# Renal Committee
## Reliability Testing Overview

Jack Kalbfleisch, PhD
Kidney Epidemiology and Cost Center
University of Michigan

*June 16, 2020*

# Measures Under Evaluation

Measures submitted in this cycle included the use of Inter-Unit Reliability (IUR) methods and Profile Inter-Unit Reliability (PIUR) methods. A summary of the results are as follows:

| Measure | IUR result | PIUR result |
|---|---|---|
| **NQF 0369**: Standardized Mortality Ratio for Dialysis Facilities | 0.50 | 0.77 |
| **NQF 2977**: Hemodialysis Vascular Access: Standardized Fistula Rate | 0.755 | N/A |
| **NQF 2978**: Hemodialysis Vascular Access: Long-term Catheter Rate | 0.76 | N/A |

# Inter-unit Reliability (IUR)

- The IUR is the proportion of the total variation in measure scores that can be attributed to between provider variation (BPV).

    - The BPV can be attributed to unmeasured confounding ~~factors~~ or to quality of care or to a combination of both.

    - If the BPV can be mainly attributed to the quality of care, the IUR can be viewed a measure of reliability (Adams, 2009).

    - If the BPV is a result mainly of unmeasured confounders, the IUR is not directly interpretable as reliability. (Kalbfleisch et al., 2018)

    - The IUR is not very sensitive to outliers (extreme values of the measure).

# Profile IUR (PIUR)

- Supplementary to IUR and more sensitive to extreme or outlying values.

- Assess a measure by its probability of identifying the same providers as extreme when data are replicated.

- The PIUR depends on
  - the method of flagging being used (e.g. empirical null, random effects, fixed effects, etc. )
  - the p-value chosen for flagging (e.g. 5%, 10%, etc)

- If the PIUR is relatively large and the IUR smaller, the measure may be most useful for identifying providers with more extreme outcomes.

- The PIUR is on the same scale as the IUR.

# Discussion

- PIUR can be a useful addition to characterizations of reliability.

- Indicates usefulness of the measure for identifying providers that are relatively extreme (e.g. better or worse than expected).

- Measures with medium to large PIUR still provide a ranking of all facilities, but most attention should be paid to the tails.

- With any quality measure, ranking providers in the center of the measure distribution is almost always difficult due to unmeasured confounders.

# Questions?