Validate study variables to reduce misclassification bias: recent tools and research needs

HYBRID WORKSHOP

27 March 2025 - 14.30-18.30

ARS Toscana - Sala Rita Dioguardi, Villa La Quiete
via Pietro Dazzi 1 - Florence, Italy

Recommendations and Practices in Validation Studies and Their Uses

Vincent Lo Re, MD, MSCE, FISPE

Center for Real-World Effectiveness and Safety of Therapeutics

Center for Clinical Epidemiology and Biostatistics

University of Pennsylvania

March 27, 2025

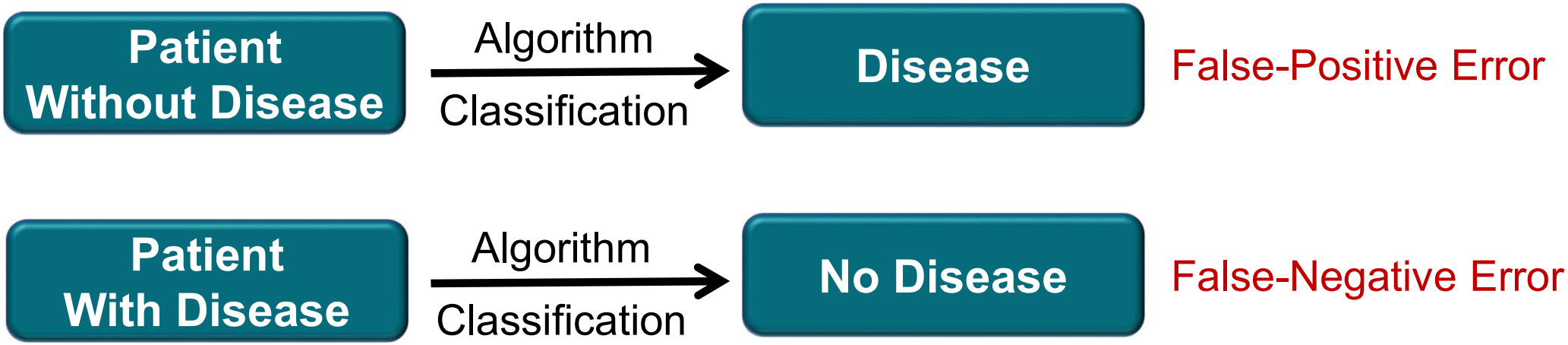
Objectives

- Rationale for validation of algorithms for health conditions in real-world data (RWD)
- Challenges in conducting validation studies
- Conventional approach to validating case-finding algorithms
- Use of validation indices to inform quantitative bias analysis
- Use of AI to identify health conditions in RWD

Case-Finding Algorithms Identify Health Conditions in RWD, But May Be Inaccurate

Study population, exposure, outcome, covariates

- **Algorithm** = definition of health condition in RWD using codes/other data
  - Analogous to diagnostic test → classify presence/absence of condition
  - Algorithms may be inaccurate → may not perfectly classify condition



Ehrenstein V. *Clin Epidemiol* 2016;8:49-51.

Validation Studies are Important to Understanding Degree of Misclassification with Algorithms

Source expected to classify individuals for condition better than algorithm

Validation Study

- Assesses accuracy of algorithm vs. reference standard
- Yields indices → sensitivity, specificity, PPV, NPV
- Inform degree of algorithm error, misclassification

Validation Approaches

- Determine condition in all persons
- Determine condition in all cases identified by algorithm
- Determine performance of algorithm in a sample

Fox MP. *Int J Epidemiol* 2020;49:1392-6.

Questions/Challenges With Validation Studies

1. Is cost, time for validation worth it? Which variables to validate?
2. Can I just use an algorithm validated in another data source?
3. How do I determine the definition of a condition used in practice?
4. What data should be used for the algorithm for the condition?
5. What indices to target? What threshold is fit-for-purpose?
6. What sampling approach should be used? How big a sample?
7. How do I use algorithm indices to inform my main study?

