Strength of Recommendation and Quality of Evidence for Recommendations in Current Infectious Diseases Society of America Guidelines

Kaitlin E. Miles,[1] **Ryan Rodriguez,**[1,2] **Alan E. Gross,**[1,2] **and Andre C. Kalil**[3]

[1]University of Illinois at Chicago, College of Pharmacy, Chicago, Illinois, USA, [2]Pharmacy Practice, University of Illinois at Chicago, College of Pharmacy, Chicago, Illinois, USA, [3]Department of Internal Medicine, Division of Infectious Diseases, University of Nebraska Medical Center, Omaha, Nebraska, USA

Background. Grading of Recommendations Assessment, Development, and Evaluation (GRADE) is a systematic approach to grading strength of recommendation (SOR) and quality of evidence (QOE) for guideline recommendations. We aimed to assess the relationship between SOR and QOE in current Infectious Diseases Society of America (IDSA) guidelines.

Methods. In this cross-sectional analysis, we analyzed the frequency of SOR-QOE pairings, including discordance (defined as strong SOR based on expert opinion, very low, or low QOE) for GRADEd recommendations in IDSA guidelines published since 2010. Data for each recommendation were extracted on SOR, QOE, the domain of disease management (one or more of diagnosis, treatment, prevention, and other categories), and relevance to drug or nondrug treatment.

Results. Seventeen eligible guidelines provided 1042 unique GRADEd recommendations (n = 237, 711, 76, and 73 pertaining to diagnosis, treatment, prevention, and other domains, respectively; n = 574 and 137 pertaining to drug and nondrug treatment). Overall, the most common SOR was strong (71.8%; n = 748) and the most common QOE was low (48.6%; n = 506). Among all strong recommendations, 47.1% (n = 352) demonstrated discordance with QOE. By domain, strong recommendations were discordant in 36.6%, 51.4%, 29.3%, and 58.1% of recommendations pertaining to diagnosis, treatment, prevention, and other domains, respectively. Similarly, 50.7% and 54.0% of strong recommendations related to drug and nondrug treatment were discordant, respectively. We identified 39.6% of discordant recommendations to be consistent with good practice statements, which are recommended to be labeled as such without formal GRADEd designations of SOR or QOE.

Conclusions. Among all IDSA guideline recommendations with strong SOR, approximately half were discordant with QOE, and this frequency varied across strata of domains of disease management.

Keywords. clinical practice guidelines; evidence-based practice; infectious diseases.

Clinical practice guidelines support effective clinical care by providing scientific evidence for recommendations to incorporate into clinical practice [1]. Qualitative assessments of each recommendation and its supporting evidence are typically provided by guidelines in a systematic framework. However, these frameworks have been inconsistent between guidelines, which may present readers with challenges in interpreting guideline recommendations [2].

Grading of Recommendations Assessment, Development, and Evaluation (GRADE) is a system for grading the strength of recommendation (SOR) and quality of evidence (QOE) for recommendations in clinical practice guidelines [2]. The GRADE framework is used by numerous organizations worldwide, allowing for standardization of evidence rating with one system that clearly addresses SOR and QOE independently [3, 4]. The GRADE framework classifies SOR as strong when the level of confidence that the recommendation's beneficial effects will outweigh the unwanted effects (or vice versa); when these tradeoffs are less certain, SOR is weak [5].

One of the determinants of SOR is QOE. The QOE is classified as high, moderate, low, or very low based on the confidence in the effects estimate [6]. The GRADE framework initially designates a body of evidence as high or low QOE when it is composed of randomized controlled trials (RCTs) or observational studies, respectively. This initial classification of QOE can then be upgraded by 1 or 2 levels based on factors such as a large magnitude of effect, dose-response gradient, or an effect of residual confounding that further supports a treatment's effect. The QOE classification may also be downgraded by 1 or 2 levels due to risk of bias, inconsistency, indirectness, imprecision, or publication bias [6–11].

