WARWICK
CLINICAL TRIALS UNIT

The secret life of the confidence interval

Simon Gates

Confidence intervals

WARWICK
CLINICAL TRIALS UNIT

Risk ratio 1.12,

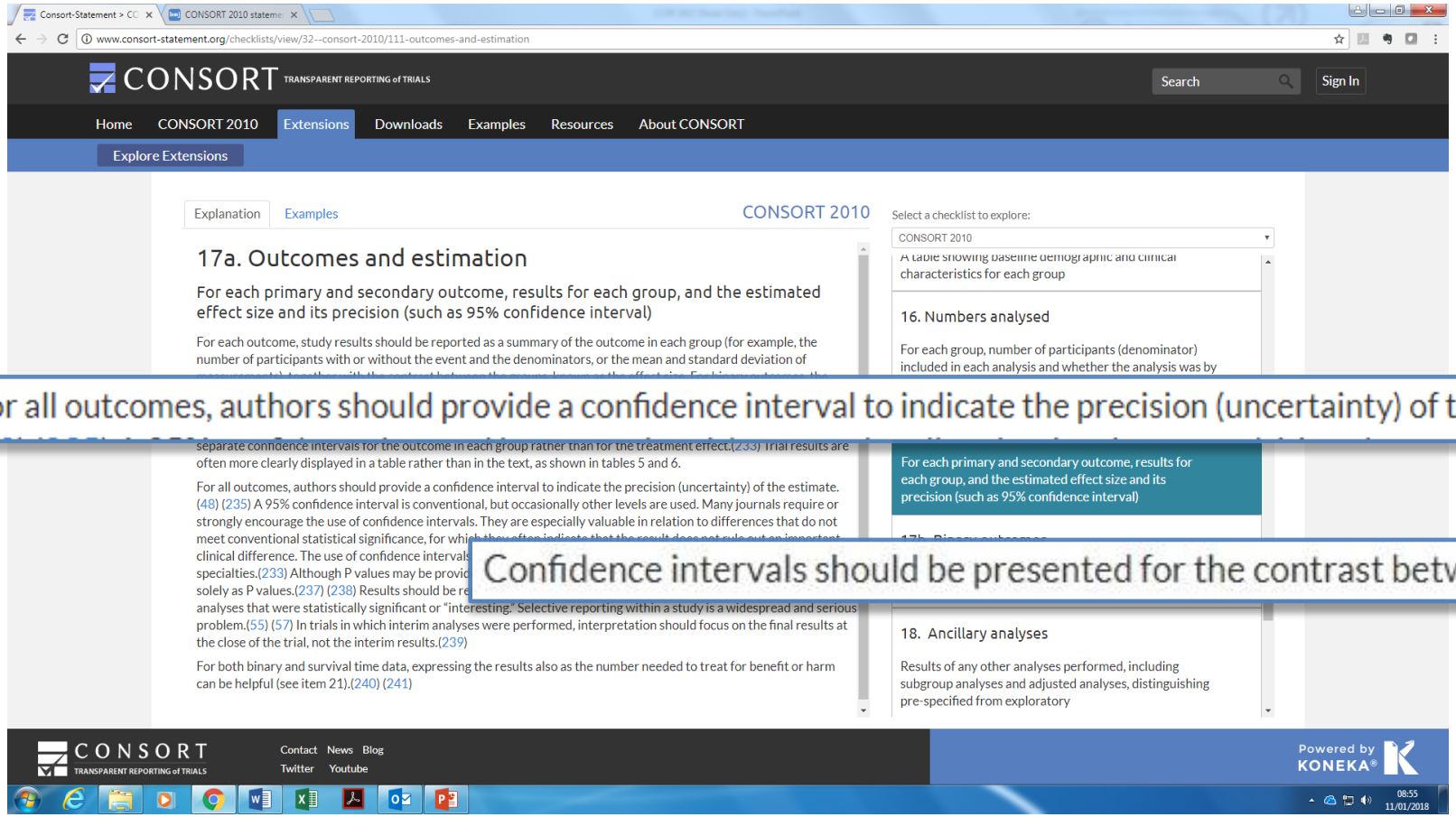
95% confidence interval 0.95, 1.32

WARWICK

CLINICAL TRIALS UNIT

WARWICK

CLINICAL TRIALS UNIT



Why have CI been promoted?

