

Applications of AI in Clinical Trials

Ryan W Stidham, MD, MS

Inflammatory Bowel Disease Program

Department of Medicine, Division of Gastroenterology

Department of Computational Medicine and Bioinformatics

University of Michigan

2022 Innovation Conference: The Future of AI in GI
September 9 - 10, 2022



Financial Disclosures

Industry Support for Investigator-Initiated Studies

Janssen, Abbvie, Bristol Myers Squibb

Consultancy or Advisory Board Membership

Abbvie, CorEvitas, Eli Lilly, Exact Sciences, Gilead, Evergreen Pharmaceuticals, Janssen, Takeda

Intellectual Property

Automated Assessment of Bowel Damage in Intestinal Diseases (US Pat. #10918326)

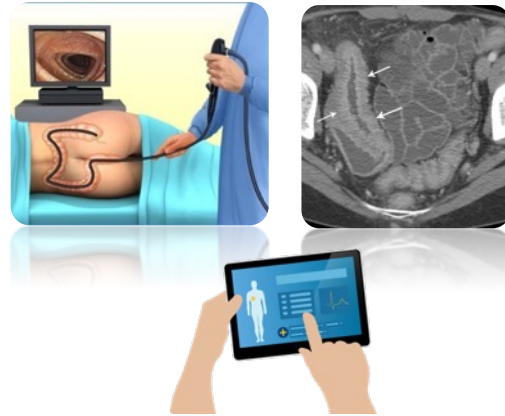
Automated Characterization of Disease Features in Endoscopic Videos (US PatPend # 62-848,724)

Technology Licenses from the University of Michigan to AMI, llc; EIQ, llc; PreNovo, llc

Conventional GI Clinical Trials



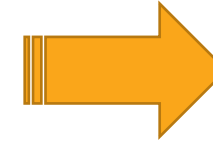
Candidate Identification



Subject Screening



Endpoint Measurement



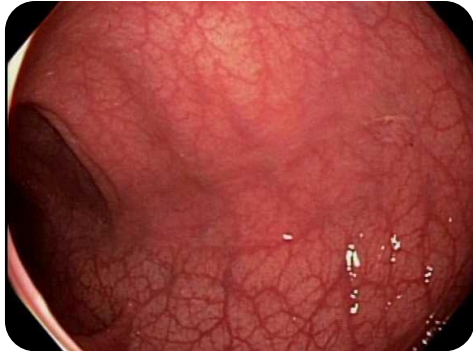
Real World Experience (RWE)

Many Pain Points

AI in Clinical Trials I: Automating Endpoint Scoring

Trials, Regulators, and Clinical Practice Relies on Expert Interpretation of Imaging for Disease Classification, Prognosis and Therapeutic Assessment

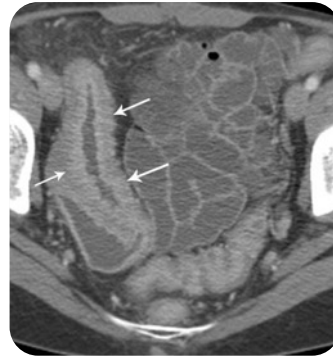
Endoscopy



IBD Endoscopic Scoring

- Mayo
- UCEIS
- SES-CD

Cross Sectional Imaging



Colorectal Cancer

- (CT) Progression Free Survival

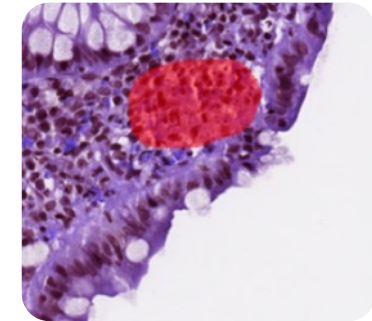
Hepatocellular Cancer

- (MRI & CT) LiRADs

IBD – Crohn’s Disease

- (MRI) MaRIA Score

Histology



IBD – UC and CD

- RHI
- Geboes Score

NAFLD

- Flip SAF
- NIH DAS

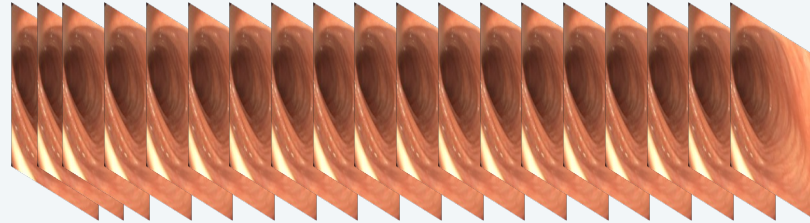
Eosinophilic Esophagitis

- Eos Counts

Neural Networks to Replicate ~~Expert~~ MES Classification in UC

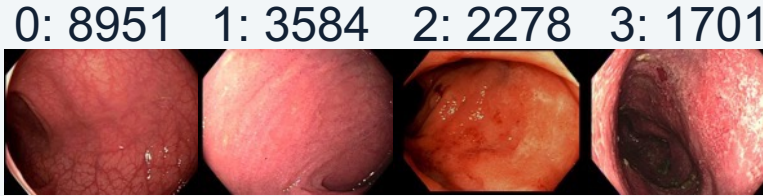
[Insert Disease Classifier of Choice]

