A comparison of interval estimates of the reliability coefficient obtained from Bayesian rejection sampling and the adjusted Searle method, in small studies

Ruth Pickering
Public Health Sciences and Medical Statistics

UNIVERSITY OF Southampton
School of Medicine

The reliability coefficient (R)

in simple replication studies

Repeated readings of variable X taken on N subjects

$$X_{ik} = \mu + \alpha_i + \varepsilon_{ik}$$

subject i= 1…N

repeat k= 1…n_i - number of repeats can vary with subject

the α_i follow a Normal $(0, \sigma^2_{subject})$
the ε_{ik} a Normal $(0, \sigma^2_{repeat})$ } independent

$$R = \frac{\sigma^2_{subject}}{\sigma^2_{subject} + \sigma^2_{repeat}}$$ - expressed as a %

R and the ICC are alternative names for the same thing

Rationale

- coverage of interval estimates for R depends on:
  - true value of R
  - N (subjects) & n (repeats)
  - distribution of α_i and ε_{ik}
- In small reliability studies
  - typically R = 60-100%, N=10-20, n=2-5
- SAS PROC MIXED
  - fits variance component models
  - doesn't directly produce estimates of R
  - includes an empirical Bayes method of rejection sampling giving credibility intervals for $\sigma^2_{subject}$, σ^2_{repeat} and thus R
- Bayesian approach extends to more complex designs
  - eg inter-observer studies
- how does the Bayesian interval perform?

Confidence intervals (CI) for R
Searle's method for balanced data: n_i=n

One-way ANOVA for variable X including subject

The ratio of $MS_{subject}$ and MS_{error} gives the F test

$$F = \frac{MS_{subject}}{MS_{error}}$$

F_L and F_U are the 2.5th and 97.5th percentiles of an $F^{N-1}_{N(n-1)}$ Distribution

Searle's CI $\left\{ \dfrac{F/F_U - 1}{n + F/F_U - 1} \;,\; \dfrac{F/F_L - 1}{n + F/F_L - 1} \right\}$ is EXACT

Adjusted Searle's CI for unbalanced data

The fixed number of repeats, n, is replaced by:

$$n_0 = \overline{n}_i - \frac{\text{Variance}(n_i)}{k \times \overline{n}_i},$$

when n_i varies over subjects to give the CI:

$$\left\{ \frac{F/F_U - 1}{n_0 + F/F_U - 1} \quad , \quad \frac{F/F_L - 1}{n_0 + F/F_L - 1} \right\}$$

Note not exact if n_i varies

Bayesian intervals from SAS PROC MIXED

```
proc mixed data=datafile method=REML;
class subject;
model varX= ;
random subject ;
prior / nsample=10000 out=rejsamples;
run;
```

- **REML estimates of $\sigma^2_{subject}$ and σ^2_{repeat} requested**
- **subject included as a random effect**
- **doesn't directly produce an interval estimate for R**
- `prior / nsample=10000 out=rejsamples;` **requests 10,000 rejection samples be drawn and output to the file 'rejsamples'**
- **subsequent calculations on 'rejsamples' need to be requested by the user**

Rejection sampling

- **Bayesian monte-carlo technique**
- **samples from the joint posterior of**

  $\sigma^2_{subject}$ **and** σ^2_{repeat}
- **specifically developed for estimating variance components**

  Wolfinger RD & Kass RE. Nonconjugate Bayesian analysis of variance component models. Biometrics 2000; 56: 768-774.
- **suggest 10,000 samples be drawn**
- **by default PROC MIXED uses Jeffrey's priors**
  - **non-informative**
  - **hyperparameters estimated from data**

Motivating example

- test-retest reliability of walking parameters in people with Parkinson's disease $\geq$ 6 months
- participants assessed twice
  - a week apart (mean 6: min-max 3-14 days)
  - same time of day on each occasion,
  - no changes in medication
- first invited 10 people, then doubled sample size to 20
- 2 unwilling to travel to the laboratory (1 in each half)
- sporadic missing values

EXAMPLE: Point (interval) estimate of R for walking parameters

parameter		# with 2 readings	Searle's	adjusted Searle's	Bayesian
Speed (meters/min)	first 9	8	74 (20, 94)	75 (21, 94)	75 (24, 94)
	all 18	17	79 (68, 95)	79 (52, 92)	79 (52, 92)
Stride Length (meters)	first 9	8	98 (90, 100)	98 (91, 100)	98 (91, 100)
	all 18	17	93 (83, 97)	93 (83, 97)	93 (83, 98)
Cadence (steps/min)	first 9	8	53 (-17, 88)	49 (-24, 86)	50 (6, 86)
	all 18	17	68 (32, 87)	67 (30, 86)	67 (31, 86)

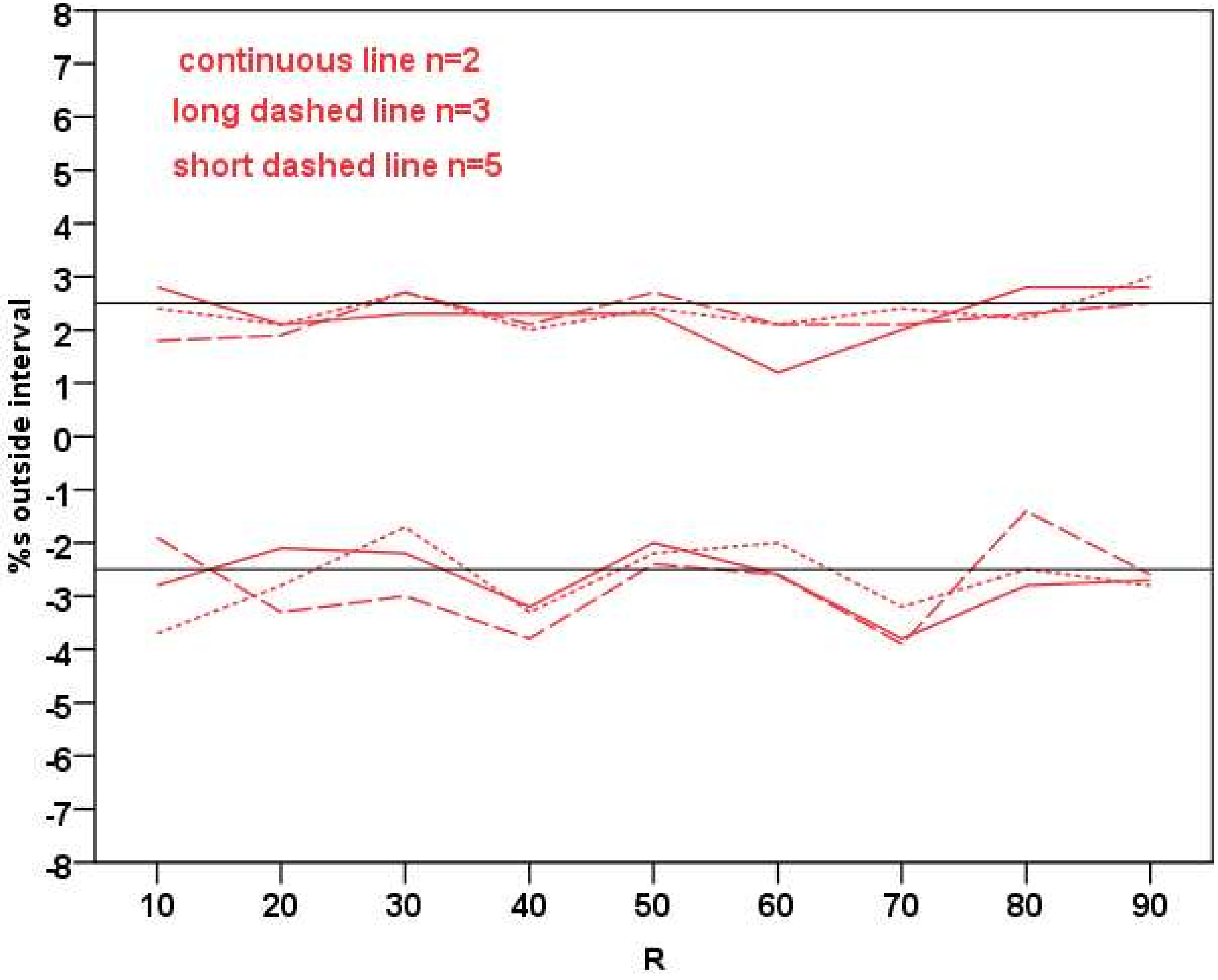
SIMULATIONS TO COMPARE THE PROPERTIES OF THE ADJUSTED SEARLE AND BAYESIAN INTERVALS

