Statistical Considerations in the Development of a Consensus Statement

Joshua J. Cook, M.S.[1], Andrew Jimenez, B.S.[2], Joshua Stillman, B.S.[3], Achraf Cohen, Ph.D. [1]

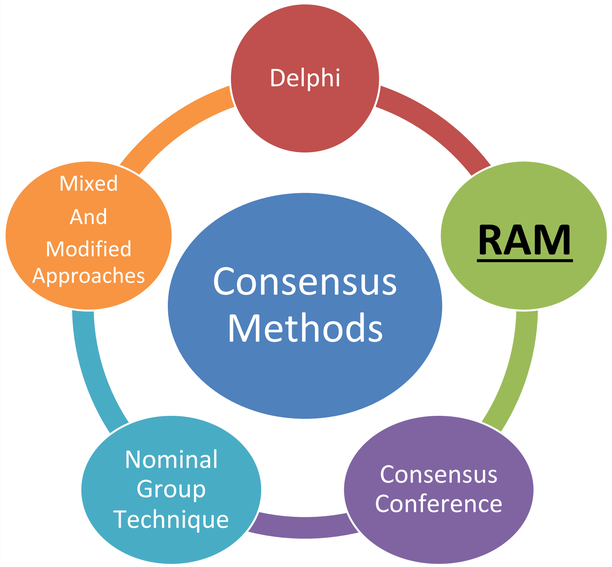
[1]University of West Florida (UWF), [2]Georgetown University School of Medicine (GU), [3]University of Florida (UF)

INTRODUCTION

Defining a **consensus statement:**

*"A comprehensive summary of the **opinions** of a panel of experts… **to provide guidance** to health care professionals, especially on controversial or poorly understood aspects of care"*

-Miller-Keane Encyclopedia



When are they used?

- **Rare / high-risk** medical conditions
- Open standards for technological development
- Economic policy guidance

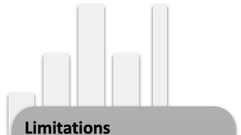
EVALUATING RAND/UCLA METHODOLOGY (RAM)

RAND/UCLA Appropriateness Method (RAM):

1. Define **expert criteria** (industry-dependent)
2. Select **panel** of experts (n = 9-15; not based on statistical power / hypothesis tests)
3. **Literature review** for initial clinical comparison round (usually PRISMA guidelines)
4. 1st round - experts independently rate appropriateness of **clinical scenarios** based on clinical judgment/appropriateness (1-9)
5. 2nd round - experts meet to **discuss** ratings, then given the opportunity to **rate again; third round if needed**
6. **Expert consensus reached** to provide recommendations for clinical indications

Benefits

- Expert Guidance
- Encourages Collaboration
- Systematic and Transparent
- Flexible
- Reduces Variability in Care



Limitations

- Subjective
- Panel Bias
- Groupthink
- Resource Intensive
- Dynamic/Evolving
- Lacks classical statistics

STATISTICAL CONSIDERATIONS

Statistical Analysis:

- *Descriptive statistics:* median and IQR rating for each scenario (resistant to extreme opinions)
- *Hypothesis testing / power analysis:* not applicable; opinion-based, try to balance backgrounds of experts
- *Consensus threshold:* based on scenario medians (appropriate, uncertain, inappropriate) and IQR (level of disagreement)

Methods to increase validity and reliability:

Method	Statistical Analogy
Balanced number and background of experts	Statistical power
Multiple, iterative rating rounds for experts with discussion and opportunities to revise	Replication
Appropriateness and agreement scales	Cohen's Kappa

CONCLUSION AND ADDITIONAL RESOURCES

Conclusion:

- **Consensus studies** (i.e., statements) are key tools to assess clinical treatment appropriateness and agreement among experts.
- However, they are **based on opinions**, and thus do not have the same statistical rigor of traditional studies.
- Care should be taken to include measures outside of traditional statistics to ensure the **validity and reliability of findings.**

GitHub Repository:

- Full **Links & Terms**
- **Literature** Review
- **Simulation of RAM in R**
- Author **Contact**

REFERENCES

1. Birko S, Dove ES, Özdemir V. Evaluation of Nine Consensus Indices in Delphi Foresight Research and Their Dependency on Delphi Survey Characteristics: A Simulation Study and Debate on Delphi Design and Interpretation. PLoS One. 2015;10(8):e0135162. Published 2015 Aug 13. doi:10.1371/journal.pone.0135162
2. Fitch, K., Bernstein, S. J., Aguilar, M. D., Burnand, B., LaCalle, J. R., Lázaro, P., ... & Kahan, J. P. (2001). The RAND/UCLA appropriateness method user's manual. RAND Corporation.
3. Kwong, J. S., Chen, H., & Sun, X. (2016). Development of Evidence-based Recommendations: Implications for Preparing Expert Consensus Statements. Chinese medical journal, 129(24), 2998–3000.
4. Miller BF, Keane CB, O'Toole MT. Encyclopedia & Dictionary of Medicine, Nursing, & Allied Health. 7th Edition. Saunders; 2003.
5. Siwiec, J., Konieczna, M., & Koperny, M. (2019). Good practice for guidance development – review of consensus methods. Zeszyty Naukowe Ochrony Zdrowia. Zdrowie Publiczne i Zarzadzanie, 17(4), 217-229. doi:https://doi.org/10.4467/2G842627OZ.19.023.12185
6. Sparks JB, Klamerus ML, Caverly TJ, et al. Planning and Reporting Effective Web-Based RAND/UCLA Appropriateness Method Panels: Literature Review and Preliminary Recommendations. J Med Internet Res. 2022;24(8):e33898. Published 2022 Aug 26. doi:10.2196/33898