



THE FORUM
For Collaborative ResearchSM

LESSONS LEARNED

Zachary Goodman, MD

Director of Hepatic Pathology Consultation and
Research, Inova Fairfax Hospital

FDA Guidance

Webinar – January 29, 2021



Histological Endpoints Based on Liver Biopsies

Used For	Challenges	Examples of Potential Solutions
<ul style="list-style-type: none">• Eligibility criteria• Stratification• Efficacy	<ul style="list-style-type: none">• Sampling issues• Pathologist reading discordance* : inadequate inter- and intra-reader concordance	<ul style="list-style-type: none">• Standard procedure for processing slides• Training of pathologists before/during the trial• ≥2 pathologists: if discordant, 3rd pathologist• Same slide read by all pathologists• Central pathology reading• Histopathology adjudication committee• <i>Sponsor should prespecify the details of liver biopsy interpretation</i>

Multicenter trials?
“Alignment?”

Median scores of 3-5
Pathologist Panels?

Digitized images?

*Source: Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials, Journal of Hepatology, Davison, B.A. et al, available online 28 June 2020.

Kappa Statistic

- **Cohen's kappa coefficient (κ)**

- Measure inter-rater and intra-rater reliability for qualitative (categorical) items
- Adjusts for chance (random) agreement
- **EXAMPLE:** How often do two psychiatrists agree on diagnosis of schizophrenia?

- **Weighted kappa**

- Adjusts for degree of agreement in ordered categories
- **EXAMPLE:** How often do two pathologists agree on Ishak fibrosis stage?

- **Fleiss' kappa**

- For more than two raters
- **EXAMPLE:** How often do nine pathologists agree on ballooning degeneration?

- **Interpretation**

- 0.8 to 1.0 = almost perfect

- 0.6 to 0.8 = substantial

- 0.4 to 0.6 = moderate

- 0.2 to 0.4 = fair

- 0 to 0.2 = slight

Kappa Statistic for Interobserver Variability

	Kappa
● Cardiac auscultation (nonexpert)	0.19
● Endoscopy for varices	0.38
● Mammogram interpretation	0.47
● Grading of breast cancer	0.43 to 0.74
● Staging of liver fibrosis	0.4 to 0.9

- Interpretation
 - 0.8 to 1.0 = almost perfect
 - 0.6 to 0.8 = substantial
 - 0.4 to 0.6 = moderate
 - 0.2 to 0.4 = fair
 - 0 to 0.2 = slight

Interobserver Variability

(Kappa or ICC)

METAVIR HCV

(10 pathologists, 30 cases)

Bedossa et al (1994)

	Kappa
● Fibrosis score	0.78
● Steatosis	0.63
● Lobular inflammation	0.38
● Portal inflammation	0.52
● Ballooning	0.34
● Acidophil bodies	0.22
● Interface hepatitis	0.36
● Bile duct injury	0.04

NASH CRN

(7 pathologists, 32 cases)

Kleiner et al (2005)

	Kappa
● Fibrosis score	0.85
● Steatosis, grade	0.83
● Lobular inflammation	0.60
● Portal inflammation	0.55
● Ballooning	0.66
● Acidophil bodies	0.28
● Megamitochondria	0.28
● Mallory's hyaline	0.64