Training Set
of Still Images



UC Patients
3,082 Patients
(16,514 Images)

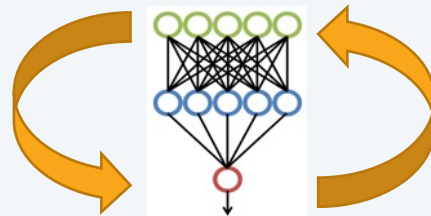
Adjudicated
MES Scoring



Test Set (10%)
304 Patients
1,652 Images



Training Set (90%)
2,778 Patients
14,862 Images



Predictive Model

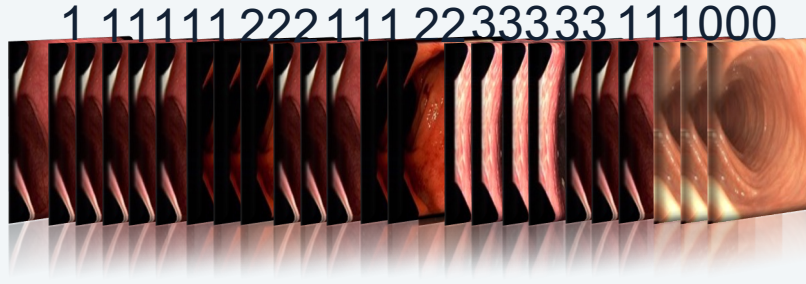
CNN
Model Development



Mayo Score Probability	
0	0.001
1	0.014
2	0.260
3	0.711

Great Success REPLICATING Expert UC Grading

Still Image UC Severity Classification



AI classification of endoscopic activity in UC Ozawa et al

Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis

Tsuyoshi Ozawa, MD, PhD,^{1,2} Soichiro Ishihara, MD, PhD,^{1,3,4} Mitsuhiro Fujishiro, MD, PhD,⁵ Hiroaki Saito, MD,⁶ Youichi Kumagai, MD, PhD,⁷ Satoki Shichijo, MD, PhD,⁸ Kazuharu Aoyama, PhD,⁹ Tomohiro Tada, MD, PhD^{1-4,9}

Ozawa et al. *GIE* 2019

JAMA Network | **Open.**

Original Investigation | Gastroenterology and Hepatology

Performance of a Deep Learning Model vs Human Reviewers in Grading Endoscopic Disease Severity of Patients With Ulcerative Colitis

Ryan W. Stidham, MD, MS, Wenshuo Liu, PhD, Shrinivas Bishu, MD, Michael D. Rice, MD, Peter D. R. Higgins, MD, PhD, Ji Zhu, PhD, MSc, Brahmajee K. Nallamothu, MD, MPH, Akbar K. Waljee, MD, MSc

Stidham et al. *JAMA Net Open* 2019

Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients With Ulcerative Colitis

Kento Takenaka,¹ Kazuo Ohtsuka,¹ Toshimitsu Fujii,¹ Mariko Negi,² Kohei Suzuki,¹ Hiromichi Shimizu,¹ Shiori Oshima,³ Shintaro Akiyama,¹ Maiko Motobayashi,¹ Masakazu Nagahori,¹ Eiko Saito,¹ Katsuyoshi Matsuoka,¹ and Mamoru Watanabe¹

Takenaka et al. *Gastroenterology* 2020

Full Motion Endoscopic Video Classification



NEW METHODS: Clinical Endoscopy

Fully automated endoscopic disease activity assessment in ulcerative colitis

Heming Yao, BS,¹ Kayvan Najarian, PhD,^{1,2,3,4,5} Jonathan Gryak, PhD,^{1,5} Shrinivas Bishu, MD,⁷ Michael D. Rice, MD,⁷ Akbar K. Waljee, MD, MSc,^{6,7,8,9} H. Jeffrey Wilkins, MD,¹⁰ Ryan W. Stidham, MD, MS^{1,6,7}

Ann Arbor, Michigan; Plymouth Meeting, Pennsylvania, USA

Yao, Stidham, Najarian et al. *GIE* 2021

CLINICAL—ALIMENTARY TRACT

Central Reading of Ulcerative Colitis Clinical Trial Videos Using Neural Networks

Klaus Gottlieb,^{1,*} James Requa,^{2,*} William Karnes,² Ranga Chandra Gudivada,¹ Jie Shen,¹ Efren Rael,² Vipin Arora,¹ Tyler Dao,² Andrew Ninh,² and James McGill¹

Gottlieb et al. *Gastroenterology* 2021

Value Proposition for Automated Endoscopic Scoring in Trials



RELIABILITY: Near perfect reproducibility and objectivity

- While incorporating bias and “imperfections” of experts, results are reproducible



EFFICENCY and SPEED:

- Avoid enrollment delays awaiting reviewer scoring
- Reduce time needed for clinical trials



LOWER COST: Less human labor

- Expect reduced costs



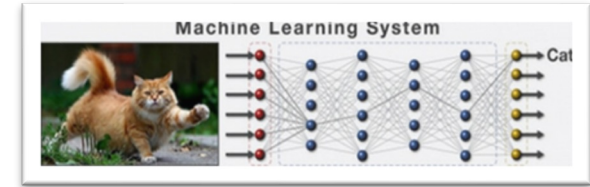
UNIFORMITY: Same scoring instrument between trials

- Big Pharma + Small Pharma + Academia + Community Practice: Same Instrument

Automated Scoring for Trials: Challenges and Solutions

WHAT IS THE AI PERFORMANCE BENCHMARK?