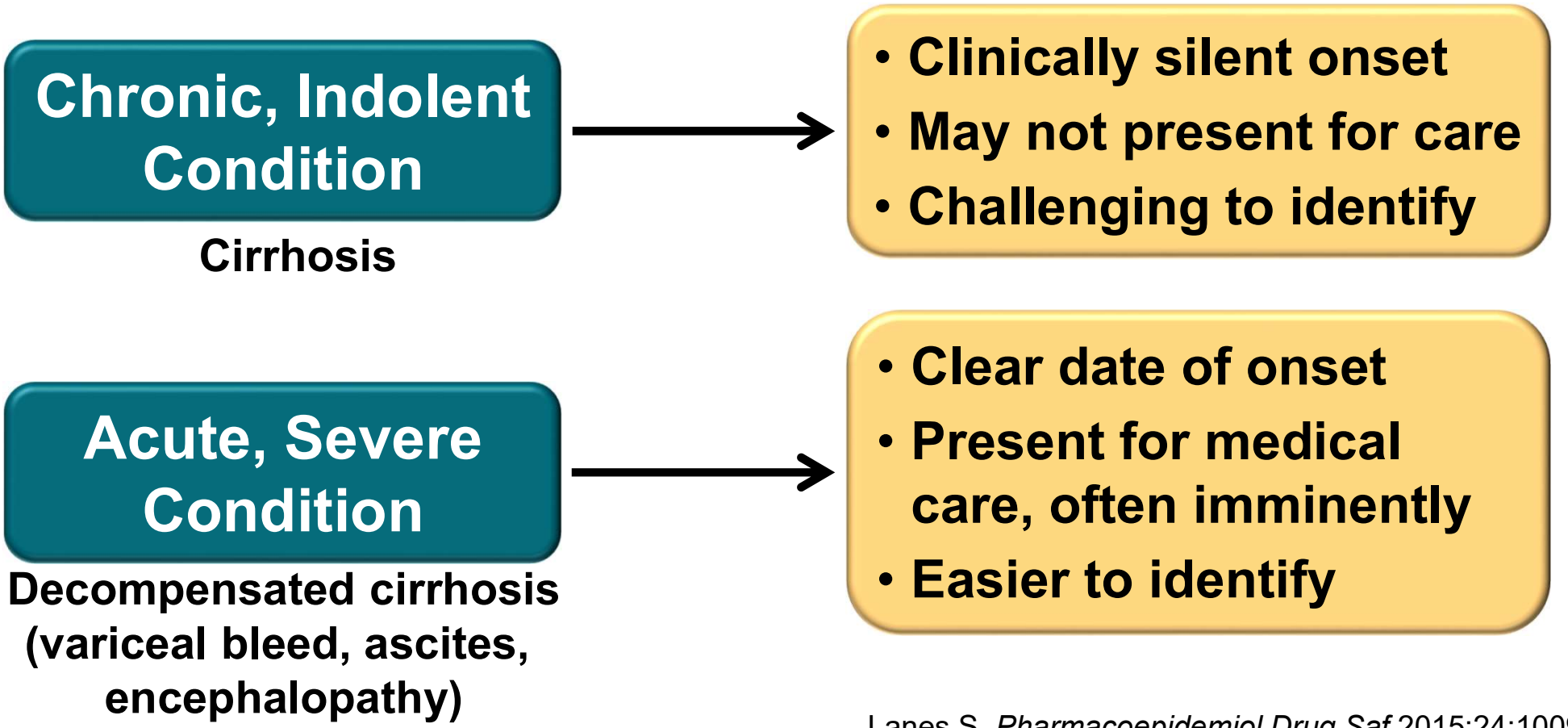
Girman CJ. *Pharmacoepidemiol Drug Saf* 2022;31:717-720.

Conventional Approach to Algorithm Validation

1. Select health condition of interest

Weinstein EJ. *Pharmacoepidemiol Drug Saf* 2023;32:1-8.

Select Health Condition for Study that Minimizes Potential for Misclassification



Lanes S. *Pharmacoepidemiol Drug Saf* 2015;24:1009-16.

Conventional Approach to Algorithm Validation

1. Select health condition of interest

2. Formulate case definition of health condition (used in practice)

- Review clinical literature; consult clinicians, patients, data experts

Weinstein EJ. *Pharmacoepidemiol Drug Saf* 2023;32:1-8.