- size of simulated datasets similar to example data
  - N - number of subjects - 10, 20, 30
  - n - number of repeats - 2, 3, 5
- balanced or unbalanced (50% with only 1-2 repeats)
- subject effects ~ normal(0,1)
  or lognormal(0, 0.25) standardised to $\sigma^2_{subject}=1$
- error terms ~ Normal $(0,\sigma^2_{repeat})$
- range of σ^2_{repeat} giving R= 10%, 20%, … , 80%, 90%
- 1000 datasets simulated for each combination
  - PROC MIXED simulated 10,000 rejection samples for each one

RESULTS - COVERAGE

upper lines - % true R >upper limit
lower lines - % true R <lower limit

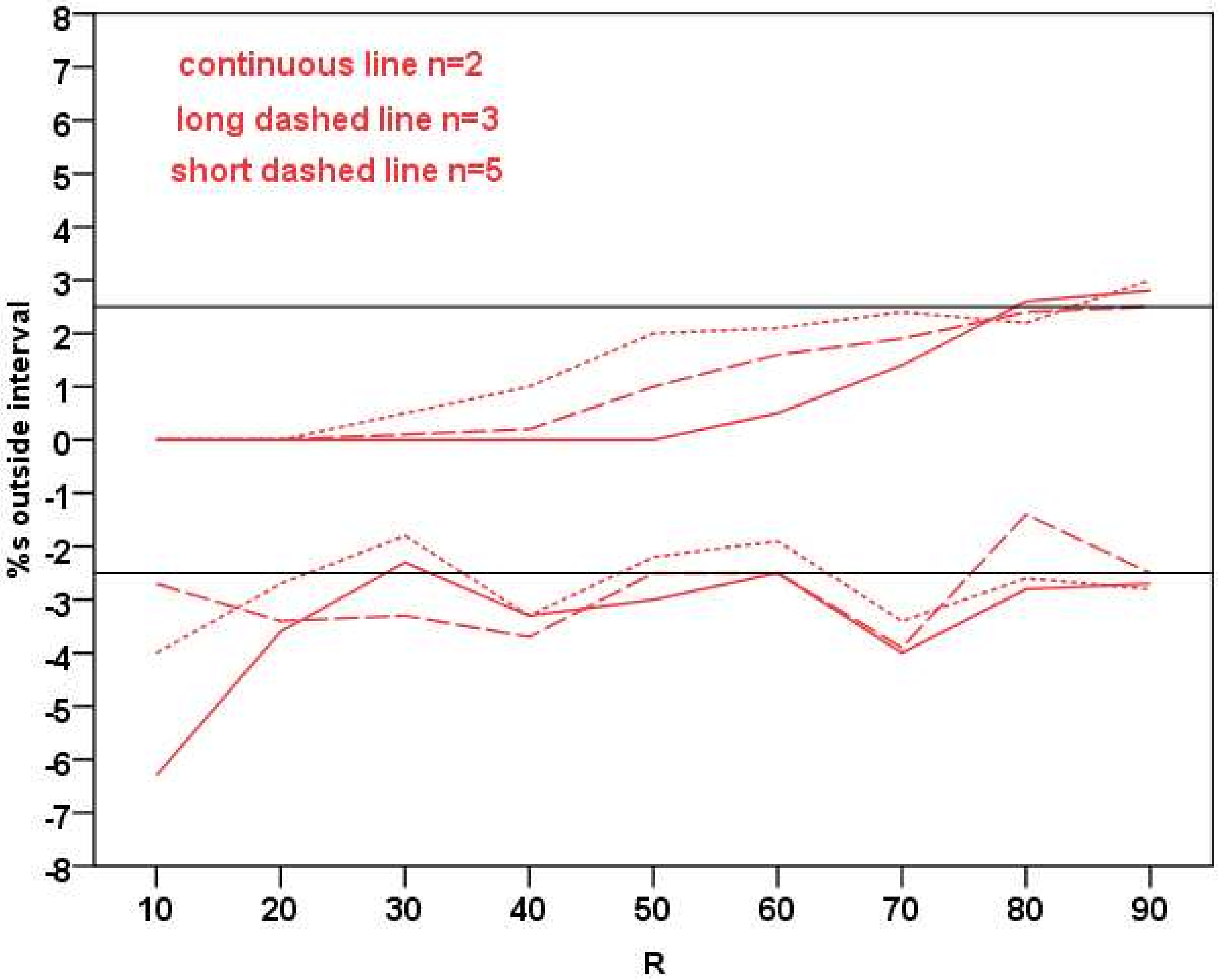
Searle's CI with balanced data: number of subjects N=10