- Matching human agreement is AI target
- Explainability > “Correctness”



Prediction: CAT
Because: FUR, WHISKERS



HOW DO WE TRAIN ?

- More Reviewers \neq More Accuracy
- Training should oversample under-represented
- VERY RIGEROUS EXPERT TRAINING AND EVAL



PROVE TRUE REPRODUCIBILITY

- Need repeated colonoscopy same patient

	Reviewer B Mayo Score, %				
Reviewer A Mayo Score, %	0	1	2	3	Rev B (n)
0	77.1	22.4	0.6	0.0	9160
1	14.6	54.7	30.3	0.4	3430
2	0.2	6.0	69.3	24.5	2405
3	0.0	0.1	14.3	85.7	1519
Rev A (n)	7563	4069	2976	1906	16514

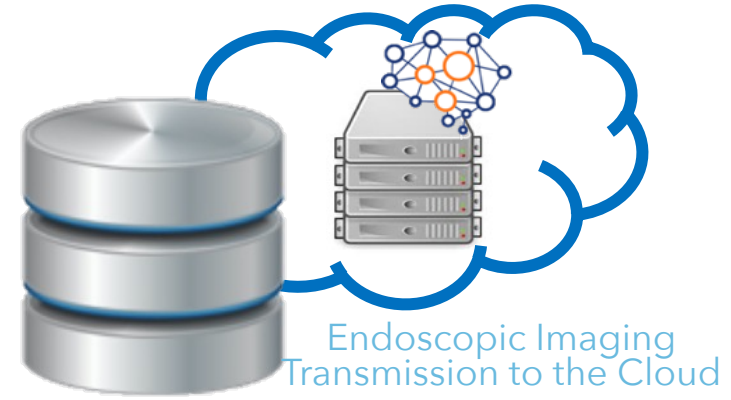
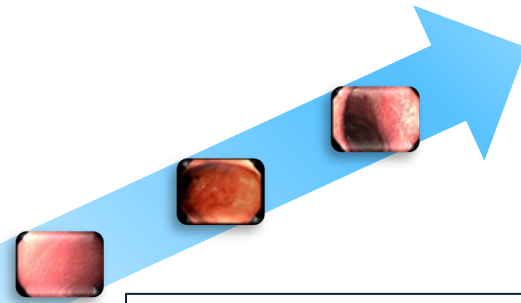
VENDOR INTEROPERABILITY

- Standardization of Minimum Quality Requirements
- Establish acceptable disagreement between AI Tools



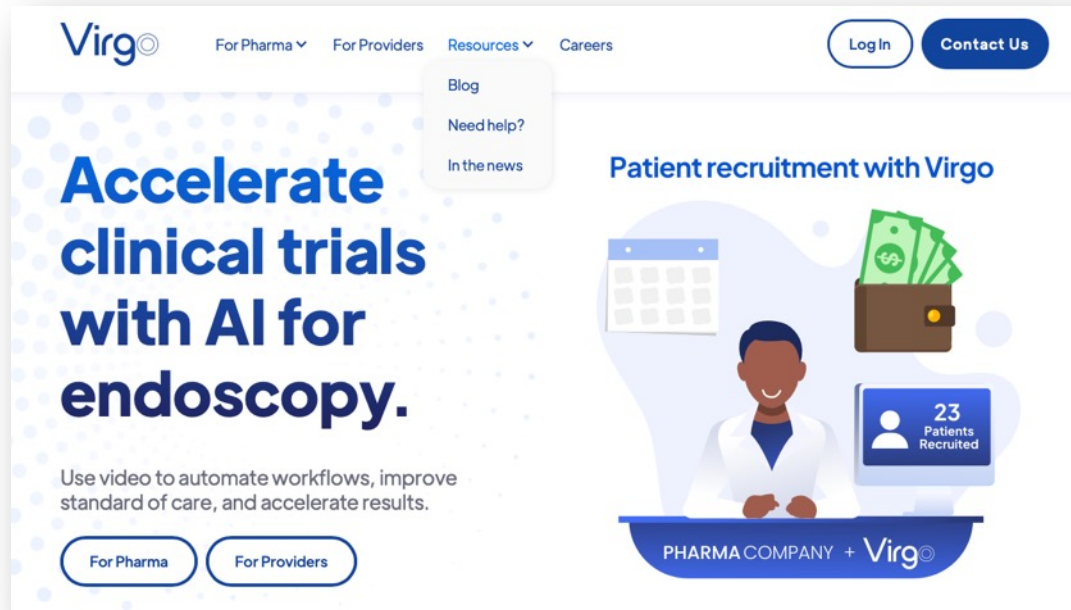
AI in Trials II: Population-Level Automated Assessments

