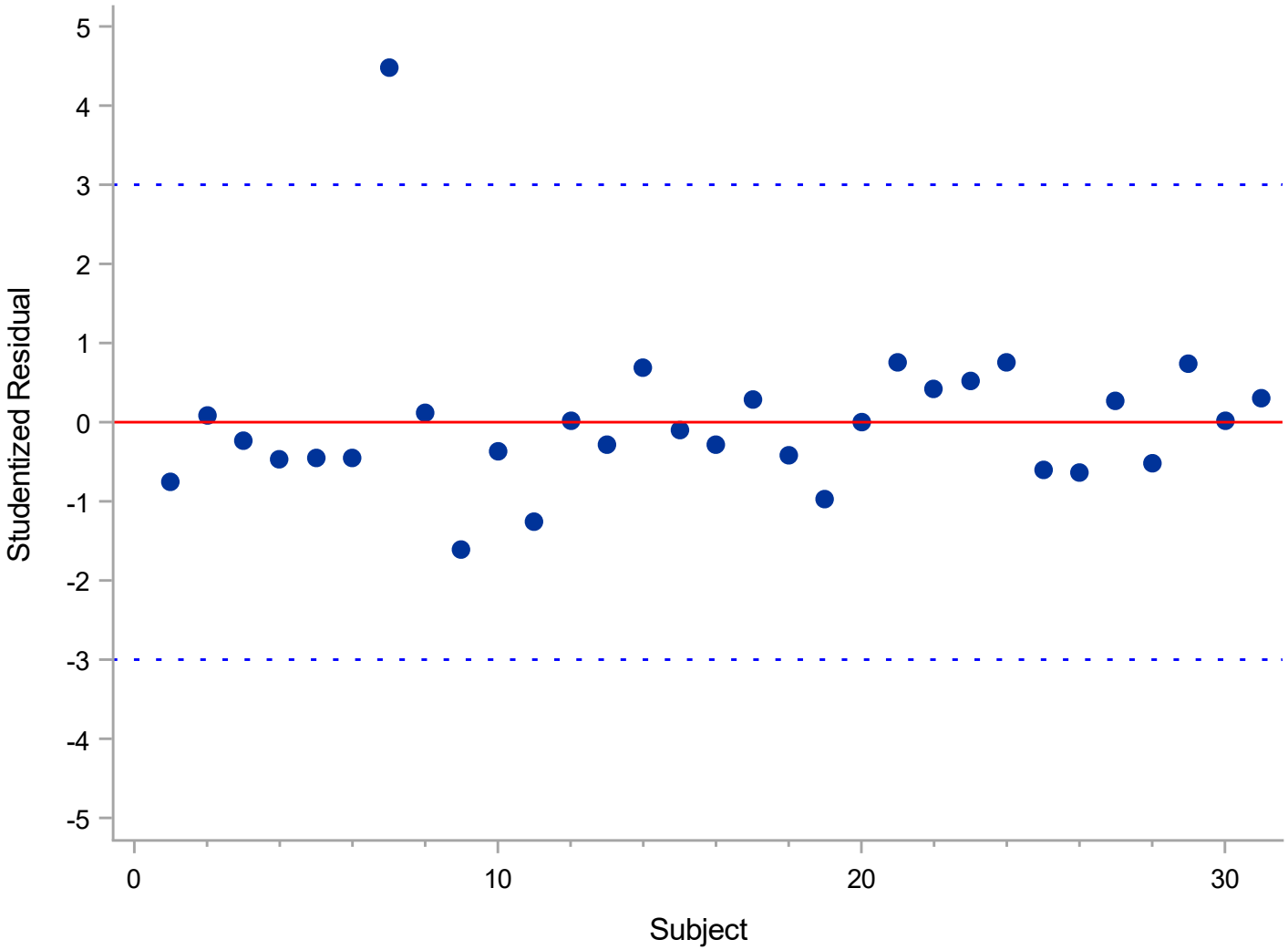


Figure: Study 1a: s_{SF} – Residuals: AUC

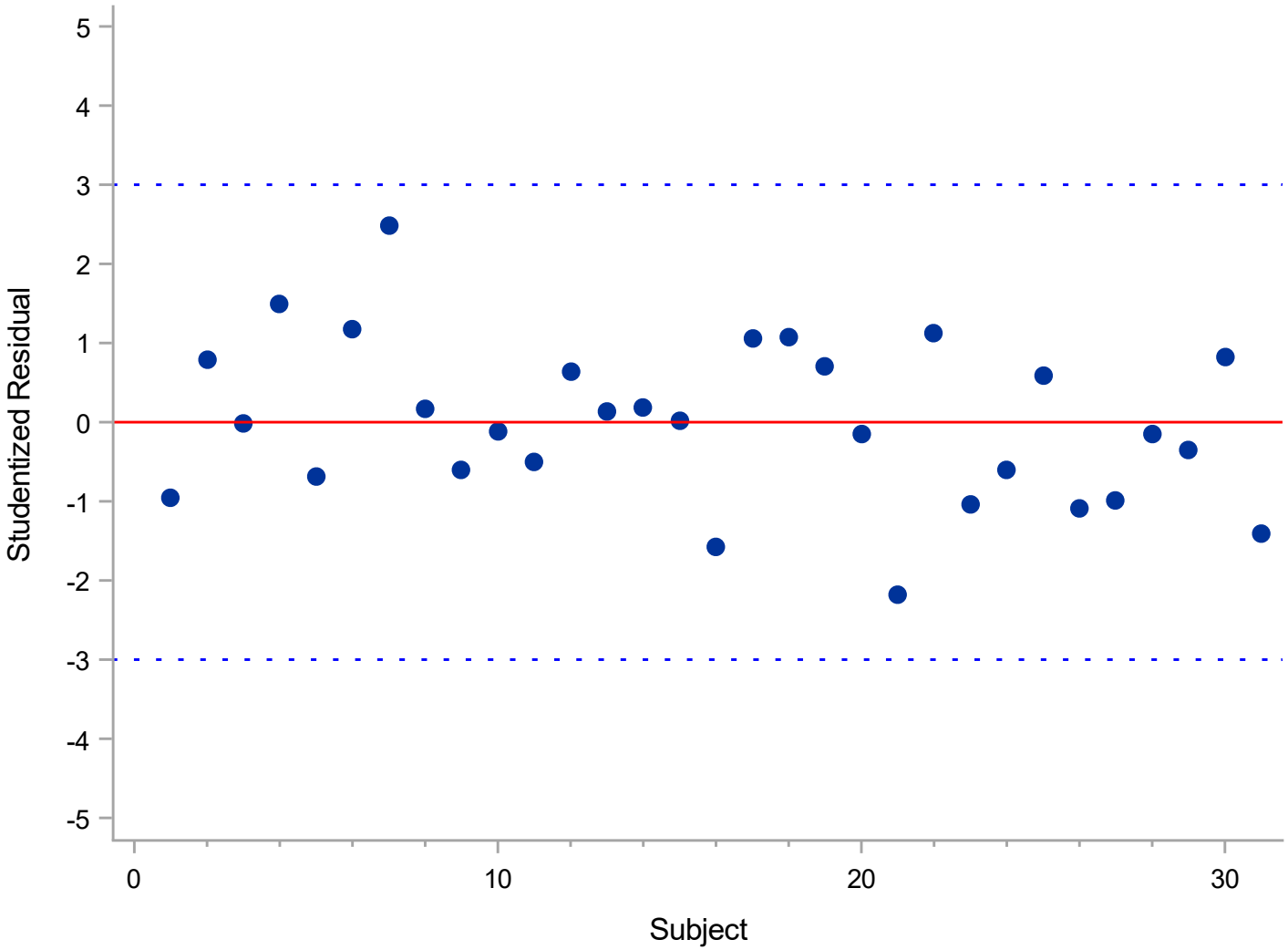