RESULTS - COVERAGE

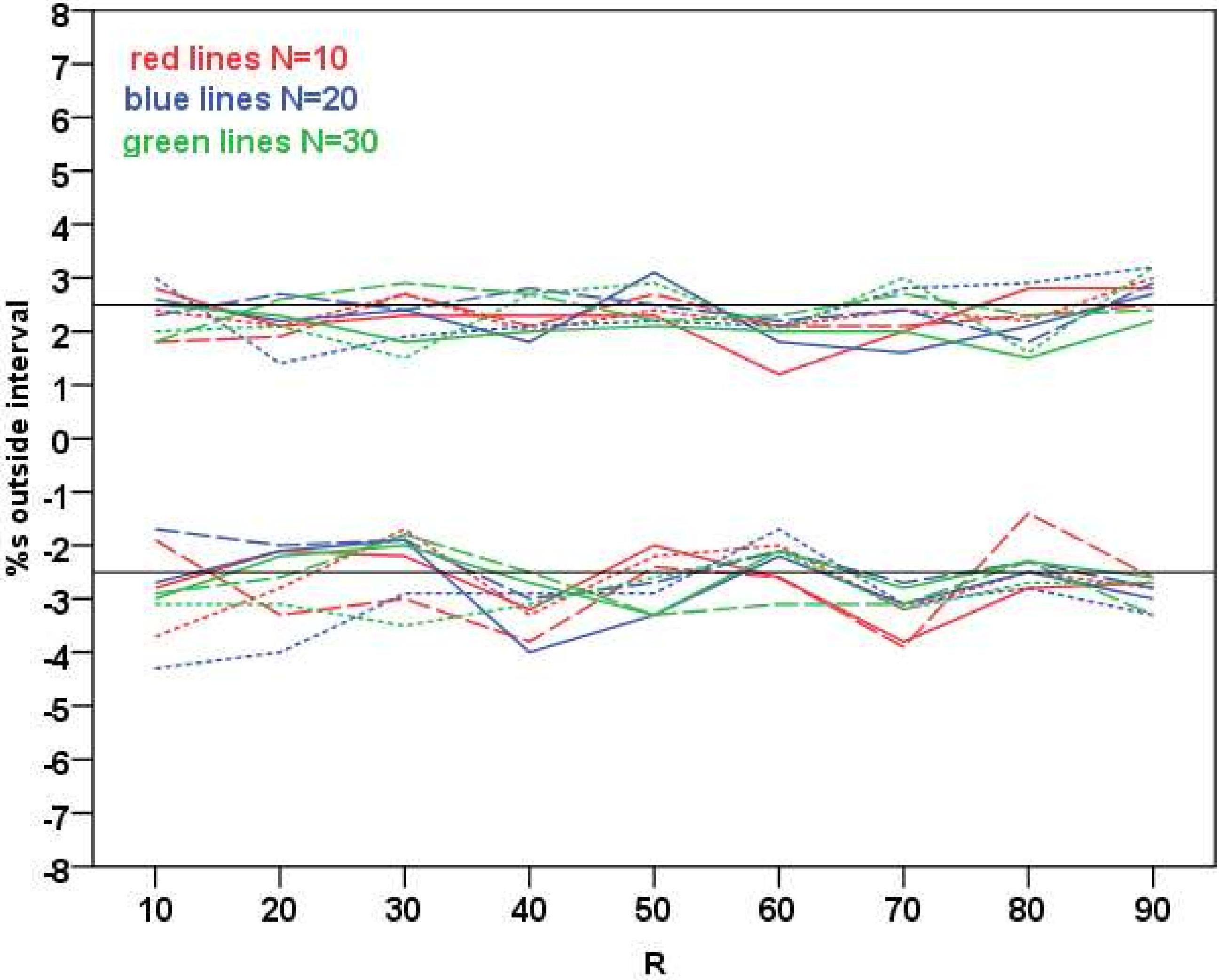
upper lines - % true R >upper limit
lower lines - % true R <lower limit

Bayesian interval with balanced data: N=10



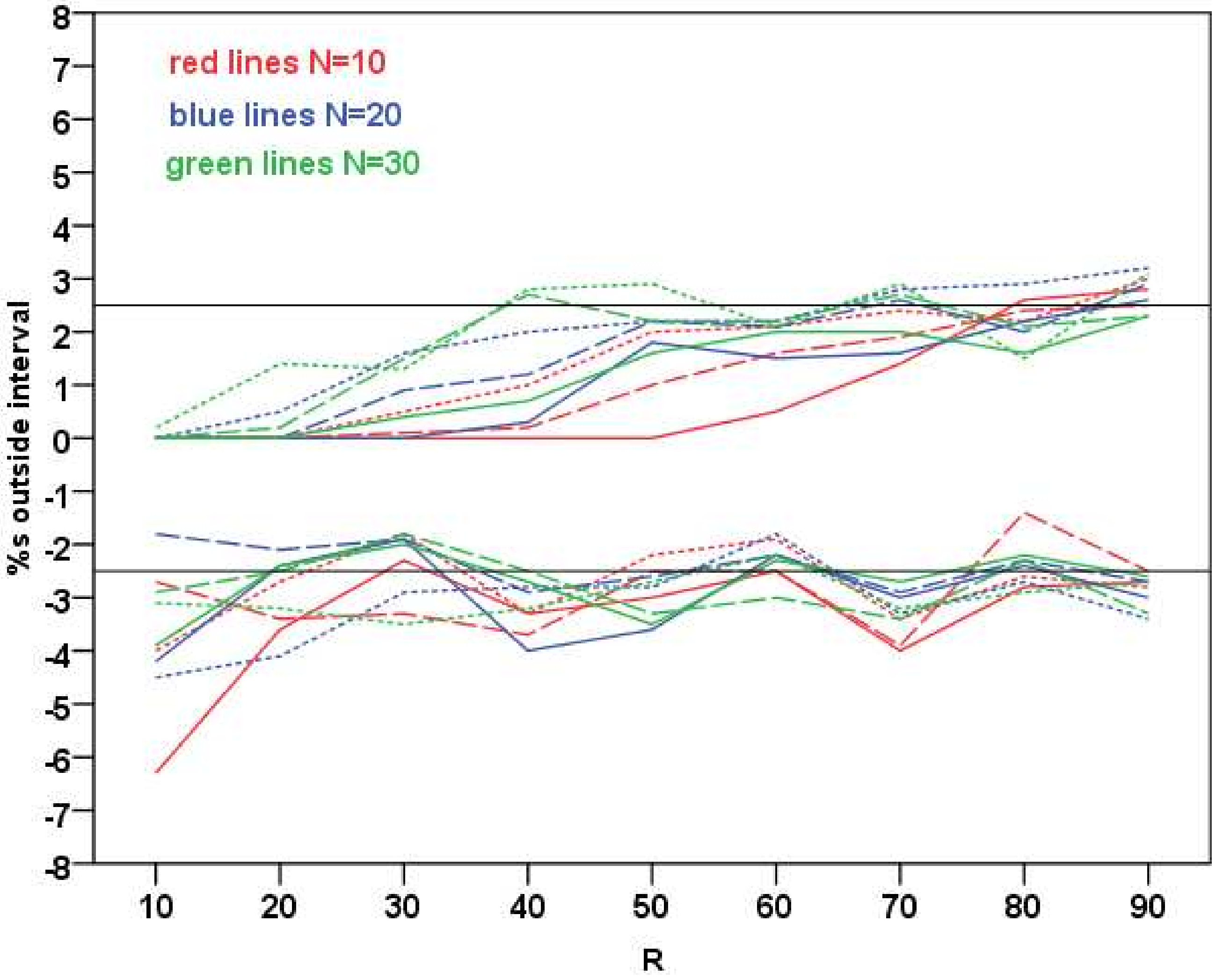
RESULTS - COVERAGE
upper lines - % true R >upper limit
lower lines - % true R <lower limit