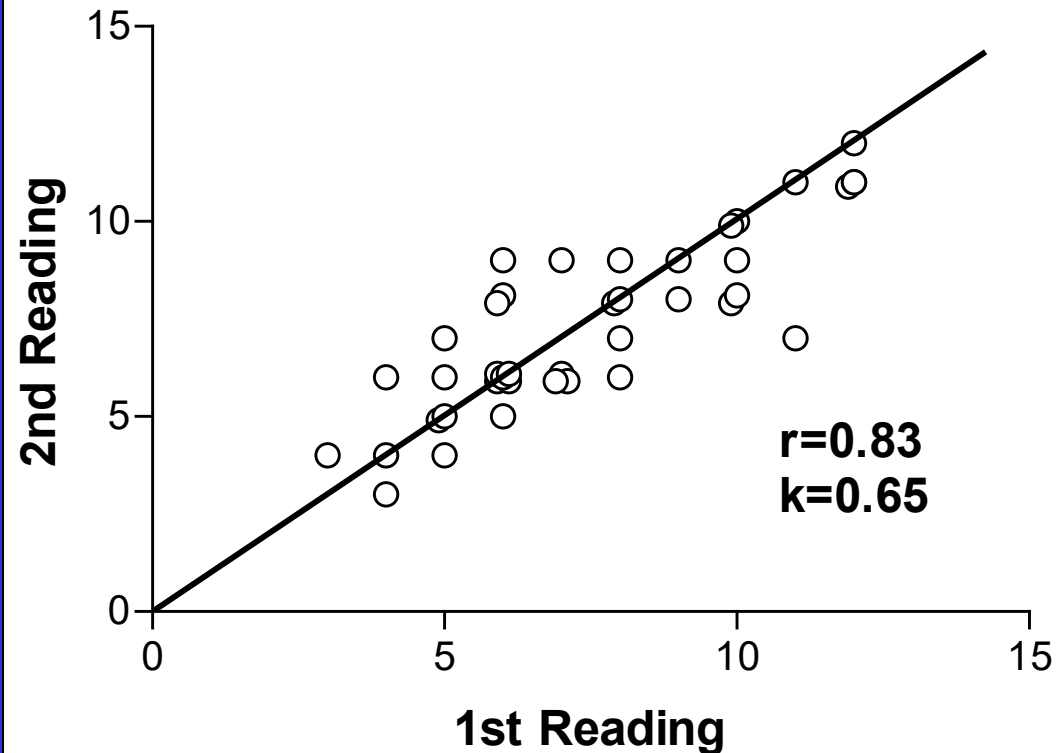
- Interpretation
 - 0.8 to 1.0 = almost perfect
 - 0.6 to 0.8 = substantial
 - 0.4 to 0.6 = moderate
 - 0.2 to 0.4 = fair
 - 0 to 0.2 = slight

Histologic Scoring of Inflammation

Reproducibility

- Metavir (30 Biopsies)
Activity (0-3)
Kappa = 0.25-0.56
- Westin (95 Biopsies)
Ishak score (0-18)
Kappa = 0.11-0.41
- Goodman (40 Biopsies)
Ishak score (0-18)
Kappa = 0.65
same = 38%
+/-1 = 65%

