

Application of a Machine Learning Model in the 2+1 Central Reading Paradigm to Assess Endoscopic Severity in Ulcerative Colitis Trials (P0307)

Chakib Battouli¹, Pavel Brodskiy², Klaus Gottlieb¹, Mohammad Haft-Javaherian², William Eastman¹, Julian Lehrer², Evan Yu², Derek Onken², Darren Thomason², Daniel Colucci², Walter Reinisch³, Shrujal Baxi²

1. Eli Lilly and Company, Indianapolis, IN, USA; 2. Iterative Health Inc., Cambridge, MA, USA; 3. Medical University of Vienna, Vienna, Austria

Disclosures: CR, MS, ML, and DD are employees of Eli Lilly and may hold stock or stock options in Eli Lilly. PG, MH, JL, ET, DT, DC, and SB are employees of Iterative Health and may hold stock or stock options in Iterative Health. WR has served as a speaker for AbbVie, Celltrion, Janssen, Dorosigma Medica, HGS, Incyte, Pfizer, Astra, Takeda, as a consultant for AbbVie, Amgen, ADC Diagnostics, Boehringer Ingelheim, Bristol Myers Squibb, Celltrion, Galapagos, Janssen, Pfizer, Teva and has received research funding from AbbVie, Janssen, Sanofi, Takeda. TBaxi is an advisory board member for AbbVie, Amgen, Boehringer Ingelheim, Bristol Myers Squibb, Celltrion, Galapagos, Janssen, Pfizer, Teva and has received research funding from AbbVie, Janssen, Sanofi, Takeda.

Introduction

- The 2+1 central reading approach is the most common workflow to assess the endoscopy subscore in ulcerative colitis (UC) trials.
- Inter-rater disagreement in expert assessments is high, impacting endoscopic outcomes in trials.
- Machine learning (ML) models have demonstrated potential to aid in the assessment standardization of the endoscopy subscore.
- Paradigms for their use in endpoint assessments are currently under investigation.

We investigate the application of a ML model to assess one of the initial two reads in the 2+1 central reading approach for evaluation of the endoscopy subscore in UC.

Methods

- Endoscopy video recordings from the mirikizumab trials and routine practice were used to develop a state-of-the-art multi-stage deep learning algorithm to assess the endoscopy subscore on full-length UC videos.
- 639 videos (~25%) from the Phase 3 LUCENT 1 (NCT03518086) induction trial (week 0 and 12 procedures) were held out to evaluate performance of the final, locked model and this analysis.
- Videos in this cohort had a 2+1 centrally read endoscopy subscore, per protocol, assessed through independent video review by the local site investigator (LR) and a central reader (CR1), with discrepancies addressed through an additional independent central read (CR2).

Results

Table 1. Performance metrics of pairwise comparisons between the ML model, LR, and CR1 in assessments of the endoscopy subscore on complete UC videos.

Assessment	Metric	LR vs CR1	LR vs ML	CR1 vs ML
Ordinal endoscopic subscore (0, 1, 2, 3)	QWK	0.711	0.687	0.723
	Agreement	62.4%	59.5%	66.5%
Endoscopic Improvement (0, 1[-] vs 2, 3 [+])	Agreement	90.6%	87.8%	88.7%
	Specificity	70.5%	69.8%	76.1%
	Sensitivity	96.2%	92.8%	91.6%
Endoscopic remission (0 [+]- vs 1, 2, 3 [-])	Agreement	93.4%	92.6%	93.3%
	Specificity	97.3%	96.0%	95.6%
	Sensitivity	31.6%	39.5%	42.9%

Abbreviations: LR, local reader; CR1, central reader 1; ML, machine learning model; QWK, quadratic weighted kappa.

Figure 1. 4x4 confusion matrix comparing the ML model to central reader 1 assessment of the endoscopy subscore on complete UC videos.

		ML model assessed endoscopy subscore			
		0	1	2	3
Central reader 1 assessed endoscopy subscore	0	12 (42.9%)	11 (39.3%)	5 (17.9%)	0 (0.0%)
	1	19 (21.3%)	47 (52.8%)	21 (23.6%)	2 (2.2%)
	2	7 (3.7%)	31 (16.3%)	111 (58.4%)	41 (21.6%)
	3	1 (0.3%)	5 (1.5%)	71 (21.4%)	255 (78.8%)

Conclusions

- We demonstrate that this ML model can be used effectively in the 2+1 central reading approach for evaluation of the endoscopy subscore in UC, and automate one of the two initial assessments with a standardized tool.
- The high discord between human reviewers creates a ceiling effect when comparing the ML model to any single human read, with performance measurements capped by the variability in the human read.