Searle's CI with balanced normal data



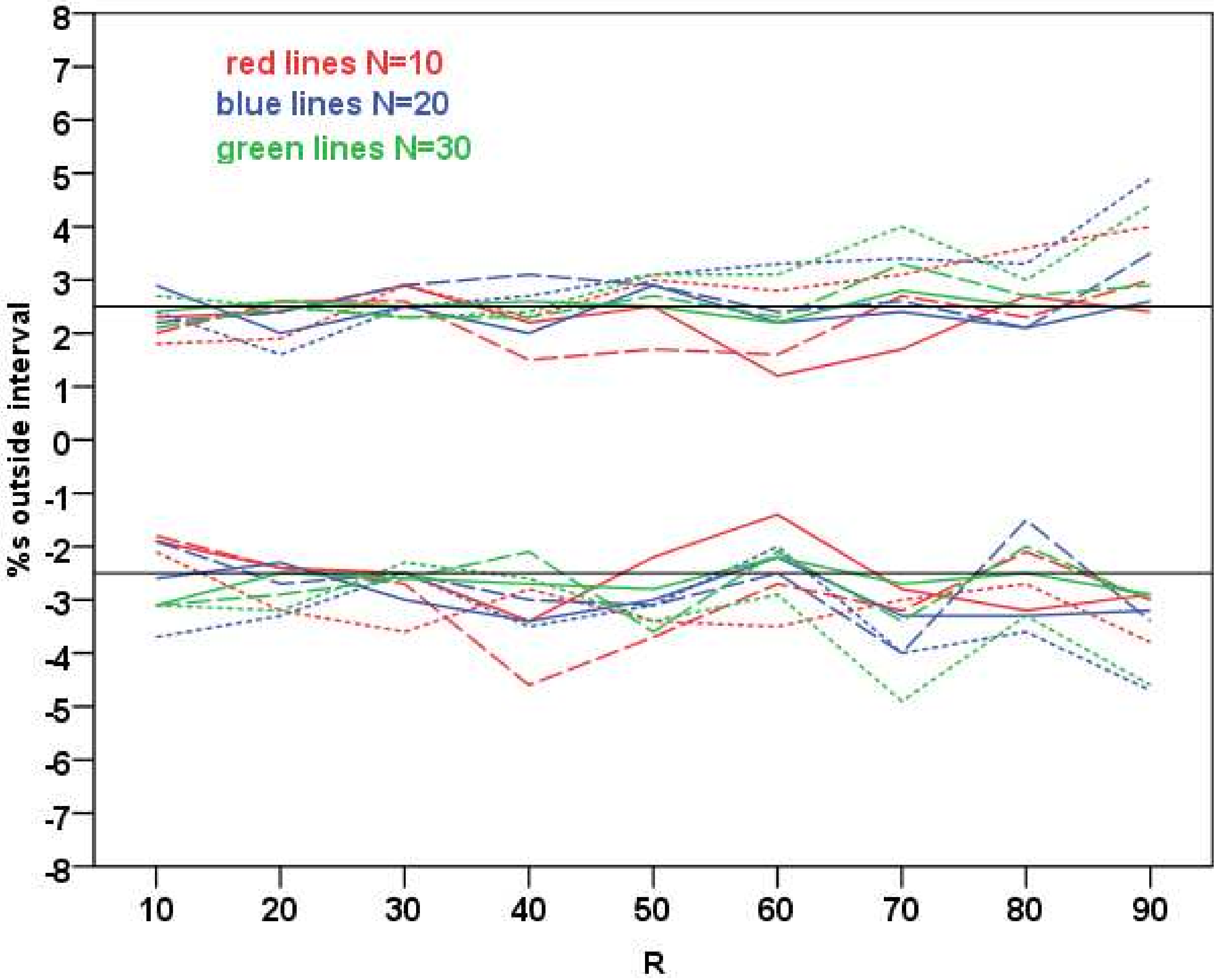
RESULTS - COVERAGE
upper lines - % true R >upper limit
lower lines - % true R <lower limit

Bayesian interval with balanced normal data



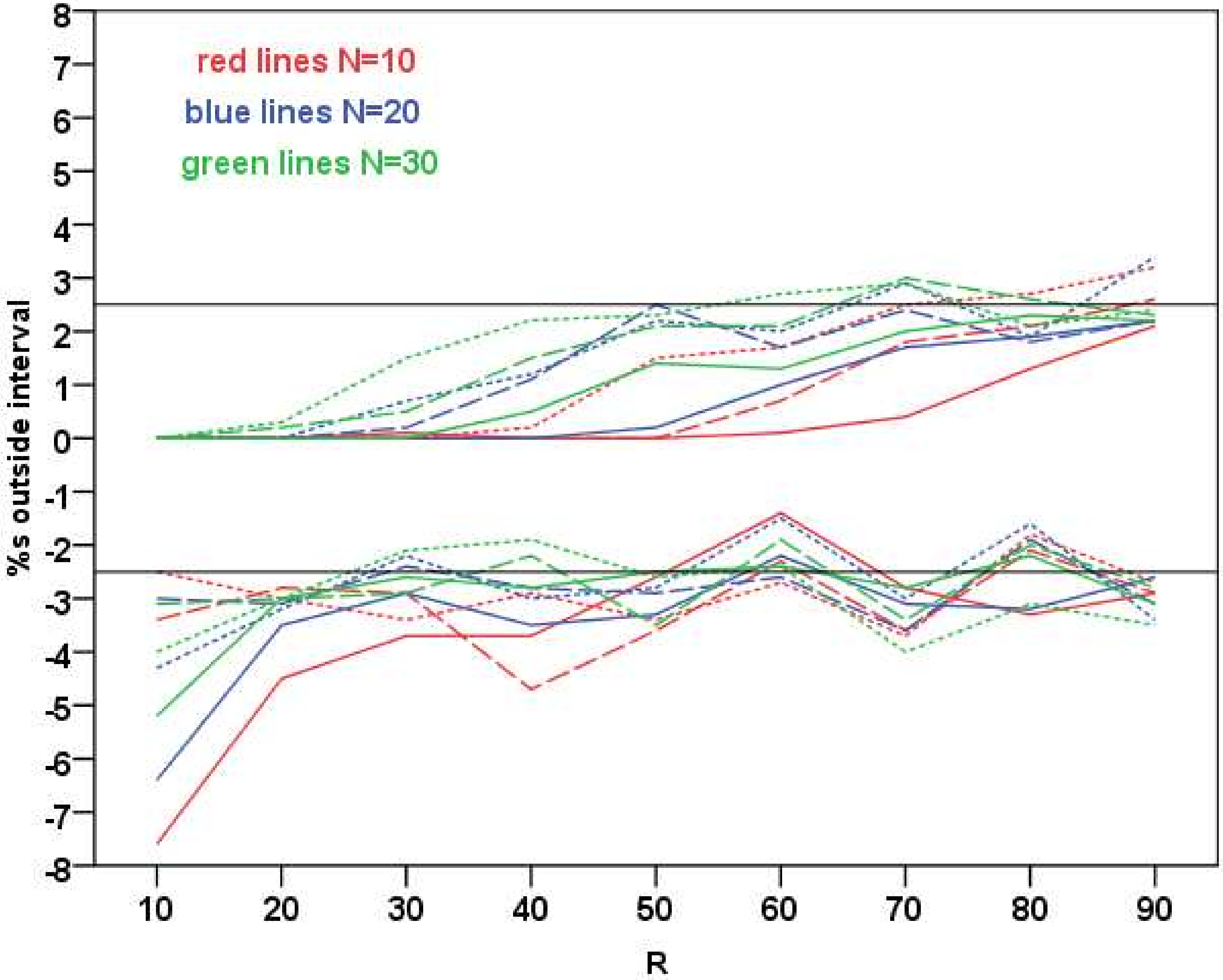
RESULTS - COVERAGE
upper lines - % true R >upper limit
lower lines - % true R <lower limit

