Evaluating Quality of Anesthesiologists' Supervision

- This talk includes many similar slides
  - Paging through produces animation
  - Use right/ left arrow keys, → and ←
- PDF viewers
  - Adobe Acrobat will open directly into Single Page
  - Presentation: Preferences, Full Screen, No Transition
- Google Chrome, Microsoft Edge, Firefox, or Safari
  - Select: "Fit to page", "Page fit", or "Single page"

Updated 03/01/24

Evaluating Quality of Anesthesiologists' Supervision of Anesthesia Residents and Nurse Anesthetists

Franklin Dexter, MD PhD FASA

Director, Division of Management Consulting

Professor, Department of Anesthesia

University of Iowa

Franklin-Dexter@UIowa.edu

www.FranklinDexter.net

Financial Disclosure

- I am employed by the University of Iowa, in part, to consult and analyze data for hospitals, anesthesia groups, and companies
- Department of Anesthesia bills for my time, and the income is used to fund our research
  - I receive no funds personally other than my salary and allowable expense reimbursements from the University of Iowa, and have tenure with no incentive program
  - I own no healthcare stocks (other than indirectly through mutual funds)

ACGME Clinician Educator Milestones

- Pillar 1: Reflective Practice
  - Evaluate teaching activities
- Pillar 3: Recognition and Mitigation of Bias
  - Mitigate the effect of bias
- Pillar 4: Professional Responsibilities
  - Exemplary professional behavior

ACGME Clinician Educator Milestones

- Pillar 1: Reflective Practice
  - ✓ Evaluate teaching activities
- Pillar 3: Recognition and Mitigation of Bias
  - ✓ Mitigate the effect of bias
- Pillar 4: Professional Responsibilities
  - ✓ Exemplary professional behavior

  - ✓ Supervision scale applied as described in this lecture

Ongoing Professional Practice Evaluation and Supervision

Ongoing Professional Practice Evaluation and Supervision

- Example of hospital accreditation standards; these from The Joint Commission

Ongoing Professional Practice Evaluation and Supervision

- Example of hospital accreditation standards; these from The Joint Commission
  - ➢ Information collected about every practitioner

Ongoing Professional Practice Evaluation and Supervision

- Example of hospital accreditation standards; these from The Joint Commission
  - Information collected about every practitioner
  - ➢ OPPE used at least annually to decide whether to continue, limit, or revoke hospital privileges

Ongoing Professional Practice Evaluation and Supervision

- Example of hospital accreditation standards; these from The Joint Commission
  - Information collected about every practitioner
  - OPPE used at least annually to decide whether to continue, limit, or revoke hospital privileges

    "Current competence in performing the requested privilege(s) is verified by peers knowledgeable about the applicant's professional performance"

TJC Standard MS.06.01.03

Ongoing Professional Practice Evaluation and Supervision

- Performance report for past 6-months sent to Chief Medical Officer in <u>undesirable</u> format

Ongoing Professional Practice Evaluation and Supervision

- Performance report for past 6-months sent to Chief Medical Officer in undesirable format

Name	Score	Evaluations
Dr. B	3.58	10
Dr. M	3.65	53
Dr. X	3.71	29
Dr. G	3.77	37
…	…	…
Dr. K	3.85	114
…	…	…
Dr. Z	4.00	6

Ongoing Professional Practice Evaluation and Supervision

- Performance report for past 6-months sent to Chief Medical Officer in undesirable format

Name	Score	Evaluations
Dr. B	3.58	10
Dr. M	3.65	53
Dr. X	3.71	29
Dr. G	3.77	37
...	...	...
Dr. K	3.85	114
...	...	...
Dr. Z	4.00	6

➢ Scores invalid and unreliable with N Evaluations

Ongoing Professional Practice Evaluation and Supervision

- Performance report for past 6-months sent to Chief Medical Officer in undesirable format

Name	Score	Evaluations	
Dr. B	3.58	10	"Worst" in department?
Dr. M	3.65	53	
Dr. X	3.71	29	
Dr. G	3.77	37	
…	…	…	
Dr. K	3.85	114	
…	…	…	
Dr. Z	4.00	6	"Best" in department?

Ongoing Professional Practice Evaluation and Supervision

- Performance report for past 6-months sent to Chief Medical Officer in undesirable format

Name	Score	Evaluations	
Dr. B	3.58	10	"Worst" in department?
Dr. M	3.65	53	
Dr. X	3.71	29	
Dr. G	3.77	37	
...	...	...	
Dr. K	3.85	114	
...	...	...	
Dr. Z	4.00	6	"Best" in department?

➢ Ranks invalid and unreliable with N Evaluations

Ongoing Professional Practice Evaluation and Supervision

- Monitoring supervision relies on anesthesia residents, nurse anesthetists, and other anesthesia practitioners' review
  - Paired daily in actual (*in situ*) clinical practice
  - Frequent ongoing sampling from many independent raters
  - Psychometrically reliable and valid

Ongoing Professional Practice Evaluation and Supervision

- Monitoring supervision relies on anesthesia residents, nurse anesthetists, and other anesthesia practitioners' review
  - Paired daily in actual (*in situ*) clinical practice
  - Frequent ongoing sampling from many independent raters
  - Psychometrically reliable and valid
- ➢ Ongoing professional practice evaluation (OPPE) is mandatory, so alternative is to rely on other quantitative metrics

Ongoing Professional Practice Evaluation and Supervision

- Monitoring supervision relies on anesthesia residents, nurse anesthetists, and other anesthesia practitioners' review
  - Paired daily in actual (*in situ*) clinical practice
  - Frequent ongoing sampling from many independent raters
  - Psychometrically reliable and valid
- ➤ Ongoing professional practice evaluation (OPPE) is mandatory, so alternative is to rely on other quantitative metrics

Alternatives to Monitoring Supervision for OPPE

- Low incidence clinical outcomes
  - Mortality
  - Post-anesthesia care unit reintubation
  - Wrong-side regional nerve block placement
- Low sensitivity to detect differences among anesthesiologists once apply appropriate statistical methods to avoid false detection

Glance LG et al. Anesth Analg 2016

Glance LG et al. Anesthesiology 2016

Dexter F, Hindman BJ. Anesthesiology 2016

Alternatives to Monitoring Supervision for OPPE

- Relatively high incidence clinical outcomes
  - Postoperative patient satisfaction
  - Post-anesthesia care unit arrival pain scores
  - Prolonged times to tracheal extubation
  - Hypotension during induction of anesthesia

Kynes JM et al. Anesth Analg 2013

Wanderer JP et al. Anesth Analg 2015

Chen Y et al. Anesth Analg 2016

Bayman EO et al. Anesthesiology 2016

Epstein RH et al. Br J Anaesth 2017

Freundlich RE et al. J Clin Anesth 2020

Alternatives to Monitoring Supervision for OPPE

- Relatively high incidence clinical outcomes
  - Postoperative patient satisfaction
  - Post-anesthesia care unit arrival pain scores
  - Prolonged times to tracheal extubation
  - Hypotension during induction of anesthesia
- ➢ Risk adjusted scores fail to discriminate among anesthesiologists and/or lack validity

Alternatives to Monitoring Supervision for OPPE

- Systems-based practice measures
  - Perioperative temperature management
  - Surgical Care Improvement Project (SCIP) antibiotic guidelines
  - Unscheduled absences
- Limited validity as measures of individual anesthesia practitioners' quality of care
  - Do not reliably differentiate either

Schonberger RB et al. Anesth Analg 2015

Epstein RH et al. Anesth Analg 2018

Epstein RH et al. J Clin Anesth 2019

Attributes of Supervision

Attributes of Supervision

- Supervision
  - Department's functional definition for purposes of evaluating anesthesiologists
    - All anesthetic activities contributing to patient care, when the anesthesiologist being evaluated is not the provider continually present with the patient

Attributes of Supervision

- Supervision incorporates several attributes
  - Each attribute is included in de Oliveira Filho et al.'s scale for measuring anesthesiologists' supervision of anesthesia residents during clinical operating room care

de Oliveira Filho GR et al. Anesth Analg 2008

Attributes of Supervision

1) The faculty provided me timely, informal, non-threatening comments on my performance and showed me ways to improve

2) The faculty was promptly available to help me solve problems with patients and procedures

3) The faculty used real clinical scenarios to stimulate my clinical reasoning, critical thinking and theoretical learning

Attributes of Supervision

4) The faculty demonstrated theoretical knowledge, proficiency at procedures, ethical behavior, and interest/compassion/respect for patients

5) The faculty was present during the critical moments of the anesthetic procedure (e.g., anesthesia induction, critical events, complications)

6) The faculty discussed with me the perianesthesia management of patients prior to starting an anesthetic procedure and accepted my suggestions, when appropriate

Attributes of Supervision

7) The faculty taught and demanded the implementation of safety measures during the perioperative period (e.g., anesthesia machine checkout, universal precautions, prevention of medication errors, etc.)

8) The faculty treated me respectfully, and strived to create and maintain a pleasant environment during my clinical activities

9) The faculty gave me opportunities to perform procedures and encouraged my professional autonomy

Answering the 9 Questions

Answering the 9 Questions

- Choices beneath each question
  1. never
  2. rarely
  3. frequently
  4. always
- Questions presented daily in same sequence
- Generally, takes < 90 seconds per evaluation
  – End of workday after patient care completed

Hindman BJ et al. Anesth Analg 2013

Dexter F et al. Anesth Analg 2014

Answering the 9 Questions

- <u>Score</u> = mean of answers to the 9 questions
- For each combination of rater (e.g., resident) and ratee (e.g., anesthesiologist), calculate mean of the scores
- For each ratee, calculate <u>average</u> of the means among all raters
  - Equally weighting each rater

Dexter F et al. Anesth Analg 2014a,b

Indications that Supervision is Single Dimension Construct

Indications that Supervision is Single Dimension Construct

- Scale designed to include all attributes
- Scale includes each attribute in residents' written comments made when providing a score below the overall average among anesthesiologists in the department

de Oliveira Filho GR et al. Anesth Analg 2008

Dexter F et al. Anesth Analg 2016

Indications that Supervision is Single Dimension Construct

- Scale designed to include all attributes
- Scale includes each attribute in residents' written comments made when providing a score below the overall average among anesthesiologists in the department
- ➤ Cronbach α in routine use 0.948 (SE 0.001)

de Oliveira Filho GR et al. Anesth Analg 2008
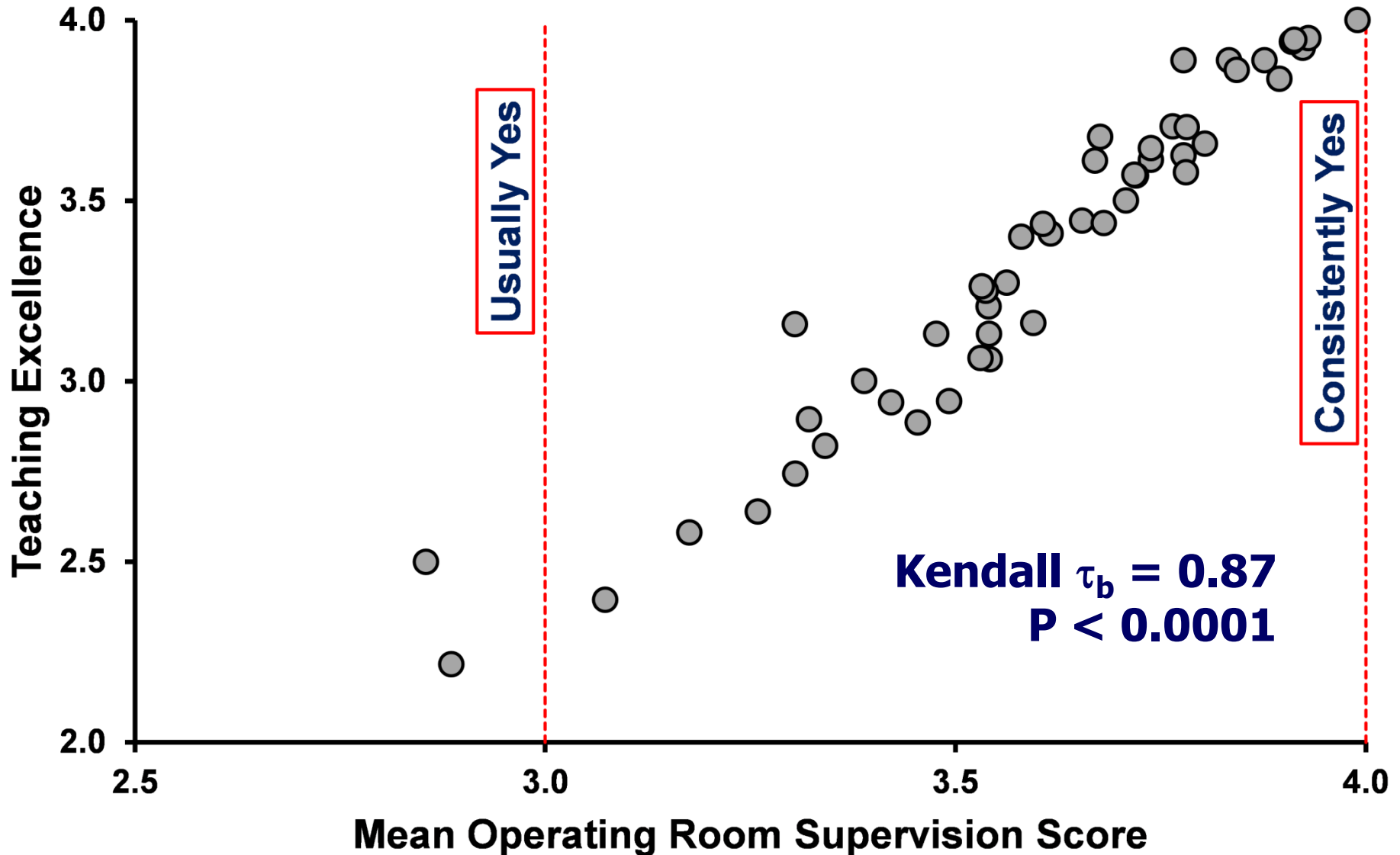Dexter F et al. Anesth Analg 2016

