Group 7's insights on the confidence intervals

Appendix

General Section

CI's used in published papers

Confidence intervals: Linking evidence to practice

Definition of Confidence Intervals in the paper

Kamper (2019) provides an accessible and clinically relevant explanation of confidence intervals (CIs), specifically designed for healthcare professionals and researchers. Confidence intervals are described as ranges around an estimated effect size, such as the difference between treatments, that represent plausible values for the true population parameter based on sample data.

The report thoroughly explains that the width of these intervals depends primarily on three factors: the chosen confidence level (commonly 95%), the variability in the observed data, and the sample size. Larger studies with smaller variability produce narrower confidence intervals, indicating more precise estimates of the true effect. Conversely, smaller sample sizes or greater variability result in wider intervals that reflect lower precision and greater uncertainty in the estimate.

A crucial clarification made by Kamper (2019) pertains to a common misconception. A 95% confidence interval does not imply a 95% probability that the true parameter lies within the interval for the current study. Instead, it signifies that if the exact same study were repeated many times, 95% of those calculated confidence intervals would be expected to contain the true population parameter. This understanding promotes correct interpretation and application of confidence intervals in clinical decision-making.

Furthermore, the paper discusses how increasing the confidence level, for example from 90% to 99%, will increase the interval's width. This means one can be more confident the interval contains the true effect but at the expense of precision.

Sterne et al. (2021) provides complementary insights emphasizing that confidence intervals represent ranges of plausible values for the true effect size and should be interpreted in the clinical context. They reiterate the dependence of interval width on sample size and variability

1

and highlight the danger of misinterpretation when confidence intervals are misconstrued as having a fixed probability of containing the true parameter in a single study. Their work supports the view that confidence intervals offer richer information than p-values alone and are essential for understanding the precision and reliability of findings.

Impact of Variability on Confidence Intervals

Variability in the data, quantified by statistics such as the standard deviation or standard error, has a direct influence on the width of confidence intervals. When individual outcomes vary widely, known as high variability, the standard error increases, causing the confidence interval to widen. This wider range means there is less precision and confidence in pinpointing the true effect.

For example, if patient responses to a treatment show minimal differences, the resulting confidence interval will be narrow, offering precise estimates for clinical application. However, if responses differ greatly owing to factors like patient characteristics or external influences, the increased variability leads to a wider confidence interval which signals caution in interpreting the results (Kamper, 2019; Sterne et al., 2021).

Consequences of Excessively Wide Confidence Intervals

Excessively wide confidence intervals are problematic because they provide a broad spectrum of possible true effect sizes, often spanning from substantial benefit to negligible or even harmful effects. Such wide intervals are typically associated with smaller sample sizes or high variability in data and suggest a high degree of uncertainty.

Kamper (2019) stresses that wide confidence intervals limit the practical utility of study findings, making it difficult for clinicians and researchers to draw confident evidence-based conclusions. Sterne et al. (2021) further emphasize that misinterpretation of wide intervals can lead to over- or underestimation of treatment effects. This uncertainty underlines the need for larger, better-controlled studies to achieve estimates with acceptable precision and meaningful clinical relevance.

Conclusion (TLDR)

Both Kamper (2019) and Sterne et al. (2021) effectively bridge the statistical concepts of confidence intervals with practical clinical interpretation. They highlight how confidence intervals can communicate the precision and reliability of study findings while emphasizing the critical roles of sample size and variability. The clear explanation of the implications of interval width makes these papers especially valuable for guiding evidence-based clinical decisions.

References

1. Kamper, S. J. (2019). Confidence intervals: Linking evidence to practice. Journal of Orthopaedic & Sports Physical Therapy, 49(10), 763–764. https://doi.org/10.2519/jospt.2019.0706

2. Sterne, J. A. C., Davey Smith, G., & Cox, D. R. (2021). The clinician's guide to p values, confidence intervals, and clinical interpretation. Eye, 35(12), 3361–3370. https://doi.org/10.1038/s41433-021-01863-w

CI's used in case-studies

Real-Life Case Study: Proxy Reporting Bias in India's Time-Use Survey (TUS 2019)

Background

The **Time Use Survey (TUS 2019)**, prepared by the National Statistical Office (NSO) of India, set out to measure the time that individuals spend on paid and unpaid work, caregiving, and leisure time. The survey used a stratified multistage random sampling design and was carried out on 1.38 lakh (138,000) household respondents from all Indian states with 95% confidence intervals for time-use estimations (NSO, 2020). But there was a subsequent finding by Sharma, Swaminathan, & Lahoti (2024), which uncovered a crucial weakness proxy reporting bias. In most families, a male family member responded on behalf of others, especially females, instead of gathering self-reporting information.

What Went Wrong: Sampling and Reporting Bias

Though the questionnaire design was random, response substitution distorted representativeness. 36% of respondents were represented by proxies. Remuneration-paid and non-pay household and care works overreported and underreported by 12–21% and 5–33%, respectively. More women did this work since men didn't respect or see the importance of jobs like childcare and cleaning. This non-sampling error biased precision of confidence intervals and made national gendered divide of work unreliable estimates.

Consequences

- **Women's unpaid and informal work was undercounted**, hiding their real contribution to the economy
- **Policy Misguidance**: Flawed data led to poor gender budgeting and underfunded social programs.
- **Reduced Data Credibility**: Showed that even large samples can be biased if data collection is flawed.

How the Issue Could Have Been Avoided

- **Self-Reporting**: The first household member should report time use without substitution bias.
- **Enumerator Training**: Enforce gender-sensitive, secluded interviews for precision enhancement.
- **Use of Technology**: Use digital diaries or phone apps for immediate self-reporting.
- **Analytical Adjustment**: Flag proxy data and correct through weighting or calibration for bias removal.

Why Adequate Sampling Methods Are Important

Proper sampling guarantees representativeness, valid confidence intervals, and non-biased inference. Even with random sampling, data collection bias (such as proxy reporting) results in systematic error. Robust response and sampling protocols enable valid statistical inference and effective policy formulation.

Risks of Bias in Sampling

- **Selection Bias**: Some categories underrepresented; few women were not personally interviewed.
- **Response Bias**: Male proxies incorrectly reported women's unwaged for social norms.
- **Measurement Bias**: Vague categories such as "household chores" overpowered specific unpaid works.
- **Nonresponse Bias**: The Data of absent members substituted with incorrect proxy responses.

Conclusion (TLDR)

The TUS 2019 example reveals that faulty reporting, and not buggy sampling design, resulted in invalid inferences regarding gendered labour. Correct random sampling, coupled with proper self-reporting, produces valid confidence intervals and reliable national measures. Without addressing bias, even large-scale surveys risk producing data that are precisely inaccurate.

References

1. National Statistical Office. (2020). Time Use in India 2019. Ministry of statistics and programme implementation, government of India. https://mospi.gov.in/publication/time-use-india-2019

2. Sharma, D., Swaminathan, H., & Lahoti, R. (2025). Does it matter who you ask for time use data? The World Bank Economic Review. https://doi.org/10.1093/wber/lhaf004

CI's used in our project dataset