Searle’s CI with unbalanced normal data



COVERAGE
lower lines - % true R <lower limit
upper lines - % true R >upper limit

Bayesian interval with unbalanced normal data

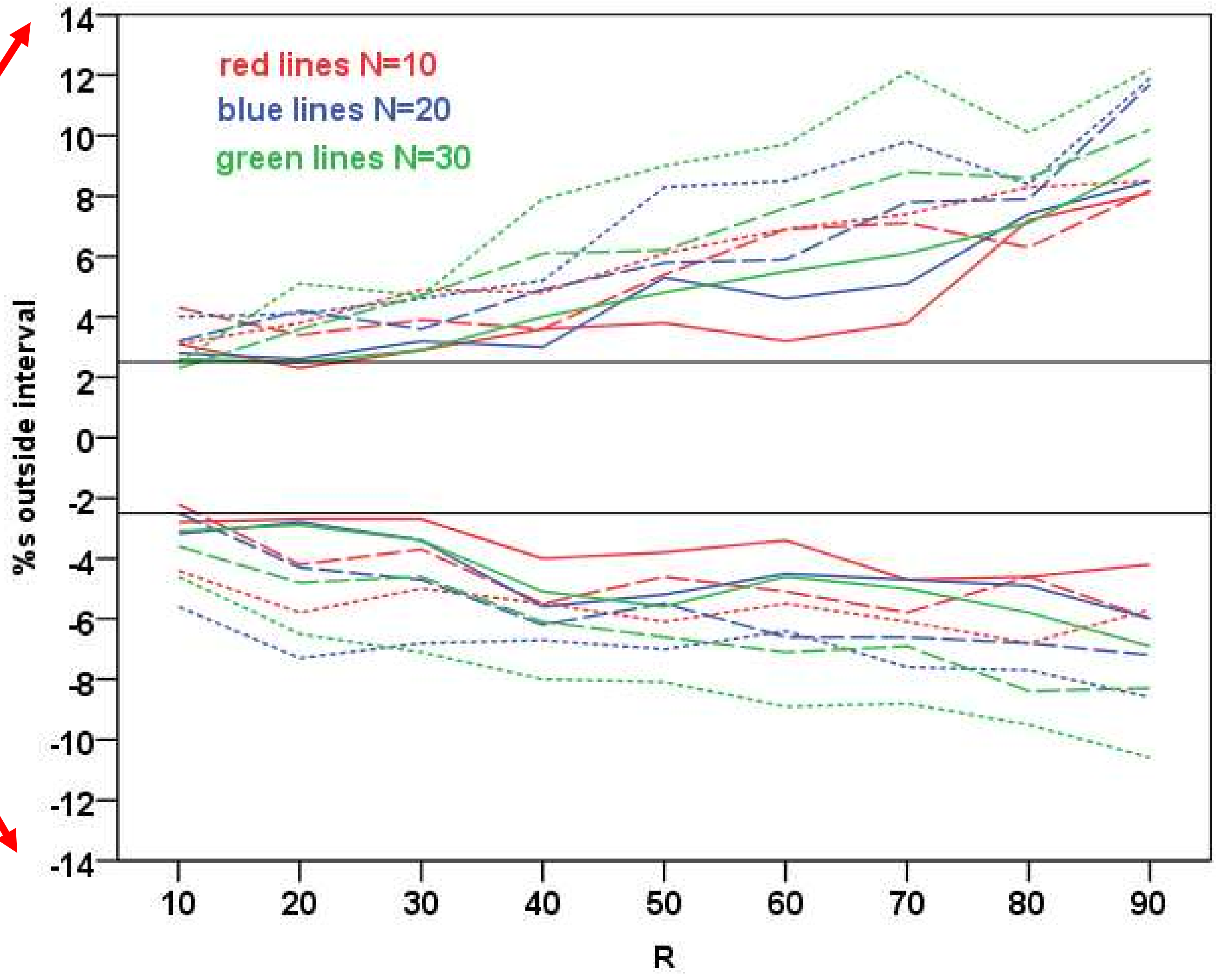


RESULTS - COVERAGE
upper lines - % true R >upper limit
lower lines - % true R <lower limit

Searle's CI with balanced non-normal data

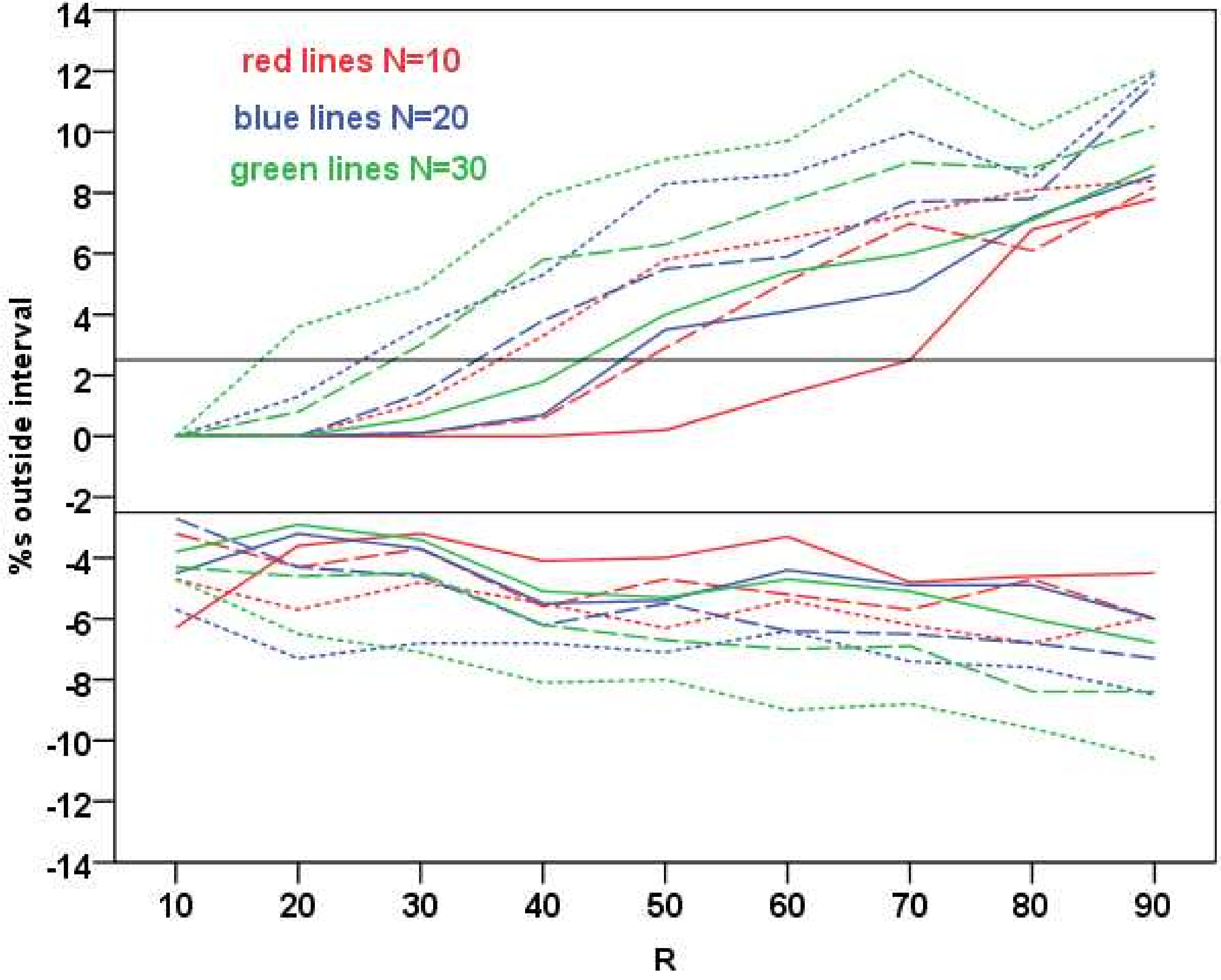
Note change of scale

14% vs 8%



RESULTS - COVERAGE
upper lines - % true R >upper limit
lower lines - % true R <lower limit