Indications that Supervision is Single Dimension Construct

- Teaching is attribute important to the supervision of residents (trainees)

Hindman BJ et al. Anesth Analg 2013

Concordance between Teaching Evaluations and Supervision Score

Indications that Supervision is Single Dimension Construct

- Each increase in the anesthesiologist's number of resident comments of the anesthesiologist teaching poorly was associated with a lower average score (P = 0.0002)

- Evaluations with comments related to teaching poorly had lower scores than other evaluations with comments (P < 0.0001)

Dexter F et al. Anesth Analg 2016

ACGME Clinician Educator Milestones

- Pillar 1: Reflective Practice
  - ✓ Evaluate teaching activities   Confirmed
- Pillar 3: Recognition and Mitigation of Bias
  - Mitigate the effect of bias
- Pillar 4: Professional Responsibilities
  - Exemplary professional behavior

ACGME. Clinician educator supplemental guide, August 2022

Indications that Supervision is Single Dimension Construct

- Each anesthesiologist evaluated not only by residents (trainees) but also by nurse anesthetists (experienced providers)
  - Averages were correlated, $P < 0.0001$
  - Cronbach α = 0.895 (SE 0.003)
  - Most common score = 4.00 for both groups, $P < 0.0001$

Dexter F et al. Anesth Analg 2014
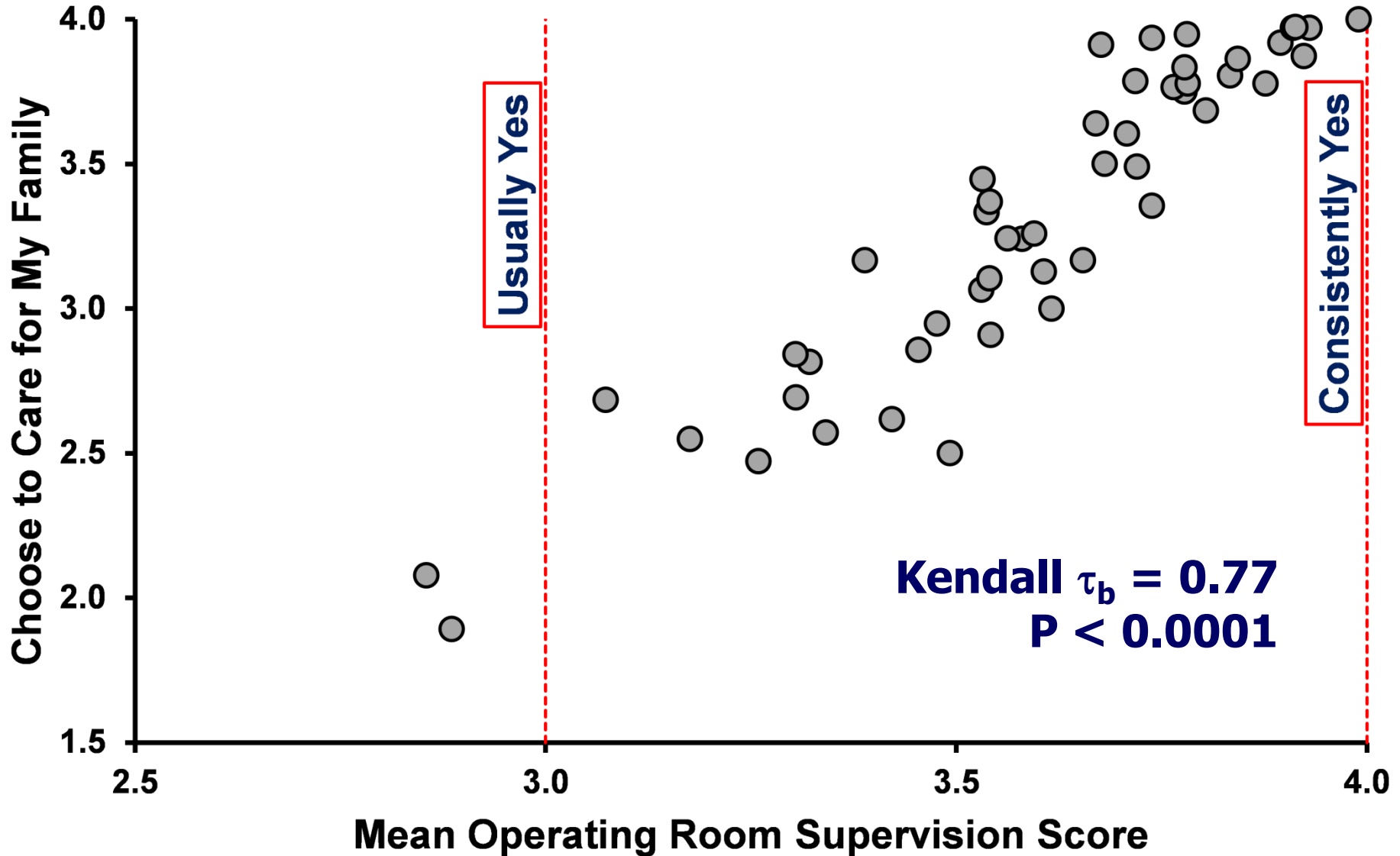Dexter F et al. Anesth Analg 2015

Indications that Quality of Supervision Matters

Hindman BJ et al. Anesth Analg 2013

"I would choose this instructor to care for … my family"

Kendall τ_b = 0.77
P < 0.0001

Usually Yes

Consistently Yes

Choose to Care for My Family

Mean Operating Room Supervision Score

Indications that Quality of Supervision Matters

- Residents reporting overall supervision of department < 3.00 ("frequent") reported making more "mistakes that had negative consequences for the patient"
  - Accuracy (area under the curve) of 89% (99% confidence interval, 77% to 95%)
- Supervision < 3.00 predicted "medication errors (dose or incorrect drug) in" last year
  - Accuracy of 93% (99% CI 77% to 98%)

De Oliveira GS Jr et al. Anesth Analg 2013

Indications that Quality of Supervision Matters

- Residents reporting overall supervision during current rotation < 3.00 ("frequent") reported 3 (75th percentile) and 6 (95th) errors in past year with negative consequences for patients
  - Residents reporting supervision ≥3.00 reported fewer errors (2 and 4; $P < .0001$)
    - Resident burnout not correlated (all $P > .134$) with numbers of reported errors while controlling for quality of supervision

De Oliveira GS Jr et al. Anesth Analg 2015

Indications that Quality of Supervision Matters

- Nurse anesthetists' written comments' theme "I did not see the anesthesiologist during the case(s) together" increased odds (48.2) of supervision score < 3.00 (P < 0.0001)

- Resident comments of insufficient presence associated with scores less than those of other evaluations with comments (P < 0.0001)

  – Anesthesiologists with ≥ 1 such comment had lower average scores than others (P = 0.0071)

Dexter F et al. Anesth Analg 2015

Dexter F et al. Anesth Analg 2016

Indications that Quality of Supervision Matters

- Positive correlations between residents' evaluation of overall departmental supervision and safety culture (all $P < 0.0001$)
  - Overall perceptions of patient safety
  - Non-punitive response to errors
  - Handoffs and transitions
  - Feedback and communication about errors
  - Communication openness
  - Teamwork within the unit

De Oliveira GS Jr et al. Anesth Analg 2015

Indications that Quality of Supervision Matters

- Among the dozens of variables studied in national survey of residents' perceptions of their current rotation, supervision score most closely predicted by same one variable using multiple types of regression trees
  - <u>Teamwork</u> within the unit

De Oliveira GS Jr et al. Anesth Analg 2015

Indications that Quality of Supervision Matters

- Among the dozens of variables studied in national survey of residents' perceptions of their current rotation, supervision score most closely predicted by same one variable using multiple types of regression trees
  - Teamwork within the unit

De Oliveira GS Jr et al. Anesth Analg 2015

Professionalism

- Departments required to provide hospitals with physician-specific metrics demonstrating competence in professional practice
  - How anesthesiologists maintain privileges
  - Preceding section of lecture on Ongoing Professional Practice Evaluation (OPPE)
- Such assessments include the core competency of professionalism

Professionalism

- Supervision scale includes 8 phrases pertaining to professionalism

- Multiple written comments provided by residents with below average supervision scores pertained to professionalism

Dexter F et al. Can J Anesth 2017

Professionalism

- Each increase in the anesthesiologist's number of resident comments of the anesthesiologist being [disrespectful] was associated with a lower average score (P = 0.0002)

- A supervision score < 3.00 ("frequent") had odds ratio of 85 for resident written comment of disrespectful faculty behavior (P < 0.0001)

Dexter F et al. Anesth Analg 2016

ACGME Clinician Educator Milestones

- Pillar 1: Reflective Practice
  - Evaluate teaching activities
- Pillar 3: Recognition and Mitigation of Bias
  - Mitigate the effect of bias
- Pillar 4: Professional Responsibilities
  - ✓ Exemplary professional behavior Confirmed

ACGME. Clinician educator supplemental guide, August 2022

Summary of Attributes of Quality Supervision

- Attributes in sequence of earlier presentation
  - Quality of clinical teaching
  - Clinical performance
  - Engagement
  - Safety culture
  - Teamwork
  - Professionalism

Influence of Reporting Scores on the Supervision Scores

Influence of Reporting Scores on the Supervision Scores

- Monitoring anesthesiologists' supervision and reporting them resulted in greater scores for both residents and nurse anesthetists
  - Multiple comparisons, all P ≤ 0.0011
- Among nurse anesthetists, increase due mostly to questions associated with teaching (e.g., "stimulate my clinical reasoning, critical thinking, and theoretical learning")

