DR CRAIG ROTHENBERG (Orcid ID : 0000-0003-3594-600X)

Article type : Non-Systematic Review

AN APPRAISAL OF EMERGENCY MEDICINE CLINICAL PRACTICE GUIDELINES: DO WE AGREE?

Authors:

Alyssa Zupon, MD[1]

Craig Rothenberg, MPH[2]

Katherine Couturier[2]

Ting-Xu Tan[2]

Gina Siddiqui[2]

Matthew James[2]

Dan Savage[3]

Edward R. Melnick, MD, MHS[2]

Arjun K. Venkatesh, MD, MBA, MHS[2,4]

Affiliations:
1. Yale University School of Medicine, New Haven, CT

2. Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT

3. Department of Emergency Medicine, University of California, San Francisco (UCSF) Fresno Medical Education Program, CA

4. Yale New Haven Hospital Center for Outcomes Research and Evaluation, New Haven, CT

Corresponding Author

Arjun Venkatesh

Department of Emergency Medicine

Yale University School of Medicine

New Haven, CT, USA

Email: arjun.venkatesh@yale.edu

Phone: 203-785-2353

Disclosures

Dr. Arjun Venkatesh reports support from the Emergency Medicine Foundation Health Policy Scholar Award and the Yale Center for Clinical Investigation KL2 TR000140 from the National Center for Advancing Translational Science (NCATS), a component of the National Institutes of Health (NIH).

ABSTRACT:

Background: Clinical practice guidelines (CPGs) have been published by the American College of Emergency Physicians (ACEP) since 1990 to advance evidence-based emergency care. ACEP clinical policies have drawn anecdotal criticism for bias, yet the overall quality of these guidelines has not previously been quantified. We sought to examine ACEP clinical policies using a recognized, validated appraisal instrument: Appraisal of Guideline for Research & Evaluation (AGREE II).

Methods: Systematic assessment of current ACEP clinical policies using the AGREE II instrument, which contains 23 appraisal items (scored on a 1-7 scale) in six domains and two overall assessments. Each policy was independently appraised by five trained appraisers. Primary outcomes were AGREE II ratings for each item, domain, and "Overall Assessment," and scores were reported as standardized percentages from all five appraisers. Secondary analyses examined associations between AGREE II

ratings and policy publication date, strength of underlying evidence, and strength of recommendations. Additional analysis examined relationships between domain and "Overall Assessment" ratings.

Results: Twenty guidelines published from April 2007 to November 2017 were included. Of the six domains, "Scope and Purpose" scored highest (mean 90%) and "Applicability" scored lowest (mean 35%). The four remaining domains ("Stakeholder Involvement," "Rigor of Development," "Clarity of Presentation," and "Editorial Independence") had mean scores of 53% - 78%. The mean "Overall Assessment" rating was 69% and was not associated with policy publication date, strength of underlying evidence, or strength of recommendations. We found positive associations between "Overall Assessment" ratings and two domains: "Rigor of Development" ($r = 0.70$) and "Clarity of Presentation" ($r = 0.70$).

Conclusions: Based on validated AGREE II criteria, ACEP clinical policies can be most improved by addressing their application in practice. ACEP clinical policies' overall quality did not improve over the assessed time period and is not explained by the quality of underlying evidence.

What's Known

- ACEP clinical policies have raised controversy, possibly due to their focus on clinical scenarios in which lower classes of evidence and expert opinion are the only available evidence.
- To date, no formal evaluation of the rigor of development and overall quality of ACEP clinical policies has been performed.

What's New

- The ratings for clinical policy quality were not associated with publication date, strength of underlying evidence, or strength of recommendations.
- ACEP clinical policies are robust in their objectives, structure, and methodological rigor but weak in addressing the factors that influence the application of their recommendations in clinical practice.

Review criteria

- We included all ACEP clinical policies listed as "current" as of May 24, 2017, from the ACEP Clinical & Practice Management website, http://www.acep.org/clinicalpolicies/.
- The data were abstracted using the electronic web tool created by AGREE II developers. All policies were reviewed independently by five appraisers.