Bayesian interval with balanced non-normal data



MEAN WIDTH OF INTERVALS
Normal balanced data

		Searle's		Bayesian
		mean width	mean +ve width	mean width
N=10 n=2	R=10	108.6	61.9	67.3
	R=90	34.4	34.2	33.8
N=10 n=5	R=10	52.3	44.3	47.0
	R=90	21.7	21.7	21.7
N=30 n=2	R=10	67.7	42.6	45.0
	R=90	15.4	15.4	15.4
N=30 n=5	R=10	29.0	25.7	25.9
	R=90	11.1	11.1	11.1

SUMMARY OF PROPERTIES OF INTERVALS

- the Bayesian interval is too high for small R
  - less of a problem for large N (subjects) & n (repeats)
  - less/no problem for R typical of reliability studies
  - evident in balanced/unbalanced, normal/non-normal data
- Searle's and Bayesian intervals similar for large R in balanced normal datasets
- unbalanced data with large R: Searle's CI undercovers for large R, N & n: the Bayesian interval has close to nominal coverage for large R
- non-normal data: both intervals undercover for large R
  - more of a problem for large N & n
- Bayesian interval wider (up to 8%) when R small
  - similar widths for large R typical of measurement studies

CONCLUSIONS

- Bayesian interval easy to obtain from PROC MIXED with properties similar to Searle's CI for R typical of reliability studies
  - probably not satisfactory for small R
- balance no issue with the Bayesian interval
- serious under-coverage for both intervals at larger R when data is not normal

REFERENCES

- Wolfinger RD & Kass RE. Nonconjugate Bayesian analysis of variance component models. Biometrics 2000; 56: 768-774.
- Donner A, Wells G. A comparison of confidence interval methods for the intraclass correlation coefficient. Biometrics 1985; 42: 401-412.
- Burdick RK, Graybill FA. Confidence intervals on variance components. Marcel Dekker: New York, 1992.
- Ukoumunne OC. A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. Statistics in Medicine 2002; 21: 3757-3774.
- Ukoumunne OC, Davison AC, Gulliford MC, Chinn S. Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient. Statistics in Medicine 2003; 22; 3805-21.
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O. SAS® for Mixed Models, second edition. SAS Publishing 2006.
- Searle SR. Linear Models. Wiley: New York, 1971.

Point estimates of R

- Searle's method : based on estimates of $\sigma^2_{subject}$ and σ^2_{repeat} incorporating n_0, if n_i vary
  - truncated at zero if $\sigma^2_{subject}$ -ve
- R calculated from the REML estimates
- the median of the 10,000 R calculated from each rejection sample

SUMMARY OF PROPERTIES OF POINT ESTIMATES

- **balanced normal datasets**
  - mean estimates ordered according to

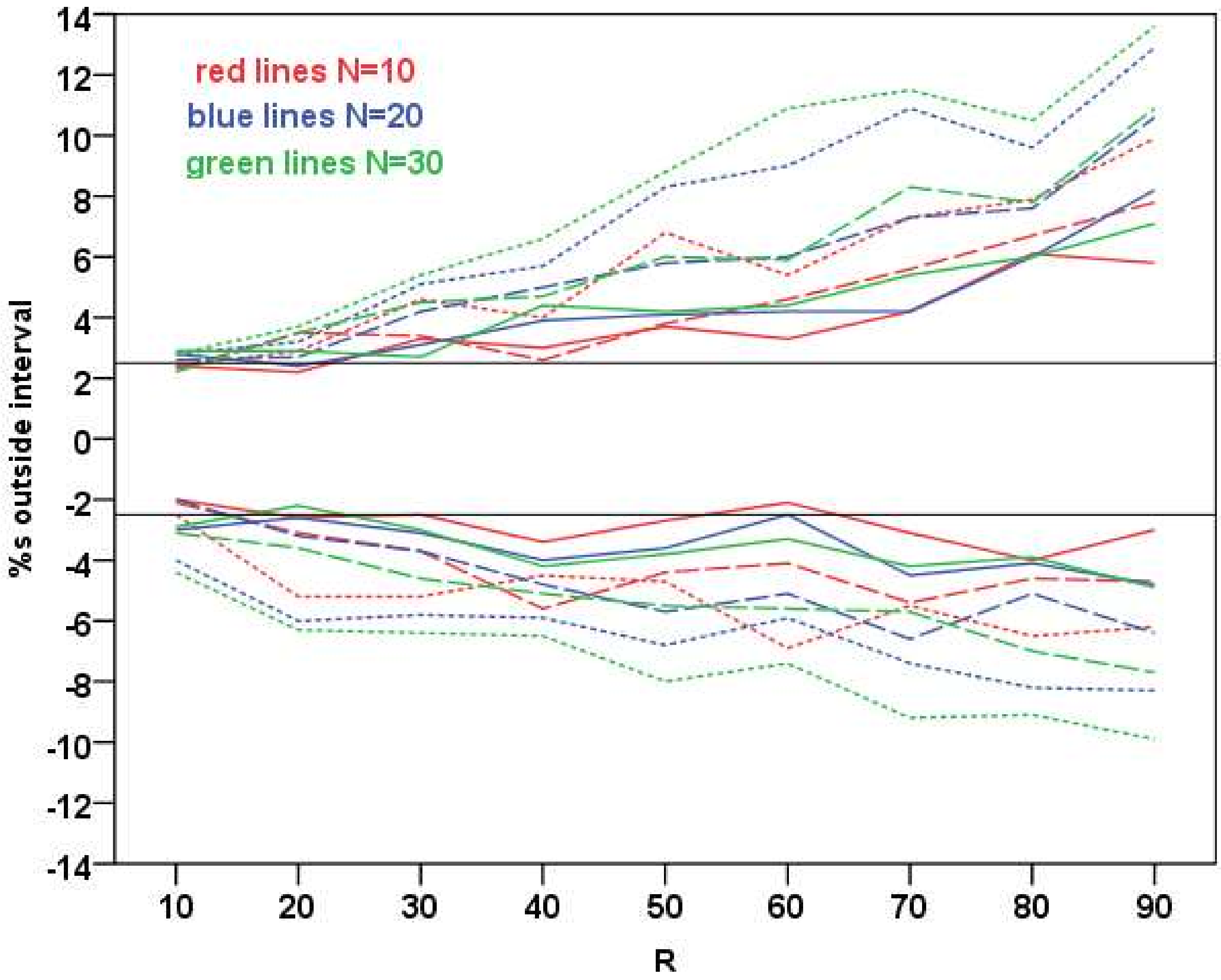
  true ≈ Searle's < truncated Searle's ≈ REML < median rejection
  - for small N, n and R
  - eg N=10, n=2, R=10% ≈ 10.3 < 18.7 ≈ 18.7 < 30.5
  - all similar for larger R
  - eg N=10, n=2, R=90% ≈ 87.7 ≈ 87.7 ≈ 87.7 ≈ 87.8
- **unbalanced normal datasets**
  - similar pattern with greater over-estimation
  - eg N=10, n=2, R=10% ≈ 10.2 < 24.5 ≈ 23.5 < 39.6
  - eg N=10, n=2, R=90% ≈ 87.7 ≈ 87.7 ≈ 87.5 ≈ 87.4
- **non-normal datasets**
  - similar to balanced datasets
  - eg N=10, n=2, R=10% ≈ 9.6 < 18.5 ≈ 18.5 < 30.3
  - eg N=10, n=2, R=90% ≈ 84.8 ≈ 84.8 ≈ 84.8 ≈ 85.1

Use of point estimates

- present R from REML estimates with the Bayesian interval?