The Infectious Diseases Society of America (IDSA) adopted the GRADE system in 2008 to improve the rating of QOE within clinical practice guidelines [3]. Critics of the previous grading system (IDSA-US Public Health Service Grading System) noted the inability of the system to account for subtleties when determining the robustness of the evidence used to grade recommendations. The GRADE framework attempts to overcome this limitation by considering the nuances among studies with the same design when rating the quality of evidence (ie, not all RCTs are of similar quality and robustness).

Principles of evidence-based medicine suggest that stronger recommendations should be supported by higher quality evidence, such as RCTs [1]. This association between SOR and QOE in clinical practice guidelines has been previously explored [12–17]. Prior assessments of IDSA guidelines produced using the IDSA-US Public Health Service Grading System, published during the years of 1994 to 2010, found that approximately one half of recommendations were supported by low-level evidence consisting of nonrandomized studies and expert opinion [12, 13, 16]. An analysis of guidelines published between 1984 and 2008 by the American College of Cardiology and the American Heart Association found that 46% of 2711 recommendations reporting QOE were based on level C evidence (derived from expert opinion, case reports, and standards of care), and a subsequent analysis found similar results [14, 17]. Sims et al [15] analyzed the correlation of SOR and QOE among 681 recommendations in various critical care guidelines that utilized GRADE methodology published between 2011 and 2017. Strength of recommendation and QOE were positively correlated (ρ = .66); however, 32.1% of recommendations were discordant (ie, a recommendation with strong SOR based on low, very low, or expert opinion QOE), a condition that is reasonable in some paradigmatic situations, but generally discouraged [18].

Discordant recommendations may also be consistent with good practice statements, which are recommended to be designated as such without formally GRADEd ratings of SOR and QOE [19, 20]. Good practice statements are recommended in place of the formal GRADE process when panelists have high confidence that indirect evidence supports net benefit and when it would be unnecessarily burdensome to collect and summarize this evidence. In general, good practice statements can be identified when this situation is present or when the unstated alternative is absurd or violates ethical norms [19]. When presented, good practice statements should be accompanied by documentation justifying their necessity [19, 20]. Previous studies have identified good practice statements to be present in 13.5%, 18.1%, and 36% of discordant recommendations from UpToDate, the World Health Organization, and the Endocrine Society, respectively [21–23]. Given the findings of these and the aforementioned studies, all authors discussed the need for higher quality evidence and more rigorous adherence to GRADE best practices [12–17, 21–23].

To our knowledge, recommendations in IDSA guidelines using the GRADE system published after the year 2010 have not been assessed. Accordingly, we sought to evaluate the correlation and discordance between SOR and QOE for recommendations in current IDSA guidelines using GRADE methodology. We also aimed to perform predetermined stratified analyses by domain of disease management (including diagnosis, prevention, treatment, or other domains) and recommendations pertaining to drug and nondrug treatment to elucidate any differences in discordance across domains of recommendations.

METHODS

Data Inclusion
We reviewed IDSA clinical practice guidelines that were categorized by the IDSA practice guidelines webpage as current from May 2010 to the end of our review as of August 7, 2019. We included guidelines that used GRADE methodology as delineated in the methodology section of each guideline. We performed our primary analysis excluding recommendations that were not GRADEd.

Data Extraction
For each GRADEd recommendation, we recorded the SOR and QOE as ordinal variables based on their assigned categories ranging from none to strong and from none/expert opinion to high, respectively. If a guideline designated a recommendation's QOE as a range (eg, low to very low), we assigned it the higher ranked category, and QOE of none and expert opinion were paired together, as performed in previous analyses [15]. Recommendations in IDSA guidelines are cataloged using a numbered system; however, in some instances, one numbered statement included multiple independently GRADEd recommendations. We recorded each individually GRADEd recommendation as a unique observation for inclusion in the full analysis set.

We classified each statement as (1) pertaining to diagnosis, treatment, prevention, or other domains of disease management and (2) whether a statement pertained to drug or nondrug treatment. For recommendations where these classifications were unclear, we resolved disagreements by consensus. Some statements contained recommendations that encompassed more than 1 domain; these were classified as pertaining to each applicable domain. For each recommendation, we recorded whether there was discordance between SOR and QOE, which we defined as strong SOR based on QOE that was none and/or expert opinion, very low, or low. We reviewed discordant statements to identify whether they were consistent with good practice statements, which were categorized using guidance described by Guyatt et al [19].