Message for the clinic

- ACEP clinical policies rated highly based on the validated AGREE II instrument with notable strengths in guideline construction and weaknesses in clinical applicability.
- Guideline quality did not improve after ACEP methodological updates and is not related to quality of underlying evidence.

Introduction

The rapid expansion of medical literature in recent decades has provided increased access to evidence that can improve health care delivery, yet much of this information has been published without curation, thereby limiting translation into practice (1-4). To address this, clinical practice guidelines (CPGs) were developed to provide synthesized and critically appraised scientific evidence to enhance medical decision-making. CPGs aim to advance the quality of health care delivery through acceleration of knowledge translation, promotion of cost-effective practices, and reduction of practice variation (1).

There has been a notable increase in the publication and use of CPGs among various medical specialty societies, health care institutions, and governmental bodies (5-8). This increase, however, has raised concern about the lack of standardization in the CPG development process and information presentation (3, 9-11), and this has motivated initiatives to formalize methods for guideline appraisal. One of the many medical specialties promulgating CPGs is emergency medicine. Guidelines specifically for emergency care were first published by the American College of Emergency Physicians (ACEP) in 1990 to answer specific, clinically relevant questions considered to be of high frequency or high risk in emergency medicine (2). However, ACEP's 2013 clinical policy on intravenous tissue plasminogen activator (tPA) use in acute ischemic stroke sparked controversy with concerns for bias in development (12 -15), resulting in substantial reevaluation of the clinical policy development process, rating methodology, and management of conflicts of interest (3). This methodological update was applied to the revised clinical policy on tPA use, published in September 2015 (16), and all ACEP clinical policies since. Prior research has shown ACEP clinical policies to be disproportionately based on expert opinion instead of higher classes of evidence such as controlled clinical trials (4), and this may be explained by ACEP's focus on clinical scenarios for which there are

controversies or emerging evidence. Another prior qualitative assessment of ACEP clinical policies found many recommendations too vague for translation into practice (17). To date, there has been no formal quantitative appraisal of ACEP's clinical policies to evaluate their quality based on validated criteria.

The Appraisal of Guidelines for Research and Evaluation (AGREE) Collaboration developed a systematic framework for assessing the methodological rigor, transparency of development, and quality of reporting in CPGs (18). The updated AGREE II instrument has been cited in over 650 publications and has been utilized to inform CPG users and developers of the quality of available guidelines and identify avenues for future improvement (18-20). AGREE II is the only appraisal tool that has been validated internationally and is formally endorsed by several organizations, including the World Health Organization (5). This instrument contains domains that address specific aspects of guideline quality and is ideal for assessing the gaps identified by prior research on ACEP clinical policies. There has been limited research on whether AGREE II ratings are linked to the strength of the CPG's underlying evidence or whether specific AGREE II domains are especially important to clinicians evaluating the CPG's overall quality.

Accordingly, we sought to assess the methodological rigor, transparency of development, and overall quality of current ACEP clinical policies using the AGREE II instrument. Secondarily, we sought to examine whether AGREE II ratings improved over time and, specifically, after ACEP methodological updates in 2015. We also evaluated whether ratings in certain domains were associated with the overall CPG quality assessment. Finally, we examined whether AGREE II ratings were associated with the strengths of underlying evidence or recommendations in these clinical policies.

Methods

Study design

Systematic assessment of ACEP clinical policies using the AGREE II appraisal instrument (6). This review follows the PRISMA guideline.