WARWICK
CLINICAL TRIALS UNIT

- Over-reliance on significance tests and assumption that “significant” = clinically important
- Estimate “uncertainty” or “precision” of point estimates
- Estimate the range of values that the parameter could take

or their differences. The implication of hypothesis testing—that there can always be a simple “yes” or “no” answer as the fundamental result from a medical study—is clearly false and used in this way hypothesis testing is of limited value.[2]

Gardner and Altman, BMJ 1986

What do confidence intervals mean?

WARWICK
CLINICAL TRIALS UNIT

Audience participation

What do confidence intervals mean?

WARWICK
CLINICAL TRIALS UNIT

The confidence interval indicates the level of uncertainty around the measure of effect (precision of the effect estimate) …
…by having an upper and lower confidence limit we can infer that the true population effect lies between these two points.
If the confidence interval crosses 1 e.g. 95%CI 0.9-1.1 this implies there is no difference between arms of the study.

(website intended for medical students)

ference between two such estimates. The CI is a range of values either side of the estimate between which we can be 95% sure that the true value lies. A series of identical studies carried out

(paper by a famous statistician 2005)

effect. A 95% confidence interval offers the range of values for which there is 95% certainty that the true value of the parameter lies within the confidence limits, i.e., the results of

(paper in a clinical journal 2017)

General beliefs

WARWICK
CLINICAL TRIALS UNIT

- 95% CI show range where true value is most likely to be (95% probability)
- 95% CI shows the range of most plausible or most likely values for the treatment effect
- True difference is more likely to be near the middle of the CI
- If experiment is repeated, 95% of estimates of treatment effect will fall in the 95% CI
- Values outside a 95% confidence interval are unlikely (probability of < 5%)

How are CI used?

WARWICK
CLINICAL TRIALS UNIT

- Often used as surrogate significance tests
  - If zero effect is outside 95% CI, difference is significant
  - If not, non-significant
- Rarely any consideration of uncertainty or range of effects consistent with the data

How are CI used?

WARWICK
CLINICAL TRIALS UNIT

- Sample of 111 “non-significant” treatment comparisons
- Four leading medical journals (Lancet, BMJ, JAMA, NEJM)
- Only FOUR comparisons (3 papers) referred to confidence interval or uncertainty when drawing conclusions
- 60% concluded “treatment not effective,” or “no difference,” and 24% “no significant difference” even if CI clearly included clinically important values

The NEW ENGLAND JOURNAL of MEDICINE

WARWICK
CLINICAL TRIALS UNIT

ORIGINAL ARTICLE

Ticagrelor versus Aspirin in Acute Stroke or Transient Ischemic Attack

S. Claiborne Johnston, M.D., Ph.D., Pierre Amarenco, M.D., Gregory W. Albers, M.D., Hans Denison, M.D., Ph.D., J. Donald Easton, M.D., Scott R. Evans, Ph.D., Peter Held, M.D., Ph.D., Jenny Jonasson, Ph.D., Kazuo Minematsu, M.D., Ph.D., Carlos A. Molina, M.D., Yongjun Wang, M.D., and K.S. Lawrence Wong, M.D., for the SOCRATES Steering Committee and Investigators*

Primary outcome (time to stroke, MI or death):
Hazard ratio 0.87, 95% CI 0.76, 1.00

CONCLUSIONS

In our trial involving patients with acute ischemic stroke or transient ischemic attack, ticagrelor was not found to be superior to aspirin in reducing the rate of stroke, myocardial infarction, or death at 90 days. (Funded by AstraZeneca;

WARWICK
CLINICAL TRIALS UNIT

Effectiveness of a 6-year multidomain vascular care intervention to prevent dementia (preDIVA): a cluster-randomised controlled trial

Eric P Moll van Charante*, Edo Richard*, Lisa S Eurelings, Jan-Willem van Dalen, Suzanne A Ligthart, Emma F van Bussel, Marieke P Hoevenaar-Blom, Marinus Vermeulen, Willem A van Gool

Dementia; intervention 121/1853, control 112/1601
HR 0.92, 95% CI 0.71, 1.19

Conclusion: “Intervention did not result in a reduced incidence of all-cause dementia.”