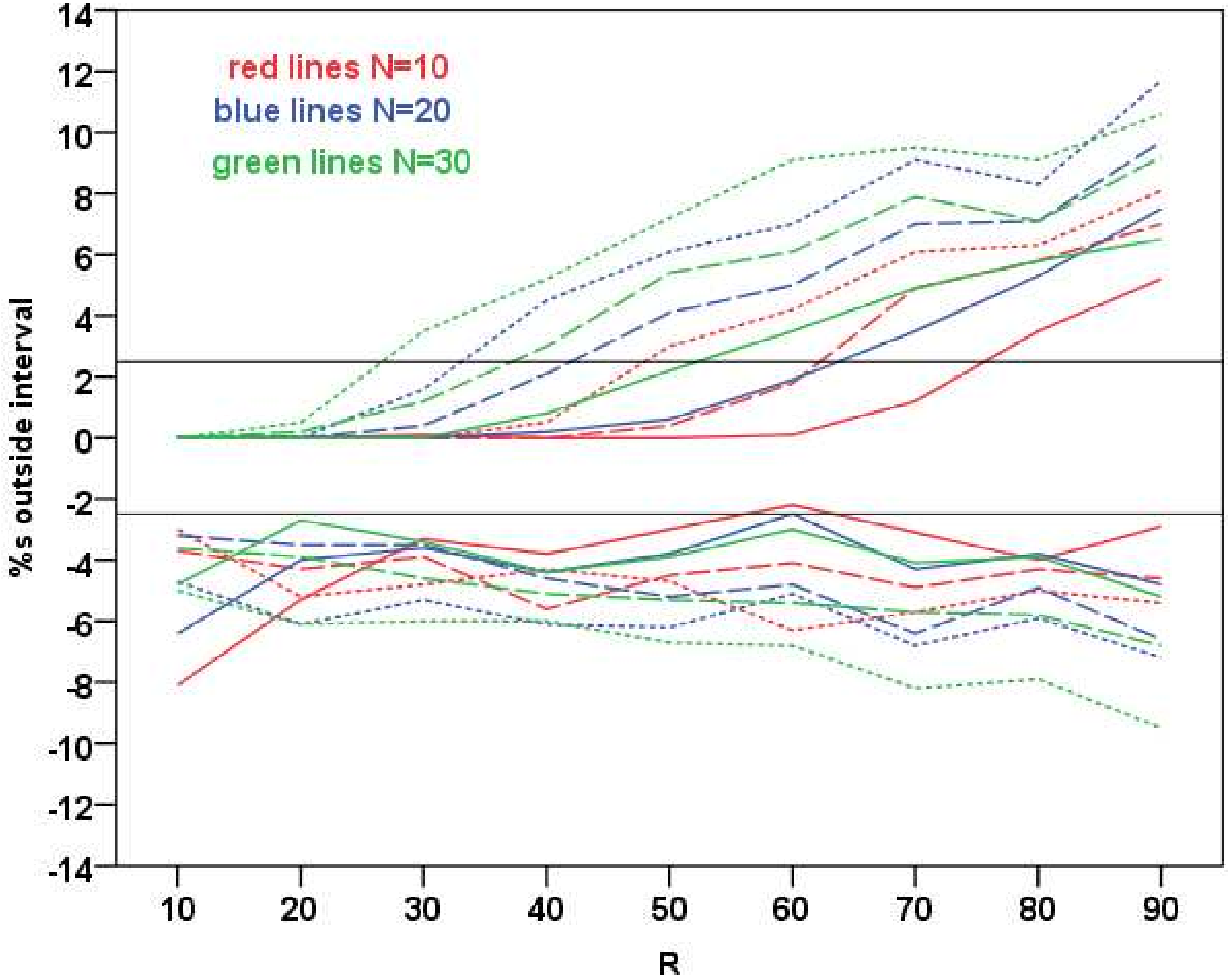
RESULTS - COVERAGE
upper lines - % true R >upper limit
lower lines - % true R <lower limit

Searle's CI with unbalanced non-normal data



RESULTS - COVERAGE
upper lines - % true R >upper limit
lower lines - % true R <lower limit

Bayesian interval with unbalanced non-normal data



MEAN WIDTH OF INTERVALS
Normal unbalanced data

		Searle's		Bayesian
		mean width	mean +ve width	mean width
N=10 n=2	R=10	181.9	70.6	75.5
	R=90	91.6	53.7	48.7
N=10 n=5	R=10	69.5	50.7	57.1
	R=90	25.2	25.2	25.6
N=30 n=2	R=10	108.8	51.4	54.8
	R=90	21.7	21.7	21.9
N=30 n=5	R=10	40.9	32.2	33.6
	R=90	12.4	12.4	12.5

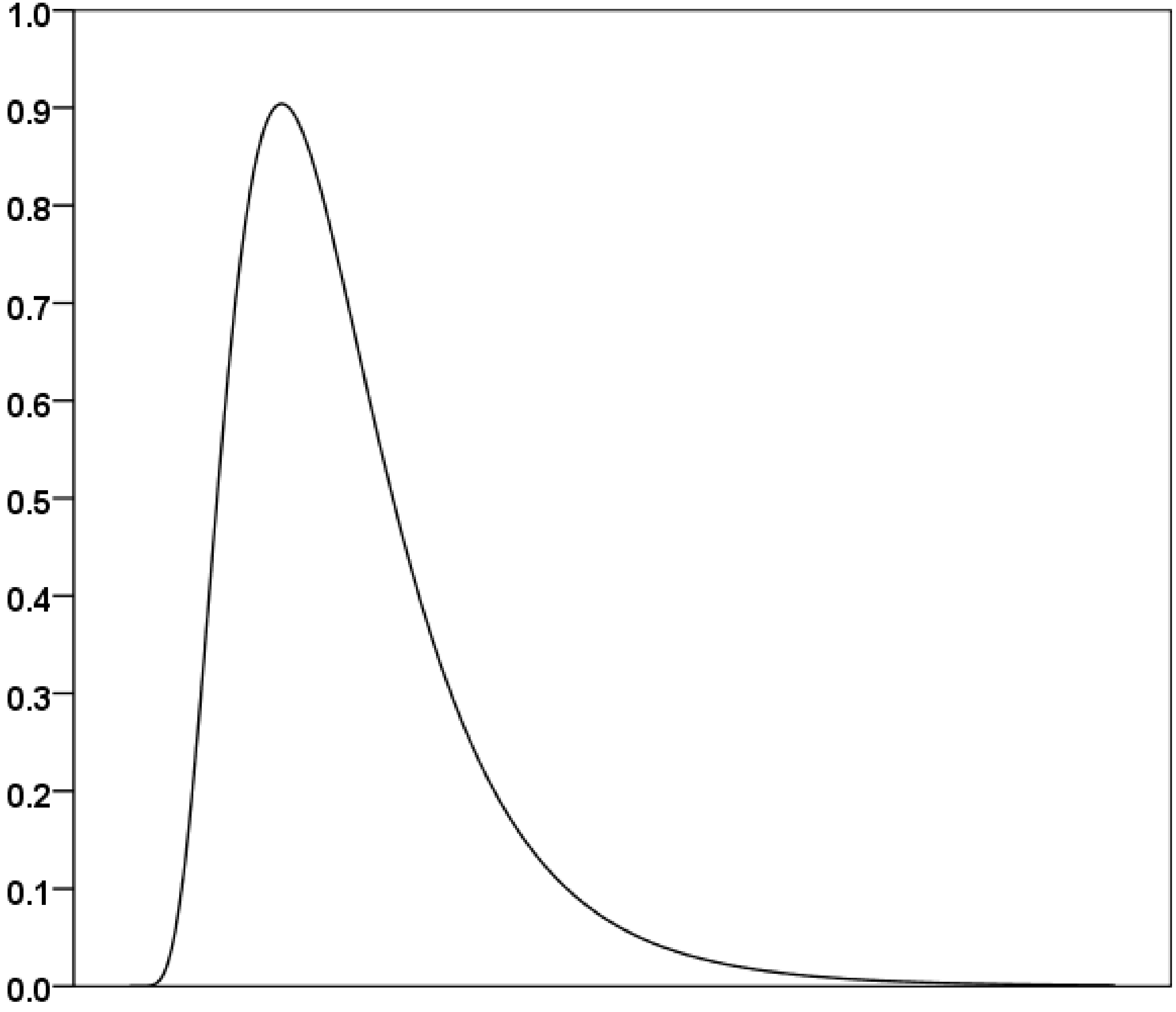
MEAN WIDTH OF INTERVALS
non-normal balanced data