We defined the domain of diagnosis as recommendations pertaining to establishing the presence or quality of infection. The domain of prevention was defined as recommendations related to primary or secondary prevention of infectious disease (eg, vaccination, pre- and postexposure prophylaxis, decolonization, and hygiene). The domain of treatment was defined as recommendations regarding initiating, ending, or modifying therapy of suspected or confirmed infection (ie, after diagnosis occurs or when diagnosis is not needed). Recommendations that did not meet the definitions for diagnosis, prevention, or treatment were categorized as other. A local institutional review board determined this research does not meet the definition of human subject research and required no oversight.

Analytic Approach

Descriptive statistics were used to report the frequency of SOR-QOE pairings, including discordance. We performed an exploratory correlation analysis using Spearman's rank-order correlation to quantify the correlation between SOR and QOE for the full analysis set (all GRADEd recommendations), by domain of disease management, and by drug and nondrug treatment, and χ^2 tests of independence between discordance and each domain of disease management and drug- and nondrug treatment. All analyses were performed in R version 4.0.0 (R Foundation for Statistical Computing).

RESULTS

Distribution of Recommendations

At the time of data collection, we identified 17 clinical practice guidelines published since 2010 that were endorsed by IDSA and were labeled as current and used GRADE methodology. These guidelines included 944 numbered statements, from which we recorded 1042 unique GRADEd recommendations that composed the full analysis set. Statements that were not GRADEd (n = 36) were excluded; 12 of these were specifically denoted by panelists as good practice statements. Most recommendations were designated with a SOR of strong (n = 748; 71.8%) and a QOE of low (n = 506; 48.6%) (Table 1).

Most recommendations pertained to the domain of treatment (n = 711; 68.2%), followed by diagnosis (n = 237; 22.7%), prevention (n = 76; 7.3%), and other domains (n = 73; 7.0%). Fifty-five recommendations were designated as pertaining to more than 1 domain. Recommendations in the other domain were related to monitoring, practice (eg, suggestion to use clinical decision support to improve antibiotic prescribing for antimicrobial stewardship), and epidemiology.

Clinical practice guidelines for candidiasis contributed the most recommendations pertaining to treatment (n = 128; 19.2% of all treatment recommendations), followed by aspergillosis (n = 75; 11.2%). Guidelines for infectious diarrhea and healthcare-associated ventriculitis and meningitis issued the most recommendations within the diagnosis domain (n = 39 and n = 36, respectively; 16.5% and 15.2%). Within the prevention domain, most recommendations originated from guidelines for aspergillosis (n = 25; 33.8%), followed by infectious diarrhea (n = 16; 21.6%). Guidelines for implementing an antimicrobial stewardship program and human immunodeficiency virus (HIV) chronic pain management provided the most recommendations pertaining to the other domain (n = 23 and n = 7, respectively; 35.9% and 10.9%).

Discordance of Strength of Recommendation and Quality of Evidence

Among 748 recommendations with strong SOR, 47.1% (n = 352) demonstrated discordance with QOE (low, n = 309; very low, n = 43) (Table 1). These discordant recommendations represented 33.8% of all GRADEd recommendations. By domain, discordance was present in 58.1%, 51.4%, 36.6%, and 29.3% of strong recommendations pertaining to other, treatment, diagnosis, and prevention domains, respectively ($P < .0001$) (Table 2). Likewise, 50.7% and 54.0% of strong recommendations related to drug and nondrug treatment demonstrated discordance ($P = .61$) (Table 3).

Stratified analysis of recommendations from each clinical practice guideline revealed a large proportion of recommendations originated in guidelines for candidiasis, aspergillosis, and skin and soft tissue infections; these recommendations composed 36.0% of all recommendations (Figure 1). Among recommendations in each guideline, we found that the following guidelines included the most discordant recommendations: leishmaniasis (59.2%), HIV

Table 1. Strength of Recommendation and Quality of Evidence for All GRADEd Recommendations

Quality of Evidence	Strength of Recommendation, No. (%)		Total, No. (%)
	Weak	Strong	
High	3 (0.3)	94 (9.0)	97 (9.3)
Moderate	49 (4.7)	302 (29)	351 (33.7)
Low	197 (18.9)	309 (29.7)	506 (48.6)
Very low	45 (4.3)	43 (4.1)	88 (8.4)
Total, No. (%)	294 (28.2)	748 (71.8)	1042 (100)

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation.