Multi-Site Video Digitization & Review



- Passive background video capture
- Automated endoscopic video analysis
- Detection of findings of interest
- Knowledge of WHO, WHEN, & WHERE disease occurs

Commercial Startups Aiming to Use AI Endoscopic Analysis and Video Collection to Help Accelerate Clinical Trials



Virgo SVS, LLC



Iterative Scopes, LLC

Addressing Pain Point of Identifying and Screening Viable Candidates for IBD Clinical Trials



Technologic AI Endoscopy Cloud Capabilities Within Academia

IBD-ACCESS

Automated Crohn's and Colitis Scoring System

VideoID	Visit	MES_Prediction	conf_est	MES_0	MES_1	MES_2	MES_3	video_qual_score
10206_20170515	Baseline	2	0.78	0.45	0	0.53	0.01	0.44
10297_20180103	Baseline	1	0.44	0.63	0.34	0.03	0	0.65
10995_20171227	Baseline	2	0.95	0.57	0.27	0.16	0	0.43
11135_20170322	Baseline	3	0.86	0.59	0.02	0.02	0.37	0.71
11135_20171024	Baseline	2	0.71	0.77	0.14	0.07	0.02	0.65
11135_20171031	Baseline	3	0.88	0.52	0.15	0.02	0.31	0.81
11135_20161114	Baseline	3	0.61	0.02	0.23	0.33	0.41	0.69
11135_20170529	Baseline	2	0.55	0.33	0.23	0.4	0.03	0.67
11135_20171212	Baseline	1	0.91	0.92	0.08	0.01	0	0.72
11135_20171227	Baseline	1	0.83	0.83	0.15	0.01	0	0.61
11135_20170118	Baseline	2	0.58	0.57	0.28	0.14	0.01	0.75



East Ann Arbor



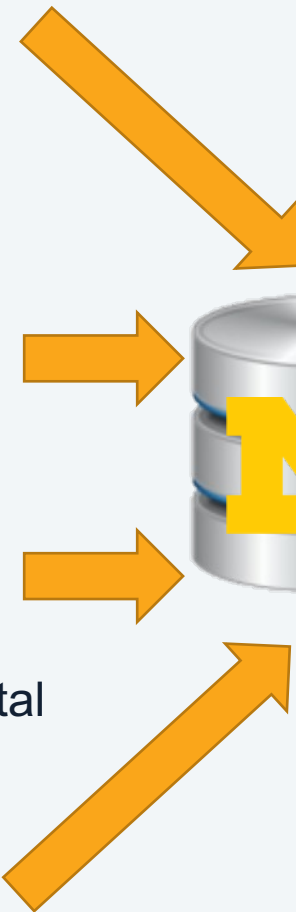
Brighton



University Hospital

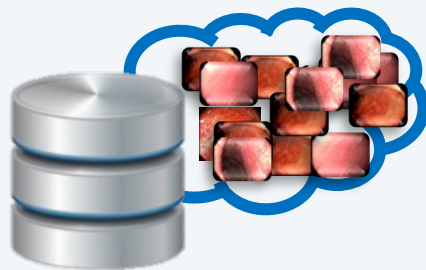
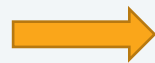


Northville



**Automated Passive/ Background Analysis
of IBD Patients Undergoing Endoscopy**

Value Proposition for Population-Level Endoscopic AI



VideoID	Visit	MES_Prediction	conf_est	MES_0	MES_1	MES_2	MES_3	video_qual_score
10206_20170515	Baseline	2	0.78	0.45	0	0.53	0.01	0.44
10297_20180103	Baseline	1	0.44	0.63	0.34	0.03	0	0.65
10995_20171227	Baseline	2	0.95	0.57	0.27	0.16	0	0.43
11135_20170322	Baseline	3	0.86	0.59	0.02	0.02	0.37	0.71
11367_20171024	Baseline	2	0.71	0.77	0.14	0.07	0.02	0.65
12419_20171031	Baseline	3	0.88	0.52	0.15	0.02	0.31	0.81
12561_20161114	Baseline	3	0.61	0.02	0.23	0.33	0.41	0.69
12826_20170529	Baseline	2	0.55	0.33	0.23	0.4	0.03	0.67
13398_20171212	Baseline	1	0.91	0.92	0.08	0.01	0	0.72
13913_20171227	Baseline	1	0.83	0.83	0.15	0.01	0	0.61
14988_20170118	Baseline	2	0.58	0.57	0.28	0.14	0.01	0.75

- **SPEED:** Rapidly Identify ELIGIBLE Clinical Trial Candidates
- **EFFICENCY:** Avoid Need to Repeat Colonoscopy for Trial
- **UNDERSERVED:** Identify Patients Distant from Tertiary Care Centers Who Typically Do Not Have ACCESS to Trials