Example Case Definition Used in Validation: Hepatic Decompensation

Diagnosis	Definition
Ascites	a) Identified on abdominal imaging **or** b) Paracentesis performed
Spontaneous Bacterial Peritonitis	a) Ascites neutrophil count ≥250 cells/mL **or** b) Bacterial growth from fluid culture
Variceal Bleed	a) Active bleeding on endoscopy report **or** b) Variceal bleed in progress note
Encephalopathy	Diagnosis recorded in gastroenterology note

Classified with decompensation if met definition for any diagnosis

Lo Re V. *Pharmacoepidemiol Drug Saf* 2011;20:689-99.

Conventional Approach to Algorithm Validation

1. Select health condition of interest

2. Formulate case definition of health condition (used in practice)

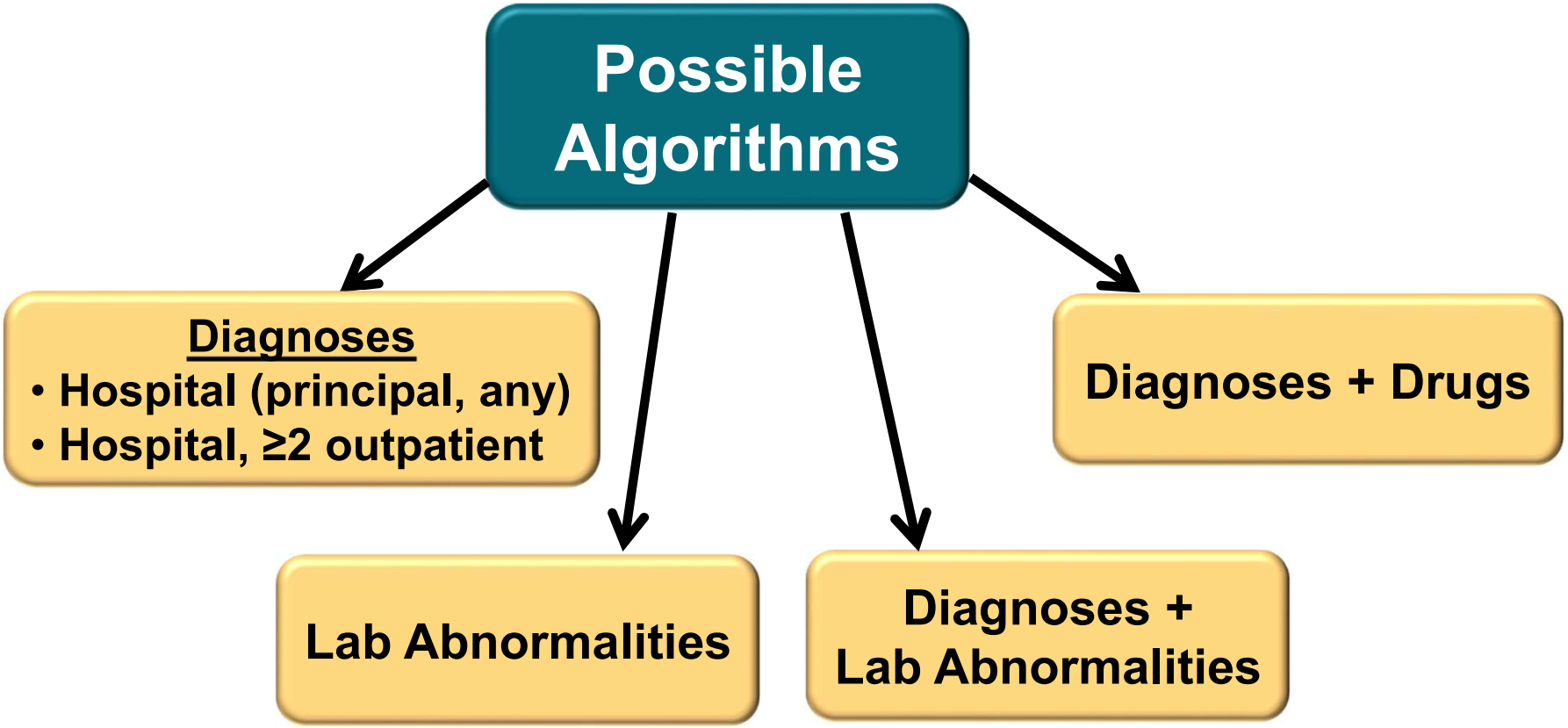
- Review clinical literature; consult clinicians, patients, data experts

3. Devise case-finding algorithm(s) to ascertain health condition

- Often use structured data: diagnoses, procedures, drugs, labs
- May use unstructured data: MD notes, radiology reports, pathology

Weinstein EJ. *Pharmacoepidemiol Drug Saf* 2023;32:1-8.