Dexter F, Hindman BJ. Anesth Analg 2015

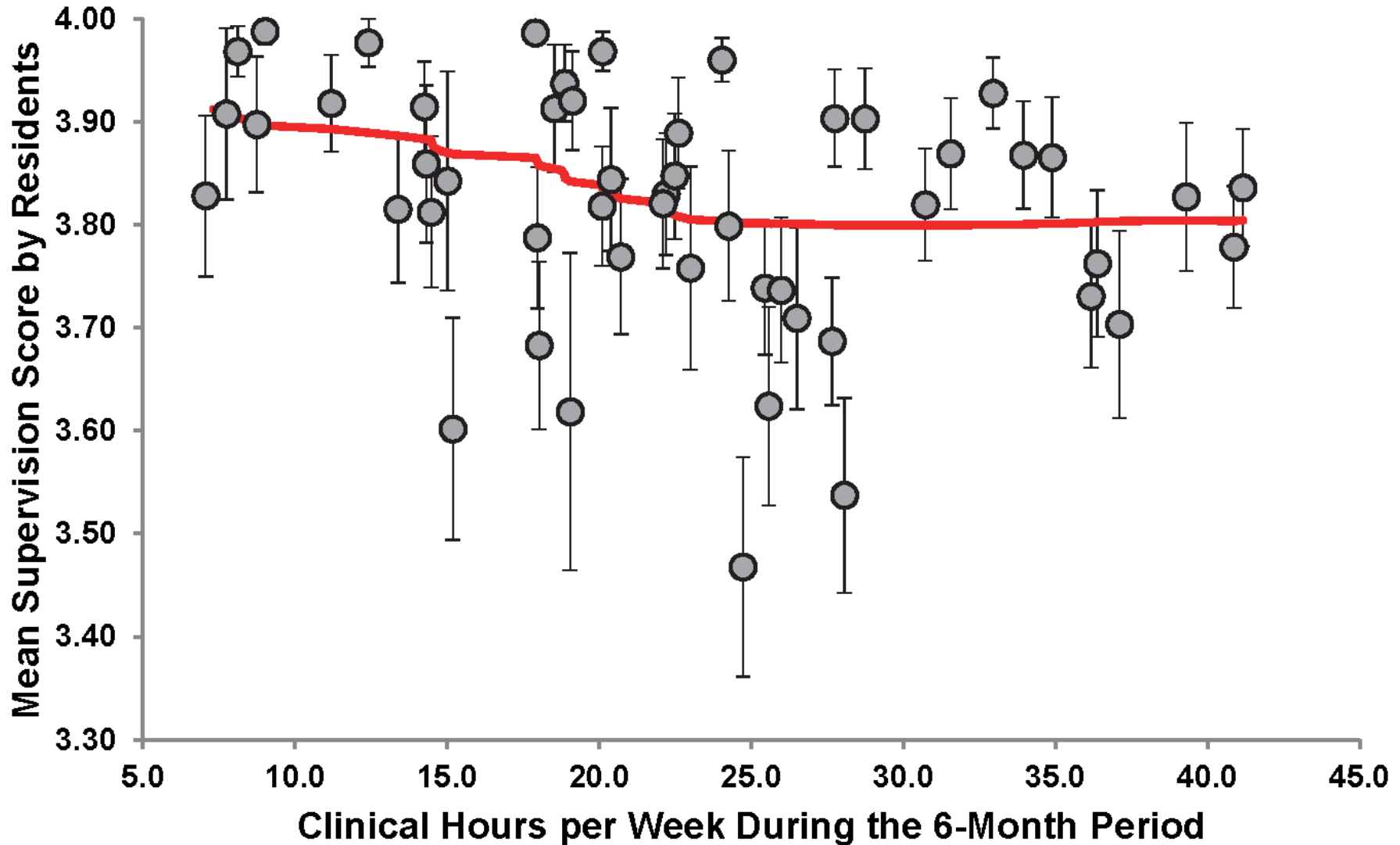# Value of Evaluating Supervision Scores for Anesthesiologists

Value of Evaluating Supervision Scores for Anesthesiologists

- Anesthesiologists' mean supervision scores provided both by residents and nurse anesthetists were not positively correlated with hours of faculty clinical activity
  - Multiple comparisons, all P > 0.65

Dexter F, Hindman BJ. Anesth Analg 2015

Value of Evaluating Supervision Scores for Anesthesiologists

Value of Evaluating Supervision Scores for Anesthesiologists

- Active anesthesiologist can provide ineffective supervision and a less frequent anesthesiologist can be very effective
  - Evaluating quality of supervision serves as independent measure of the value each anesthesiologist adds to care of the patients

Dexter F, Hindman BJ. Anesth Analg 2015

Value of Evaluating Supervision Scores for Department

Value of Evaluating Supervision Scores for Department

- Anesthesiologists' supervision of residents is mandatory and evaluated for reaccreditation

- Residents' mean ± SD of daily supervision score meeting expectations is 3.40 ± 0.30

- Evaluations of department and of individual anesthesiologists using their averages are correlated (Kendall τ_b = 0.35, P = 0.0032)
  - Median ratio 86% (SE 1%)

Dexter F et al. Anesth Analg 2013
Hindman BJ et al. Anesth Analg 2015

Value of Evaluating Supervision Scores for Department

- Anesthesiologists' supervision of residents is mandatory and evaluated for reaccreditation

- Residents' mean ± SD of daily supervision score meeting expectations is 3.40 ± 0.30

- Evaluations of department and of individual anesthesiologists using their averages are correlated (Kendall τ_b = 0.35, P = 0.0032)
  - Median ratio 86% (SE 1%)

➢ Achieve departmental score ≥ 3.00 by achieving individual average ≥ 3.40

Evaluating Supervision Less Often than Daily?

- Instead of daily evaluation, maximum one evaluation of ratee by a rater weekly?

- Many fewer requests (14%), but evaluations then must be completed on day was requested

- However, most (78%) evaluations are completed after the day requested

- In practice, then, very small (2%) reduction in evaluation requests can be achieved in practice

Dexter F et al. Cureus 2023

Covariates

Covariates Not Important

- Residency class
  - No association between residents' perception of supervision by anesthesiologists that meets expectations and years since start of training (P = 0.77)
  - Small differences among classes in scores
    - Mean differences $\leq$ 0.07 units

Dexter F et al. Anesth Analg 2013

Hindman BJ et al. Anesth Analg 2013

Covariates Not Important

- Negligible differences in residents' scores when
  - Resident had more units of work that day with the anesthesiologist (τ_b = +0.083 [SE 0.014])
  - Anesthesiologist had more units of work that day with other providers (τ_b = −0.057 [SE 0.014])
- No association between residents' scores and
  - Patients cared for together (τ_b = +0.01, P=0.71)
  - Days worked together (τ_b = −0.01, P=0.46)

Dexter F et al. Anesth Analg 2014

Hindman BJ et al. Anesth Analg 2013

Covariates Not Important

- Absence (P > 0.10) of correlation between residents' ratings of their rotations and:
  - Residents'   age
                    hours worked per week
                    gender
  - Program   size (number of residents)
                    rotation (specialty)

De Oliveira GS Jr et al. Anesth Analg 2013

Covariates Not Important

- Absence (P > 0.10) of correlation between residents' ratings of supervision quality and:
  - Cases performed at night,
  - Cases of high physiological complexity,
  - Cases with high physical status of patients,
  - Cases with pediatric patients,
  - Cases with elderly patients

Dexter F et al. Cureus 2024

Covariates Not Important

- Specialization of anesthesiologist
  - Calculate Herfindahl of distribution of each anesthesiologist's anesthesia CPT codes
  - Herfindahl^{-1} = number of common procedures
  - No association between specialization and quality of supervision of residents (P = 0.31)
  - Specialization is associated with <u>lesser</u> quality scores among nurse anesthetists (P = 0.0001), but differences are small

Dexter F et al. Anesth Analg 2016
Dexter F et al. Anesth Analg 2017
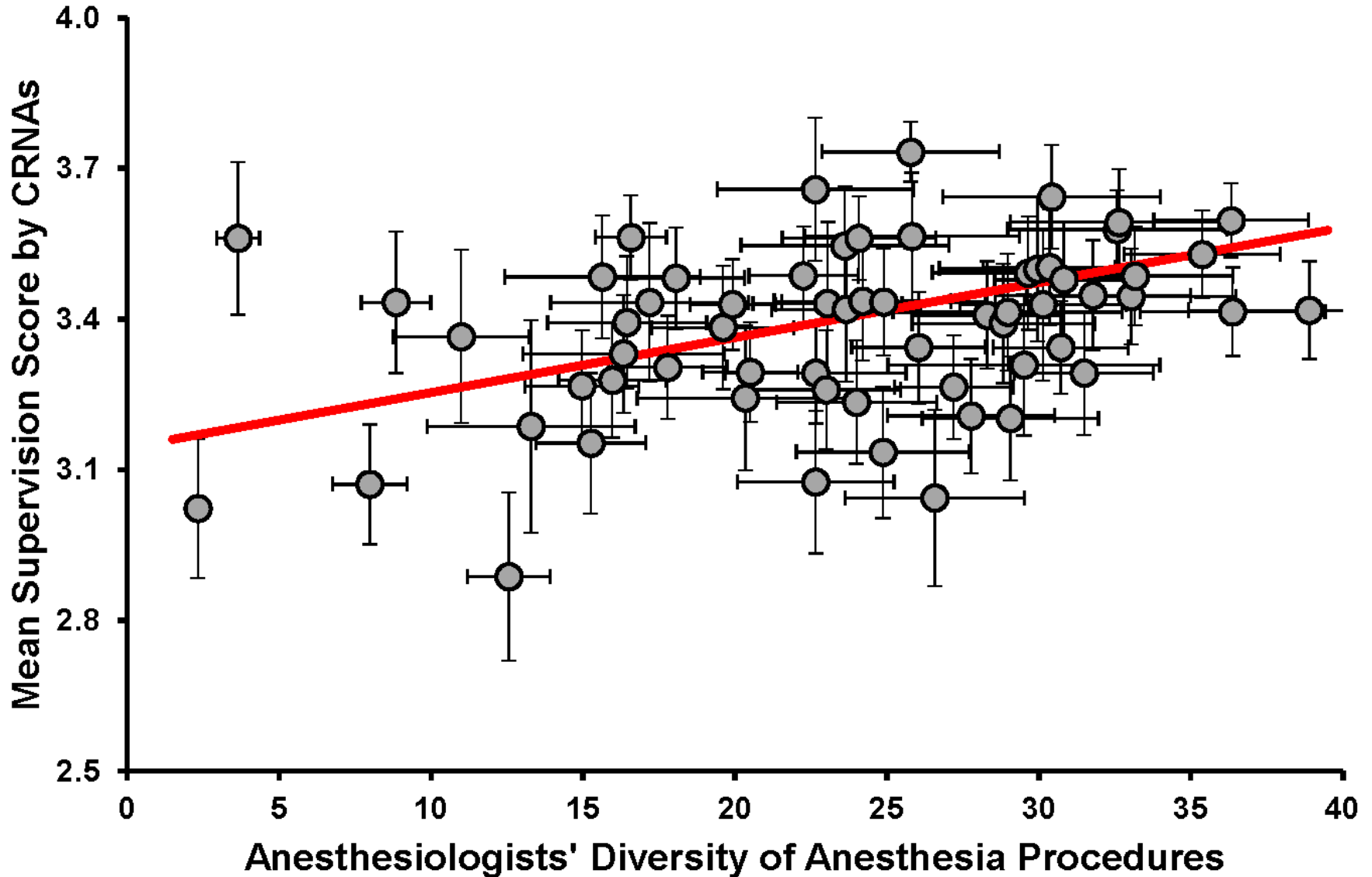
Covariates to Include

- Control for resident vs. nurse anesthetist
  - Scores provided by residents greater than by nurse anesthetists (P < 0.0001)
  - Pairwise differences by anesthesiologist greater than zero too (P < 0.0001)

Dexter F et al. Anesth Analg 2014
Dexter F et al. Anesth Analg 2015

Covariates to Include

- Leniency of the resident (or nurse anesthetist)
  - Scientific term for heterogeneity among raters

Dexter F et al. Health Care Manag Sci 2020

Dexter F et al. Can J Anesth 2017

Covariates to Include

- Leniency of the resident (or nurse anesthetist)
  - Scientific term for heterogeneity among raters
  - ➤ Unless adjust for rater leniency, evaluations of anesthesiologists' clinical performance <u>will</u> be biased, even if evaluations qualitative