- **Phase IV SURVEILLANCE:** Post-Marketing Real-World Drug Monitoring Feasible.
- **POPULATION HEALTH:** Manage Large Populations & Monitor Quality

AI Population Screening for Trials: Challenges and Solutions

EXPENSIVE

- Requires new hardware – Who Pays?
- Service vs. Local Hardware / PACS Purchase Models



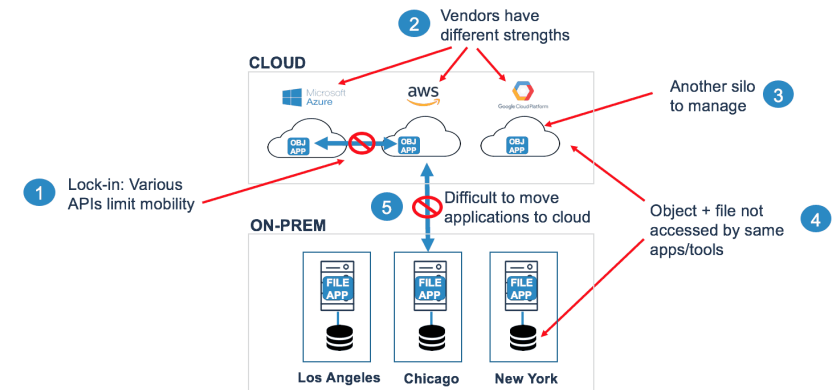
GDPR, Privacy and Data Ownership Considerations

- Expect US to Eventually Resemble EU
- 21st Century Consent Process
- Transparency, Proactive Trust, Ethical Behavior



Data SECURITY

- Multi-cloud vs. On-Prem Storage and Compute
- Chain of Custody / Chain of Use



VENDOR INTEROPERABILITY

- Standardization of Interchange Languages
- Standardization of Minimum Quality Requirements



AI in Clinical Trials III: AI-Enhanced Disease Measurement

Human Scoring Instruments Require Simplicity for Practicality

Mayo Endoscopic Score (MES) for UC

Normal: 0



Intact vascular pattern

Mild: 1



Erythema, decreased vascularity, mild friability

Moderate: 2



Marked erythema, absent vascularity, friability, erosions

Severe: 3



Bleeding, ulceration

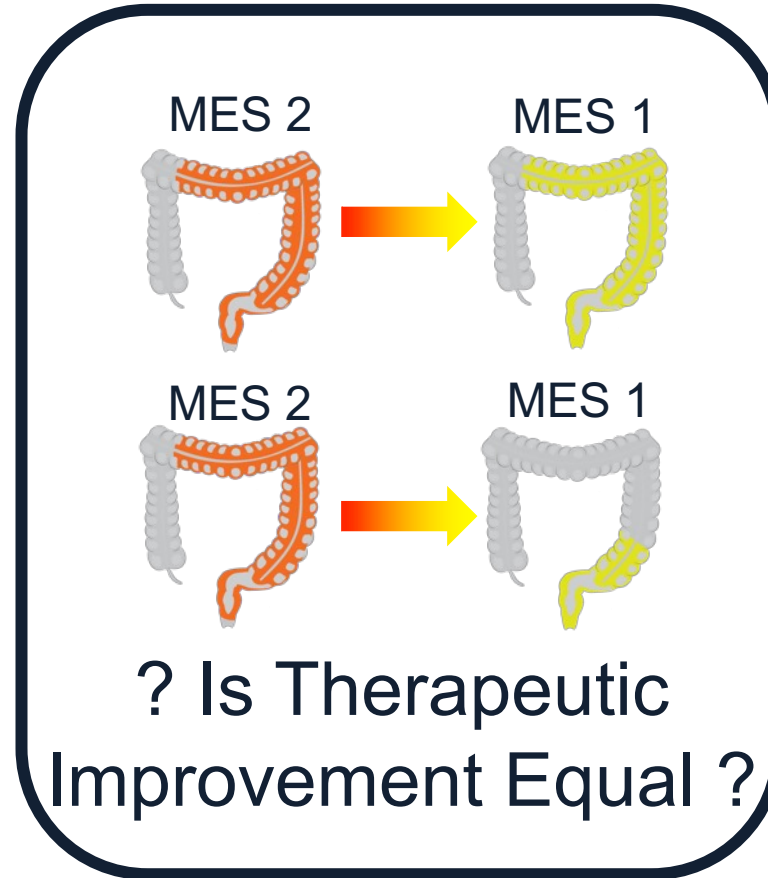
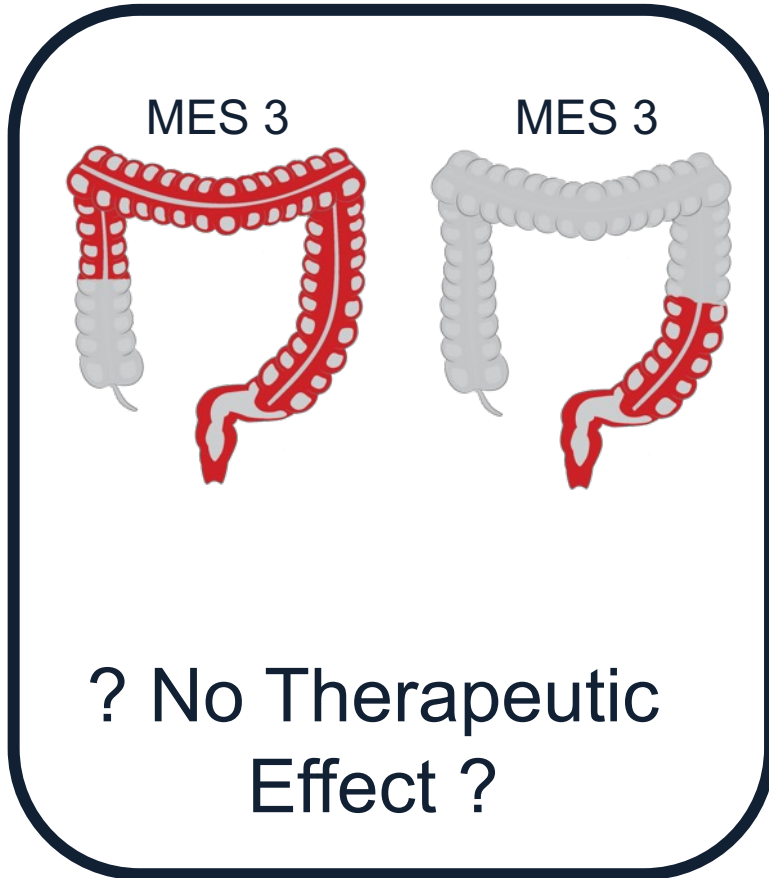
Remission

