ACEP Clinical Policies Inter-rater Reliability of Quality of Evidence Assessment

Michael Brown, MD[1]; Amy Kaji, MD, PhD[2,3];
Stephen Wall, MD, MSc, MAEd[4]; Richard Byyny, MD, MSc[5,6];
Heemun Kwok, MD, MSc[7]; Jason Haukoos, MD, MSc[5,6,9]

1. Department of Emergency Medicine, Michigan State University College of Human Medicine, Grand Rapids, MI, USA; 2. Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, CA, USA; 3. Department of Emergency Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 4. Department of Emergency Medicine, Bellevue Hospital Center, New York University School of Medicine, New York, NY, USA; 5. Department of Emergency Medicine, Denver Health Medical Center, Denver, CO, USA; 6. University of Colorado School of Medicine, Aurora, CO, USA; 7. Division of Emergency Medicine, Harborview Medical Center, University of Washington, Seattle, WA, USA; 9. Department of Epidemiology, Colorado School of Public Health, Aurora, CO, USA

Disclosures

All authors are members of the Clinical Policies Committee,

American College of Emergency Physicians (ACEP)

Background and Aims

- IOM standard - panels rate the level of confidence in the evidence

- Uncertainty regarding the reliably of assessing the quality of a body of evidence[1]

- Inter-rater agreement using GRADE[2] good; kappa = 0.68

- Aim: assess **agreement of grading** among methodologists

1. Norris S, Bero L. GRADE Methods for Guideline Development: Time to Evolve? *Ann Intern Med*. 2016; 165:810-811.
2. Kumar A, et al. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines panel members with limited experience with GRADE. *J Clin Epidemiol*. 2016;75:115-118.

Methods

- Retrospective observational study of 6 clinical policies
- 5 methodologists
  - Initial independent grades compared
- Levels I, II, III, and X
  - Classes of Evidence grading using a structured process
- Measures of agreement:
  - Raw agreement
 - Weighted kappa (k)
  - Linear regression for trend

Results

- 515 articles graded
  - Dec 2015 to Oct 2016
  - median of 90 articles per policy (range: 7 to 140)

- Median weighted k = 0.3 (range: 0.1 to 0.4)

- Median raw agreement = 0.70 (range: 0.2 to 1.0)

Results

- Pairwise sets
  - median number of pairwise graded articles = 10 (range: 1 to 50)
  - median pairwise weighted k = 0.32 (range -0.05 to 0.90)
  - median raw agreement = 0.72 (range 0.50 to 0.94)
- No trend was identified for k or raw agreement across policies
  - beta = -0.05, p = 0.07 for kappa
  - beta = -0.02, p = 0.28 for raw agreement