Bayman EO et al. Perioper Care Oper Room Manag 2017

Dexter F et al. Perioper Care Oper Room Manag 2022

Dexter F et al. J Clin Anesth 2017

Dexter F et al. J Clin Anesth 2020

Dexter F et al. Health Care Manag Sci 2020

Dexter F et al. Can J Anesth 2017

Dexter F et al. Anesth Analg 2020

Covariates to Include

- Leniency of the resident (or nurse anesthetist)
  - Scientific term for heterogeneity among raters
  - Unless adjust for rater leniency, evaluations of anesthesiologists' clinical performance will be biased, even if evaluations qualitative
  - ➤ Raw scores are not just hypothetically influenced by <u>implicit bias</u>, subsequent results show that <u>they are</u> biased
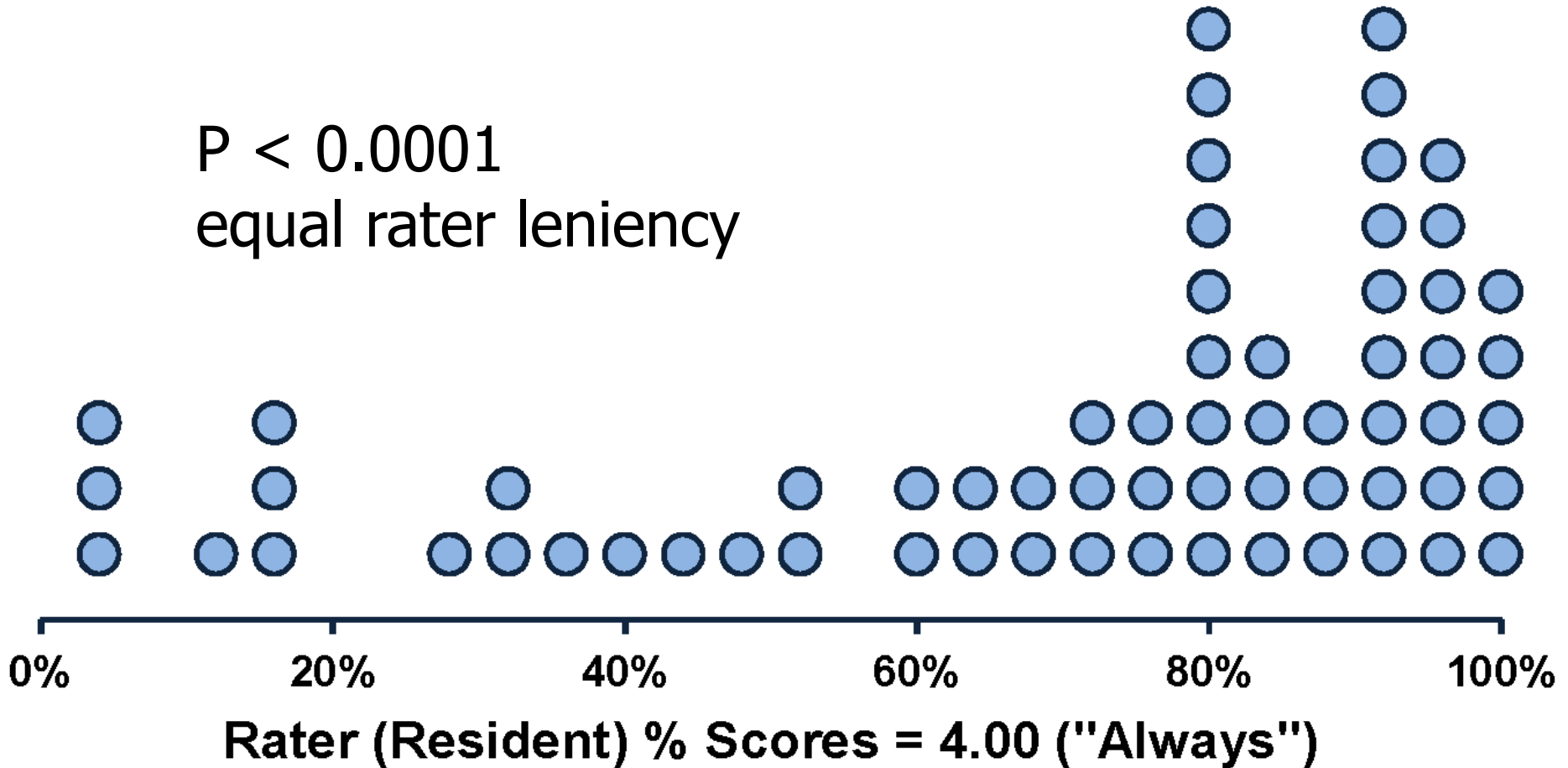
Covariates to Include

- Leniency of the resident (or nurse anesthetist)
  - Scientific term for heterogeneity among raters
  - From cumulative effect of all questions
    - For each rater, calculate mean answer to each of the 9 questions among all ratees
    - Cronbach α = 0.98, very large
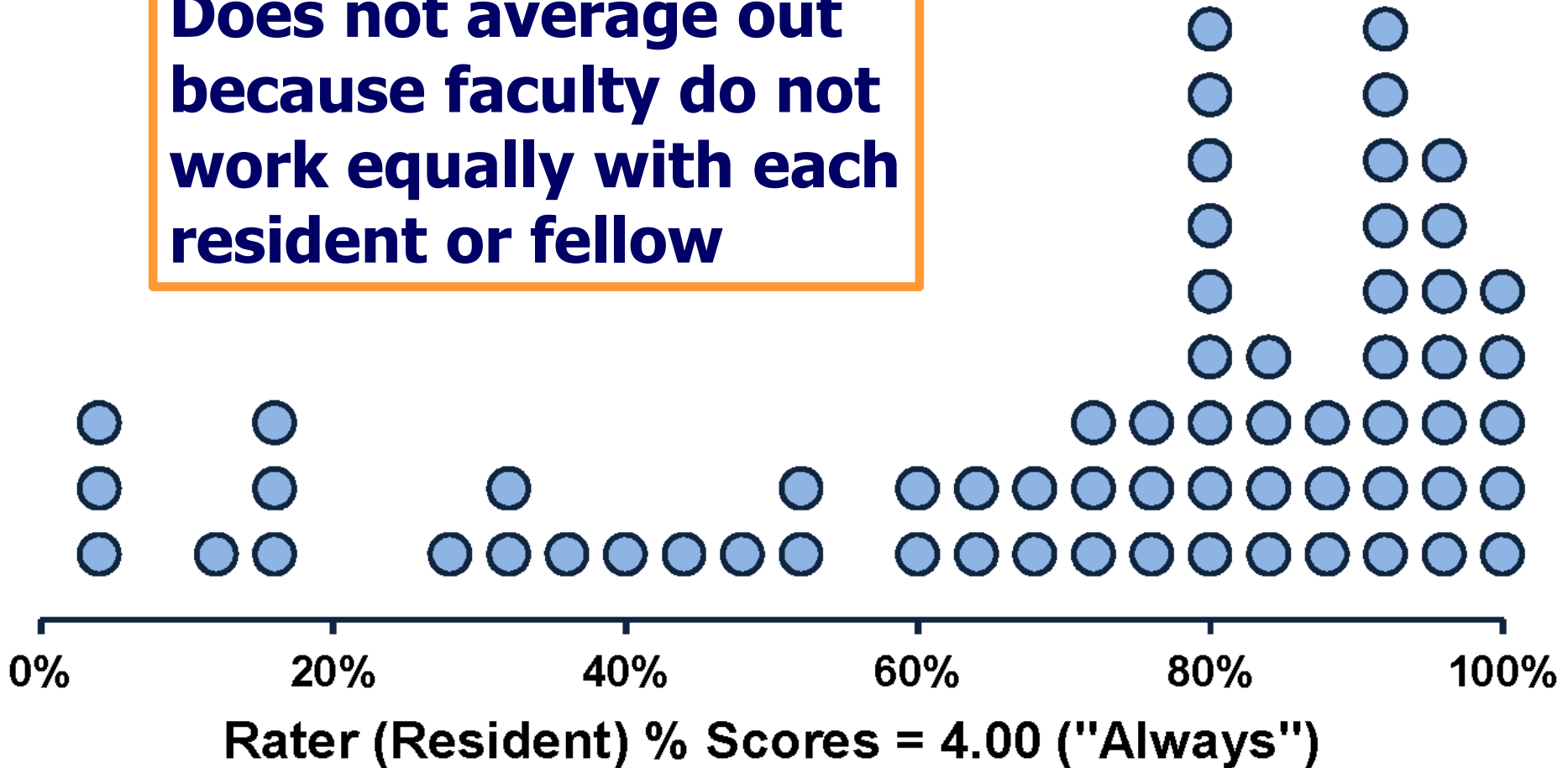
Dexter F et al. Can J Anesth 2017

Covariates to Include

P < 0.0001
equal rater leniency

Rater (Resident) % Scores = 4.00 ("Always")

Dexter F et al. Can J Anesth 2017

Covariates to Include

Does not average out because faculty do not work equally with each resident or fellow

Rater (Resident) % Scores = 4.00 ("Always")
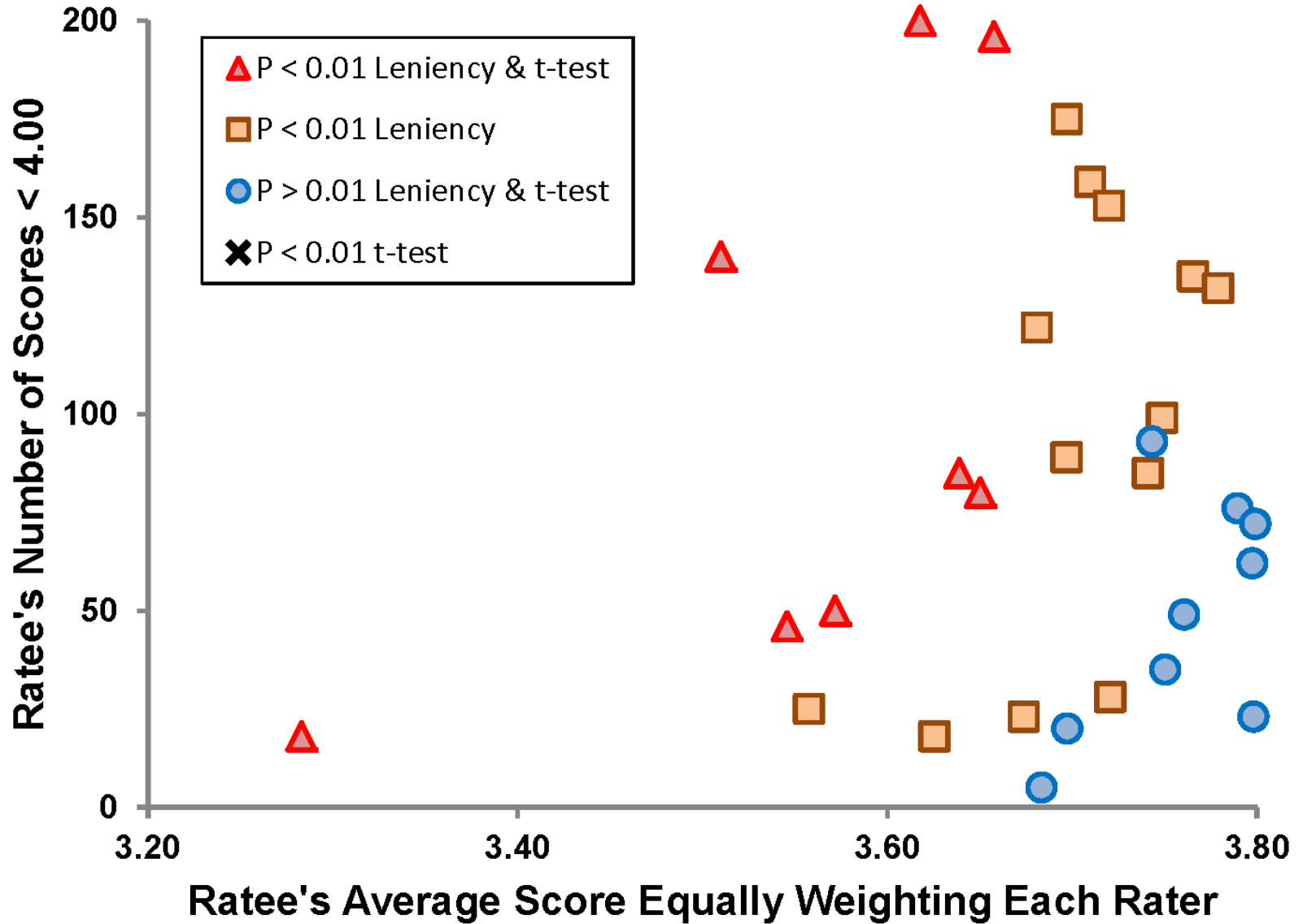
Dexter F et al. Can J Anesth 2017