Table 2. Strength of Recommendation and Quality of Evidence for GRADEd Recommendations Stratified by Domain of Disease Management

| | Strength of Recommendation, n (%)[a,b] | | | | | | | |
| | Diagnosis | | Treatment | | Prevention | | Other | |
Quality of Evidence	Weak	Strong	Weak	Strong	Weak	Strong	Weak	Strong
High	1 (0.4)	20 (8.4)	2 (0.3)	54 (7.6%)	0	20 (26.3)	0	2 (2.7)
Moderate	13 (5.5)	82 (34.6)	29 (4.1)	204 (28.7)	5 (6.6)	21 (27.6)	3 (4.1)	16 (21.9)
Low	48 (20.3)	52 (21.9)	117 (16.5)	238 (33.5)	12 (15.8)	15 (19.7)	27 (37.0)	22 (30.1)
Very low	14 (5.9)	7 (3.0)	32 (4.5)	35 (4.9)	1 (1.4)	2 (2.6)	0	3 (4.1)
Total	76 (32.1)	161 (67.9)	180 (25.3)	531 (74.7)	18 (23.7)	58 (76.3)	30 (41.1)	43 (58.9)

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation.

[a]Percentages represent overall percentages within each domain of disease management.

[b]Total number of observations exceeds 1042 because some recommendations were designated as pertaining to more than 1 domain.

chronic pain management (55.2%), coccidioidomycosis (53.6%), and candidiasis (50.3%) (Figure 1).

Correlation of Strength of Recommendation and Quality of Evidence

Correlation between SOR and QOE was weakly positive among all recommendations (ρ = .33; 95% confidence interval [CI], 0.28 to 0.39; P < .001). When stratified by domain, the strength of correlation varied for recommendations pertaining to prevention (0.45; 95% CI, 0.24 to 0.61), diagnosis (0.43; 95% CI, 0.32 to 0.53), treatment (0.29; 95% CI, 0.22 to 0.36), and other domains (0.27, 95% CI, 0.02 to 0.48). A weak correlation remained in stratified analyses of recommendations for nondrug (0.29; 95% CI, 0.22 to 0.37) and drug treatment (0.30; 95% CI, 0.13 to 0.46).

Good Practice Statements

We identified 139 (39.6%) of all discordant statements to be consistent with good practice statements. None of these recommendations were accompanied by formal documentation of the necessary conditions to issue a good practice statement, as recommended in guidance by Guyatt et al [19].

DISCUSSION

In this review of IDSA recommendations produced using GRADE methodology, SOR was often discordant with QOE.

Among all strong recommendations, 47.1% were discordant, being supported by low or very low QOE. Novel findings from this research revealed that this amount of discordance varied across domains of disease management, more so than for recommendations pertaining to drug and nondrug treatment.

Our findings generally agree with those of previous analyses. For example, Lee et al [12] and Khan et al [13, 16] found that IDSA guideline recommendations published before 2010 are largely based on lower QOE. These publications demonstrated that approximately 55% of recommendations produced using the IDSA-US Public Health Service Grading System were supported by level III evidence (derived from expert opinion), 31% by level II evidence (derived from observational studies), and 14% by level I evidence (derived from ≥1 RCT). Although these findings cannot be directly compared with our own because of the use of different systems across sets of guidelines, our analysis of current IDSA guideline recommendations reveals a similar prevalence of lower quality evidence underlying recommendations.