Figure 1. Pairwise weighted kappa for each reviewer-pair by policy. Policies are listed in sequential order from left to right.
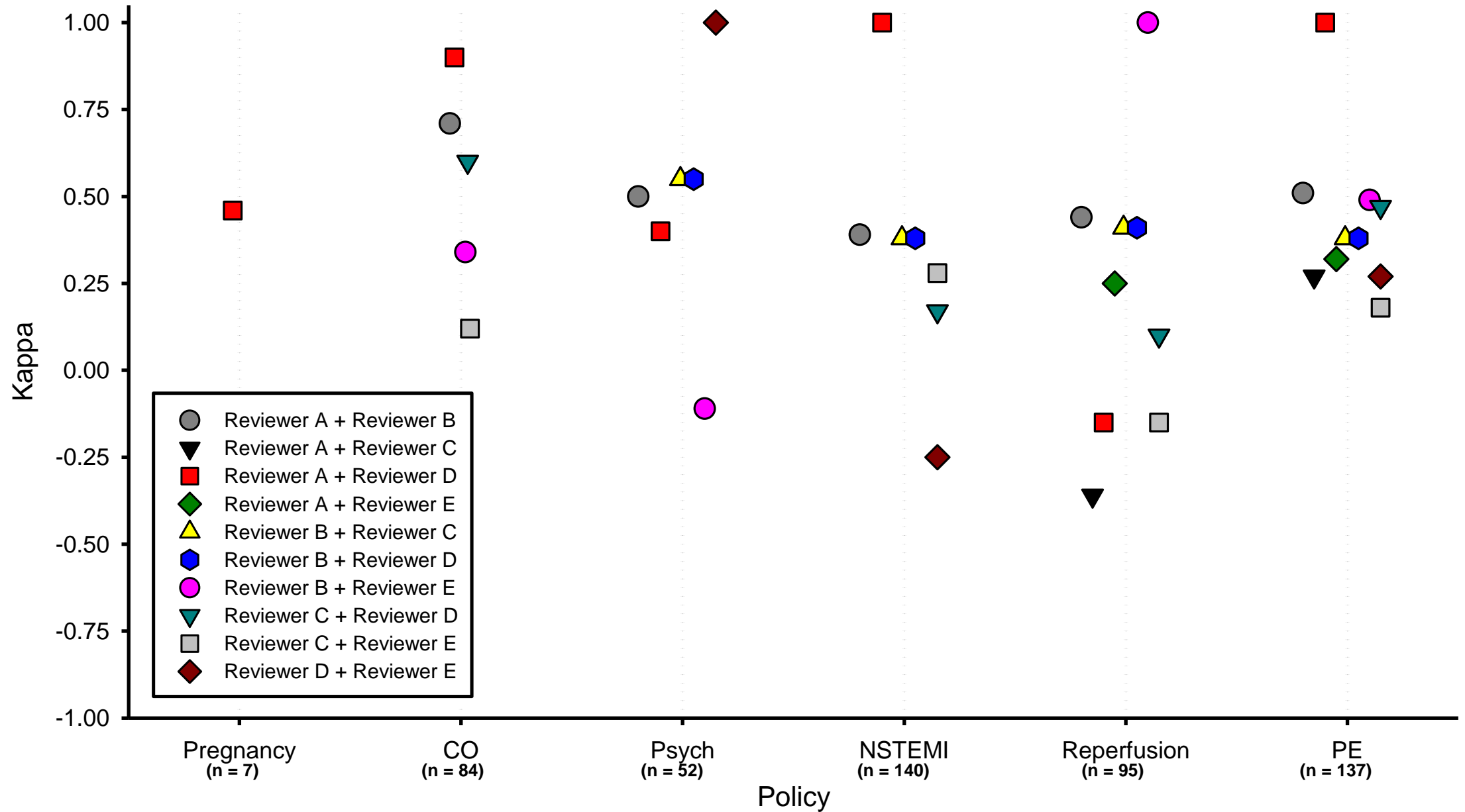
Figure 2. Pairwise agreement for each reviewer-pair by policy. Policies are listed in sequential order from left to right.
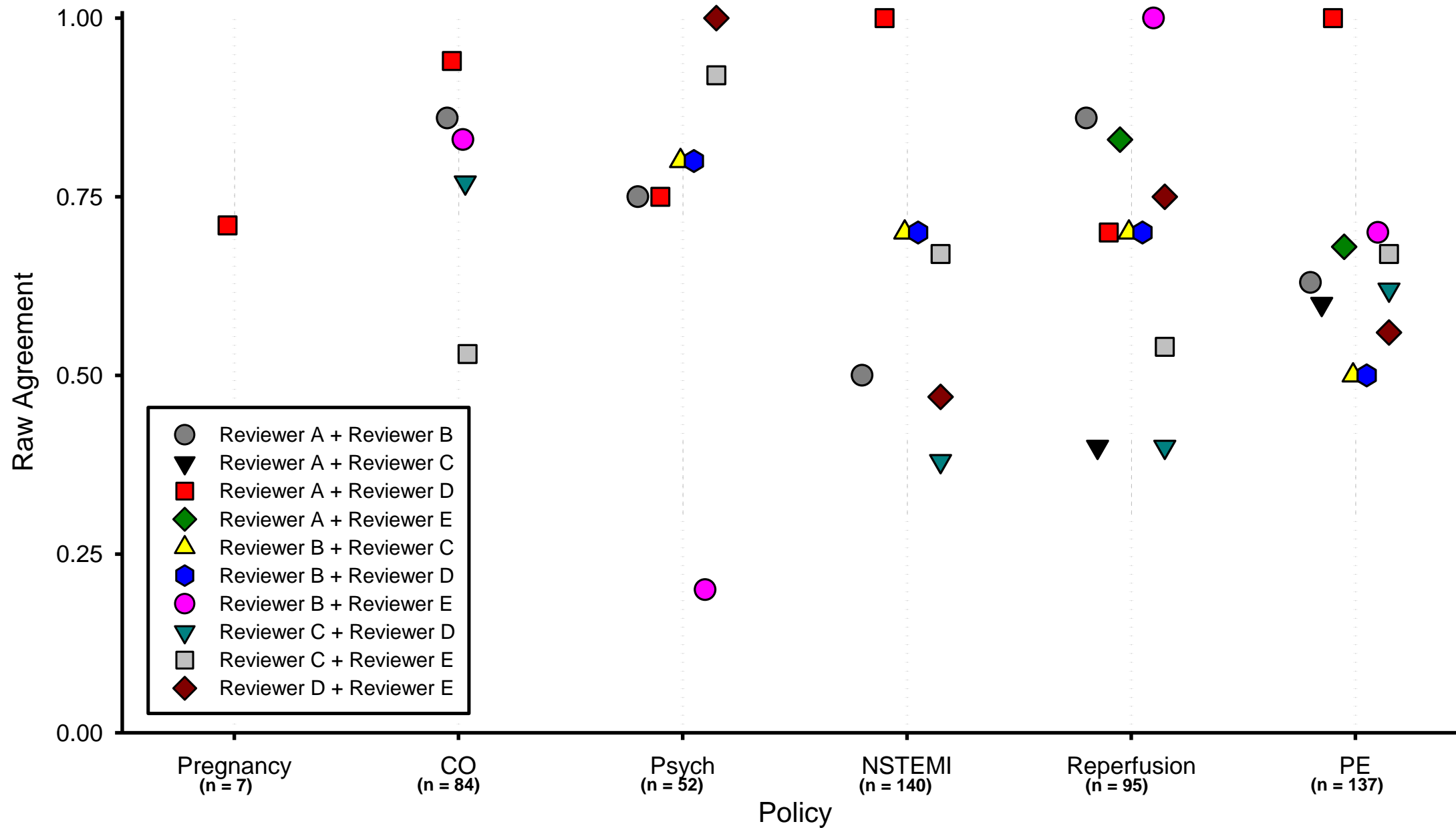
Limitations