Covariates to Include

- For assessment and progressive quality improvement within a department, use logistic regression of % scores = 4.00, treating the rater as a covariate
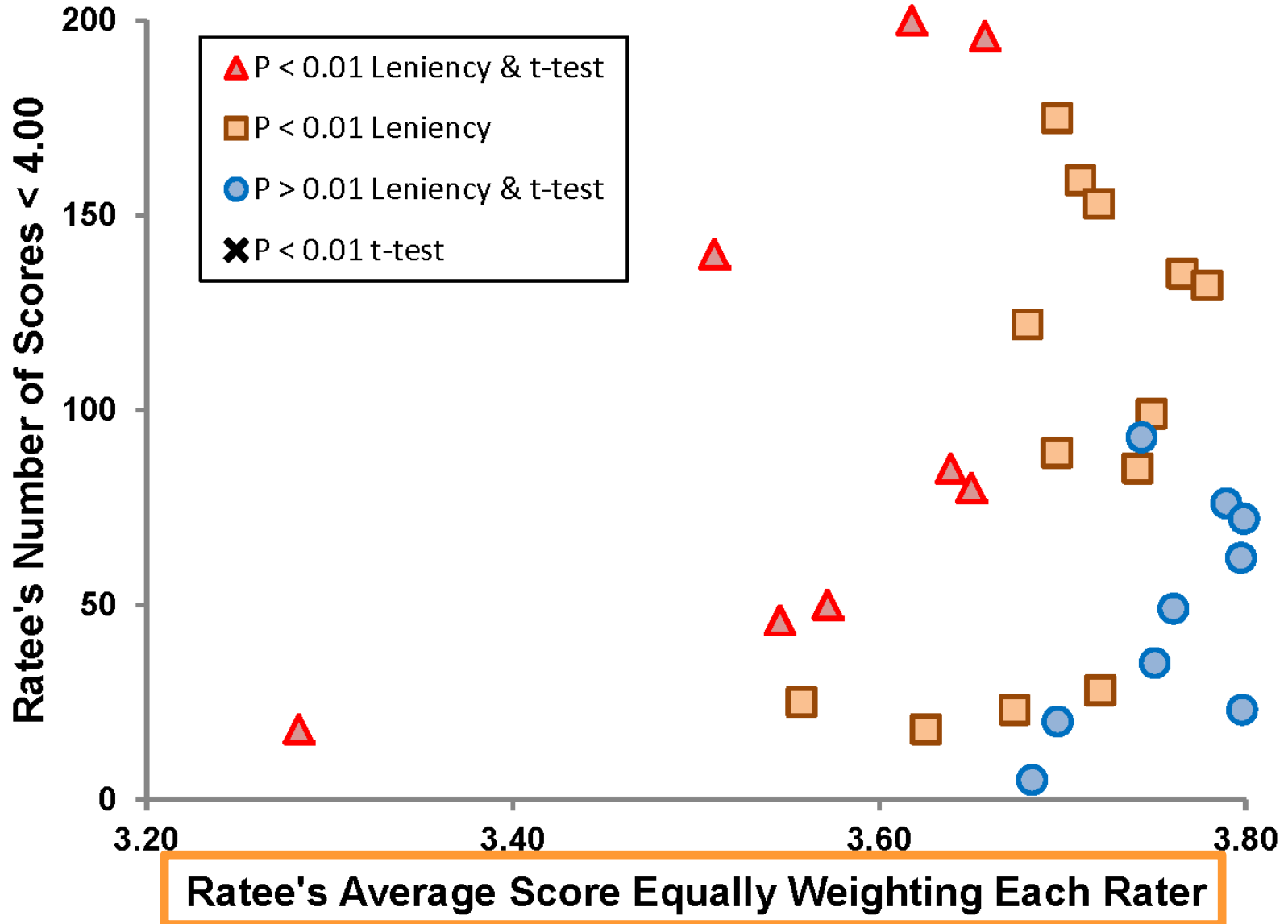
Dexter F et al. Can J Anesth 2017

Covariates to Include

Ratee's Number of Scores < 4.00 (y-axis)

Ratee's Average Score Equally Weighting Each Rater (x-axis)

Legend:
- △ P < 0.01 Leniency & t-test
- □ P < 0.01 Leniency
- ● P > 0.01 Leniency & t-test
- ✖ P < 0.01 t-test

Covariates to Include

Legend:
- △ P < 0.01 Leniency & t-test
- ☐ P < 0.01 Leniency
- ○ P > 0.01 Leniency & t-test
- ✖ P < 0.01 t-test

Y-axis: Ratee's Number of Scores < 4.00

X-axis: Ratee's Average Score Equally Weighting Each Rater

Covariates to Include

Ratee's Number of Scores < 4.00 (y-axis)

Ratee's Average Score Equally Weighting Each Rater (x-axis)

Legend:
- △ P < 0.01 Leniency & t-test
- □ P < 0.01 Leniency
- ○ P > 0.01 Leniency & t-test
- ✘ P < 0.01 t-test

Covariates to Include

Scatter plot. X-axis: Ratee's Average Score Equally Weighting Each Rater (3.20 to 3.80). Y-axis: Ratee's Number of Scores < 4.00 (0 to 200).

Legend:
- △ P < 0.01 Leniency & t-test
- □ P < 0.01 Leniency
- ○ P > 0.01 Leniency & t-test
- ✖ P < 0.01 t-test

Covariates to Include

Covariates to Include

Legend:
- △ P < 0.01 Leniency & t-test
- □ P < 0.01 Leniency
- ○ P > 0.01 Leniency & t-test
- ✕ P < 0.01 t-test

X-axis: Ratee's Average Score Equally Weighting Each Rater

Y-axis: Ratee's Number of Scores < 4.00

Covariates to Include

Legend:
- △ P < 0.01 Leniency & t-test
- ▢ P < 0.01 Leniency
- ● P > 0.01 Leniency & t-test
- ✖ P < 0.01 t-test

P = 0.0005
Leniency better detector

Y-axis: Ratee's Number of Scores < 4.00
X-axis: Ratee's Average Score Equally Weighting Each Rater

Example OPPE Report
to Department Chair

Ratee	Odds ratio (99% interval) (max score / raw count)
20207057	0.08 (< 0.26) (15/31)
20202896	0.17 (< 0.38) (51/78)
20203404	0.24 (< 0.64) (54/70)
20205255	0.31(< 0.76) (45/65)
20200888	4.81 (> 1.26) (58/63)
20203008	4.93 (> 1.29) (52/60)
20203267	13.44 (> 1.93) (25/28)
20205510	0.17 (< 1.71) (6/11)
20207156	0.19 (< 1.03) (13/20)
20201272	1.15 (> 0.29) (35/41)

Example OPPE Report to Department Chair

Ratee	Odds ratio (99% interval) (max score / raw count)
20207057	0.08 (< 0.26) (15/31)
20202896	0.17 (< 0.38) (51/78)
20203404	0.24 (< 0.64) (54/70)
20205255	0.31(< 0.76) (45/65)
20200888	4.81 (> 1.26) (58/63)
20203008	4.93 (> 1.29) (52/60)
20203267	13.44 (> 1.93) (25/28)
20205510	0.17 (< 1.71) (6/11)
20207156	0.19 (< 1.03) (13/20)
20201272	1.15 (> 0.29) (35/41)

Benefit of Adjusting for Rater Leniency with 1 Year Data

Unadjusted logistic regression failed to detect that anesthesiologist significantly in lower half, but mixed effects model found odds ratio less than 1.00	7% (5/73)
Unadjusted logistic regression falsely detected that anesthesiologist significantly in lower half, while mixed effects model found odds ratio not significantly different than 1.00	3% (2/73)
Unadjusted logistic regression failed to detect that anesthesiologist significantly in upper half, but mixed effects model found odds ratio greater than 1.00	11% (8/73)
Overall misclassification (just for above or below average) using unadjusted analysis	21% (15/73)

Dexter F et al. J Clin Anesth 2020

Benefit of Adjusting for Rater Leniency with 1 Year Data

Unadjusted logistic regression failed to detect that anesthesiologist significantly in lower half, but mixed effects model found odds ratio less than 1.00	7% (5/73)
Unadjusted logistic regression falsely detected that anesthesiologist significantly in lower half, while mixed effects model found odds ratio not significantly different than 1.00	3% (2/73)
Unadjusted logistic regression failed to detect that anesthesiologist significantly in upper half, but mixed effects model found odds ratio greater than 1.00	11% (8/73)
Overall misclassification (just for above or below average) using unadjusted analysis	**21%** (15/73)

Dexter F et al. J Clin Anesth 2020

Benefit of Adjusting for Rater Leniency with 1 Year Data

Unadjusted logistic regression failed to detect that anesthesiologist significantly in lower half, but mixed effects model found odds ratio less than 1.00	7% (5/73)
Unadjusted logistic regression falsely detected that anesthesiologist significantly in lower half, while mixed effects model found odds ratio not significantly different than 1.00	3% (2/73)
Unadjusted logistic regression failed to detect that anesthesiologist significantly in upper half, but mixed effects model found odds ratio greater than 1.00	11% (8/73)
Overall misclassification (just for above or below average) using unadjusted analysis	**21% (15/73)**

Dexter F et al. J Clin Anesth 2020

ACGME Clinician Educator Milestones

- Pillar 1: Reflective Practice
  - Evaluate teaching activities
- Pillar 3: Recognition and Mitigation of Bias
  - ✓ Mitigate the effect of bias   Yes
- Pillar 4: Professional Responsibilities
  - Exemplary professional behavior

ACGME. Clinician educator supplemental guide, August 2022

Feedback to Raters to Increase Information from Evaluations

- None of the following raters has provided incremental information about ratees
  - 100 evaluation requests, all completed, all ratees given maximum score of 4.00
  - 100 evaluation requests, 50 completed, no ratee given maximum score of 4.00
  - 100 evaluation requests, 0 completed

Dexter F et al. Perioper Care Oper Room Manag 2022

Feedback to Raters to Increase Information from Evaluations

- None of the following raters has provided incremental information about ratees
  - 100 evaluation requests, all completed, all ratees given maximum score of 4.00
  - 100 evaluation requests, 50 completed, no ratee given maximum score of 4.00
  - 100 evaluation requests, 0 completed

➢ Completing evaluations shows raters' work habits and conscientiousness, but objective of evaluation is to learn about the ratees

Dexter F et al. Perioper Care Oper Room Manag 2022

Feedback to Raters to Increase Information from Evaluations

- None of the following raters has provided incremental information about ratees
  - 100 evaluation requests, all completed, all ratees given maximum score of 4.00
  - 100 evaluation requests, 50 completed, no ratee given maximum score of 4.00
  - 100 evaluation requests, 0 completed

Scores, but provide no information

Dexter F et al. Perioper Care Oper Room Manag 2022

Feedback to Raters to Increase Information from Evaluations

- Insight from anesthesiologists' 40,027 evaluations of nurse anesthetists' work habits
  - Few (12%) consecutive ratings by same rater included >10 ratings with all scores the same

Dexter F et al. Perioper Care Oper Room Manag 2022

Feedback to Raters to Increase Information from Evaluations

- Insight from anesthesiologists' 40,027 evaluations of nurse anesthetists' work habits
  - Few (12%) consecutive ratings by same rater included >10 ratings with all scores the same
    - ➤ Those runs mattered, nonetheless, because ratees best evaluated by multiple raters

Dexter F et al. Perioper Care Oper Room Manag 2022

Feedback to Raters to Increase Information from Evaluations

- Insight from anesthesiologists' 40,027 evaluations of nurse anesthetists' work habits
  - Few (12%) consecutive ratings by same rater included >10 ratings with all scores the same
    - Those runs mattered, nonetheless, because ratees best evaluated by multiple raters
    - ➤ Those runs continued, median 13 additional ratings with scores the same