Likewise, our review of discordant recommendations identified 39.6% as being consistent with good practice statements. This is in line with previous findings by Brito et al [23], which found that 36% of recommendations in guidelines by the Endocrine Society comprised good practice statements. Our findings add that none included formal documentation of

Table 3. Strength of Recommendation and Quality of Evidence for GRADEd Recommendations for Drug and Nondrug Treatment[a]

| | Strength of Recommendation, n (%) | | | |
| | Drug Treatment | | Nondrug Treatment | |
Quality of Evidence	Weak	Strong	Weak	Strong
High	2 (0.3)	50 (8.7)	0	4 (2.9)
Moderate	27 (4.7)	156 (27.2)	2 (1.5)	48 (35.0)
Low	103 (17.9)	186 (32.4)	14 (10.2)	52 (38.0)
Very low	24 (4.2)	26 (4.5)	8 (5.8)	9 (6.6)
Total	156 (27.2)	418 (72.8)	24 (17.5)	113 (82.5)

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation.

[a]Percentages represent overall percentages within each domain of treatment.
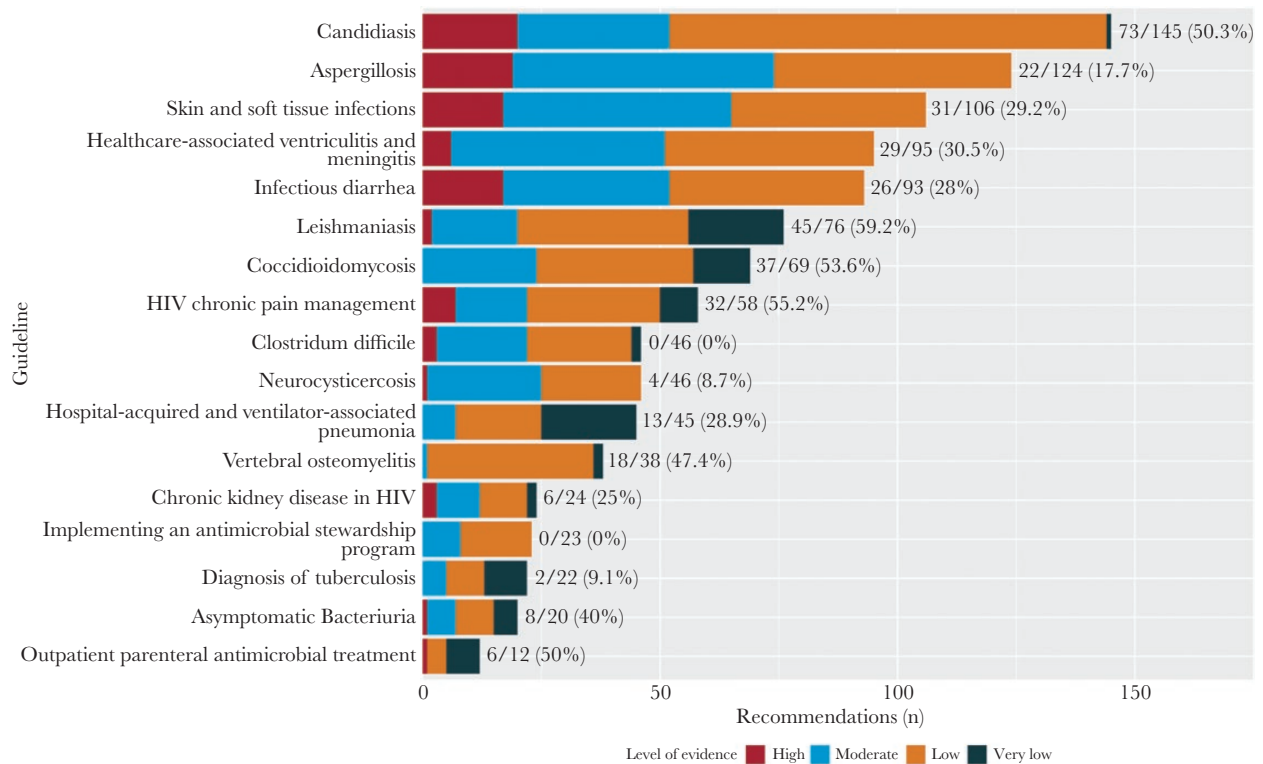
Figure 1. Quality of evidence for GRADEd recommendations stratified by guideline. Fractions and proportions reflect the number of discordant recommendations among total GRADEd recommendations for each guideline. GRADE, Grading of Recommendations Assessment, Development, and Evaluation; HIV, human immunodeficiency virus.

conditions necessary to issue good practice statements. Finally, our research found a lower overall correlation between SOR and QOE in IDSA clinical practice guidelines (ρ = .33) compared with critical care guidelines reviewed by Sims et al [15] (ρ = .66).