		Searle's		Bayesian
		mean width	mean +ve width	mean width
N=10 n=2	R=10	108.5	61.4	67.0
	R=90	39.5	38.9	38.0
N=10 n=5	R=10	51.7	43.3	46.4
	R=90	24.9	24.9	25.0
N=30 n=2	R=10	67.7	42.4	44.8
	R=90	16.8	16.8	16.8
N=30 n=5	R=10	28.8	25.3	25.6
	R=90	12.0	12.0	12.0

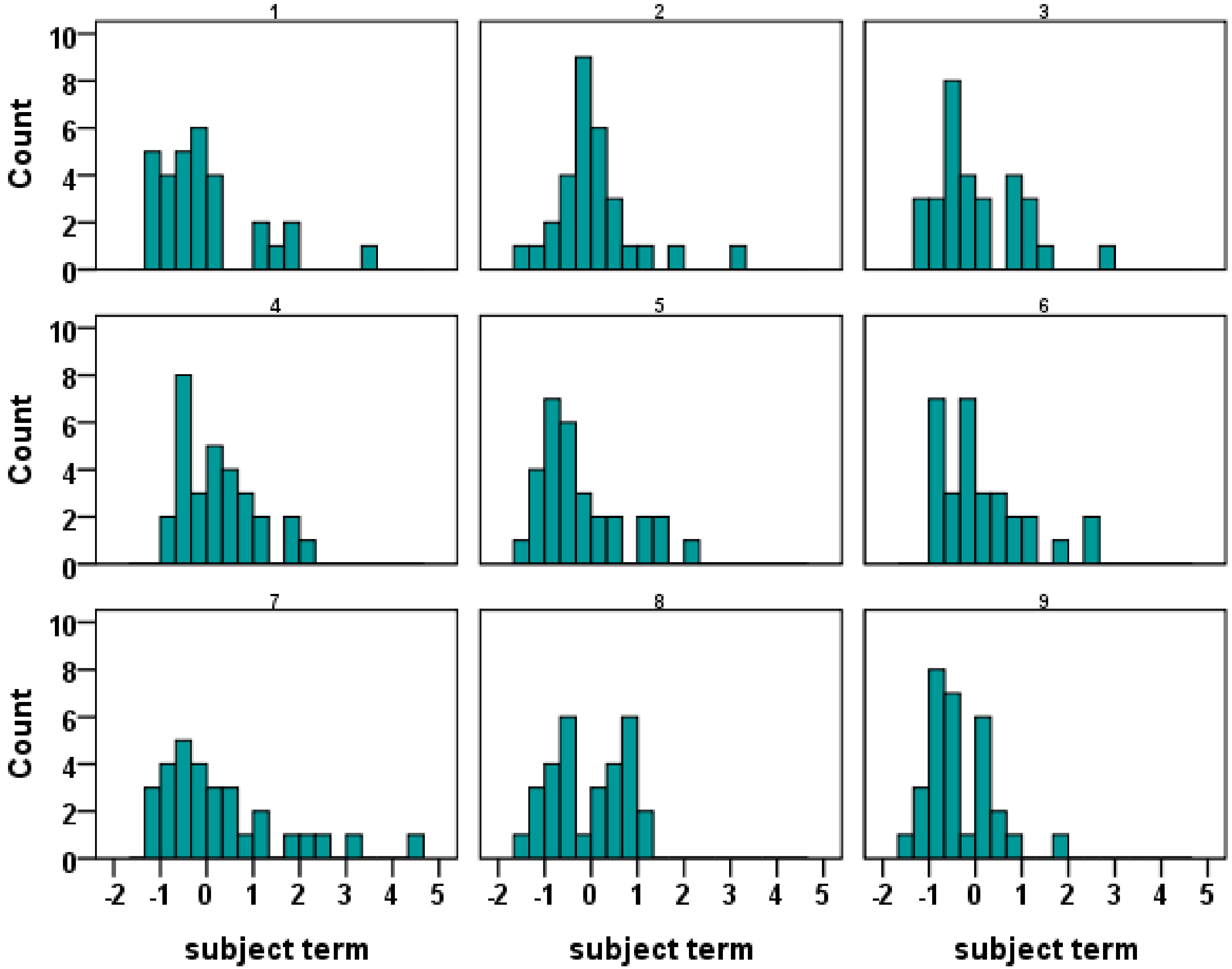
MEAN WIDTH OF INTERVALS
non-normal unbalanced data

		Searle's		Bayesian
		mean width	mean +ve width	mean width
N=10 n=2	R=10	182.0	70.3	75.4
	R=90	66.1	57.5	52.0
N=10 n=5	R=10	69.3	50.2	54.8
	R=90	29.0	29.0	29.5
N=30 n=2	R=10	108.9	51.3	54.8
	R=90	23.7	23.7	23.9
N=10 n=5	R=10	40.7	31.7	33.1
	R=90	13.6	13.6	13.6

pdf of a lognormal(0,0.25) skewness=1.75



Simulated non-normal subjects effects for the first 9 datasets, N=30, n_i=2, R=10%



When is Searle's lower limit <0?

lower limit = $\dfrac{F/F_U - 1}{n + F/F_U - 1}$

-ve if F= $\dfrac{MS_{subject}}{MS_{error}} < F_U$

For samples of N=9, n=2 $F^{N-1=8}_{N(n-1)=9}$= 4.102

For cadence $MS_{subject}$ = 442.9, MS_{serror} = 134.0

For 'reasonable' values of N & n, 2< $F^{N-1}_{N(n-1)}$ <3

N and n $\rightarrow \infty$ $F^{N-1}_{N(n-1)} \rightarrow 1$