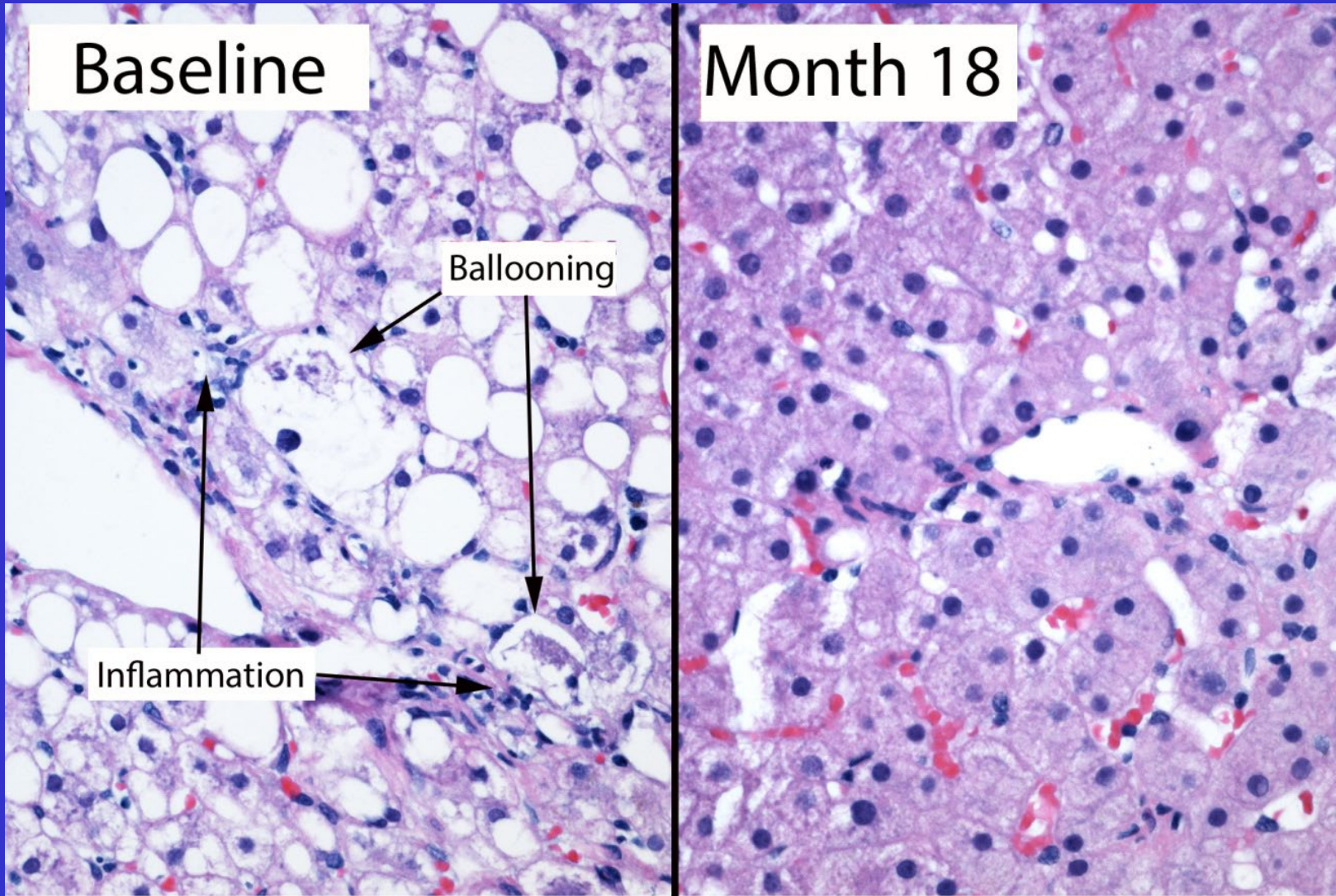
Ishak Inflammatory Grade (0-18)
Reproducibility - 40 HCV Biopsies



“Histologic Improvement” in HBV Trial

	Adefovir (n=172)	Placebo (n=170)
• Mean HAI		
— Knodell	- 2.5	- 0.1
— Ishak	- 2.4	- 0.1
• \geq 2-point \downarrow		
— Knodell	62%	28%
— Ishak	64%	30%
• 50% Improvement	36%	7%
• Metavir 1-point \downarrow	60%	23%
• Almost normal (0-3)	48%	10%