Non-Remission

Price of Simplicity Can Be Instruments that Lack Granularity

Progress in AI Analysis of IBD for Clinical Trials

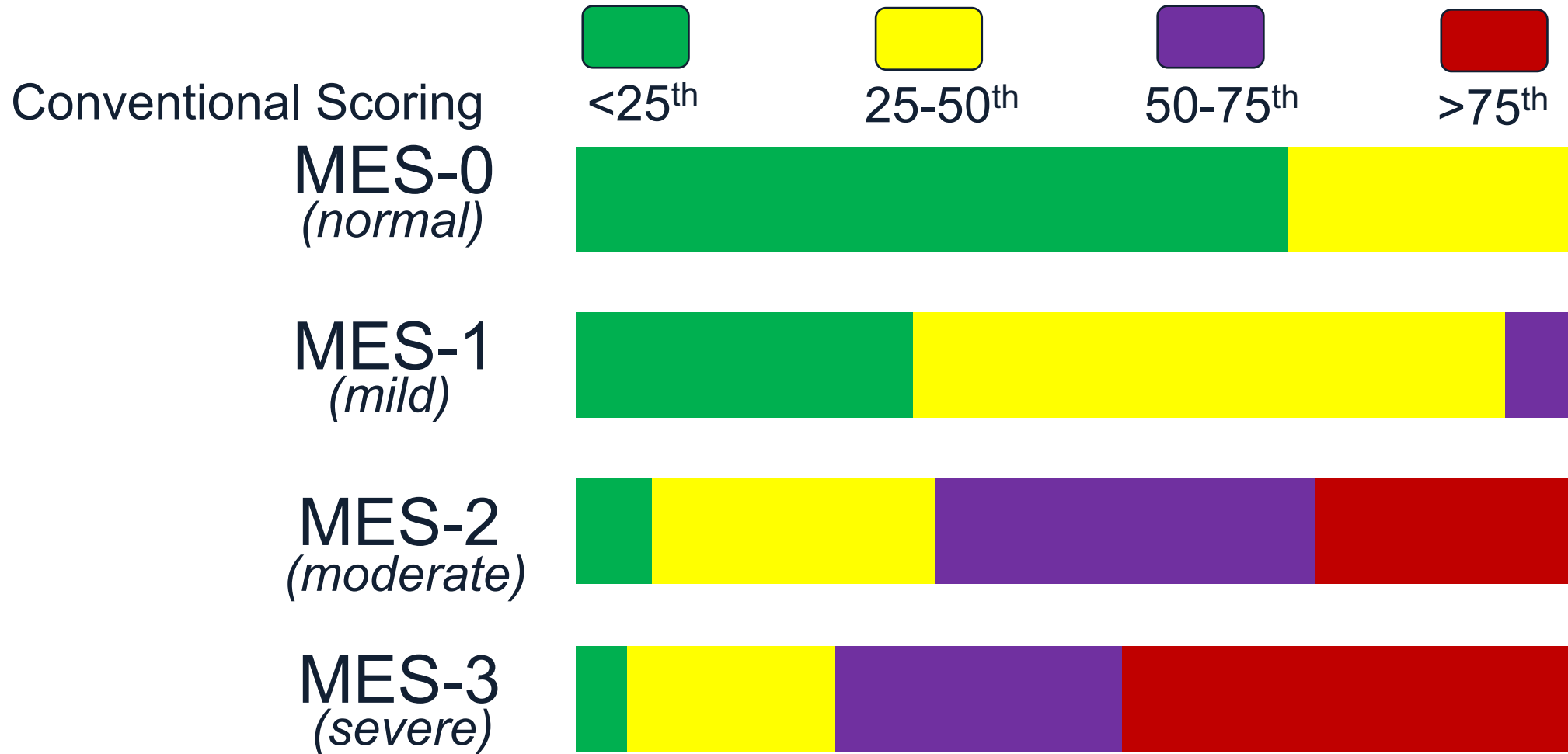
Computational Video Analysis to Address Current Scoring Limitations