Options for Case-Finding Algorithms to Identify Health Conditions in RWD



Many options to consider…

Conventional Approach to Algorithm Validation

1. Select health condition of interest

2. Formulate case definition of health condition (used in practice)

- Review clinical literature; consult clinicians, patients, data experts

3. Devise case-finding algorithm(s) to ascertain health condition

- Often use structured data: diagnoses, procedures, drugs, labs
- May use unstructured data (e.g., notes, radiology reports, pathology)

4. Determine the reference standard

- Reference: Chart, disease registry, patient/clinician questionnaire

Weinstein EJ. *Pharmacoepidemiol Drug Saf* 2023;32:1-8.

Conventional Approach to Algorithm Validation

5. Prioritize validity measures, select sample, collect data to classify

Ideal → Maximize all validity measures

Reality → Must accept trade-offs

Weinstein EJ. *Pharmacoepidemiol Drug Saf* 2023;32:1-8.

Prioritizing Algorithm Validity Measures

- Consider goal of algorithm, main accuracy concern

Goal of Algorithm	Main Concern	Validity Measures of Interest
Risk ratios of condition for different drugs	False-positives	↑ PPV, specificity
Absolute risk, risk difference of condition with drugs	False-negatives	↑ sensitivity
Identify patients with condition for eligibility	False-positives	↑ PPV, specificity

Ehrenstein V. *Am J Epidemiol* 2024;193:1612-24.

Conventional Approach to Selection / Abstraction

Selection of Patients

- Seek PPV: sample who met algorithm
- Seek NPV: sample who did not meet algorithm
- Seek sensitivity: sample true cases
- Seek specificity: sample true non-cases

Usually Irrespective of Exposure Group

Sample Size

Based on:

- Desired width of 95% CI (+/-10%)
- Expected estimate of parameter (e.g., ≥80%)

Chart Abstraction

- Select components, data, time window of review
- Decouple abstraction/adjudication → blind to drug
- Train abstractor, pilot abstraction

Conventional Approach to Algorithm Validation

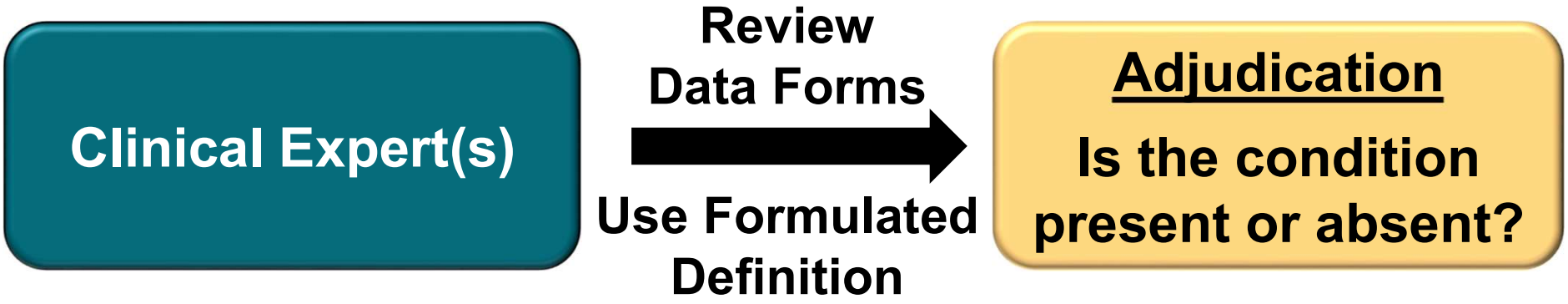
5. Prioritize validity measures, select sample, collect data to classify

6. Adjudicate the health condition

- Agreement with disease registry, questionnaire
- Expert review of charts or structured forms to confirm condition

Weinstein EJ. *Pharmacoepidemiol Drug Saf* 2023;32:1-8.