$p < 0.0001$

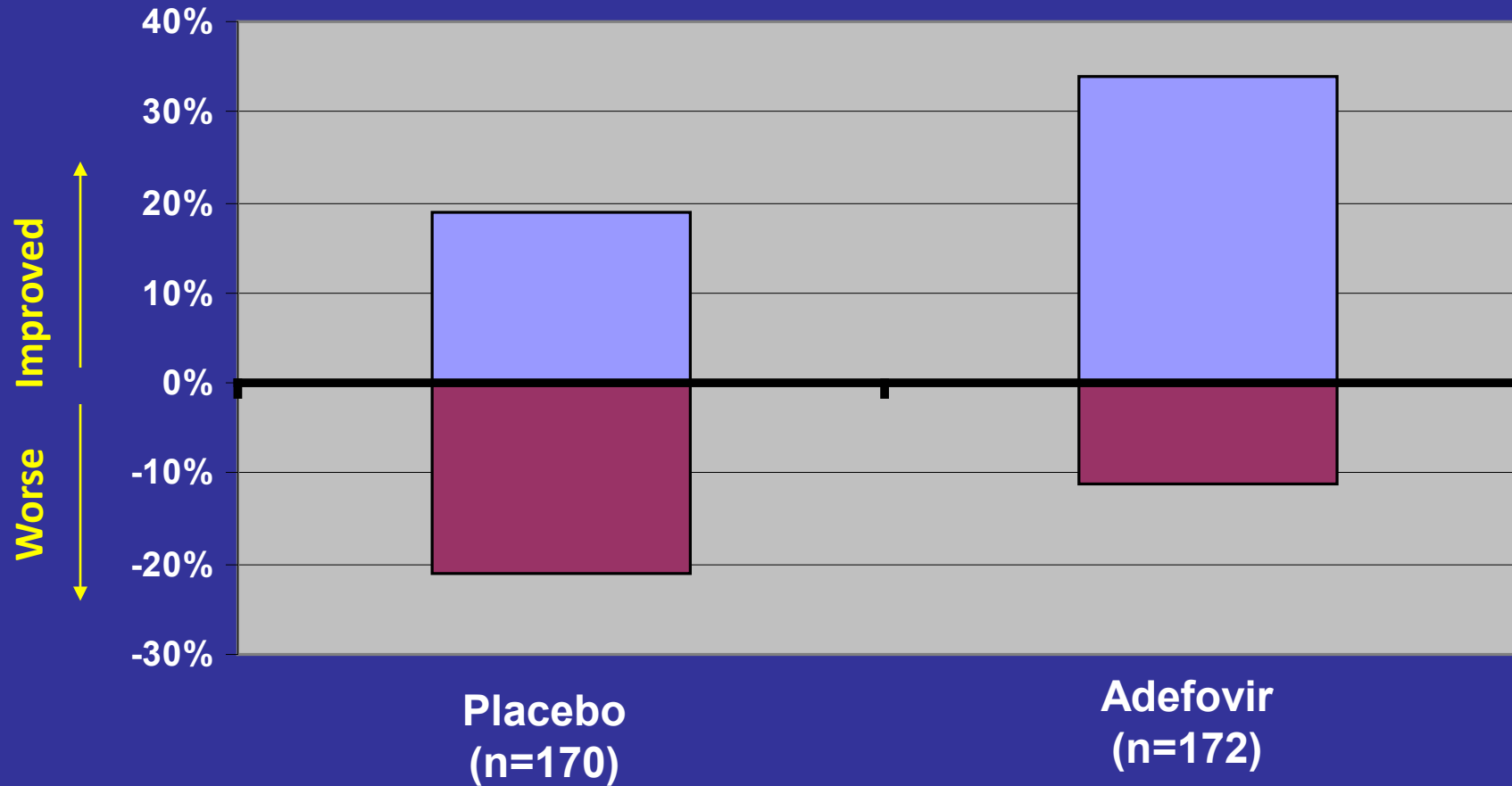


Fibrosis Scoring Reproducibility

- Ishak Fibrosis Score (0-6) in HCV

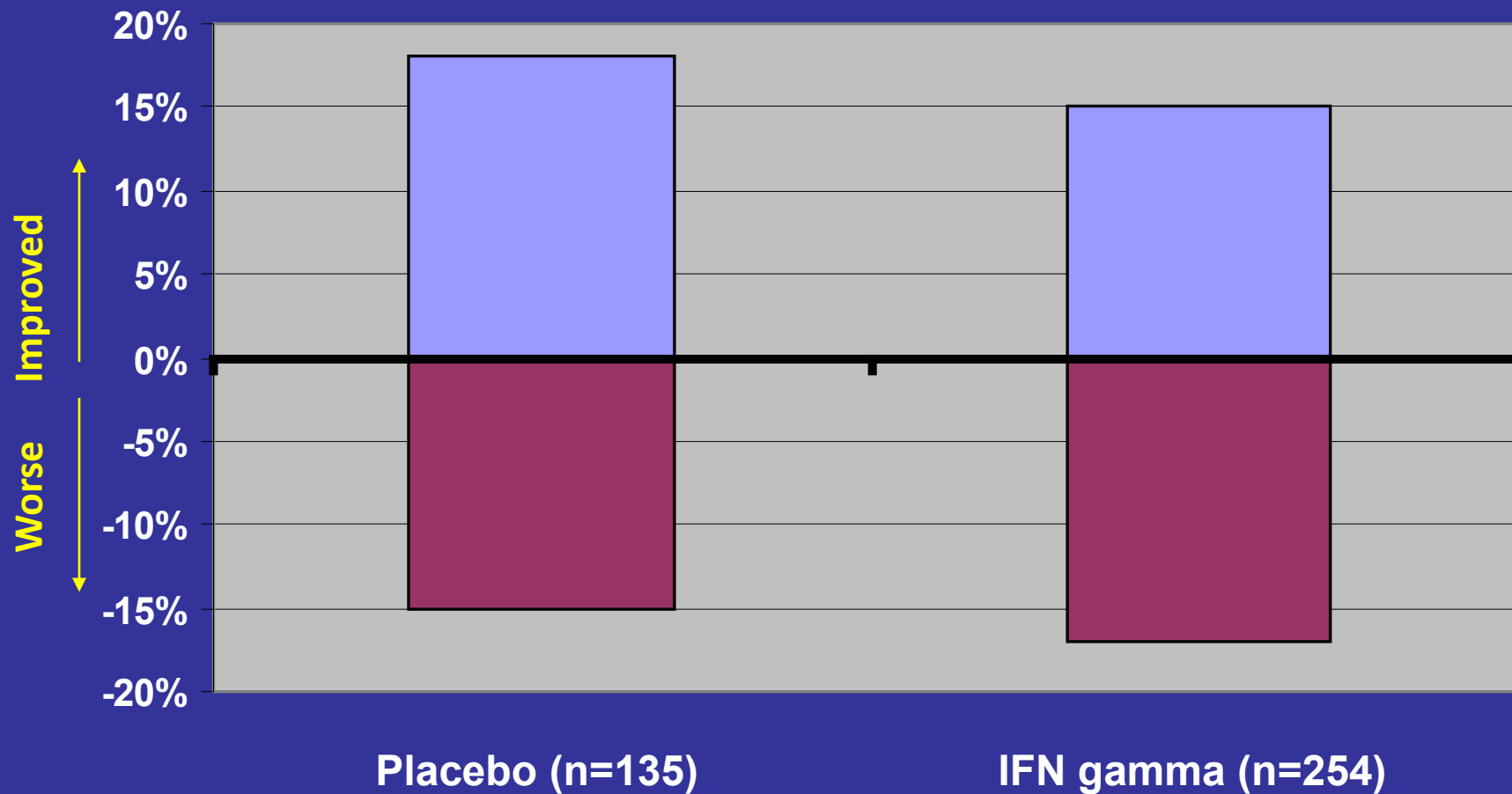
	Same	+/- 1
Westin (95 bxs)	29%	84%
ZG/JC (654 bxs)	48%	86%
ZG/ZG (40 bxs)	72%	100%
HALT C pathologists (200 biopsies – 70 large, 70 small 60 fragmented)	50%	91%