Figure: Study 1a: s_S – Residuals: AUC

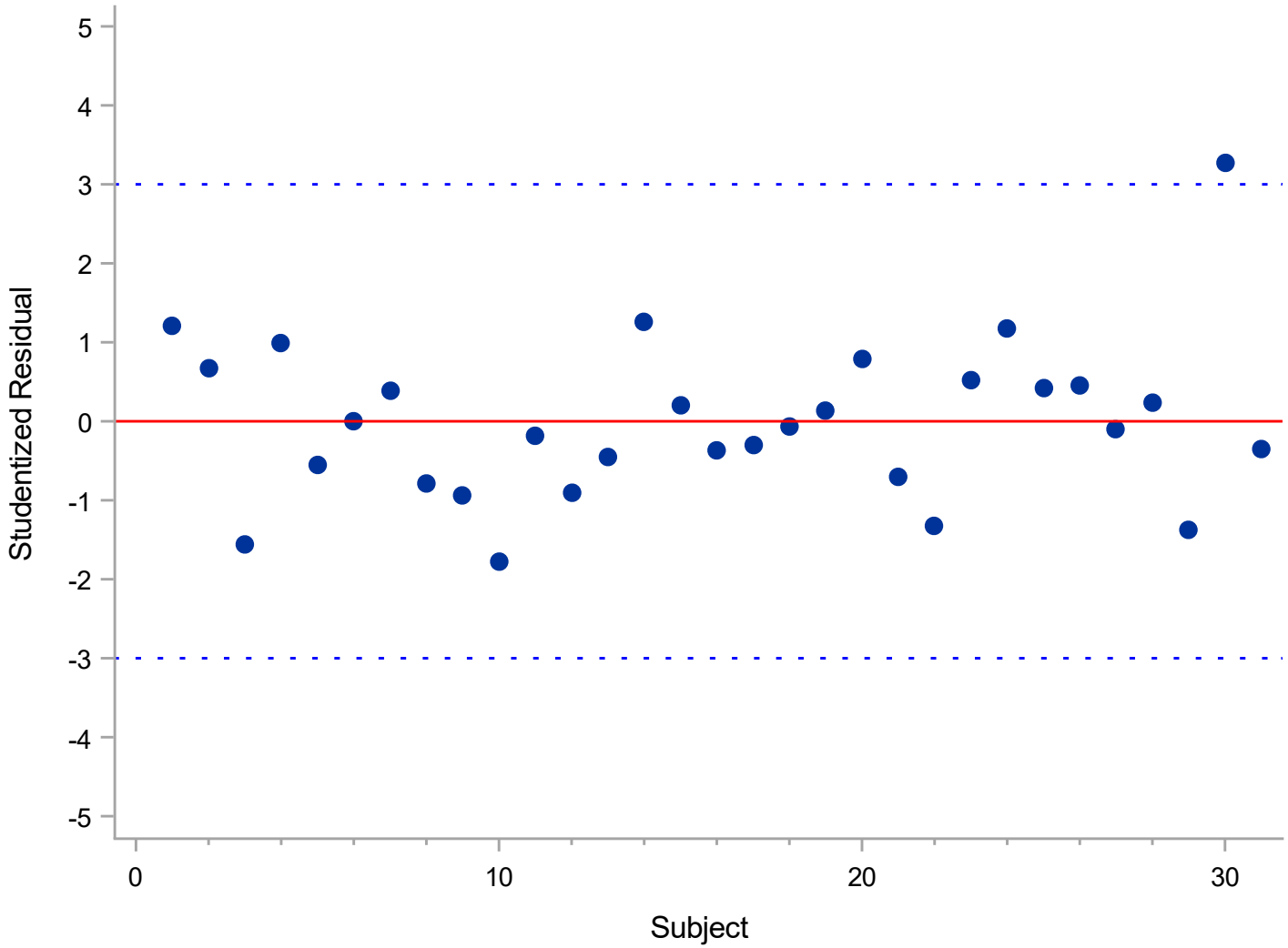


Figure: Study 7 – Conditional Residuals: C_{max} (Slide 58)