Selection of clinical practice guidelines

We included all ACEP clinical policies listed as "current" as of May 24, 2017, from the ACEP Clinical & Practice Management website, http://www.acep.org/clinicalpolicies/. During data collection, one clinical policy was replaced with a revised version ("*Emergency Department Management of Patients Needing Reperfusion Therapy for an ST-Segment Elevation Acute Myocardial Infarction*") and thus our study included the revised policy and excluded the prior version. Also, after data collection, one clinical policy ("*Critical Issues in the Evaluation and Management of Adult Patients Presenting to the Emergency Department with Syncope*") was removed from the website's "current" list but remained in this study, as it was current at the time of initial guideline selection. All clinical policies are authored by ACEP and follow the ACEP clinical policy development process, which includes expert review from medical specialists and societies relevant to the clinical topic. ACEP clinical policies are specific to emergency care in the United States, are regularly published and maintained, and are sponsored by ACEP. While other professional organizations publish guidelines for, or relevant to, emergency care, few other groups have a regular process or a committee responsible for guideline maintenance. This study did not include CPGs either published by other professional organizations in emergency medicine or primarily authored by other organizations and co-signed or endorsed by ACEP.

Data abstraction

The data were abstracted using the electronic web tool created by the AGREE II developers, available at http://www.agreetrust.org/. This instrument consists of 23 key items organized within six quality domains and two additional global assessments. Each item is rated on a Likert scale between 1

(strongly disagree) and 7 (strongly agree). Each domain captures a unique dimension of guideline quality, specifically "Scope and Purpose," "Stakeholder Involvement," "Rigor of Development," "Clarity of Presentation," "Applicability," and "Editorial Independence" (Appendix).

"Scope and Purpose" addresses the overall aim of the guideline, the specific health questions, and the target population (patients, the public, etc.). "Stakeholder Involvement" focuses on the extent to which the guideline development group includes individuals from all relevant professional groups, represents the views of its intended users, and clearly defines its target population. "Rigor of Development" relates to the process utilized to gather and synthesize the evidence, the methods utilized to formulate the recommendations, and the criteria used to update them. "Clarity of Presentation" addresses the language, structure, and format of the guideline. "Applicability" pertains to factors affecting guideline implementation, strategies to improve uptake, and resource implications of applying the recommendations in practice. "Editorial Independence" assesses whether the views of the funding body have influenced the content of the guideline and whether competing interests of guideline group members have been recorded and addressed.

After completing assessments for each of the six domains, the instrument prompts the reviewer for an "Overall Assessment" of the guideline (using the same 1 to 7 Likert scale) and a categorical recommendation for use in clinical practice ("yes," "yes with modification," or "no"). These two global assessments are based on the reviewer's overall impression of the guideline and are not calculated from item or domain ratings.

Group appraisal process

The appraisal process was performed to exceed recommendations of the AGREE II instrument developers. The AGREE II developers recommend a minimum of two appraisers and optimally four appraisers for stable estimates of CPG quality. In our study, all twenty ACEP clinical policies were reviewed by five appraisers (AZ, TT, GS, KC, and MJ). Prior to data abstraction, each appraiser completed a standardized online training module specifically for use of AGREE II (7), and a group session was conducted after reviewing the first three guidelines to ensure consistent use of

definitions and to optimize inter-rater reliability. All appraisals were performed independently between May 2017 and September 2017. Each appraiser was assigned a unique and random order to perform the data abstraction to minimize bias due to increased familiarity with the instrument over time.

Data abstraction for underlying evidence and recommendations

Each clinical policy contains clinical recommendations based on medical literature to address critical questions faced by emergency physicians. For each recommendation found in a clinical policy, we recorded the proportion of recommendations that were Level C (the weakest level of recommendation) (Supplemental Table I). Level C recommendations are based on evidence from Design Class III studies or expert consensus. We also recorded the proportion of references within each clinical policy that were graded as Design Class III evidence. Studies considered as Class III evidence are case series, case reports, consensus or review papers, and studies with a higher level of design but were downgraded by the Committee based on issues with study quality.

Outcomes

The primary outcomes were AGREE II ratings for each item, domain, "Overall Assessment," and recommendation for use in clinical practice.

Analysis

For the primary descriptive analysis, domain and "Overall Assessment" ratings were standardized as a percentage according to the following formula recommended by AGREE II developers (6):

$$\frac{(Obtained\ score - \ Minimum\ possible\ score)}{(Maximum\ possible\ score - \ Minimum\ possible\ score)}\ X\ 100$$

The standardized score using this formula was determined for each domain and "Overall Assessment" for each reviewer. The standardized percentages from all five reviewers were combined and reported as the mean, standard deviation, coefficient of variation, and range.