Adjudication of Health Condition Based on Medical Record Data

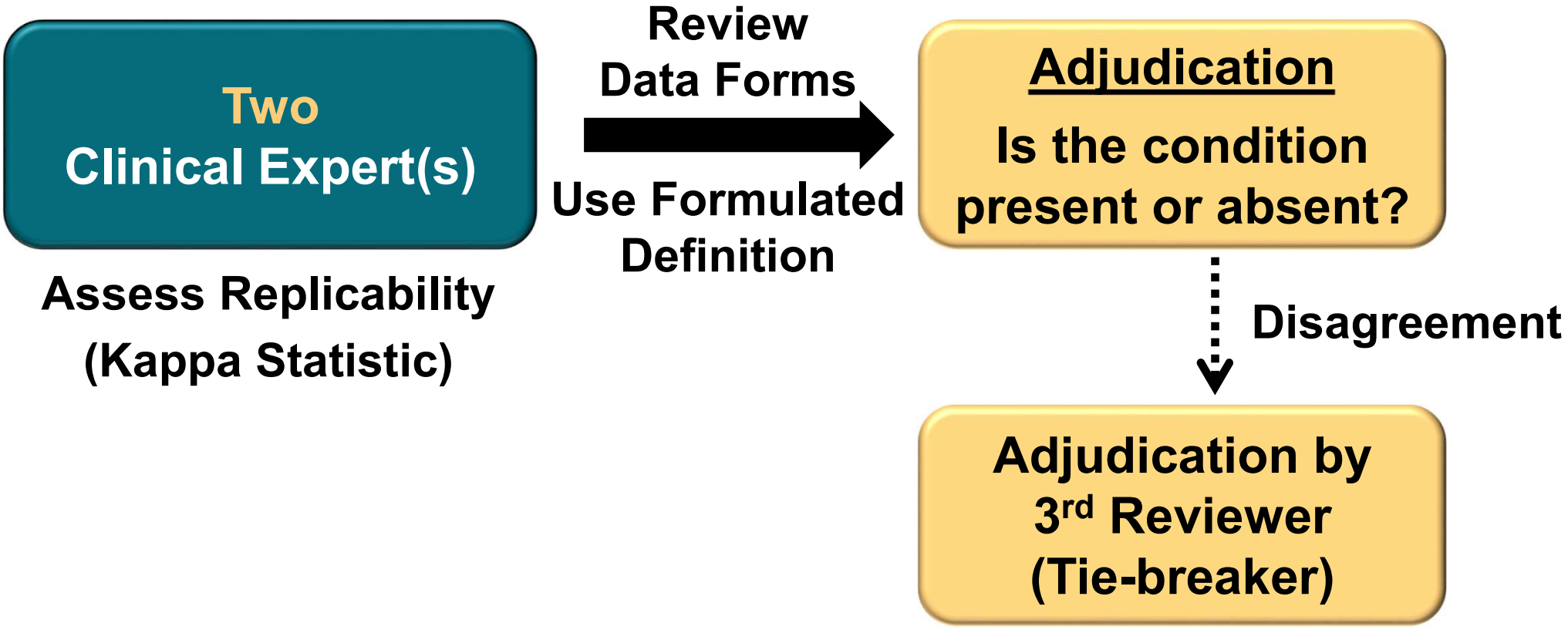


Standardized, reproducible process: ↓ intra- and inter-rater error

Assess adjudication procedures after 5-10 chart reviews

Nissen F. *Breathe* 2019;15:64-8.

Adjudication of Health Condition Based on Medical Record Data



Nissen F. *Breathe* 2019;15:64-8.

Conventional Approach to Algorithm Validation

5. Prioritize validity measures, select sample, collect data to classify

6. Adjudicate the health condition

- Agreement with registry, questionnaire
- Expert review of charts or structured forms to confirm condition

7. Determine the algorithm accuracy (validity measures)

- PPV ≥80% → fit-for-purpose
- PPV <80% → refine/re-validate algorithm; adjudicate all cases
- Threshold of algorithm accuracy considered fit-for-purpose is unclear

Weinstein EJ. *Pharmacoepidemiol Drug Saf* 2023;32:1-8.

Considerations With Case-Finding Algorithms

- Sensitivity analyses with different algorithm definitions
- Algorithms are usually not transportable
  - Different populations, disease prevalence, coding, practices
  - Accuracy differs by database → re-validate in new database
- Conventional validation → single validity measure (often PPV)
  - Masks differences in algorithm accuracy by exposure group
  - If PPV or sensitivity differ in groups, misclassification can be large

[1]Lanes S, Beachler DC. *Pharmacoepidemiol Drug Saf* 2023;32:700-3.