**Calculating the Cumulative Disease Severity (CDS)
Showing Promise for More PRECISE Quantitation of IBD Activity**

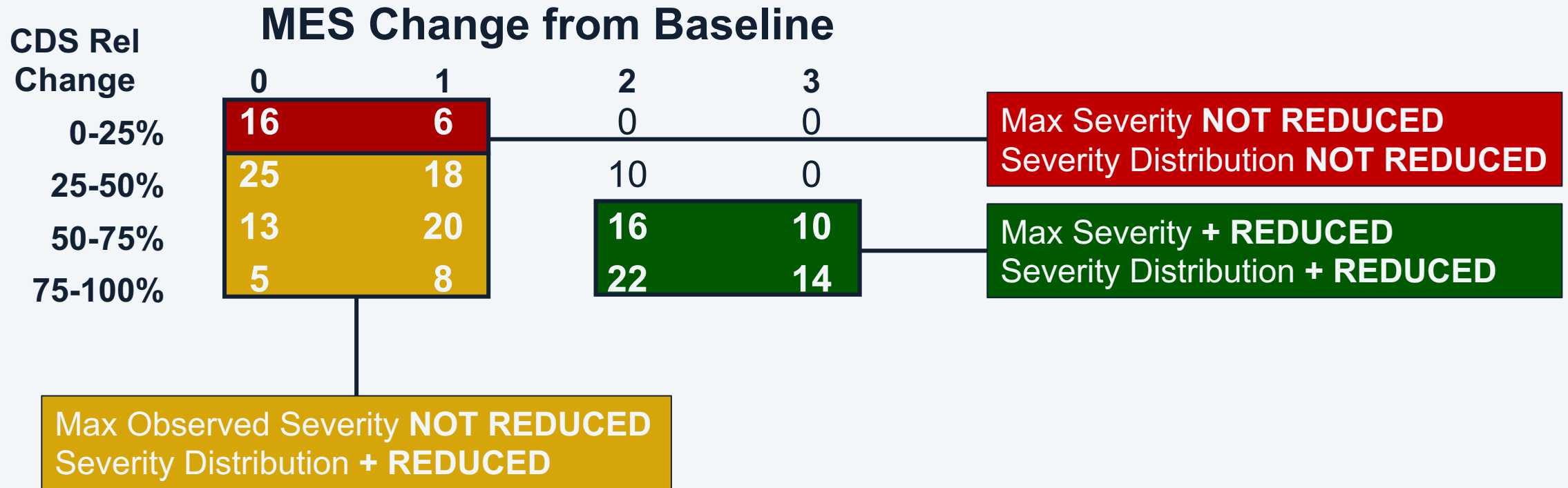
AI Scoring Highlights *Population Heterogeneity* Using Conventional FDA Approved Scoring

Colonoscopy CDS Severity Percentile



A Different Perspective on Endoscopic Response

AI Scoring Better Captures Patients with Tissue Healing



AI WILL Redefine Our Concept of Quantifying Tissue Disease & Healing

Value Proposition for AI Disease Grading in Trials

REDUCED TRIAL SAMPLE SIZES

- Improved Disease Quantification Improves Power
- Trial Completed More Quickly, Less Patients Exposed

IMPROVED PATIENT SELECTION

- AI-Enhanced Scoring Helps Predict Responders Pre-Treatment
- Better Information from Pre-Clinical and Phase I & II Studies

BETTER MEASURES OF PARTIAL RESPONSE

- Overcome lack of conventional instrument granularity
- Until we have curative medications, need to consider partial response

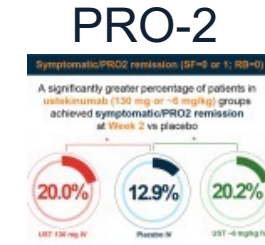
DREAMING.....AI GENERATED DISEASE MEASURES

- Forthcoming MOAs May Be Better Assessed With New Metrics
- AI Methods Will Inform New Measure Development

Barriers and Challenges

QUESTIONING ENDPOINTS

- Is linkage to clinical outcome needed ?
- Separate biologic vs. clinical outcome ?
- Plurality of good FDA Approved AI Endpoints ?