A novel approach to this study was our categorization of recommendations into domains of disease management and drug and nondrug treatment. Although discordance was not uncommon across all strata, the numerically higher prevalence of discordance in treatment and other domains may indicate areas warranting comparatively greater attention. Similar cross-sectional analyses of practice guidelines might be used to systematically identify and prioritize specific areas for research, funding, guideline development, and clinical decision making in settings where practice is guided by lower quality evidence.

There may be explanations for our finding that discordant recommendations were common. The GRADE framework methodology permits upgrading or downgrading QOE of bodies of evidence based on several factors, including indirectness of research [4]. In the case of infectious disease research, evidence may have been downgraded because of, for example, indirect evidence that does not directly compare the specific interventions, populations, or outcomes of interest [24]. For instance, discordant recommendations are provided for coverage of *Staphylococcus aureus* and *Pseudomonas aeruginosa* in patients

treated empirically for hospital-acquired pneumonia [25]. This guideline states that QOE underlying these recommendations was based on indirect evidence, because the causative pathogen was identified in only one third of cases. This suggests that incentives for direct comparative effectiveness research in infectious diseases could be valuable, especially for generic drug treatments, where research is unlikely to be funded by industry sponsors.

Discordant recommendations may also be a natural result of the rare incidence of some infectious diseases. This could lead to challenges in their study and low QOE, which may be reflected in our finding that the guideline for leishmaniasis issued the highest proportion of recommendations that were discordant. With 69 cases of cutaneous leishmaniasis reported in the United States from 2007 to 2017, this rare infection is difficult to research [26]. Such challenges may be insurmountable and may reflect a natural plateau in QOE for some conditions.

More importantly, discordant recommendations may have been appropriately issued if they were consistent with paradigmatic situations defined by GRADE methodology [18]. Although this is generally discouraged, it may be justified. Among these are situations in which low-quality evidence suggests benefit in a life-threatening situation. For example, authors of the leishmaniasis guideline issued a discordant recommendation endorsing treatment for clinically manifest visceral

leishmaniasis in certain special populations such as pregnant patients [27]. The authors justify this based on the maternal and fetal risks of untreated clinically manifest visceral leishmaniasis reported by case reports and case series. In this life-threatening situation, the benefits of treatment outweigh the fetal risks of antileishmanial agents. Similar to previous reports, we found inconsistency in the availability and detail of accompanying information to categorize discordant recommendations into one of the paradigmatic situations, which requires assumptions by reviewers that may not be reflective of the panelists' determinations [15, 21, 23]. Therefore, we did not record these assessments, and we also suggest that guidelines could benefit by documenting which paradigmatic situation applies to discordant recommendations to clarify panelists' decision making, adherence to GRADE methodology, and use by clinicians [15].

Limitations of our study include a large proportion of discordant recommendations that originated from a small group of guidelines, which may influence our overall findings. In addition, the heterogeneity in number of discordant recommendations across guidelines should be considered. Our classification of recommendations into domains of disease management and drug and nondrug treatment introduces some subjectivity; however, this was mitigated by consensus review. As mentioned, we did not attempt to categorize discordant recommendations into defined paradigmatic situations; thus, we cannot determine from this analysis whether discordance was appropriate. However, our findings related to recommendations that were consistent with good practice statements provide helpful insight into the improper use of GRADE for some recommendations. Future research could more formally investigate reasons for discordance in IDSA guidelines and the presence or absence of sufficient documentation to categorize the corresponding paradigmatic situation.

CONCLUSIONS

Approximately half of all strong recommendations in current IDSA guidelines using GRADE methodology from 2010 to 2019 demonstrated discordance between QOE and SOR, and this frequency varied across domains of disease management.

