DEFINING THE REFERENCE STANDARD FOR BIOMARKERS IN NASH: HISTOLOGY VERSUS CLINICAL OUTCOMES

Kathleen M Donohue MD, MSc
Clinical Team Lead
Liver & Inborn Errors
Division of Gastroenterology & Inborn Errors Products
FDA

www.fda.gov
Disclaimers

• Views expressed in this presentation are those of the speaker and do not necessarily represent an FDA official position

• I do not have any financial disclosures regarding pharmaceutical drug products
NASH reference standard?

**Accelerated Approval: Histology**
- Resolution of steatohepatitis AND no worsening of liver fibrosis
- Improvement in liver fibrosis AND no worsening of steatohepatitis
- Both resolution of steatohepatitis and improvement in fibrosis

**Full Clinical Approval: Clinical Outcomes**
- Progression to cirrhosis on histopathology
- Hepatic decompensation events
- ↑ MELD from ≤12 to > 15
- Transplant
- All-cause mortality
Histology Standard

Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease, IOM, 2010
Histology Standard

Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease, IOM, 2010
Histology Standard

Adapted from Fleming et al. Annals of Internal Medicine 1996
CGD Study Group NEJM 1991
Histology Standard

NASH

NEW BIOMARKER

Fibrosis

Inflammation

DRUG

CIRRHOSIS DECOMPENSATION TRANSPLANT MORTALITY

Novel Pathway
Histology Standard

- **NASH**
  - NEW BIOMARKER
  - Inflammation
- **DRUG**
  - Novel Pathway
  - Fibrosis
- **HISTOLOGY**
- **CIRRHOSIS DECOMPENSATION TRANSPLANT MORTALITY**
Benchmarking to histology

CIRRHOSIS

DECOMPENSATION

TRANSPLANT

MORTALITY

Inflammation

Fibrosis

Off Target Effects

DRUG

NEW

BIOMARKER

NASH

HISTOLOGY

CIRRHOSIS

DECOMPENSATION

TRANSPLANT

MORTALITY
NASH reference standard?

Accelerated Approval: Histology
- Resolution of steatohepatitis AND no worsening of liver fibrosis
- Improvement in liver fibrosis AND no worsening of steatohepatitis
- Both resolution of steatohepatitis and improvement in fibrosis

Full Clinical Approval: Clinical Outcomes
- Progression to cirrhosis on histopathology
- Hepatic decompensation events
- ↑ MELD from ≤12 to > 15
- Transplant
- All-cause mortality
Clinical outcome standard: Composite endpoint?
Composite Endpoints: Strengths

- ↑ statistical efficiency
- ↓ sample size, cost, time
- Avoids arbitrary choice among several important outcomes for same disease
Composite Endpoints: Requirements

• components
  – clinically meaningful
  – of similar importance to patients
• expected effects on each component are similar
• the clinically most important components should at least not be affected negatively

Freemantle et al. JAMA 2003
Montori et. al BMJ 2005

13
Composite Endpoints: Limitations

• Treatment effect often smallest on the most important outcome
• Inconsistent treatment effect on components – e.g. ↓ chest pain, ↑ deaths
• Composite with nonfatal events (i.e. excluding death) is invalid – censors worst outcome
Toward better biomarkers
FDA NASH Pipeline

2016

- Pre-submission Meetings: 31%
- INDs: 28%
- Expedited Program Requests: 17%
- initial Pediatric Study Plans: 6%
- Inter-Center Consultations: 8%
- Biomarker Qualification Program: 3%

2017

- Pre-submission Meetings: 32%
- INDs: 13%
- Expedited Program Requests: 36%
- initial Pediatric Study Plans: 10%
- Inter-Center Consultations: 8%
- Biomarker Qualification Program: 17%
- EMA Collaborations: 6%
Reporting Biomarker Studies

Challenges and Standards in Reporting Diagnostic and Prognostic Biomarker Studies