Secondary analyses were performed to further assess ACEP clinical policies. First, we examined whether AGREE II domain and "Overall Assessment" ratings changed over the ten-year span (2007-2017) that the clinical policies were published. We further evaluated AGREE II ratings before and after the ACEP clinical policy methodological update in September 2015 using t-tests for each domain and "Overall Assessment." Second, we evaluated the association between domain rating and "Overall Assessment" rating by calculating correlation coefficients between each domain and the "Overall Assessment" rating. Third, we examined the association between AGREE II ratings and the strengths of the policies' underlying evidence and recommendations. We report correlation coefficients between "Overall Assessment" ratings and 1) the proportion of Class III evidence within an ACEP clinical policy and 2) the proportion of Level C recommendations within an ACEP clinical policy. These measures have been previously utilized to describe the strength of emergency medicine clinical practice guidelines (6).

In addition to utilizing multiple appraisers, we also assessed agreement across the five appraisers. We first considered calculation of the intra-class coefficients (ICC) to measure inter-rater reliability but found ICC values unrepresentative of raw agreement given the narrow distribution of ratings and therefore used an alternative method to assess inter-rater agreement. We classified each appraiser's domain rating into three categories based on sentiment (1-3, "disagree"; 4, "neutral"; 5-7 "agree"). As a strict assessment of agreement, we calculated the percentage of appraisals for each item for which all appraisers reported the same sentiment category rating. This provides an easily interpretable measure of agreement between appraisers. We observed varying degrees of agreement in sentiment categories by domain: "Scope and Purpose" 97%, "Stakeholder Involvement" 62%, "Rigor of Development" 44%, "Editorial Independence" 28%, and "Applicability" 1%. The chi-squared statistic was statistically significant at $p < 0.0001$ for each domain, indicating that the rate at which all raters agreed was significantly greater than would be expected if ratings were random.

For all analysis, we considered alpha equal to or less than 0.05 to be statistically significant, and we accounted for multiple comparisons using a Bonferroni correction where appropriate (8). Data analysis was performed using R statistical software (version 3.4.2). This was not considered human subjects research.

Results

This study included twenty current clinical policies published by ACEP between April 2007 and November 2017. Of all included policies, 13 were published prior to methodological updates in September 2015 and seven (35%, ACEP 1- 6 and ACEP 20) were published after the update (Supplemental Table II).

Of the six AGREE II domains, "Scope and Purpose" had the highest mean rating and the lowest variability (mean 90%, coefficient of variation (CV) 0.03) (Table I). "Applicability" had the lowest mean rating and highest variability (mean 35%, CV 0.16). The four remaining domains, from highest to lowest rating, were "Rigor of Development," "Clarity of Presentation," "Editorial Independence," and "Stakeholder Involvement." The mean results from the standardized scores of all five reviewers are provided in Supplemental Table III.

For the "Overall Assessment," the mean rating for all twenty clinical policies from the five appraisers was 69% with CV of 0.13 (Table I). The categorical assessment of recommendation for clinical use was evaluated by practicing emergency physicians, and the recorded responses were considered to be summary recommendations based on the overall impression of the policy. The vast majority of responses were "yes" or "yes with modifications" (64% and 30%, respectively). There were, however, "no" responses for clinical policies on asymptomatic elevated blood pressure, prescribing opioids for adult patients, suspected appendicitis, and acute carbon monoxide poisoning. These four clinical

policies also had the four lowest mean "Overall Assessment" ratings of all clinical policies in this study (50%, 53%, 50% and 63%, respectively).

There was no significant relationship between "Overall Assessment" rating and date of policy publication (Supplemental Figure 1). Further, there was no significant improvement in the "Overall Assessment" ratings or in any of the six domains after updates to ACEP clinical policy development process in September 2015 (Table II).

