Discrete Voting Thresholds in Revision ACL Consensus: An Arithmetic and Evidence Audit

Cole Richardson, Ian Matthews, Grant Douglas

Abstract

Percentage agreement changes in discrete steps when an expert panel is small. This secondary arithmetic audit examines the agreement rule reported in an adult revision ACL consensus statement. Conditional on 23 valid binary responses, a 75 percent threshold requires 18 affirmative votes: 17 votes yield 73.913 percent and 18 yield 78.261 percent. Increasing the nominal rule to 80 percent requires 19 votes. These calculations describe the operational label rather than clinical certainty or the distribution of opinions among all surgeons. The accompanying methodological synthesis separates panel composition, item wording, response denominators, evidence appraisal, and research priorities. No individual ballots, new patient outcomes, or treatment effects are analyzed. The contribution is a reproducible rule audit that clarifies what an agreement percentage can establish and identifies which additional process records would make consensus interpretation more transparent.

Keywords: expert consensus; discrete thresholds; revision ACL; agreement; evidence certainty; research priorities.

1. A percentage is generated by a count

An agreement percentage can appear continuous even when it comes from a small number of votes. In a 23-person panel, one vote changes the proportion by approximately 4.348 percentage points. A threshold therefore identifies a set of possible vote counts rather than a finely measured boundary in clinical knowledge. The difference between a displayed percentage just below the threshold and a consensus label can be one person's response. That fact does not invalidate the process. It clarifies the precision with which the agreement decision can be interpreted.

Cao and colleagues' APKASS revision ACL statement describes a 23-member panel and a predefined 75 percent agreement threshold (Cao et al., 2026). These parameters supply the input for the present arithmetic audit. We do not reconstruct individual responses or assume every item necessarily has the same valid-response denominator. The calculation asks what the declared rule implies when all 23 contribute a binary answer. Its interpretation remains attached to that condition. Agreement on a clinical proposition does not identify its treatment effect or make the panel a probability sample of clinicians.

This distinction matters because the term consensus can be read as a statement about certainty. Agreement records how a specified group responded to a specified item under a specified process. It does not by itself establish the size of a treatment effect, the quality of the supporting studies, or the prevalence of the opinion outside the panel. The audit therefore keeps exact arithmetic separate from clinical interpretation. It is a small reproducible analysis of a decision rule, accompanied by an evidence-synthesis framework for using its output responsibly.

2. Inputs and calculation

For panel size n and threshold p, the minimum affirmative count is the smallest integer k for which k divided by n is at least p. Equivalently, k is the ceiling of n multiplied by p. We apply this rule to n equal to 23 and thresholds from 70 through 90 percent in five-percentage-point steps. The calculation is deterministic. It requires no model fit, sampling distribution, or participant-level data. The code and resulting table retain the panel size, nominal threshold, required count, and attainable agreement at that count.

The common-denominator assumption is explicit. If an item permits abstention, multiple selections, missing responses, or another voting rule, its applicable denominator may differ. The threshold must then be applied to the rule actually used. A published percentage that does not correspond to an integer count over 23 is a reason to inspect the item-specific denominator and calculation, not automatic proof of a reporting error. The present analysis does not classify particular survey items as incorrectly reported and does not recover unavailable ballot-level data.

There is also no sampling confidence interval in the audit. The expert panel is not treated as a random probability sample of all relevant clinicians. A binomial interval could answer a mathematical question under a sampling model, but attaching one without that model would create an unsupported appearance of population inference. The result instead reports exact attainable percentages and the sensitivity of the rule to a vote or a threshold change. Its uncertainty concerns the applicability and documentation of the voting procedure, not the arithmetic once those inputs are fixed.

3. Results of the rule audit

At the 75 percent threshold, 18 affirmative votes are required. Seventeen of 23 is 73.913 percent, while 18 of 23 is 78.261 percent. There is no attainable agreement between those two values under a complete binary vote with denominator 23. The apparent proximity of a rounded percentage to 75 should therefore be interpreted in terms of the underlying count. Rounding 73.913 to 73.9 does not change whether the predefined rule is met. Nor should a threshold be changed retrospectively simply because an item fell near it.

Table 1. Exact threshold calculations conditional on 23 valid binary responses. The attained percentage is the lowest agreement satisfying each nominal rule.