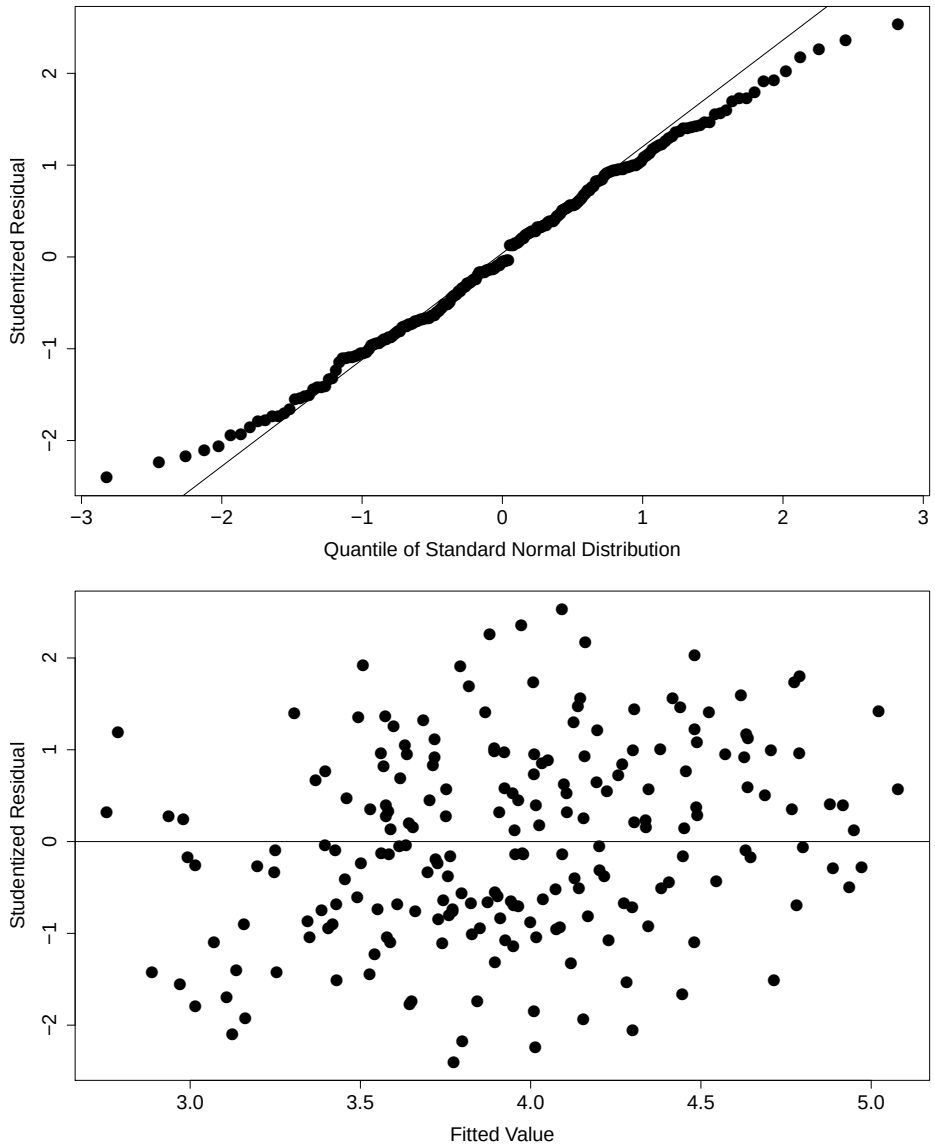


Figure: Study 7: s_{DT} – Residuals: C_{max}

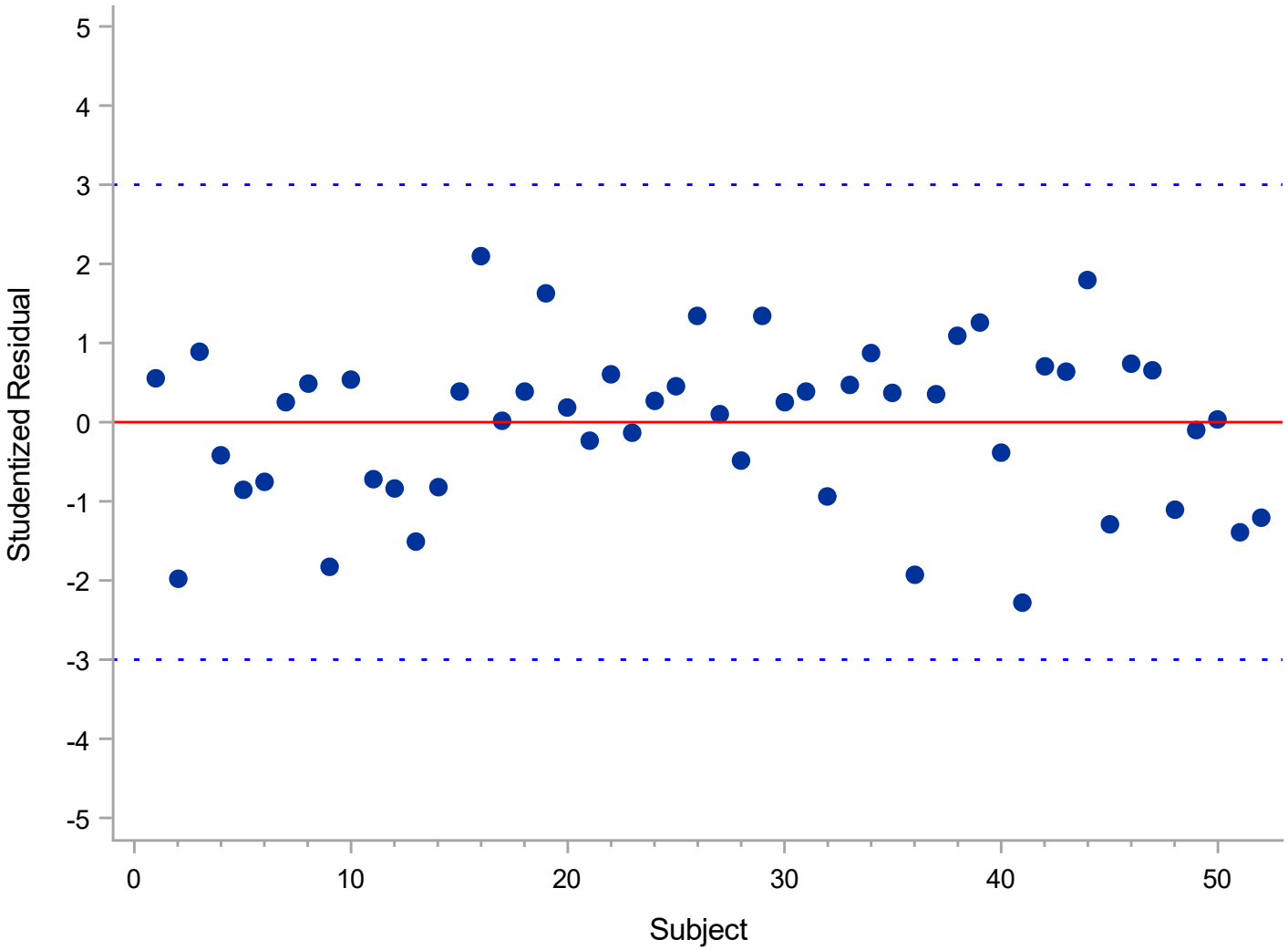


Figure: Study 7: s_{DR} – Residuals: C_{max}

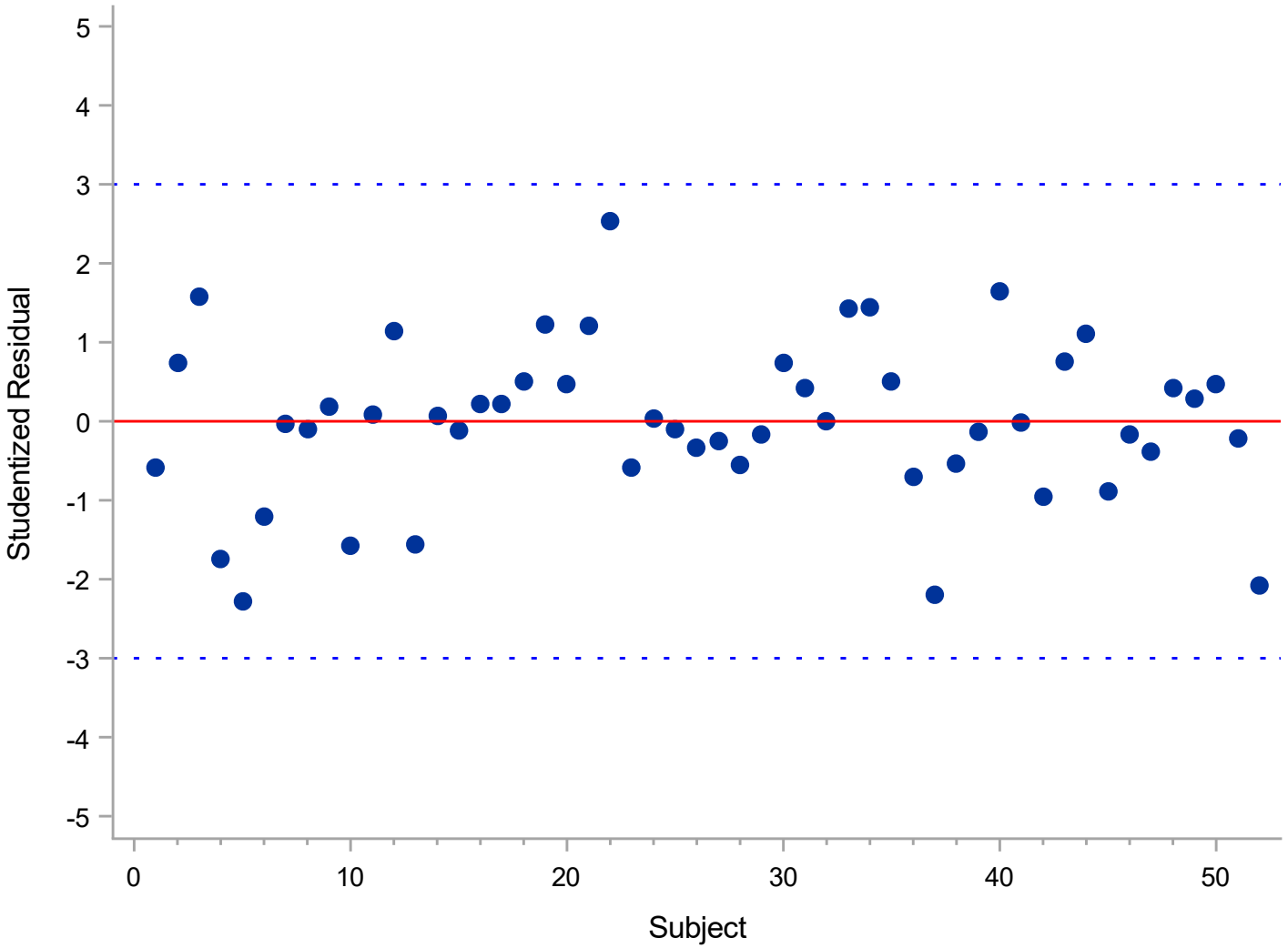


Figure: Study 7: s_{SF} – Residuals: C_{max}

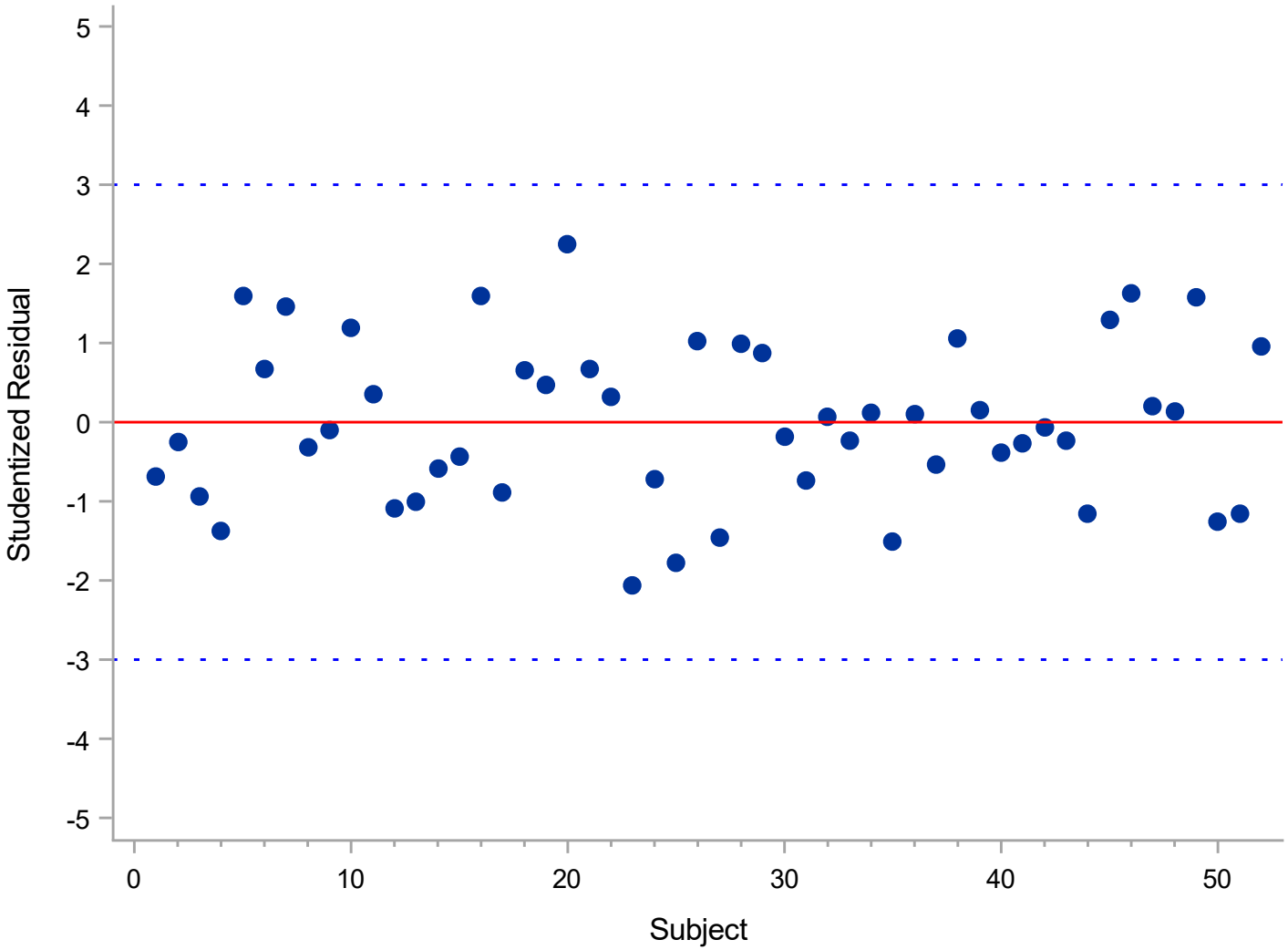
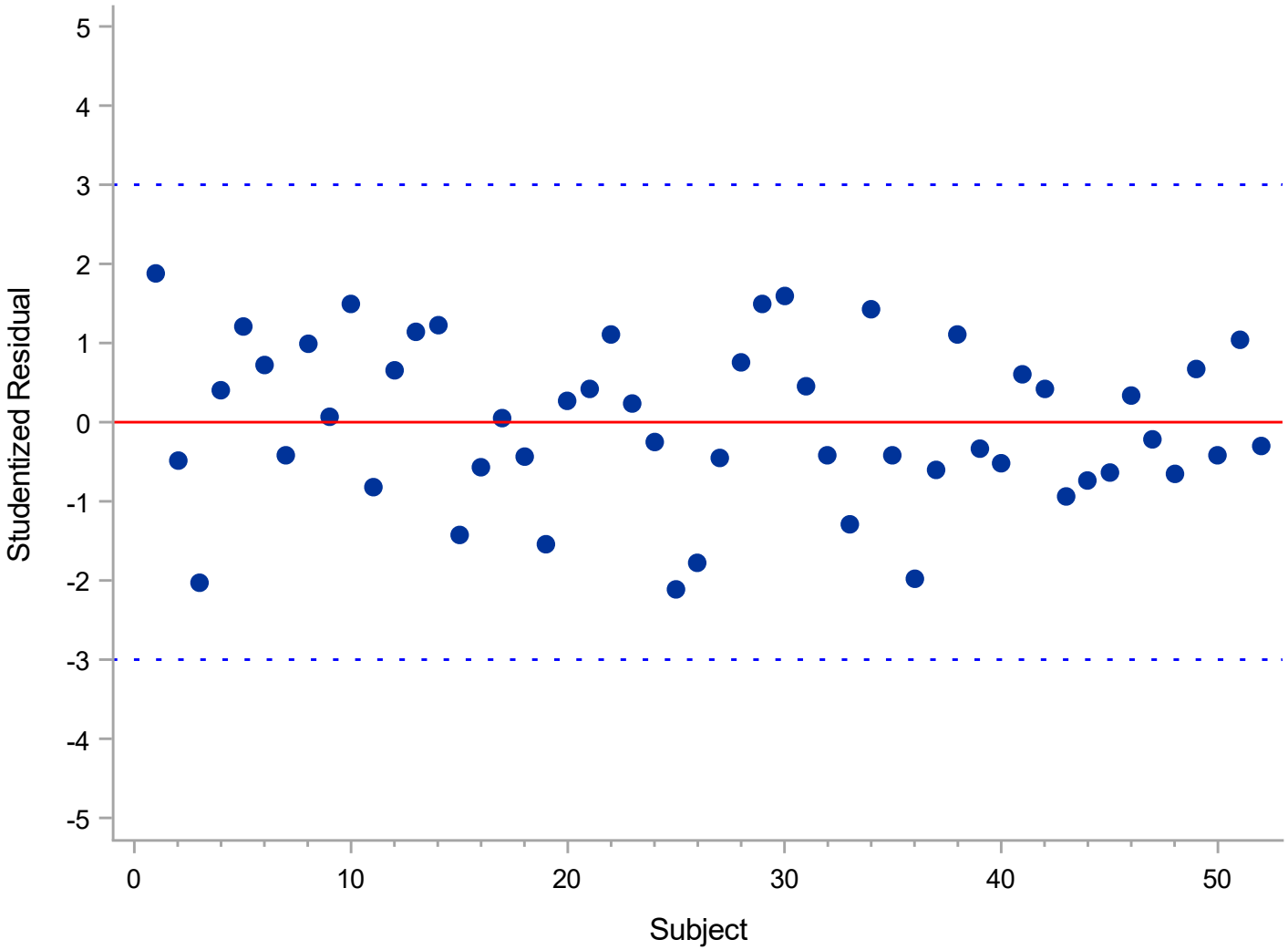


Figure: Study 7: s_S – Residuals: C_{max}



Robust methodology

- Achieve robustness by:
  1. Use standard model for replicate-design crossover study (see above)
  2. But specify heavy-tailed distributions for residuals (and subject effects)
  3. Here, use Student t distributions with low degrees of freedom
  4. (Again: Preferable to estimate the degrees of freedom)
- Fit the model using a Bayesian approach
- Compare Student t with normal model (both Bayes)

Robust methodology

- Achieve robustness by:
  1. Use standard model for replicate-design crossover study (see above)
  2. But specify heavy-tailed distributions for residuals (and subject effects)
  3. Here, use Student t distributions with low degrees of freedom
  4. (Again: Preferable to estimate the degrees of freedom)
- Fit the model using a Bayesian approach
- Compare Student t with normal model (both Bayes)

Robust methodology

- Achieve robustness by:
  1. Use standard model for replicate-design crossover study (see above)
  2. But specify heavy-tailed distributions for residuals (and subject effects)
  3. Here, use Student t distributions with low degrees of freedom
  4. (Again: Preferable to estimate the degrees of freedom)

- Fit the model using a Bayesian approach
- Compare Student t with normal model (both Bayes)