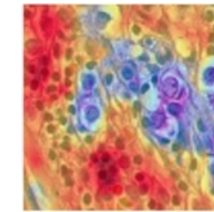


≠

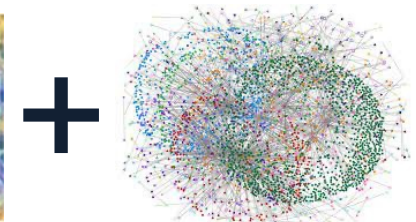


CAN WE HANDLE GRANULARITY ?

- Humans think with decision trees, not probabilities
- Need to establish score thresholds
- Move towards within-individual change vs. absolute?
- Will we end up simplifying/compressing AI granularity?



Max Data Extraction



Complex Interaction Networks



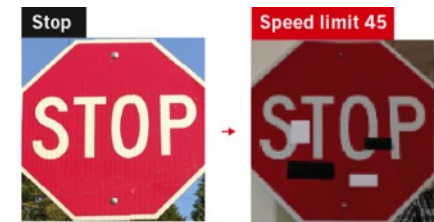
Mind Blown

WHAT AI MEASURES CAN WE TRUST ?

- Abstract vs. Understandable
- Innovation vs. Experience
- AI Analysis Tracible to Conventional Understanding?



Cuddle With It ?
Or Eat it?



Stop or Go?



The Future State of Clinical Trials in GI

Automated Screening at Population-Level

- Faster Recruitment
 - Underserved Access
 - Quicker Trial Completion
 - Phase 4 Monitoring
-
- Privacy
 - Medico-Legal
 - Data Ownership
-
- Updated Consent
 - Next Gen Data Storage
 - Block-Chain of custody

Prediction of Higher Probability Responder

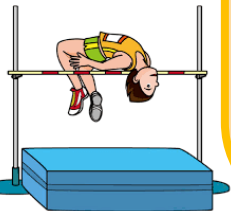
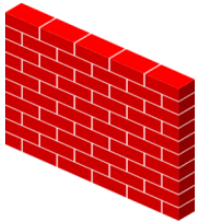
- Targeted Enrollment
 - Reduce Tx Failures
 - Less Subjects Needed
-
- Labeling Restrictions
 - Narrow Use Window
-
- FDA Approval Flexibility
 - Payors Use RWE for expanded indications

Automated Established Disease Scoring

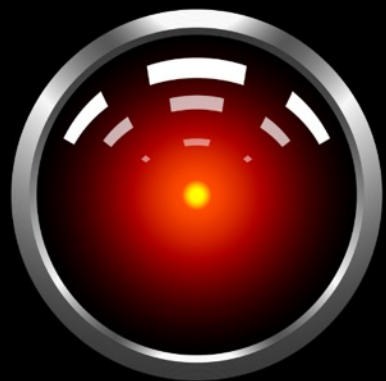
- Reduce Trials Costs
 - Uniformity of Studies
 - Small Pharma/Academia
-
- Plurality of Vendors
 - +++ FDA Review Time
-
- Streamline Automated Scoring Approval
 - Avoid Vendor Primacy

Better Measures of Disease Activity

- Bespoke MOA Measures
 - Trials Efficiency
 - New Biologic Insights
-
- +++ New Measures
 - ? Unique to MOA
 - Trust/Explainability
-
- Encourage Exploration
 - Burden of Traceability for Primary Endpoints



COLON 9000



Thank You

ryanstid@med.umich.edu

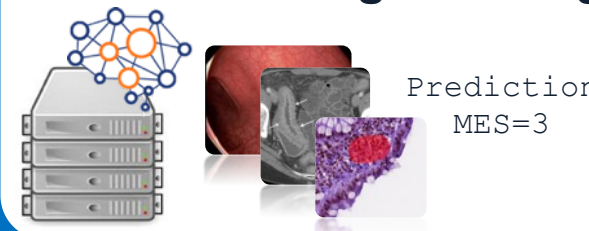
 **aga** American
Gastroenterological
Association



Operational Needs of GI Clinical Trials

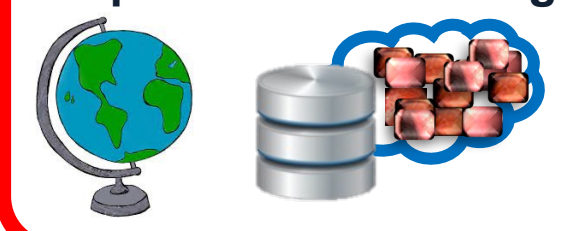
AI Solutions for Clinical Trials

Automated Image Scoring

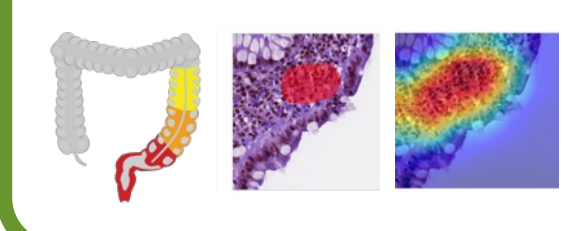


Prediction
MES=3

Population Monitoring



AI-Enhanced Measures



Measure Standardization



Rapid Recruitment



Community Representation



Improved Subject Selection



Increased Power & Sample Size Reduction



Improved RWE for Approved Tx