Nominal threshold	Required votes	Attained percent	Votes below requirement
70 percent	17	73.913	16
75 percent	18	78.261	17
80 percent	19	82.609	18
85 percent	20	86.957	19
90 percent	21	91.304	20

Changing the rule from 75 to 80 percent raises the minimum count by one. A sensitivity table can make this dependence visible without implying that one threshold is scientifically correct. It can also reveal when two nearby nominal thresholds produce the same required count in another panel size. The purpose is transparency about the operational label. The calculation does not show how any expert would vote under revised wording, another meeting format, or additional evidence. Those are separate empirical questions requiring observations of the process.

4. The voting process is part of the result

Consensus-method guidance such as CREDES emphasizes reporting how a process is conducted (Jünger et al., 2017). That guidance was developed in a particular Delphi context; using its reporting principles here does not relabel a single-round meeting as a Delphi study. Relevant features include panel selection, item development, response options, opportunities for discussion, and the handling of nonresponse. A reader needs this information to understand what the agreement count represents. The same numerical threshold can operate differently in a confidential multiround process and an open discussion followed by one vote.

Item wording also determines what is being endorsed. A broad proposition can attract agreement while leaving practical implementation unresolved. An item combining several conditions can make disagreement difficult to interpret because different respondents may object to different parts. Reporting only the final percentage loses that information. A future process could preserve item versions and reasons for revision, while clearly distinguishing responses to the original and revised wording. Such documentation makes changes in agreement interpretable without assuming that an increase necessarily reflects stronger clinical evidence.

Abstention should have a declared meaning. It may indicate uncertainty, lack of relevant experience, or a choice not to endorse either available option. Excluding abstentions from the denominator increases the proportion among remaining respondents and changes the population whose agreement is reported. Counting them as nonaffirmative answers makes another assumption. Neither convention should be introduced after seeing its effect on the label. Item-level numerators and denominators let readers inspect the rule directly and avoid relying on a rounded percentage alone.

5. Agreement and evidence certainty

GRADE distinguishes assessment of evidence from the strength of a recommendation (Guyatt et al., 2008). This distinction matters when interpreting an adult revision-management consensus. An expert statement can identify a consequential decision while the underlying comparative evidence remains uncertain. Agreement does not fill an unobserved biological or clinical link. The report should therefore retain the voting result and the appraisal of supporting studies as separate fields. Their relationship can be discussed explicitly rather than inferred from a common favorable conclusion or a numerical label above the agreement threshold.

The supporting literature also needs a transparent synthesis. PRISMA 2020 provides reporting guidance for systematic reviews (Page et al., 2021). A review accompanying a consensus should identify its scope, search, eligibility, and synthesis methods if it makes a systematic-review claim. A well-documented vote cannot compensate for an opaque evidence search, and a well-reported search cannot reveal the panel's unrecorded reasoning. The two processes provide complementary information. Reporting their connection helps identify whether an item rests on direct comparative evidence, indirect observations, or expert judgment under uncertainty.

Nonrandomized clinical studies require attention to bias, as organized by ROBINS-I (Sterne et al., 2016). A panel may reasonably need to make decisions while such evidence remains limited. The scientific interpretation should preserve that limitation instead of treating agreement as a conversion of biased evidence into a precise effect estimate. The more useful next step is to identify which uncertainty would change the decision and what study could resolve it. Consensus can then support a research agenda without being presented as the endpoint of empirical investigation.

6. Turning disagreement into a research question

An item below the threshold can contain valuable information. The distribution of alternatives may reveal uncertainty about a population, an outcome, or an implementation detail. A single plurality percentage cannot distinguish these possibilities. Recording the alternatives and their rationales can help specify a feasible comparative study. The research question should name the eligible population, the strategies, and the outcome that would discriminate among the alternatives. It should not simply ask which opinion is correct without translating the disagreement into observable consequences.

Multiplicity creates a further reporting issue when many statements or subgroup comparisons are examined. False discovery rate methods concern defined families of statistical tests under stated assumptions (Benjamini and Hochberg, 1995). They should not be applied mechanically to deterministic agreement labels, but they illustrate the need to identify the family of claims in any accompanying inferential analysis. Selecting only the strongest agreements or most striking subgroup differences can distort the apparent coherence of the panel. Complete item reporting is therefore useful even when no hypothesis tests are appropriate for the voting counts.

Heteroskedasticity-robust standard errors would not solve the panel-selection problem (White, 1980). A regression of item responses could be a useful descriptive exercise if the data and question supported it, but a covariance correction would not make the experts a random sample or recover unrecorded responses. This observation reinforces the scope of the present audit. The exact ceiling calculation is informative because it matches its available inputs. Adding statistical machinery without an identified inferential target would make the report more complicated without making its clinical interpretation stronger.