Dexter F et al. Perioper Care Oper Room Manag 2022

Bernoulli CUSUM Monitoring for Prompt Recognition Low Scores

Bernoulli CUSUM Monitoring for Prompt Recognition Low Scores

- Daily monitoring by server to detect changes in supervision scores promptly

Dexter F et al. Anesth Analg 2014

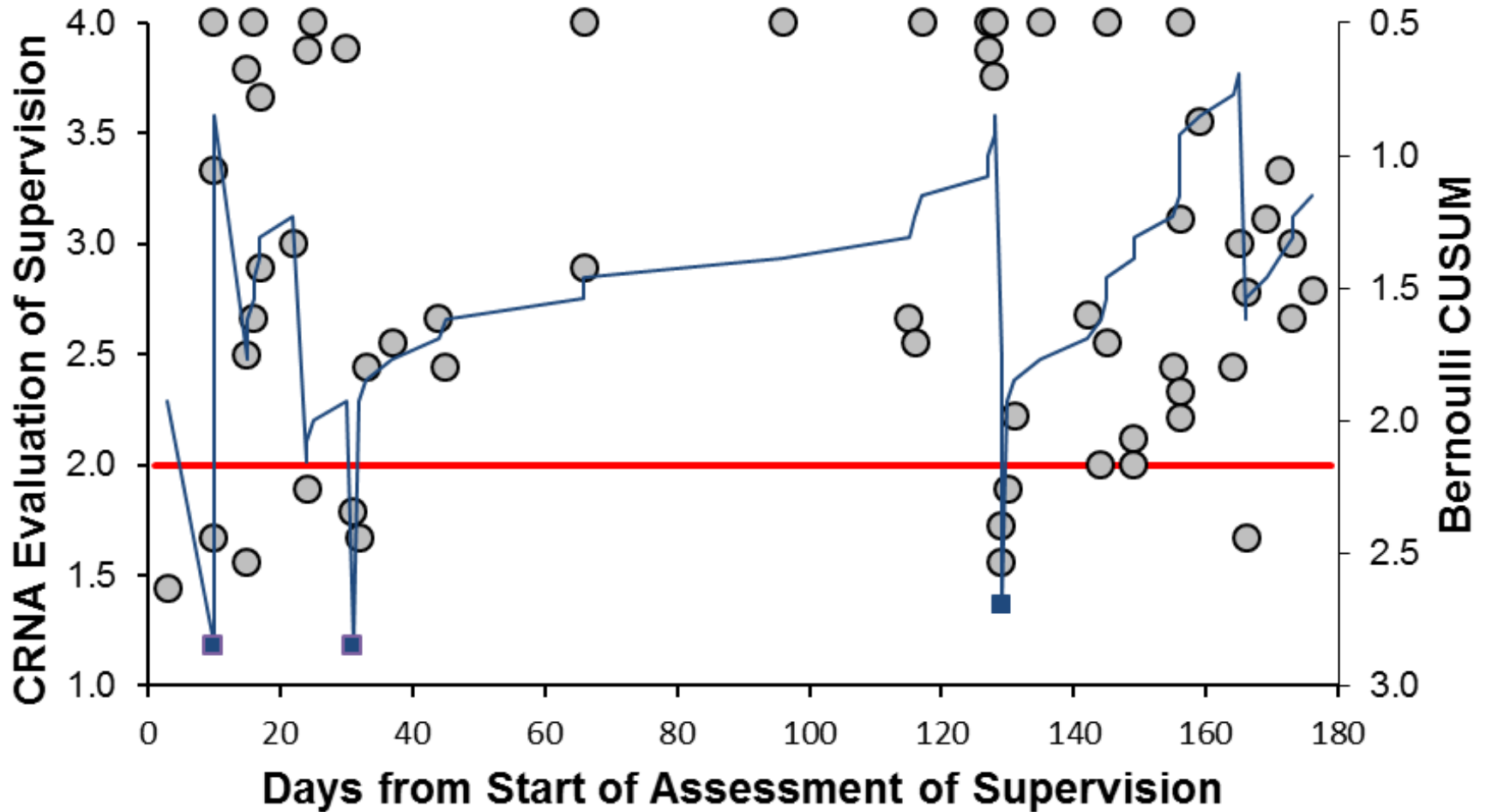Dexter F et al. Can J Anesth 2017

Bernoulli CUSUM Monitoring for Prompt Recognition Low Scores

- Example for nurse anesthetists
  - Bernoulli CUSUM starting value = 1 − 1/13
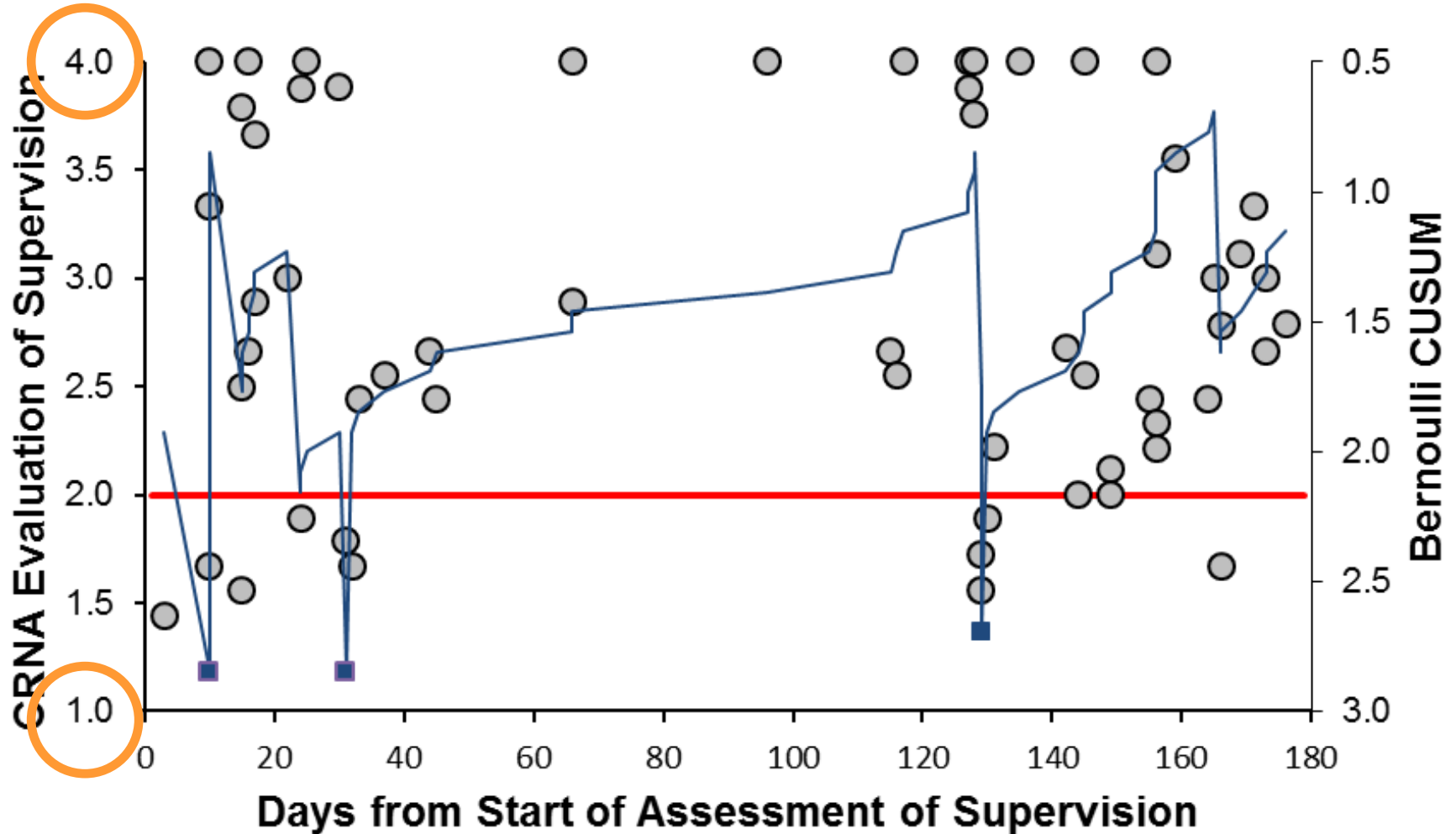 - Add (1 − 1/13) if score < 2.00 ("rarely") or subtract (1/13) otherwise
  - Bernoulli CUSUM alert when > 2.32 and restart
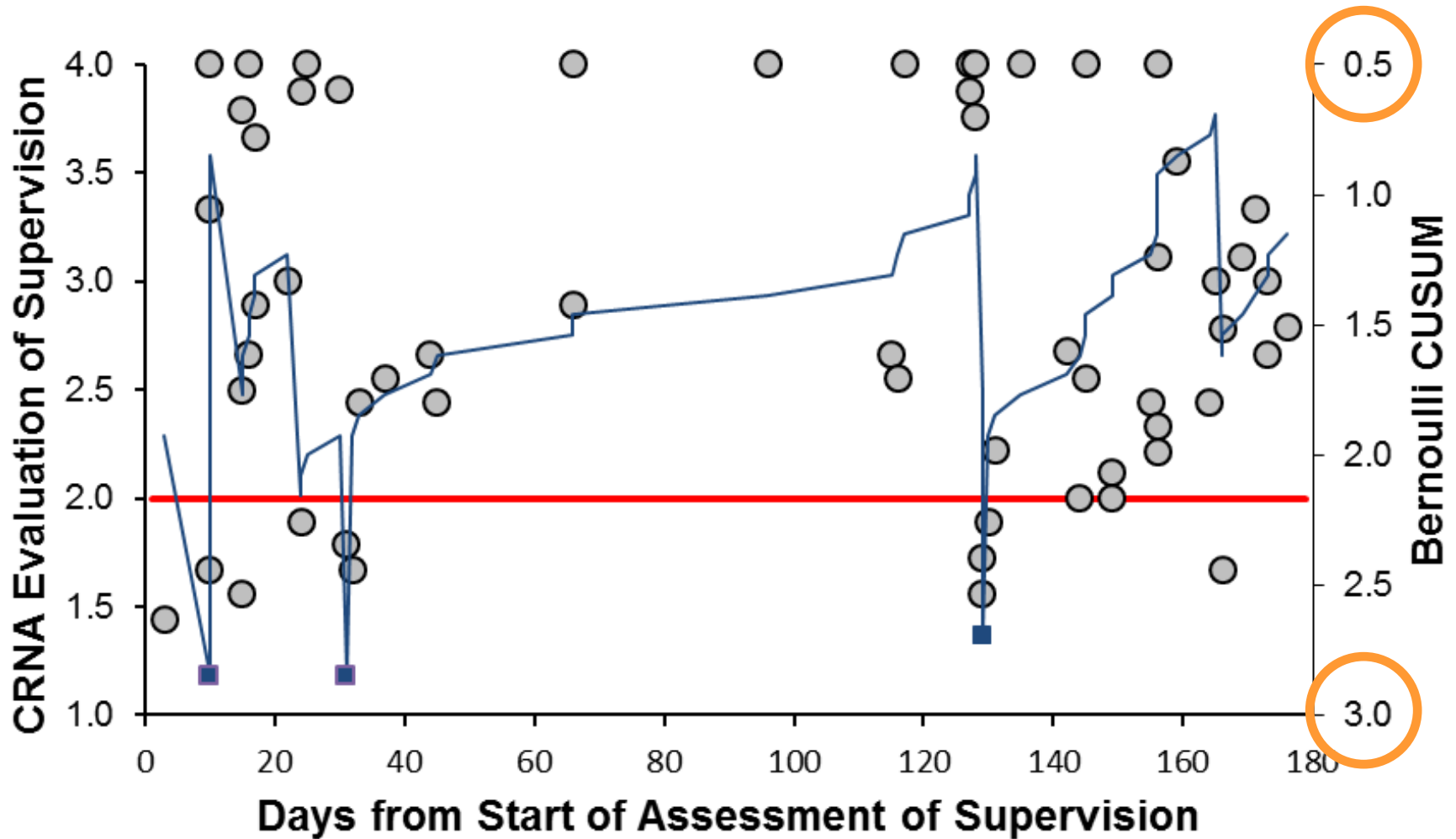
Dexter F et al. Anesth Analg 2014

Bernoulli CUSUM Monitoring for Prompt Recognition Low Scores

Bernoulli CUSUM Monitoring for Prompt Recognition Low Scores

# Bernoulli CUSUM Monitoring for Prompt Recognition Low Scores

Bernoulli CUSUM Monitoring for Prompt Recognition Low Scores

- Evaluation by anesthesia residents
  - Among upper half of anesthesiologists (27/55), based on their average scores, zero of 27 was detected (flagged) during the 6 months by the Bernoulli CUSUM
  - Among the lower quartile of anesthesiologists (13/55), 12 of 13 were detected

Dexter F et al. Anesth Analg 2014

Bernoulli CUSUM Monitoring for Prompt Recognition Low Scores

- Evaluation by nurse anesthetists
  - Among upper half of anesthesiologists (29/58) based on their average scores, only 1 of 29 was detected (flagged) during the 6 months by the Bernoulli CUSUM
  - Among the lower quartile of anesthesiologists (14/58), 13 of 14 were detected