Robust methodology

- Achieve robustness by:
  1. Use standard model for replicate-design crossover study (see above)
  2. But specify heavy-tailed distributions for residuals (and subject effects)
  3. Here, use Student t distributions with low degrees of freedom
  4. (Again: Preferable to estimate the degrees of freedom)
- Fit the model using a Bayesian approach
- Compare Student t with normal model (both Bayes)

Residuals and random effects

- Bayes$_T$
  - Student t distribution accommodates heavy tails/outliers
  - ... $e_{ijk} \sim t\left(0, \sigma^{2*}_{Wj}, \nu_{Wj}\right)$ where $\sigma^2_{Wj} = \frac{\nu_{Wj}}{\nu_{Wj}-2}\sigma^{2*}_{Wj}$
  - ... $\boldsymbol{s}_i \sim t\left(\mathbf{0}, \boldsymbol{\Sigma}^*_B, \nu_B\right)$ where $\boldsymbol{\Sigma}_B = \frac{\nu_B}{\nu_B-2}\boldsymbol{\Sigma}^*_B$
  - ... ν_{Wj}, $\nu_{Bj} \Rightarrow$ degrees of freedom
- Bayes$_N$
  - Normal distribution – not robust to outliers
  - ... $e_{ijk} \sim \text{Normal}\left(0, \sigma^{2*}_{Wj}\right)$ where $\sigma^2_{Wj} = \sigma^{2*}_{Wj}$
  - ... $\boldsymbol{s}_i \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}^*_B, \nu_B)$ where $\boldsymbol{\Sigma}_B = \boldsymbol{\Sigma}^*_B$

Residuals and random effects

- Bayes$_T$
  - Student t distribution accommodates heavy tails/outliers
  - ... $e_{ijk} \sim t\left(0, \sigma^{2*}_{Wj}, \nu_{Wj}\right)$ where $\sigma^2_{Wj} = \frac{\nu_{Wj}}{\nu_{Wj}-2}\sigma^{2*}_{Wj}$
  - ... $\boldsymbol{s}_i \sim t\left(\mathbf{0}, \boldsymbol{\Sigma}^*_B, \nu_B\right)$ where $\boldsymbol{\Sigma}_B = \frac{\nu_B}{\nu_B-2}\boldsymbol{\Sigma}^*_B$
  - ... ν_{Wj}, $\nu_{Bj} \Rightarrow$ degrees of freedom
- Bayes$_N$
  - Normal distribution – not robust to outliers
  - ... $e_{ijk} \sim \text{Normal}\left(0, \sigma^{2*}_{Wj}\right)$ where $\sigma^2_{Wj} = \sigma^{2*}_{Wj}$
  - ... $\boldsymbol{s}_i \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}^*_B, \nu_B)$ where $\boldsymbol{\Sigma}_B = \boldsymbol{\Sigma}^*_B$

Prior specification

- Vague priors
- ... μ, ζ_h, π_m, $\tau_j \sim N(0, 10000)$
- ... $\sigma_{Wj}^{-2*} \sim \text{Gamma}(0.0001, 0.0001)$
- ... v_{Wj}, $v_B \sim \text{Normal}(0, 10000)\, T(2, \infty)$ (half-normal)
- ... $\boldsymbol{\Sigma}^* \sim \text{MGH-}t\,(10000, 2)$ (matrix generalized half-t)

Implementation

- Fit model using `JAGS` via `R` package `runjags`
- Student t: Mixture of normal & gamma distribution
  - Speeds up convergence
  - Matrix generalized half-t for covariance matrix: Huang & Wand (2013)
  - Most priors are conjugate $\Rightarrow$ fast convergence
- RSABE $\Rightarrow$ calculate:
  - Posterior estimate of θ
  - Upper limit of one-sided 95% Bayesian credibility (BCI) interval for θ
- DIC statistic: Discriminate between Bayes_N & Bayes_T

Huang & Wand (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.*

Implementation

- Fit model using `JAGS` via `R` package `runjags`
- Student t: Mixture of normal & gamma distribution
  - Speeds up convergence
  - Matrix generalized half-t for covariance matrix: Huang & Wand (2013)
  - Most priors are conjugate $\Rightarrow$ fast convergence
- RSABE $\Rightarrow$ calculate:
  - Posterior estimate of θ
  - Upper limit of one-sided 95% Bayesian credibility (BCI) interval for θ
- DIC statistic: Discriminate between Bayes_N & Bayes_T

Huang & Wand (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.*

Implementation

- Fit model using `JAGS` via `R` package `runjags`
- Student t: Mixture of normal & gamma distribution
  - Speeds up convergence
  - Matrix generalized half-t for covariance matrix: Huang & Wand (2013)
  - Most priors are conjugate $\Rightarrow$ fast convergence
- RSABE $\Rightarrow$ calculate:
  - Posterior estimate of θ
  - Upper limit of one-sided 95% Bayesian credibility (BCI) interval for θ
- DIC statistic: Discriminate between Bayes_N & Bayes_T

Huang & Wand (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.*

Implementation

- Fit model using `JAGS` via `R` package `runjags`
- Student t: Mixture of normal & gamma distribution
  - Speeds up convergence
  - Matrix generalized half-t for covariance matrix: Huang & Wand (2013)
  - Most priors are conjugate $\Rightarrow$ fast convergence
- RSABE $\Rightarrow$ calculate:
  - Posterior estimate of θ
  - Upper limit of one-sided 95% Bayesian credibility (BCI) interval for θ
- DIC statistic: Discriminate between Bayes_N & Bayes_T

Huang & Wand (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.*

Pool of datasets from 2-sequence, 2-treatment, 4-period replicate designs:

- 7 studies
- ... for both AUC and C_{max}
- ... Study 1a: Schall, Ring, Endrenyi (2010)
- ... Study 1a: R contains outliers, whereas T does not
- ... Study 1b: Study 1a's T and R labels reversed

Table: AUC and C_{max}: UCLs of RSABE and Estimates of DF

		UCL of RSABE			Estimates of DF		
					Residual		
Study	Parameter	REML	$Bayes_N$	$Bayes_T$	Test	Reference	Random
1a	AUC†	-0.0294	-0.0270	-0.0126	71.6	3.2	59.3
	C_{max}†	-0.0262	-0.0155	-0.0108	68.3	4.0	78.4
1b	AUC†	-0.0091	-0.0065	-0.0107	3.2	71.5	63.2
	C_{max}†	-0.0190	-0.0087	-0.0201	4.0	68.9	76.9
2	AUC†	-0.1388	-0.1283	-0.1245	39.0	53.8	71.6
	C_{max}†	-0.1328	-0.1249	-0.1191	18.5	44.4	72.0
3	AUC†	-0.1382	-0.1295	-0.1228	17.8	52.9	70.7
	C_{max}†	-0.1321	-0.1251	-0.1204	31.8	39.6	66.1
4	AUC†	-0.1757	-0.1572	-0.1510	38.3	4.2	71.2
	C_{max}†	-0.2234	-0.2003	-0.1926	31.2	2.9	55.5
5	AUC†	-0.0085	-0.0066	-0.0065	13.1	69.0	73.5
	C_{max}†	-0.0070	-0.0055	-0.0059	7.9	76.4	75.5
6	AUC†	-0.0013	0.0009	-0.0007	29.8	4.6	55.3
	C_{max}†	-0.0009	0.0011	0.0009	28.5	65.5	59.9
7	AUC	-0.0517	-0.0471	-0.0474	62.6	63.6	68.1
	C_{max}	0.0821	0.1055	0.1060	79.1	76.8	79.2

†DIC statistic prefers $Bayes_T$ over $Bayes_N$.

Summary (need for robust methodology):

- Degrees of freedom estimates for 2/8 studies (R treatment!) are small for both AUC and C_{max}
- DIC prefers $Bayes_T$ to $Bayes_N$ in most cases
- Bayesian CIs are generally somewhat wider than REML CIs

Reason:

- Outliers/heavy tails for R treatment inflate the residual variance of R
- Leads to smaller scaling factor when robust methodology is applied

Summary (need for robust methodology):

- Degrees of freedom estimates for 2/8 studies (R treatment!) are small for both AUC and C_{max}
- DIC prefers $Bayes_T$ to $Bayes_N$ in most cases
- Bayesian CIs are generally somewhat wider than REML CIs

Reason:

- Outliers/heavy tails for R treatment inflate the residual variance of R
- Leads to smaller scaling factor when robust methodology is applied

Summary (need for robust methodology):

- Degrees of freedom estimates for 2/8 studies (R treatment!) are small for both AUC and C_{max}
- DIC prefers $Bayes_T$ to $Bayes_N$ in most cases
- Bayesian CIs are generally somewhat wider than REML CIs

Reason:

- Outliers/heavy tails for R treatment inflate the residual variance of R
- Leads to smaller scaling factor when robust methodology is applied

Summary (need for robust methodology):

- Degrees of freedom estimates for 2/8 studies (R treatment!) are small for both AUC and C_{max}
- DIC prefers $Bayes_T$ to $Bayes_N$ in most cases
- Bayesian CIs are generally somewhat wider than REML CIs

Reason:

- Outliers/heavy tails for R treatment inflate the residual variance of R
- Leads to smaller scaling factor when robust methodology is applied

Summary (need for robust methodology):

- Degrees of freedom estimates for 2/8 studies (R treatment!) are small for both AUC and C_{max}
- DIC prefers $Bayes_T$ to $Bayes_N$ in most cases
- Bayesian CIs are generally somewhat wider than REML CIs

Reason:

- Outliers/heavy tails for R treatment inflate the residual variance of R
- Leads to smaller scaling factor when robust methodology is applied

Table: AUC and C_{max}: RSABE Outcomes

		Method									
		REML				$Bayes_N$				$Bayes_T$					
		Pass		Fail		Pass		Fail		Pass		Fail		Total	
Method	**Outcome**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**
REML	Pass					13	(86.7)	2	(13.3)	14	(93.3)	1	(6.7)	15	(100.0)
	Fail					0	(0.0)	1	(100.0)	0	(0.0)	1	(100.0)	1	(100.0)
$Bayes_N$	Pass	13	(100.0)	0	(0.0)					13	(100.0)	0	(0.0)	13	(100.0)
	Fail	2	(66.7)	1	(33.3)					1	(33.3)	2	(66.7)	3	(100.0)
$Bayes_T$	Pass	14	(100.0)	0	(0.0)	13	(92.9)	1	(7.1)					14	(100.0)
	Fail	1	(50.0)	1	(50.0)	0	(0.0)	2	(100.0)					2	(100.0)
	Total	15	(93.8)	1	(6.3)	13	(81.3)	3	(18.8)	14	(87.5)	2	(12.5)	16	(100.0)

The DIC statistic preferred $Bayes_T$ over $Bayes_N$ in 14 out of 16 cases.

Summary (agreement between methods):

- Good agreement between REML and Bayes methods, in particular REML with Bayes_T