7. A reproducible consensus record

Denominator sensitivity can be shown without imputing any missing ballot. At a 75 percent threshold, 22 valid responses require 17 affirmative votes, whereas 21 require 16. These are different rules from the complete 23-response case, even though all are described with the same nominal percentage. The example illustrates why the number invited, number attending, and number answering an item should be separate fields. It does not assert that the motivating panel had either of these alternative denominators. Their role is to identify the information a reader would need if an item had nonresponse or abstention.

A change in the item can be as consequential as a change in the denominator. Combining two propositions may lower agreement because respondents disagree about only one component. Splitting the item can reveal that structure, but the resulting counts answer new questions and should not replace the original record silently. An evidence archive can retain both versions and their relationship. This allows a later panel or study to understand whether a disputed issue concerns the biological claim, the eligible population, or the proposed action. The resulting transparency is a stronger basis for research prioritization than moving a numerical threshold until a desired consensus label appears.

STROBE's general emphasis on transparent populations, variables, and observation processes offers useful reporting principles for the descriptive components (von Elm et al., 2007). Application should remain adapted to the consensus setting rather than claiming that an expert meeting is a patient cohort. The record should preserve panel eligibility, attendance, valid responses per item, the voting rule, and the complete distribution of choices. A change in the denominator should be visible where it occurs. This record lets a reader distinguish a process limitation from a numerical calculation problem.

The FAIR principles motivate persistent identifiers for item versions, evidence summaries, and permissible response records (Wilkinson et al., 2016). Public reporting need not expose identifiable

individual ballots if confidentiality was part of the process. Aggregated counts and a documented calculation can still make the main agreement labels reproducible. The key is that the public table should contain enough information to recover the label under the declared rule. Where it does not, the missing field should be identified explicitly rather than replaced with an assumed denominator.

Dependence also matters if a future process analyzes several responses from each expert. Cluster-robust methods concern specified within-group dependence under stated sampling assumptions (Cameron and Miller, 2015). They would not make a selected panel representative or recover an unrecorded ballot. A quantitative extension should define whether its unit is a person, an item, or a meeting before estimating uncertainty. The present deterministic audit requires none of those estimates. It retains the vote-count arithmetic and leaves process-level inference to a study with the appropriate response records.

The audit's numerical result is exact under its condition: with 23 valid binary responses, a 75 percent rule requires 18 votes, corresponding to 78.261 percent. Its broader contribution is to distinguish attainable agreement, evidence certainty, and clinical effect. The APKASS statement can support a focused research agenda when those distinctions remain visible. A transparent consensus label helps organize uncertainty and identify priorities, but it cannot substitute for the observations needed to compare alternative revision strategies. The present work reports a parameter audit and no new treatment recommendation.

References

Cao, Mingde, et al. "APKASS 2024 consensus statement on anterior cruciate ligament reconstruction, Part II: Management of anterior cruciate ligament revision in adults." *Asia-Pacific Journal of Sports Medicine, Arthroscopy, Rehabilitation and Technology* 45 (2026): 74-81.

Jünger, Saskia, et al. "Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review." *Palliative Medicine*, vol. 31, no. 8, 2017, pp. 684-706. https://doi.org/10.1177/0269216317690685.

Guyatt, Gordon H, et al. "GRADE: an emerging consensus on rating quality of evidence and strength of recommendations." *BMJ*, vol. 336, no. 7650, 2008, pp. 924-926. https://doi.org/10.1136/bmj.39489.470347.ad.

Page, Matthew J, et al. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." *BMJ*, 2021, pp. n71. https://doi.org/10.1136/bmj.n71.

Sterne, Jonathan AC, et al. "ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions." *BMJ*, 2016, pp. i4919. https://doi.org/10.1136/bmj.i4919.

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data*, vol. 3, 2016, article 160018. https://doi.org/10.1038/sdata.2016.18.

Benjamini, Yoav and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 57, no. 1, 1995, pp. 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

von Elm, Erik, et al. "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies." *PLoS Medicine*, vol. 4, no. 10, 2007, pp. e296. https://doi.org/10.1371/journal.pmed.0040296.

White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. https://doi.org/10.2307/1912934.

Cameron, A. Colin, and Douglas L. Miller. “A Practitioner’s Guide to Cluster-Robust Inference.” *Journal of Human Resources*, vol. 50, no. 2, 2015, pp. 317-372. https://doi.org/10.3368/jhr.50.2.317.