References

1. File TM, Jr. Guiding in the face of minimal evidence: a strength, not a weakness, of graded clinical practice guidelines. Infect Dis Clin Pract **2010**; 18:151.
2. Guyatt GH, Oxman AD, Vist GE, et al.; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ **2008**; 336:924–6.
3. Deresinski S. Guiding clinical care through evidence-free zones. Clin Infect Dis **2010**; 51:1157–9.
4. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol **2011**; 64:383–94.
5. Guyatt GH, Oxman AD, Kunz R, et al.; GRADE Working Group. Going from evidence to recommendations. BMJ **2008**; 336:1049–51.
6. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol **2011**; 64:401–6.
7. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence–study limitations (risk of bias). J Clin Epidemiol **2011**; 64:407–15.
8. Guyatt GH, Oxman AD, Sultan S, et al.; GRADE Working Group. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol **2011**; 64:1311–6.
9. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence–publication bias. J Clin Epidemiol **2011**; 64:1277–82.
10. Guyatt GH, Oxman AD, Kunz R, et al.; GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence–inconsistency. J Clin Epidemiol **2011**; 64:1294–302.
11. Guyatt GH, Oxman AD, Kunz R, et al.; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence–indirectness. J Clin Epidemiol **2011**; 64:1303–10.
12. Lee DH, Vielemeyer O. Analysis of overall level of evidence behind Infectious Diseases Society of America practice guidelines. Arch Intern Med **2011**; 171:18–22.
13. Khan AR, Baddour LM, Tleyjeh IM. Evaluation of IDSA clinical practice guidelines: a call to re-GRADE underlying evidence. Arch Intern Med **2011**; 171:1403–4; author reply 1403–4.
14. Tricoci P, Allen JM, Kramer JM, et al. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA **2009**; 301:831–41.
15. Sims CR, Warner MA, Stelfox HT, Hyder JA. Above the GRADE: evaluation of guidelines in critical care medicine. Crit Care Med **2019**; 47:109–13.
16. Khan AR, Khan S, Zimmerman V, et al. Quality and strength of evidence of the Infectious Diseases Society of America clinical practice guidelines. Clin Infect Dis **2010**; 51:1147–56.
17. Fanaroff AC, Califf RM, Windecker S, Smith SC, Jr., Lopes RD. Levels of evidence supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008–2018. JAMA **2019**; 321:1069–80.
18. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol **2013**; 66:726–35.
19. Guyatt GH, Alonso-Coello P, Schünemann HJ, et al. Guideline panels should seldom make good practice statements: guidance from the GRADE Working Group. J Clin Epidemiol **2016**; 80:3–7.
20. Guyatt GH, Schünemann HJ, Djulbegovic B, Akl EA. Guideline panels should not GRADE good practice statements. J Clin Epidemiol **2015**; 68:597–600.
21. Agoritsas T, Merglen A, Heen AF, et al. UpToDate adherence to GRADE criteria for strong recommendations: an analytical survey. BMJ Open **2017**; 7:e018593.
22. Alexander PE, Brito JP, Neumann I, et al. World Health Organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with GRADE guidance. J Clin Epidemiol **2016**; 72:98–106.
23. Brito JP, Domecq JP, Murad MH, et al. The Endocrine Society guidelines: when the confidence cart goes before the evidence horse. J Clin Endocrinol Metab **2013**; 98:3246–52.
24. Akl E, Mustafa R, Santesso N, Wiercioch W. GRADE Handbook. In: Schünemann H, Brożek J, Guyatt G, Oxman A. London, UK: The GRADE Working Group, **2013**. Available at: https://gdt.gradepro.org/app/handbook/handbook.html
25. Kalil AC, Metersky ML, Klompas M, et al. Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society. Clin Infect Dis **2016**; 63:e61–111.
26. McIlwee BE, Weis SE, Hosler GA. Incidence of endemic human cutaneous leishmaniasis in the United States. JAMA Dermatol **2018**; 154:1032–9.
27. Aronson N, Herwaldt BL, Libman M, et al. Diagnosis and treatment of leishmaniasis: clinical practice guidelines by the Infectious Diseases Society of America (IDSA) and the American Society of Tropical Medicine and Hygiene (ASTMH). Clin Infect Dis **2016**; 63:e202–64.