Dexter F et al. Anesth Analg 2014

Do Need to Use Mathematics
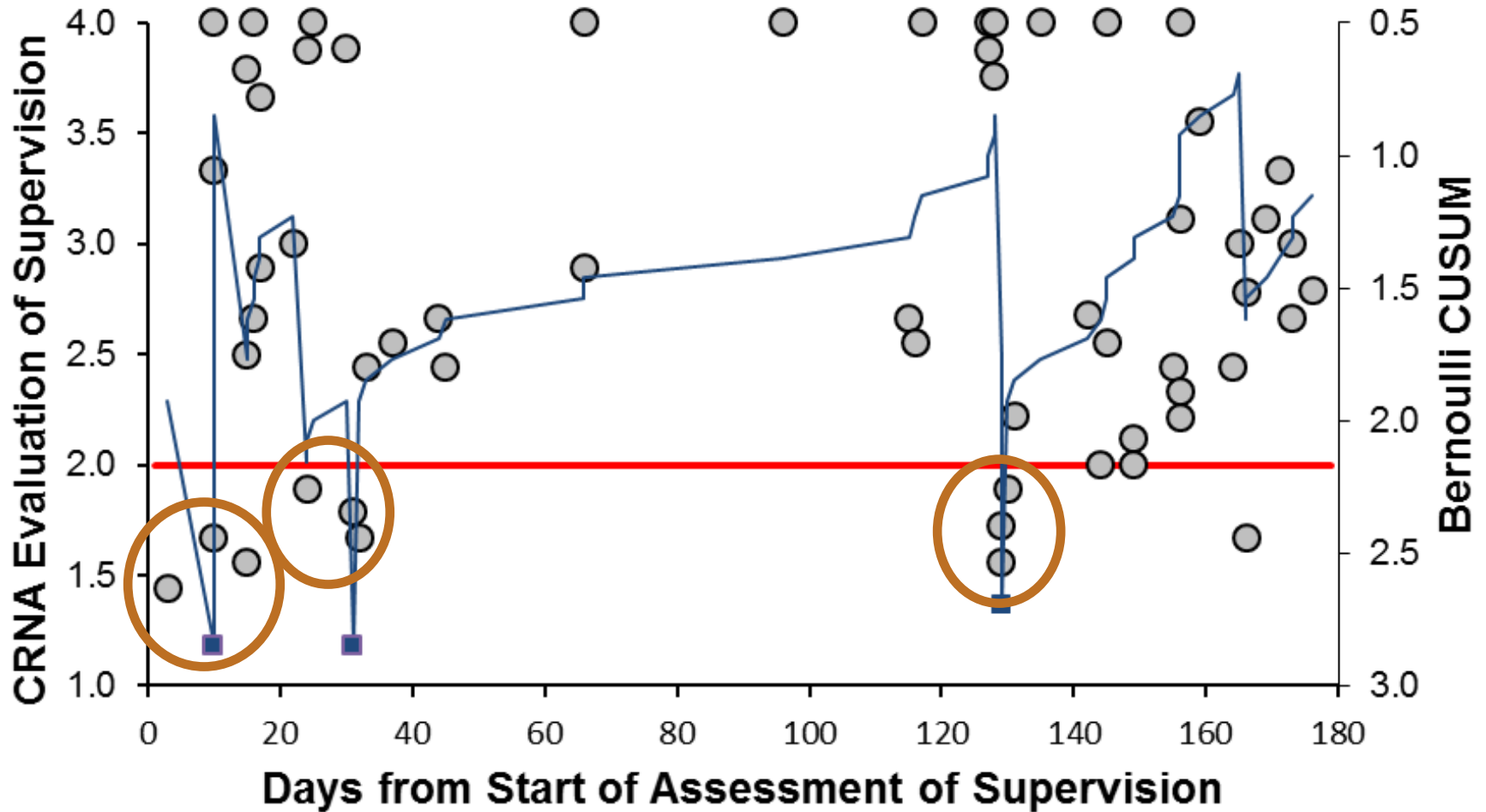
<u>Do</u> Need to Use Mathematics

- Assumption of statistical independence
  - If no correlation among evaluations, and with p representing pooled estimate for low score, then among days with 2 evaluations, p^2 would be probability both scores are low
  - Among the nurse anesthetists' 1182 evaluations on days with 2 evaluations by nurse anesthetists, p = 5.92%
  - There were 4.34-fold more days with 2 low scores than expected at random (P < 0.0001)

Do Not Spend Substantial Time Maintaining Process: 1/Month

Date	Count of Scores	Mean of all Scores	% Scores < 3.00	Count CUSUM alerts
2014 Jan-Jun	4108	3.79	2.35%	13
2014 Jul-Dec	3777	3.82	1.53%	17
2015 Jan-Jun	4003	3.85	1.45%	10
2015 Jul-Dec	4492	3.86	0.70%	7
2016 Jan-Jun	3975	3.90	0.68%	3
2016 Jul-Dec	4356	3.91	0.89%	6
2017 Jan-Jun	4078	3.93	0.37%	1
2017 Jul-Dec	4334	3.94	0.84%	6

Bernoulli CUSUM Workflow for Who Receives the E-mail

Bernoulli CUSUM Workflow for Who Receives the E-mail

- If anesthesiologist works today with a resident, and this evening Bernoulli CUSUM alerts, likely the resident's evaluation indicated less than desirable supervision

- E-mail directly to the rated anesthesiologist would result in loss of confidentiality of the resident's evaluation

Dexter F et al. Anesth Analg 2014

Bernoulli CUSUM Workflow for Who Receives the E-mail

- Bernoulli CUSUM is process for detection
- Detection prompts e-mail notification of the relevant human resources professional, not the rated anesthesiologist
- Vice Chair for Faculty Development receives e-mail with hyperlink but without identifiers
  - Logs in
  - Sees name of anesthesiologist and evaluations from past 9 different raters

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Same principles to evaluate the quality of supervision of resident physicians and pain medicine fellows by pain medicine faculty

Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Same principles to evaluate the quality of supervision of resident physicians and pain medicine fellows by pain medicine faculty

➤ Cronbach α very large, 0.975 (SE 0.001)

Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Same principles to evaluate the quality of supervision of resident physicians and pain medicine fellows by pain medicine faculty

- Cronbach α very large, 0.975 (SE 0.001)

➤ G coefficient $\cong 0.90$ with 18 raters, practical because mean 19 raters per 6-months

Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Same principles to evaluate the quality of supervision of resident physicians and pain medicine fellows by pain medicine faculty

- Cronbach α very large, 0.975 (SE 0.001)

- G coefficient $\cong 0.90$ with 18 raters, practical because mean 19 raters per 6-months

➢ Concurrent validity based on lesser scores on weeks with more pain procedures performed (i.e., more supervision expected)

Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Concurrent validity based on correlation with fellows' evaluations using a 21-item graduate medical education scale, Kendall's $\tau_b = 0.45$, $P < 0.0001$

Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Concurrent validity based on correlation with fellows' evaluations using a 21-item graduate medical education scale, Kendall's $\tau_b = 0.45$, P < 0.0001

➢ Concurrent validity based on correlation with anesthesia residents' evaluations when faculty supervising OR cases, $\tau_b = 0.38$, P = 0.0002

Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Just like for supervision in ORs, important to adjust for rater leniency

Dexter F et al. Health Care Manag Sci 2020

Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Just like for supervision in ORs, important to adjust for rater leniency

  ➢ Large systematic heterogeneity of scores among raters, $\eta^2 = 0.48$, P < 0.0001
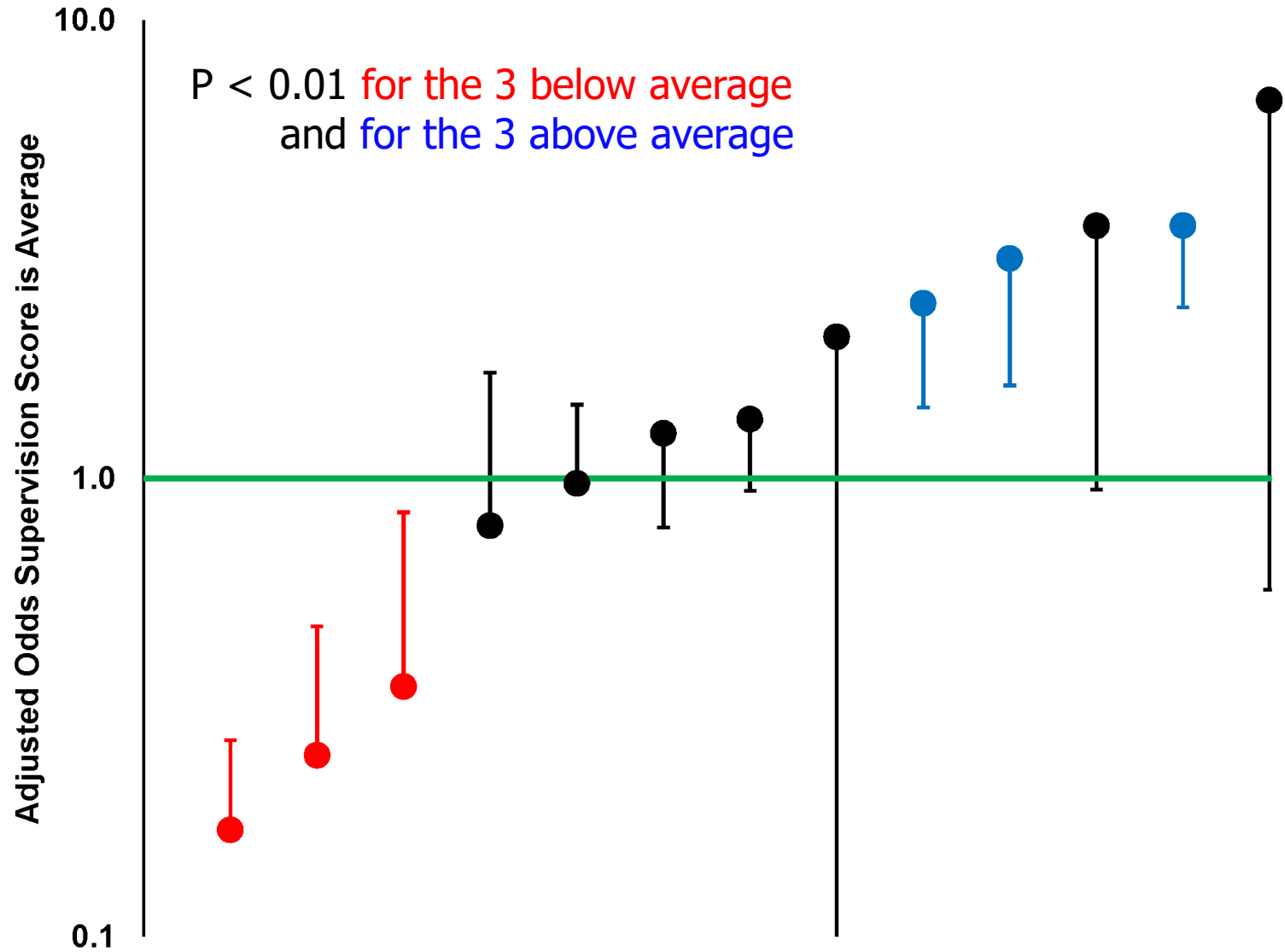
Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Just like for supervision in ORs, important to adjust for rater leniency
  - Large systematic heterogeneity of scores among raters, $\eta^2 = 0.48$, P < 0.0001
  - Pairings of ratee and rater decidedly non-random, Cramér's V = 0.349, P < 0.0001

Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

- Just like for supervision in ORs, important to adjust for rater leniency
    - Large systematic heterogeneity of scores among raters, $\eta^2 = 0.48$, $P < 0.0001$
    - Pairings of ratee and rater decidedly non-random, Cramér's V = 0.349, $P < 0.0001$
    - Raters' means nested by ratee not normally distributed, 10 $P < 0.0001$ and 11th $P = 0.0098$