Two domains were statistically associated with the "Overall Assessment" rating: "Rigor of Development" (correlation (r) = 0.70, $p < 0.001$) and "Clarity of Presentation" ($r = 0.77$, $p < 0.0001$) (Table III). The other four domains' ratings were not strongly associated with "Overall Assessment" rating ($r = 0.01$ - 0.51).

There was no significant association between "Overall Assessment" and either the proportion of Level C recommendations ($r = 0.06$, $p = 0.96$) (Figure 1a) or proportion of Class III evidence ($r = 0.01$, $p = 0.79$) (Figure 1b).

Discussion

Overall, ACEP clinical policies rated highly based on the validated AGREE II instrument for guideline quality. There was, however, variability based on AGREE II domains, with strengths in "Scope and Purpose," "Rigor of Development," and "Clarity of Presentation," and weaknesses in "Applicability," "Stakeholder Involvement," and "Editorial Independence." There were no significant improvements in any domain or "Overall Assessment" ratings over time or after methodological updates in 2015. We also found no relationship between AGREE II ratings and the strength of underlying evidence or recommendations. These findings carry important implications for emergency medicine guideline developers, the broader clinical practice guideline community, and physicians using these clinical policies to provide emergency care.

ACEP clinical policies showed highest domain ratings in "Scope and Purpose." This result is consistent with prior research utilizing the AGREE II instrument to evaluate guidelines (24-27) and can be explained by ACEP's standardized and formulaic guideline development process centered on clear clinical questions that comply with the PICO format. "Rigor of Development" was also rated highly due to clear presentation of the level of each recommendation and thorough description of the clinical policies' underlying evidence. These policies also performed well in "Clarity of Presentation," as they use clear language, are formatted logically, provide specific recommendations, and consider alternative options for management.

ACEP clinical policies were weakest in "Applicability" (domain related to implementation and adoption), and this is consistent with prior studies evaluating both guidelines in other specialties (9) and ACEP clinical policies (17, 26, 27). Poor performance in this domain could reflect a belief that guideline development and implementation are separate activities and that the organizational barriers and cost implications are better discussed among local administrators who can make individualized decisions based on local settings or institutional priorities. However, improvement could be made within the current ACEP clinical policy development process by providing some discussion of both the resource implications and barriers to policy implementation and by focusing on formulating actionable recommendations (28). If frequent revisions are infeasible, regular updates in a less formal manner could be utilized to disseminate new influential evidence that becomes available, as modeled the American Heart Association's mobile application, AHA Guidelines On-The-Go.

ACEP clinical policies also showed modest weaknesses in "Stakeholder Involvement" and "Editorial Independence." ACEP clinical policies do not explicitly mention seeking the views and preferences of the target populations such as patients or the public. Given the increasing emphasis on patient partnership in research and policy making, ACEP guideline developers could look to examples such as the National Institute for Clinical Excellence in the U.K. for models on including patient representation in the development process (5, 29). Editorial independence has been a common source of controversy in several specialties (24, 27), and our results indicate that ACEP clinical policies have a relative weakness in this domain compared to other domains but relatively strong ratings in

comparison to guidelines developed by other medical societies (10). To achieve higher ratings in "Editorial Independence," more explicit and thorough disclosure about both financial and intellectual conflicts of interest among guideline developers should be included. Further, explanation should be provided about how these conflicts of interest are managed and accounted for (i.e., whether certain competing interests excluded members from specific aspects of guideline development). While our results did not demonstrate improvement in AGREE II ratings over time, this should not be interpreted as a barrier to developing more effective guidelines. ACEP could consider these targeted improvements in the development process to improve the quality of its clinical policies.