What is the problem?

WARWICK
CLINICAL TRIALS UNIT

- CIs don't mean what people usually think they mean
- Real meaning is a bit different, and not what people expect.

  - 95% CI show range where true value is most likely to be (95% probability)
  - 95% CI shows the range of most plausible or most likely values for the treatment effect
  - True difference is more likely to be near the middle of the CI
  - If experiment is repeated, 95% of estimates of treatment effect will fall in the 95% CI
  - Values outside a 95% confidence interval are unlikely (probability of < 5%)

Correct definition of 95% CI

WARWICK
CLINICAL TRIALS UNIT

An interval for a parameter constructed by a method that ensures that, in a long run of replications, 95% of the intervals will include the true value of the parameter.

Correct definition of 95% CI

WARWICK
CLINICAL TRIALS UNIT

"Method that ensures that, in a long run of replications, 95% of the intervals will include the true value" is a CONFIDENCE PROCEDURE

Intervals it produces are CONFIDENCE INTERVALS.

Correct definition of 95% CI

WARWICK
CLINICAL TRIALS UNIT

95% probability of including the true value is a property of the population of all possible CIs

NOT

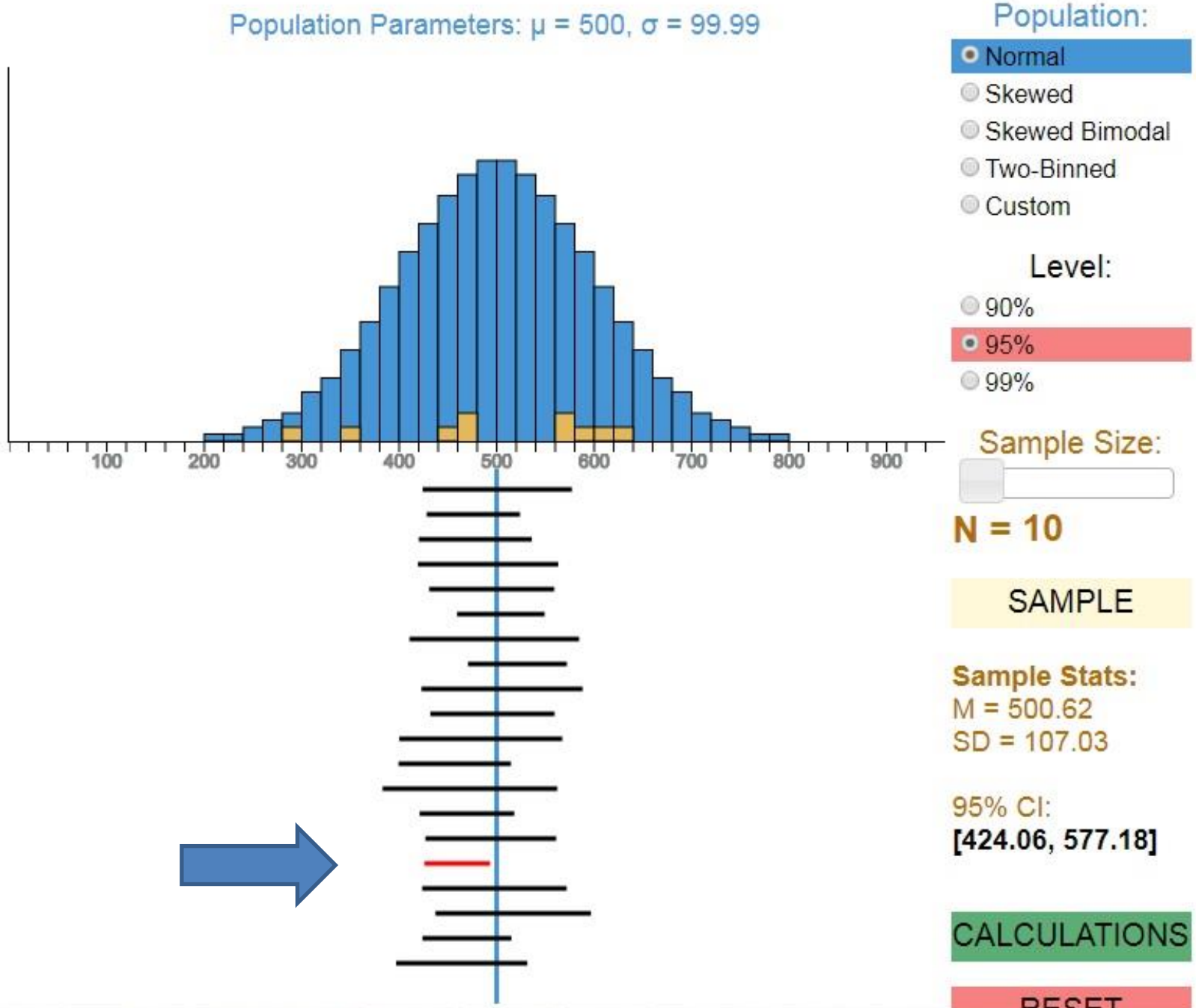
any individual CI that results from a single experiment