Dexter F et al. Anesth Analg 2020

Residents & Fellows Evaluate Pain Medicine Clinical Faculty

P < 0.01 for the 3 below average
and for the 3 above average

Adjusted Odds Supervision Score is Average

10.0

1.0

0.1

Combine Individual Evaluations for Overall Department Quality

- Metric for internal and external use
  - Annual reporting to College of Medicine
  - Departmental reviews of training program
- Data used for development of methodology
  - Resident evaluations July 2013 to June 2022
  - 48,788 evaluations by 202 distinct raters
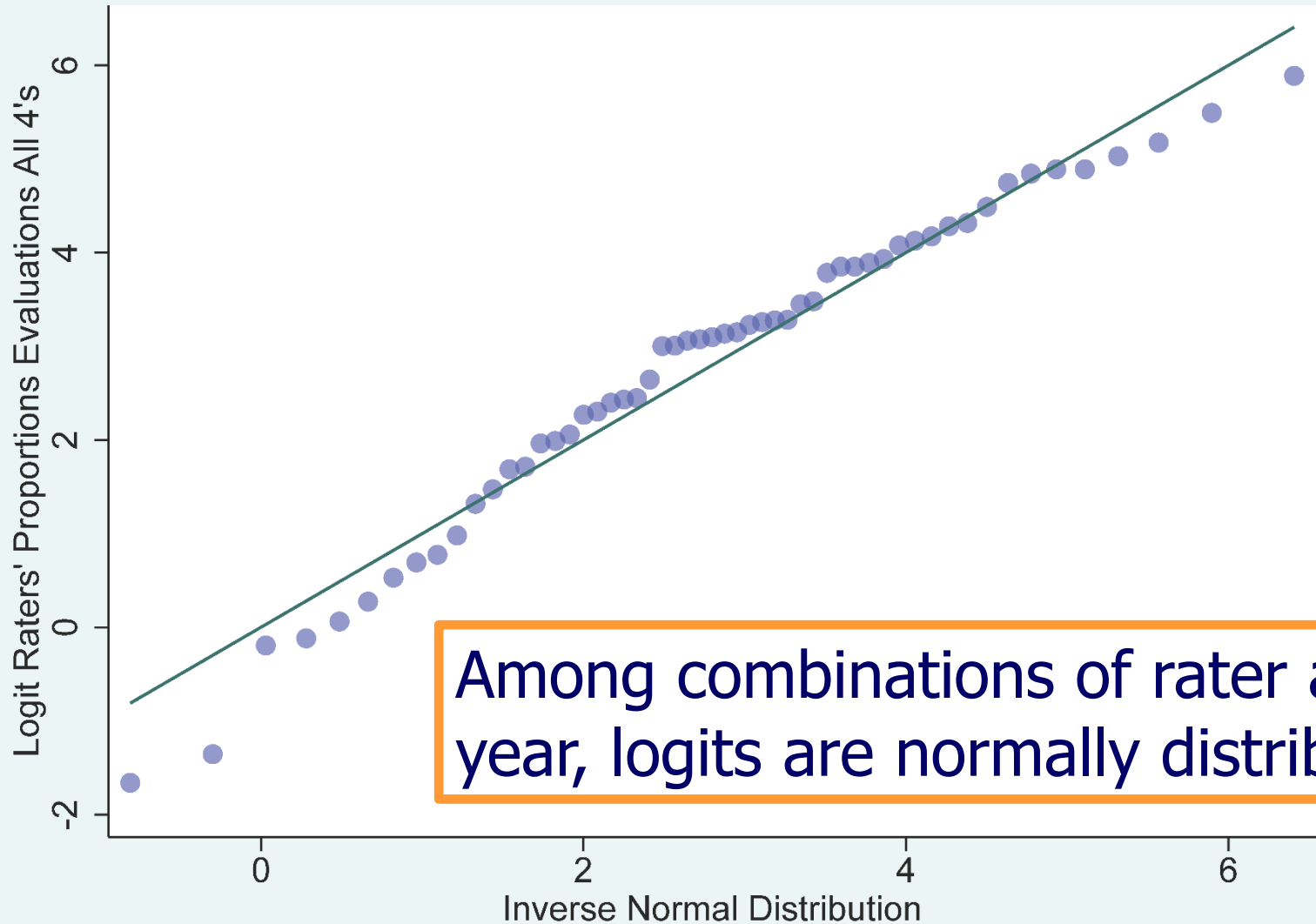
Dexter F et al. J Clin Anesth 2023

Combine Individual Evaluations for Overall Department Quality

- Mean scores equally weighting each rater?
  - For each of the 9 years, raters' mean scores negatively skewed, median G1 = 3.0
    - Lilliefors' tests all nine P $\leq$ 0.00001
  - Confidence intervals calculated using Chen's method vastly too wide (unreliable) for use

Dexter F et al. J Clin Anesth 2023

Combine Individual Evaluations for Overall Department Quality

Among combinations of rater and year, logits are normally distributed

Combine Individual Evaluations for Overall Department Quality

- Mixed effects logistic regression model for each academic year treating raters as random effect, because raters differ in leniency
  - Intercept only model
  - From estimated intercept and its 95% confidence interval, take inverse logit to obtain overall departmental quality as proportion
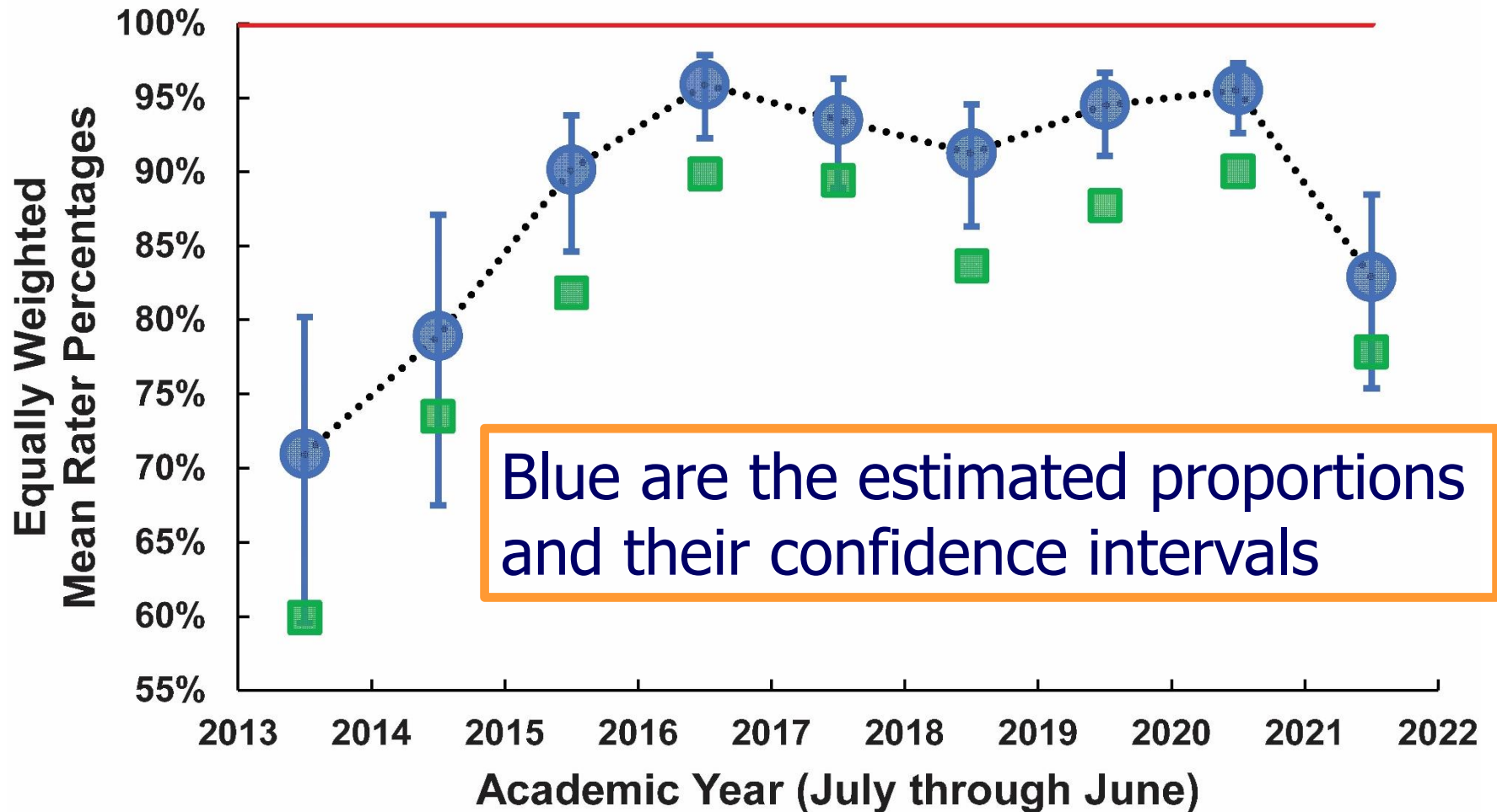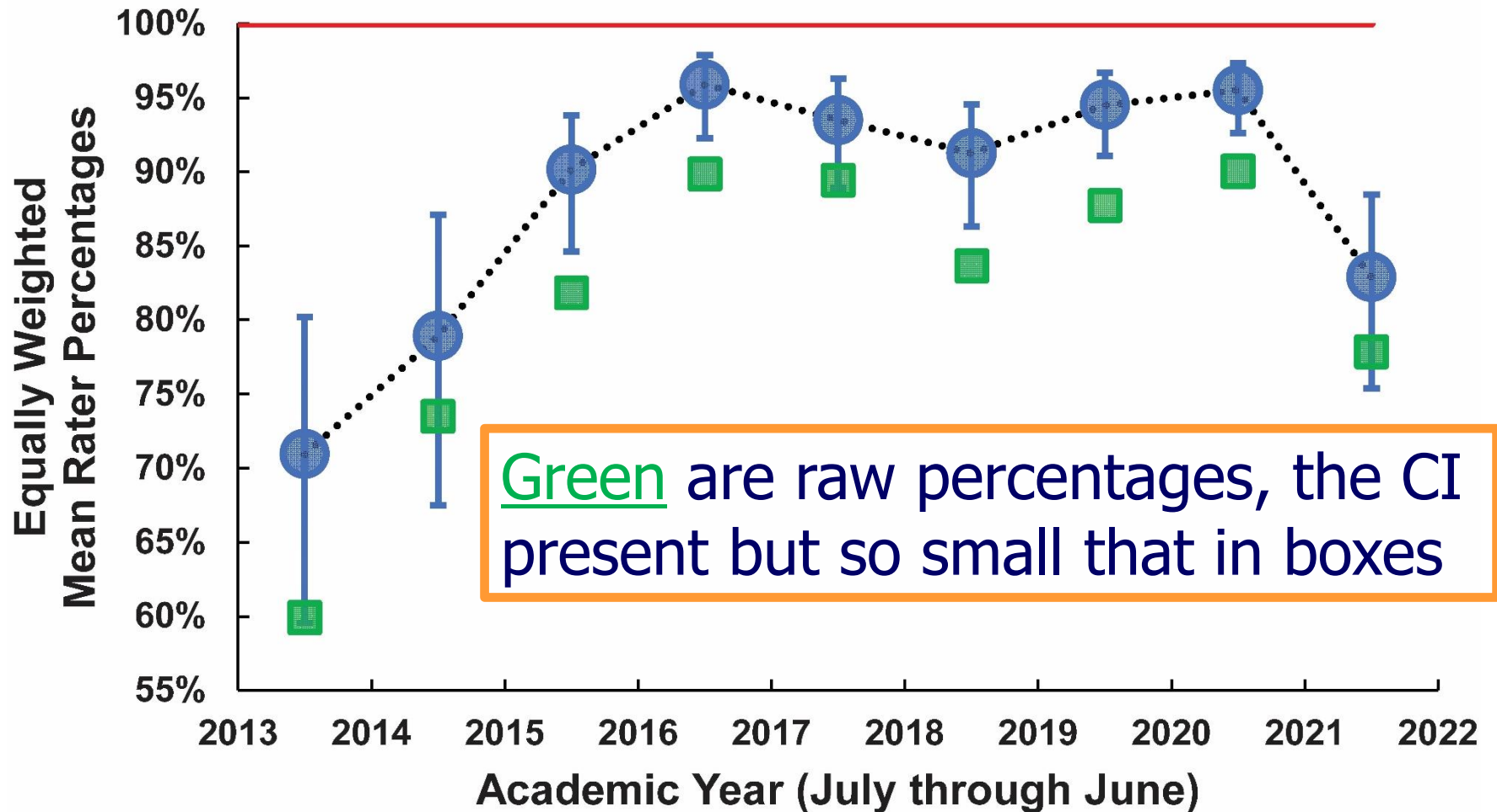
Dexter F et al. J Clin Anesth 2023

Combine Individual Evaluations for Overall Department Quality

- Mixed effects logistic regression model for each academic year treating raters as random effect, because raters differ in leniency
  - Intercept only model
  - From estimated intercept and its 95% confidence interval, take inverse logit to obtain overall departmental quality as proportion
  - ➤ Same as meta-analysis of proportions using generalized linear mixed model
    - Each study (i.e., rater) gives different estimate of the suboptimal quality of supervision

Combine Individual Evaluations for Overall Department Quality

Blue are the estimated proportions and their confidence intervals

Combine Individual Evaluations for Overall Department Quality

Green are raw percentages, the CI present but so small that in boxes

Combine Individual Evaluations for Overall Department Quality

- Mixed effects logistic regression model for each academic year treating raters as random effect, because raters differ in leniency
  - Intercept only model
  - From estimated intercept and its 95% confidence interval, take inverse logit to obtain overall departmental quality as proportion
  - ➢ Departments skipping random effects modeling will give incorrect impression that overall performance is worse than that being provided by average ratee

University of Iowa Department of Anesthesia

- Nearly every paper in this lecture was work with contribution of the University of Iowa

- Department's faculty, residents, and nurse anesthetists have contributed to scholarship in understanding the evaluation of anesthesia faculty in clinical practice

Evidence-Based Anesthesia Group Management

- www.FranklinDexter.net/education.htm
  - Example reports with calculations
  - Lectures on preoperative evaluation clinics, day of surgery decision making, PACU staffing, OR allocation and staffing, anesthesia staffing, financial analysis, comparing surgical services among hospitals, and strategic decision making
- www.FranklinDexter.net
  - Comprehensive bibliography of peer reviewed articles in operating room and anesthesia group management