In addition to domain ratings, the mean "Overall Assessment" rating for all 20 clinical policies was quite high, indicating the global strength of the ACEP clinical policy development methodology. In contrast to the domains ratings, however, the "Overall Assessment" does not include detailed criteria and is based on the appraiser's general impression of the policy. Given the inclusion of both formulaic and subjective elements in this instrument, it should be noted to those interpreting AGREE II ratings that "Overall Assessment" scores are distinct from the item-defined domain ratings. The association we found between "Overall Assessment" ratings and both "Rigor of Development" and "Clarity of Presentation" suggest that strong performance on these two domains may be influential on the appraiser's overall impression of the policy's quality. This extends the findings of Hoffmann-Esser et al. (11), which examined AGREE II ratings for 1453 guidelines and found that "Rigor of Development" had the strongest correlation with the "Overall Assessment" rating. Guideline developers can use this information for targeted improvements in these two domains to best meet the needs of those utilizing their guidelines.

Our finding that AGREE II ratings were not sensitive to the policy's strengths of underlying evidence or recommendations is consistent with the design of the AGREE II instrument. While AGREE II evaluates methodological rigor and quality of reporting, the instrument does not evaluate the quality of evidence or recommendations in the guideline. This can be considered a strength because the instrument specifically examines elements performed by, and in the control of, guideline developers. However, this is a weakness because AGREE II ratings do not indicate whether the

recommendations can be followed reliably. Including criteria that considers the strength of underlying evidence would create a more comprehensive instrument that could provide physicians with more confidence in implementing guidelines with high AGREE II ratings. Finally, evidence is still lacking about whether high AGREE II ratings translate to substantial benefits for guidelines' stakeholders. Future research should explore how utilization of the AGREE II instrument can affect the implementation of guidelines, knowledge translation, and clinical outcomes.

Strengths and Limitations

The findings of this work should be interpreted within the confines of the strengths and weaknesses of its design. This study exceeded the number of appraisers recommended by the developers of the AGREE II instrument. However, our results did demonstrate moderate reliability, which may suggest that future research on utilization of this instrument should raise this threshold. All appraisers were from the same academic center and medical specialty, and this may result in an institutional and specialty-specific bias, though AGREE II developers do not outline any requirements for appraisers' qualifications. The qualitative assessment of clinical use may be unique to select emergency physician appraisers who are not evidence-based methodologists, reflecting the practical clinical implications of this work but likely of limited generalizability. Finally, our analysis was limited to guidelines developed by ACEP and did not include any guidelines published by other specialty societies, in other countries, or in other languages; therefore, many guidelines that are pertinent to emergency care but developed by other specialties or organizations are absent from this work.

Conclusions

ACEP clinical policies rated highly and had notable strengths and weaknesses based on validated criteria provided by the AGREE II instrument. Guideline quality did not improve over time or after ACEP methodological updates in 2015 and is not related to the quality of underlying

evidence. ACEP clinical policies can be improved by including patient representation in the guideline development process, enhancing editorial independence and transparency, and addressing factors that influence the application of these policies in clinical practice.

Author Contributions:

AV and AZ established the purpose and scope of this project. AZ developed the draft data collection tool and study protocol with supervision by AV. AZ, TT, GS, KC, and MJ utilized the AGREE II instrument to assess all ACEP clinical policies in this study. AZ, AV, and CR analyzed and interpreted the data. All authors were involved in further interpreting the data and discussing the manuscript. AZ primarily drafted the manuscript with critical review and revision by all other authors.

Acknowledgements:

Alyssa Zupon would like to thank the Office of Student Research at the Yale School of Medicine for providing grant support for this project.

References

1. R G, M M, D MW, S G, and E S. Washington DC; 2011.
2. Schriger DL, Cantrill SV, and Greene CS. The origins, benefits, harms, and implications of emergency medicine clinical policies. *Ann Emerg Med.* 1993;22(3):597-602.
3. Radecki R. ACEP Clinical Policy on Intravenous Tissue Plasmogen for Stroke Continues to Evolve. http://www.acepnow.com/article/acep-clinical-policy-on-intravenous-tissue-plasmogen-for-stroke-continues-to-evolve/. Accessed December 17, 2017.
4. Venkatesh AK, Savage D, Sandefur B, Bernard KR, Rothenberg C, and Schuur JD. Systematic review of emergency medicine clinical practice guidelines: Implications for research and policy. *PLoS ONE.* 2017;12(6):e0178456.
5. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *J Clin Epidemiol.* 2010;63(12):1308-11.
6. Consortium ANS. The AGREE II Instrument [Electronic version]. 2009.