Some interesting consequences…

WARWICK
CLINICAL TRIALS UNIT

There is not a unique correct confidence procedure

- Many different procedures give valid confidence intervals
- (for 50% CI) half of the time make a CI with zero length, half the time with infinite length
- What would this mean for the usual interpretations of CIs?


Population Parameters: μ = 500, σ = 99.99
100
200
300
400
500
600
700
800
900
Population:
Normal
Skewed
Skewed Bimodal
Two-Binned
Custom
Level:
90%
95%
99%
Sample Size:
N = 10
SAMPLE
Sample Stats:
M = 500.62
SD = 107.03
95% CI:
[424.06, 577.18]
CALCULATIONS
RESET

WARWICK
CLINICAL TRIALS UNIT

95% of all possible 95% confidence intervals include the true value

But a single study gives us only one confidence interval

Why can't we say that a single 95% CI shows a range with 95% probability of including the true value?

Why can't we say that?

WARWICK
CLINICAL TRIALS UNIT

Probability that an interval contains the true value

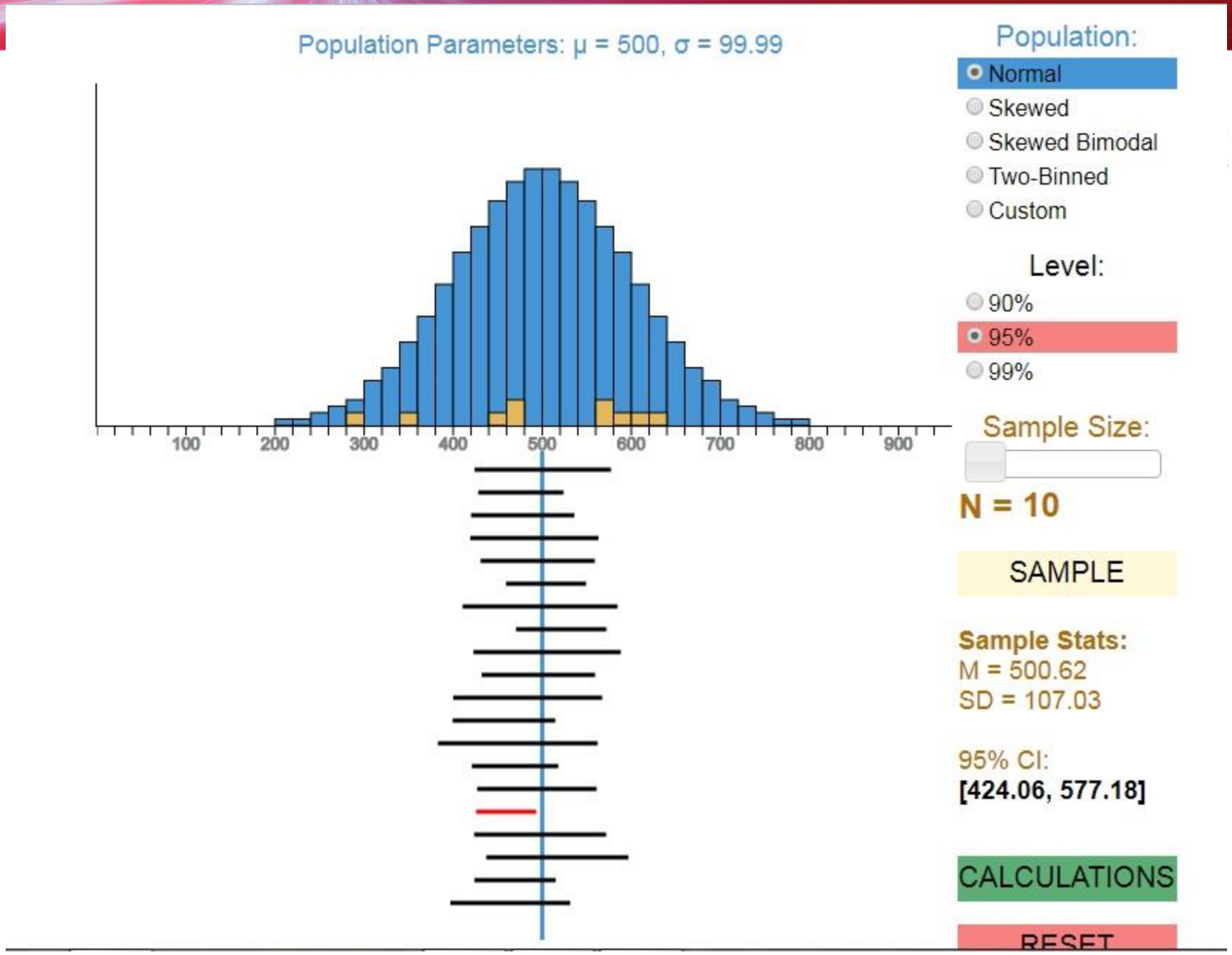
is not the same as

Probability that the true value is between the limits of a specific interval

Probabilities of different things:

- any interval containing true value; all have different limits

- true value being in a specific range defined by one interval


Population Parameters: μ = 500, σ = 99.99
100
200
300
400
500
600
700
800
900
Population:
Normal
Skewed
Skewed Bimodal
Two-Binned
Custom
Level:
90%
95%
99%
Sample Size:
N = 10
SAMPLE
Sample Stats:
M = 500.62
SD = 107.03
95% CI:
[424.06, 577.18]
CALCULATIONS
RESET

Another argument

WARWICK
CLINICAL TRIALS UNIT

- There is not one unique confidence procedure
- Many procedures can give valid confidence intervals

[confidence procedure produces intervals with the property that true value is in the interval with specified probability]

- Compare two confidence procedures

With acknowledgement to Richard Morey

WARWICK
CLINICAL TRIALS UNIT

Suppose we draw $N = 10$ normal observations with known variance.