- Conception and analyses were retrospective

- Continual refinement of ACEP clinical policies
  - Minor modifications to grading process over the time of study likely

- Precision limited by small numbers

Bottom Line

- Among a team of trained EM physician methodologists, agreement of quality of evidence assessment was fair

- Highlights the importance of using 2 or more methodologists to grade each article, and the need for a formal adjudication process to optimize final class of evidence grading

ACEP Clinical Policies Inter-rater Reliability of Quality of Evidence Assessment

Michael Brown, MD[1]; Amy Kaji, MD, PhD[2,3];
Stephen Wall, MD, MSc, MAEd[4]; Richard Byyny, MD, MSc[5,6];
Heemun Kwok, MD, MSc[7]; Jason Haukoos, MD, MSc[5,6,9]

1. Department of Emergency Medicine, Michigan State University College of Human Medicine, Grand Rapids, MI, USA; 2. Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, CA, USA; 3. Department of Emergency Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 4. Department of Emergency Medicine, Bellevue Hospital Center, New York University School of Medicine, New York, NY, USA; 5. Department of Emergency Medicine, Denver Health Medical Center, Denver, CO, USA; 6. University of Colorado School of Medicine, Aurora, CO, USA; 7. Division of Emergency Medicine, Harborview Medical Center, University of Washington, Seattle, WA, USA; 9. Department of Epidemiology, Colorado School of Public Health, Aurora, CO, USA