Francisco Azuaje, Ph.D.¹, Yvan Devaux, Ph.D.¹, and Daniel Wagner, Ph.D., M.D.¹,²

BMJ Open
STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration

Jérémie F Cohen,¹,² Daniël A Korevaar,¹ Douglas G Altman,³ David E Bruns,⁴ Constantine A Gatsonis,⁵ Lotty Hooft,⁶ Les Inwig,⁷ Deborah Levine,⁸,⁹ Johannes B Reitsma,¹⁰ Henrica C W de Vet,¹¹ Patrick M M Bossuyt¹

Annals of Internal Medicine
QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies

Penny F. Whiting, PhD; Anne W. S. Rutjes, PhD; Marie E. Westwood, PhD; Susan Mallett, PhD; Jonathan J. Deeks, PhD; Johannes B. Reitsma, MD, PhD; Mariska M. G. Leeflang, PhD; Jonathan A. C. Sterne, PhD; Patrick M. M. Bossuyt, PhD; and the QUADAS-2 Group*
<table>
<thead>
<tr>
<th>Domain</th>
<th>Patient Selection</th>
<th>Index Test</th>
<th>Reference Standard</th>
<th>Flow and Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Describe methods of patient selection</td>
<td>Describe the index test and how it was conducted and interpreted</td>
<td>Describe the reference standard and how it was conducted and interpreted</td>
<td>Describe any patients who did not receive the index tests or reference standard or who were excluded from the 2 × 2 table (refer to flow diagram)</td>
</tr>
<tr>
<td></td>
<td>Describe included patients (previous testing, presentation, intended use of index test, and setting)</td>
<td></td>
<td></td>
<td>Describe the interval and any interventions between index tests and the reference standard</td>
</tr>
<tr>
<td>Signaling questions (yes, no, or unclear)</td>
<td>Was a consecutive or random sample of patients enrolled?</td>
<td>Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td>Is the reference standard likely to correctly classify the target condition?</td>
<td>Was there an appropriate interval between index tests and reference standard?</td>
</tr>
<tr>
<td></td>
<td>Was a case–control design avoided?</td>
<td>If a threshold was used, was it prespecified?</td>
<td>Were the reference standard results interpreted without knowledge of the results of the index test?</td>
<td>Did all patients receive a reference standard?</td>
</tr>
<tr>
<td></td>
<td>Did the study avoid inappropriate exclusions?</td>
<td></td>
<td></td>
<td>Did all patients receive the same reference standard?</td>
</tr>
<tr>
<td>Risk of bias (high, low, or unclear)</td>
<td>Could the selection of patients have introduced bias?</td>
<td>Could the conduct or interpretation of the index test have introduced bias?</td>
<td>Could the reference standard, its conduct, or its interpretation have introduced bias?</td>
<td>Were all patients included in the analysis?</td>
</tr>
<tr>
<td>Concerns about applicability (high, low, or unclear)</td>
<td>Are there concerns that the included patients do not match the review question?</td>
<td>Are there concerns that the index test, its conduct, or its interpretation differ from the review question?</td>
<td>Are there concerns that the target condition as defined by the reference standard does not match the review question?</td>
<td></td>
</tr>
</tbody>
</table>

*Table 1. Risk of Bias and Applicability Judgments in QUADAS-2*
Figure 2. Sample of a study flow diagram.

The diagram is based on a diagnostic cohort study on using B-type natriuretic peptide levels to diagnose heart failure. Based on data obtained from Smith H, Pickering RM, Struthers A, Simpson I, Mant D. Biochemical diagnosis of ventricular dysfunction in elderly patients in general practice: observational study. BMJ. 2000;320:906-8.
Acknowledgements

Slides from Erica Lyons & Stephanie Omokaro
Lara Dimick-Santos
Ruby Mehta
Veronica Pei
Suna Seo
Frank Anania
Roni Cherry-France

Evangela Covert
Anissa Davis-Williams
Lisa Soule
Dragos Roman
Victor Crentsil
Julie Beitz
Mark Avigan
John R. Senior