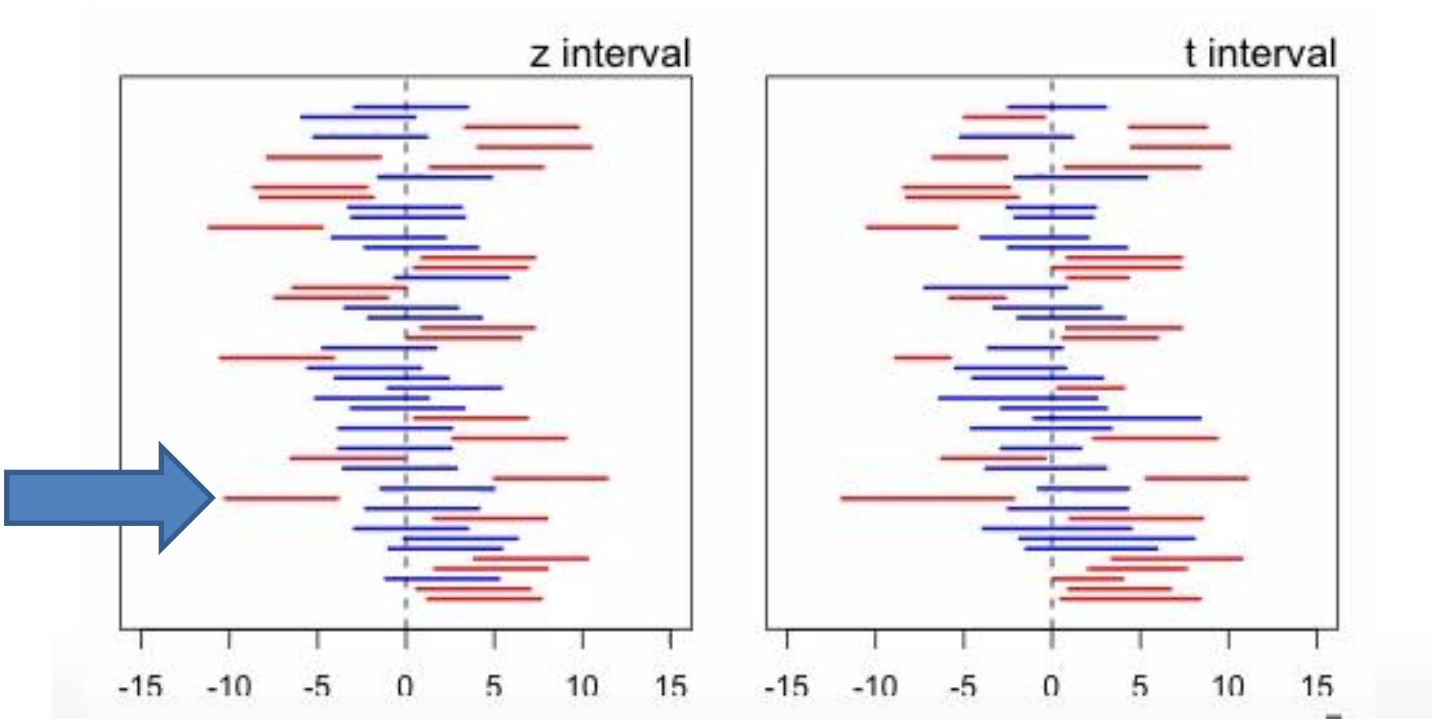
$$y_i \sim \text{Normal}(\mu, 15^2)$$

- $CI_1: \bar{y} \pm 0.674 \frac{15}{\sqrt{10}}$ (50% z interval)
- $CI_2: \bar{y} \pm 0.688 \frac{s}{\sqrt{10}}$ (50% t interval)

Both are valid 50% confidence procedures that produce 50% confidence intervals

WARWICK

CLINICAL TRIALS UNIT



Why are CIs like this?

WARWICK
CLINICAL TRIALS UNIT

Originally formulated by Neyman (1930s)

Aim was to calculate an interval such that, if you assert that the true value is in this interval, you will be correct a specified percentage of the time, in a long run of repetitions of the experiment.

Similar to hypothesis testing; a basis for making decisions while controlling the probability of errors. Not necessary or possible to interpret meaning of a specific interval.

WARWICK
CLINICAL TRIALS UNIT

Neyman 1941

"it is not suggested that we can `conclude' that [the interval contains θ], nor that we should 'believe' that [the interval contains θ]...[we] decide to behave as if we actually knew that the true value [is in the interval]. This is done as a result of our decision and has nothing to do with 'reasoning' or 'conclusion'. The reasoning ended when the [CI procedure was derived]. The above process [of using CIs] is also devoid of any 'belief ' concerning the value [] of [θ]."
(1941)

Where does this leave us?

WARWICK
CLINICAL TRIALS UNIT

- Confidence intervals don't mean what they are usually supposed to mean.
- How should we interpret a single confidence interval?
  - confidence intervals closely related to significance tests
  - 95% CI contains all the values that are not significantly ($p<0.05$) different from the data
  - Hence use of CIs as significance tests
- Range of values that are compatible (not significantly different from) the data

Irony?