Ishak Fibrosis Scores (0-6) after Therapy for Hepatitis B

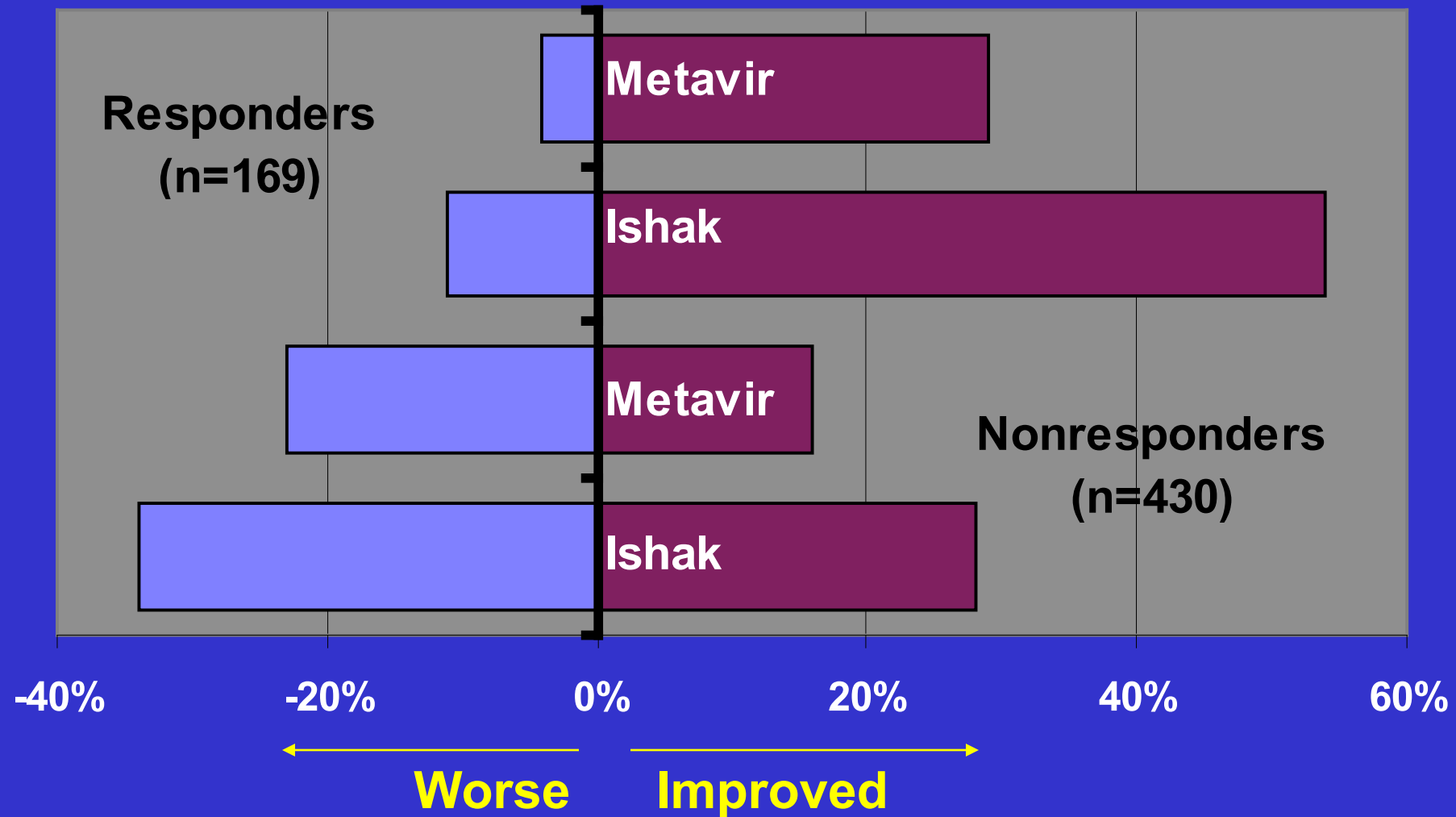


Marcellin et al:NEJM 2003

Ishak Fibrosis Scores (0-6) after 48 Weeks IFN γ for Hepatitis C



Fibrosis Scores after Peg-Interferon Rx for Hepatitis C



Recommendations for Histologic Endpoints

- If the treatment works, any reasonable way of assessing the histology will work
- Cohort effect, not individual patients
- Placebo control group helpful
- Confounding factors
 - Quality of biopsies (not small or fragmented)
 - Severity of disease (not too many with minimal disease)
 - Size of cohort (not too small)
- One pathologist is sufficient
 - If > 1 , then each pathologist is a separate experiment