Algorithm Validity Indices Can Be Used to Show If/How Misclassification Affects Study Findings

- **Methods to correct effect measures for imperfect algorithm accuracy**
  - Using **only PPV**[1] → assumes nondifferential sensitivity
  - Using **range of PPV, sensitivity** in exposure groups → **App**[2]

- **Quantitative bias analysis**
  - **Magnitude of bias** of association due to algorithm inaccuracy
  - Need range of algorithm **sensitivity, specificity in exposure groups**
    - Algorithm validation in each group → ↑ sensitivity "screening" algorithm[3,4]
    - If have 1 validity measure + disease prevalence, or 2 validity measures → **derive other validity indices**[5]

[1]Brenner H, Gefeller O. *Am J Epidemiol* 1993;138:1007-15.
[2]Hall GC. *Pharmacoepidemiol Drug Saf* 2020;29:1450-5.
[3]Lanes S, Beachler DC. *Pharmacoepidemiol Drug Saf* 2023;32:700-3.
[4]Limoncella G. *Am J Epidemiol* 2024. *In Press*.
[5]Bollaerts K. *PLOS One* 2020;15:e0231333.

Use of AI to Identify Health Conditions in RWD

- Machine learning, natural language processing
  - Incorporate unstructured, qualitative data within EMR
  - Radiology, pathology, operative reports → health condition
  - Could make validation less costly, reduce misclassification

Received: 10 November 2022 | Revised: 6 May 2023 | Accepted: 31 May 2023
DOI: 10.1002/pds.5648

ORIGINAL ARTICLE

WILEY

Performance of an automated deep learning algorithm to identify hepatic steatosis within noncontrast computed tomography scans among people with and without HIV

Jessie Torgersen[1,2,3] | Scott Akers[3] | Yuankai Huo[4] | James G. Terry[5] | J. Jeffrey Carr[5] | Alexander T. Ruutiainen[3] | Melissa Skanderson[6,7] | Woody Levin[6,7] | Joseph K. Lim[6,7] | Tamar H. Taddei[6,7] | Kaku So-Armah[8] | Debika Bhattacharya[9] | Christopher T. Rentsch[6,7,10] | Li Shen[2] | Rotonya Carr[11] | Russell T. Shinohara[2,12,13] | Michele McClain[14] | Matthew Freiberg[15] | Amy C. Justice[6,7,16] | Vincent Lo Re III[1,2]

Received: 3 January 2024 | Accepted: 19 April 2024
DOI: 10.1097/HC9.0000000000000468

AASLD

ORIGINAL ARTICLE

OPEN

Identification of hepatic steatosis among persons with and without HIV using natural language processing

Jessie Torgersen[1,2] | Melissa Skanderson[3,4] | Farah Kidwai-Khan[3,4] | Dena M. Carbonari[2] | Janet P. Tate[3,4] | Lesley S. Park[5] | Debika Bhattacharya[6] | Joseph K. Lim[3,4] | Tamar H. Taddei[3,4] | Amy C. Justice[3,4,7] | Vincent Lo Re III[1,2]

Torgersen J. *Pharmacoepidemiol Drug Saf* 2023;32:1121-30.
Torgersen J. *Hepatol Commun* 2024;8(7):e0468.

Summary

- Algorithms for health conditions may be inaccurate
- Validation study assesses potential for algorithm misclassification
- Important steps in conventional algorithm validation
- Can use validation indices to inform quantitative bias analysis
- AI could facilitate identification of health conditions in RWD

Considerations for Xabi and Question for Ersilia/Robert

Considerations for Xabi during protocol development:

- Assemble experts in methods, statistics, data, clinicians to inform protocol
- Include validation of key variables (e.g., eligibility, outcome), if not done
- Use validity indices of outcome to inform bias analysis
- Consider AI to aid abstraction, ↓ measurement error during validation

Methodologic questions for Ersilia and Robert:

- What threshold of algorithm accuracy to consider as fit-for-purpose?
- Should validation of health outcomes routinely be performed by exposure group to assess for differential misclassification?
- What is preferred approach to address outcome algorithm inaccuracy → statistically correct effect measures vs. quantitative bias analysis vs. both?

Thank You!

Philadelphia, PA, USA.