WARWICK
CLINICAL TRIALS UNIT

- Confidence intervals promoted to overcome problems of significance tests.
- But their most reasonable interpretation is … as significance tests.

What do we want?

WARWICK
CLINICAL TRIALS UNIT

- Confidence intervals don't do the things that we would like
  - Indicate range with specified probability
  - Show most likely values
- If those are things that we want, Bayesian methods are the way to get them
- We should use the appropriate methods for what we want to do rather than interpreting conventional methods wrongly

Evidence of misunderstanding

WARWICK
CLINICAL TRIALS UNIT

Psychon Bull Rev
DOI 10.3758/s13423-013-0572-3

BRIEF REPORT

Robust misinterpretation of confidence intervals

Rink Hoekstra · Richard D. Morey · Jeffrey N. Rouder · Eric-Jan Wagenmakers

Gave 562 participants (students and researchers) six (incorrect) statements about confidence intervals

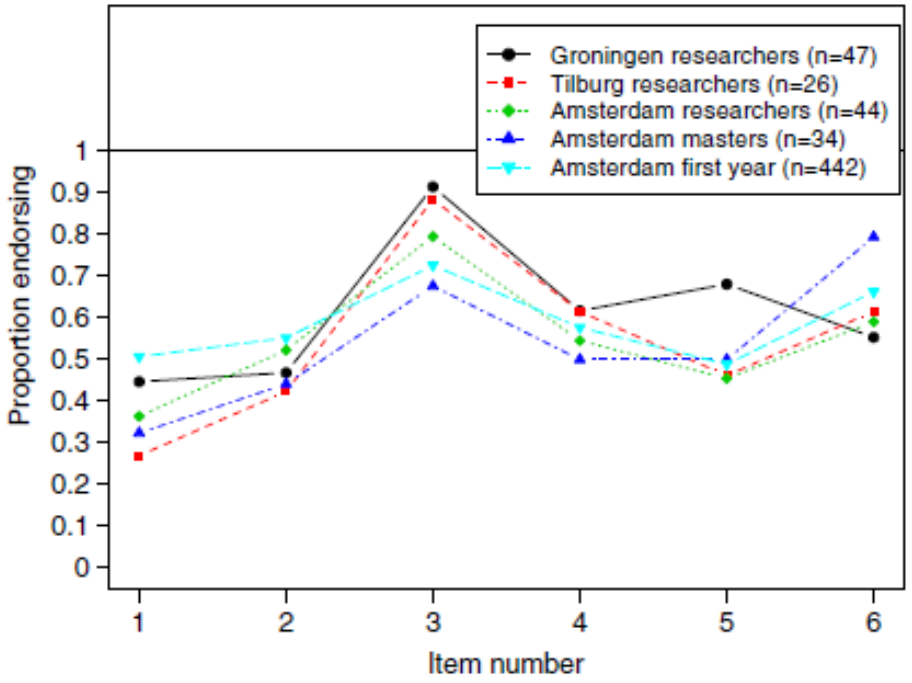
An experiment finds that the 95% CI for the mean is from 0.1 to 0.4:

1. The probability that the true mean is greater than 0 is at least 95%.
2. The probability that the true mean equals 0 is smaller than 5%.
3. The “null hypothesis” that the true mean equals 0 is likely to be incorrect.
4. There is a 95% probability that the true mean lies between 0.1 and 0.4.
5. We can be 95% confident that the true mean lies between 0.1 and 0.4.
6. If we were to repeat the experiment over and over, then 95 % of the time the true mean falls between 0.1 and 0.4.

Evidence of misunderstanding

WARWICK
CLINICAL TRIALS UNIT

Proportion of respondents saying each statement was true ranged from about 30% to about 90%



"You must unlearn what you have learned."

Summary

WARWICK
CLINICAL TRIALS UNIT

- Confidence intervals are usually misinterpreted
- 95% CI don’t show a range with 95% probability of including the true value
- The do show a range of values that are “compatible” (nor “significantly” different from) the data
- If we want probabilities of the most likely values, we need to use Bayesian methods