7. . AGREE II Online Training Tools. https://www.agreetrust.org/resource-centre/agree-ii-training-tools/.
8. Armstrong RA. When to use the Bonferroni correction. *Ophthalmic Physiol Opt.* 2014;34(5):502-8.
9. Wang Y, Luo Q, Li Y, Wang H, Deng S, Wei S, and Li X. Quality assessment of clinical practice guidelines on the treatment of hepatocellular carcinoma or metastatic liver cancer. *PLoS One.* 2014;9(8):e103939.
10. Alonso P, Irfan A, Solà I, Gich I, Delgado-Noguera M, Rigau D, Tort S, Bonfill X, Burgers J, and Schunemann H. The quality of clinical practice guidelines over the last two decades: a systematicreview of guideline appraisal studies. *Qual Saf Health Care.* 2010;19(6):e58(
11. Hoffmann-Eßer W, Siering U, Neugebauer EA, Brockhaus AC, Lampert U, and Eikermann M. Guideline appraisal with AGREE II: Systematic review of the current evidence on how users handle the 2 overall assessments. *PLoS One.* 2017;12(3):e0174831.

Figures & Tables

Table I. AGREE II Ratings for Each Domain and Overall Assessment							
	Scope and Purpose	**Stakeholder Involvement**	**Rigor of Development**	**Clarity of Presentation**	**Applicability**	**Editorial Independence**	**Overall Assessment**
Mean ± SD	90 ± 2.8	53 ± 4.2	78 ± 2.7	75 ± 5.2	35 ± 5.7	68 ± 9.2	69 ± 9.1
Coefficient of Variation (CV)	0.03	0.08	0.03	0.07	0.16	0.14	0.13
Range	84 - 96	46 - 63	73 - 82	59 - 82	25 - 46	37 - 77	50 – 83

Table II. AGREE II Domain Ratings Before and After ACEP Clinical Policy Methodology Updates			
Domain	**Pre-Sept 2015 Rating**	**Post-Sept 2015 Rating**	**p-value[A]**
Scope and Purpose	90%	90%	0.67
Stakeholder Involvement	53%	53%	0.94
Rigor of Development	77%	80%	0.02
Clarity of Presentation	74%	77%	0.18
Applicability	33%	38%	0.09
Editorial Independence	66%	73%	0.03

[A]Statistical significance was defined as $p < 0.008$ using the Bonferroni correction to account for multiple comparisons

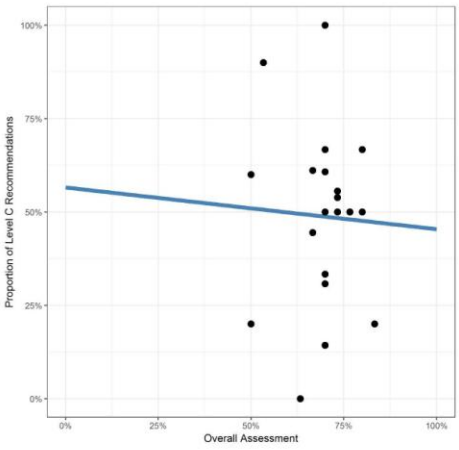
Table III. Correlation Table Comparing Six AGREE II Domains and Overall Assessment

Domain	Correlation with Overall Assessment	p-value
Scope and Purpose	0.51	0.0160
Stakeholder Involvement	0.04	0.8720
Rigor of Development	0.70	0.0010
Clarity of Presentation	0.77	0.0001
Applicability	0.45	0.0488
Editorial Independence	0.02	0.9210

[A]Statistical significance was defined as $p < 0.007$ using the Bonferroni correction to account for multiple comparisons.

Figure 1. Comparison of Overall Assessment Rating and Either Proportion of Level C Recommendations (2A) or Proportion of Class III Evidence (2B)

A:



B: