Capsule endoscopy with artificial intelligence-assisted technology: Real-world usage of a validated AI model for capsule image review

OPEN ACCESS

Authors
Fintan John O'Hara[1,2], Deirdre Mc Namara[1,2]

Institutions
1   Gastroenterology, Tallaght University Hospital, Dublin, Ireland
2   Medicine, Trinity College Dublin School of Medicine, Dublin, Ireland

Corresponding author
MB BCh BAO Fintan John O'Hara, Tallaght University Hospital, Gastroenterology, Dublin, Ireland
fintan.ohara@gmail.com

ABSTRACT

Background and study aims Capsule endoscopy is a time-consuming procedure with a significance error rate. Artificial intelligence (AI) can potentially reduce reading time significantly by reducing the number of images that need human review. An OMOM Artificial Intelligence-enabled small bowel capsule has been recently trained and validated for small bowel capsule endoscopy video review. This study aimed to assess its performance in a real-world setting in comparison with standard reading methods.

Patients and methods In this single-center retrospective study, 40 patient studies performed using the OMOM capsule were analyzed first with standard reading methods and later using AI-assisted reading. Reading time, pathology identified, intestinal landmark identification and bowel preparation assessment (Brotz Score) were compared.

Results Overall diagnosis correlated 100% between the two reading methods. In a per-lesion analysis, 1293 images of significant lesions were identified combining standard and AI-assisted reading methods. AI-assisted reading captured 1268 (98.1%, 95% CI 97.15–98.7) of these findings while standard reading mode captured 1114 (86.2%, 95% confidence interval 84.2–87.9), $P < 0.001$. Mean reading time went from 29.7 minutes with standard reading to 2.3 minutes with AI-assisted reading ($P < 0.001$), for an average time saving of 27.4 minutes per study. Time of first cecal image showed a wide discrepancy between AI and standard reading of 99.2 minutes ($r = 0.085$, $P = 0.68$). Bowel cleansing evaluation agreed in 97.4% ($r = 0.805$ $P < 0.001$).

Conclusions AI-assisted reading has shown significant time savings without reducing sensitivity in this study. Limitations remain in the evaluation of other indicators.

Introduction

Capsule endoscopy (CE) has revolutionized the management of small bowel (SB) diseases owing to its convenience and non-invasiveness [1]. Its use is endorsed by international guidelines for the investigation and diagnosis of suspected SB bleeding, iron deficiency anemia, suspected Crohn's disease, polyposis syndromes, and refractory celiac disease. Capsule reading is a time-consuming and often tedious process with average reading times in the literature ranging between 30 to 120 minutes [2]. It demands a high level of concentration without distractions to avoid missing pathology [3].

Use of artificial intelligence (AI) in CE is an attractive solution for reducing reading time by removing redundant images and

identifying suspicious abnormalities [4]. While several previous studies have demonstrated impressive sensitivity and specificity in created datasets of capsule images [4, 5, 6, 7, 8], real-world data are lacking, which is a barrier to adoption of AI in CE in clinical practice primarily due to concern about missing a lesion without human review of the entire video, despite the knowledge that human reviewers are subject to a high lesion miss rate and early fatigue [9].

The capsule journey along the gastrointestinal tract remains out the control of the operator and views of any abnormality may be fleeting and only partial. In the future, when AI will select a limited number of frames to be reviewed by the endoscopist, it is likely that over 99% of frames will never get a human check, thus the primary aim of AI in CE is currently to reduce reading times while maintaining a high sensitivity for detection of abnormalities and possibly reducing the unknown but inevitable missed lesion rate.

Reflecting the arrival of AI in endoscopy, the European Society of Gastrointestinal Endoscopy has released a position statement [10] outlining the expected value of AI in a number of areas of endoscopy including CE (▶Table 1). For acceptance, AI-assisted reading should be comparable to that of experienced endoscopists for lesion detection without increasing reading time. To date the evidence for this is lacking.

The OMOM HD CE system (Jinshan Science & Technology (Group), Yubei, China) introduced in 2020 with AI technology includes redundancy deletion, lesion detection, and classification software. It comprises three components: the capsule itself, data recorder, and computer workstation with software for interpretation and reporting. The reporting software includes a convolutional neural network (CNN)-based computer-aided detection (CADe) algorithm that can identify SB abnormalities and filter the video files for these. After downloading the video file, the software automatically processes the file.

The operator can choose to either view the entire video as standard or use the AI software, which displays only the filtered images selected by the algorithm. The filtered images are played in a video format familiar to experienced capsule readers, or as a collection of still images with suggested findings.

Xie et al recently published a paper on the development and validation of this AI software for small bowel capsule endoscopy (SBCE) video review in a Chinese study encompassing 51 CE centers [11]. It was trained using CNNs to detect 17 types of capsule endoscopy structured terminology (CEST) findings including venous structure, nodule, mass or tumor, polyps, angioectasia, red and white plaques, red spot, abnormal villi, lymphangiectasia, erythematous, edematous erosion, ulcer, aphtha, blood, and parasite [12].

During the validation phase, they reported an extremely low miss rate of 1.0% using the AI algorithm when compared with the conventional reading group. In contrast, the conventional reading group had a miss rate of 9.6% when compared with the AI-assisted group.

Our primary aim was to evaluate the diagnostic yield of the AI software versus standard reading in the detection of clinically significant SB pathology when both are performed by experienced readers in a real-world setting. Secondary aims were assessment of reading (SR) times, bowel cleansing and identification of passage to the colon for AI vs SR.

Patients and methods

Study population

In this retrospective single-center study, we looked at 40 sequential patient procedures performed using the OMOM HD capsule. These were performed in a single capsule referral center over an 8-week period. All patients referred for SBCE had met our clinically established pathways for CE to investigate suspected SB bleeding, iron deficiency anemia, Crohn's disease assessment, suspected SB Crohn's disease or SB polyps. Patients were screened for risk factors for capsule retention and had functional patency confirmed using a patency capsule if indicated.

Procedure

The capsule was administered as per normal departmental protocol. Patients fasted from the night before and came for capsule placement in the morning. No bowel prep was administered as is our local protocol. The recording device was worn for 12 hours and returned for download and analysis the following day. Recordings were downloaded to the OMOM HD capsule software platform. Standard reports were issued for all cases and clinical follow up was arranged as standard.

The studies were retrospectively analyzed using the AI software following a 3-month time delay from initial reporting with each filtered image evaluated for findings.

▶Table 1 ESGE position statement expected values for AI in capsule endoscopy recommendations.

AI task	Description	Expected value
Quality of bowel cleansing	AI-assisted endoscopist scoring according to validated scales	Comparable to scoring as adequate/inadequate by experienced endoscopists (100% agreement)
Completeness of SBCE procedure	AI-assisted identification of the cecum/colon landmarks	Comparable to cecum/colon identification by experienced endoscopists (100% agreement)
SBCE reading and lesion detection	Reading time Detection of clinically significant small bowel lesions	No increase and possible reduction in reading time Agreement > 95% with experienced endoscopist for lesion detection.

ESGE, European Society of Gastrointestinal Endoscopy; AI, artificial intelligence; SBCE, small bowel capsule endoscopy.

Recording modalities

Following download and analysis of the video by the software, the files were read in two modalities: 1) SR which was performed as per ESGE standards [2] and was reading at a maximum of 10 fps single-image mode; and 2) AI reading, which was read at 5 fps after image processing by AI-assisted technology.

SR was performed first by two gastroenterologists experienced in CE who read approximately 150 to 200 capsules per year. This was performed at the time of data capture without reference to AI findings with data examined and results recorded. The findings and report generated by this method were reviewed by our expert group (consisting of two senior gastroenterologists reading > 200 capsules per year for more than 15 years combined and three experienced readers reading 50 to 100 capsules per year) to agree on image interpretation and the text of a final report.

AI reading was performed by the same gastroenterologists 3 months later. The AI algorithm preread the examination and filtered out suspected CEST findings, each selected image was then evaluated for findings with data examined, results recorded, and a report generated.

Time of review began at examination of first SB image and ended at time of first colonic image and was recorded using a validated stopwatch. The reader was free to move between images as per their normal reviewing procedure including pausing the video and moving backward and forward through the images.

Statistics and data analysis

Data analysis was performed from July 2022 to September 2022 using Excel and SPSS. For each study reading time, pathology identified, number of images for review in the AI mode, intestinal landmark identification timings, and bowel preparation assessment (Brotz Score) were recorded.

Qualitative measures were described by mean, median, maximum, and minimum values. Pearson's coefficient was used to compare bowel prep and identification of first colonic images. The Wilcoxon signed-rank test was used to compare reading times and number of images in standard and AI-assisted reading.

A paired chi-squared test was performed with the difference in accurate detection rate of finding types between the two reading modes. Where the value of the indicators reached 100%, the 95% confidence interval (CI) calculation was based on the modified Wald method with all P values from two-sided tests and results deemed statistically significant at $P < 0.05$.

Results

Patient demographics

Forty small-bowel capsule endoscopies were included for analysis (26 male [65%]); mean age 50.0 years (standard deviation [SD] 20.0). The indications for CE were suspected SB bleeding 50% (n = 20), suspected Crohn's disease 37.5% (n = 15), Crohn's disease assessment in 7.5% (n = 3) and follow up of SB polyps in

▶ Table 2 Demographics.

Characteristic	Number or years	Percentage or SD
Male (n, %)	26	65.0%
Age (years, SD)	50.0	20.0
Indication (n, %)		
Suspected small bowel bleeding	20	50.0%
Suspected Crohn's disease	15	37.5%
Crohn's disease assessment	3	7.5%
Polyp surveillance	2	5.0%

SD, standard deviation.

▶ Table 3 Patient diagnosis.

Overall findings	AI mode (n, %)		SR mode (n, %)	
Normal	20	51.3%	20	51.3%
Ileitis	7	17.9%	7	17.9%
Angiodysplasia	4	10.3%	4	10.3%
Blood/melena	3	7.7%	3	7.7%
Meckel's diverticulum	1	2.6%	1	2.6%
Polyp	1	2.6%	1	2.6%
Tumor	1	2.6%	1	2.6%
Submucosal bulge	2	5.1%	2	5.1%
Gastric retention	1	2.6%	1	2.6%

AI, artificial intelligence; SR, standard reading.

5% (n = 2) (▶ Table 2). One patient had gastric retention of the capsule for the entire duration of the study and was excluded, leaving 39 studies included for analysis. No capsule retention was recorded.

Overall patient diagnoses

For both modalities, 19 studies were reported with abnormalities giving a 48.7% diagnostic yield (▶ Table 3) when normal variants such as lymphangiectasia were excluded. This is within the range seen in published data [13, 14]. Seven patients had ileitis to varying degrees, four cases of angiodysplasia were identified, with blood/melaena seen on a further three studies. A polyp was identified on a further study and a likely tumor was identified on another (confirmed as an adenocarcinoma on follow up enteroscopy).

There was excellent agreement between both modalities for overall diagnosis with 100% correlation. Sensitivity, specificity, and positive and negative predictive values of AI reading compared to SR were 100% on overall diagnosis.

▶ Table 4 Lesion detection sensitivity.

	AI (n)	Sensitivity	SR (n)	Sensitivity	Combined findings (n)	P value (AI vs SR)
Venous bleb	16	100.0%	14	87.5%	16	0.14
Mass/tumor	68	100.0%	65	95.6%	68	0.08
Polyp	10	100.0%	10	100.0%	10	
Angiodysplasia	81	100.0%	76	93.8%	81	0.02
Red spot	84	95.5%	60	68.2%	88	< 0.01
Lymphangiectasia	361	97.8%	296	80.2%	369	< 0.01
Ulcer/erosion	629	98.6%	570	89.3%	638	< 0.01
Blood	14	100.0%	14	100.0%	14	
Endoscopic clip	0	0.0%	4	100.0%	4	< 0.01
Tattoo	5	100.0%	5	100.0%	5	
Overall		98.1% (95.5 – 100%)		86.2% (68.2 -100%)		

AI, artificial intelligence; SR, standard reading.

Per significant lesion analysis

In the 39 studies included, 1293 images of significant lesions were identified (lymphonodular hyperplasia was excluded from analysis) when both reading modes were combined. These included angiodysplasia, ulcers/erosions, polyps, submucosal bulges, lymphangiectasia and p1 vascular lesions. AI captured 1268 (98.1%, 95% CI 97.15–98.7) of these findings while SR captured 1114 (86.2%, 95% CI 84.2–87.9). This result is significant with a $P < 0.001$ (▶ Table 4).

The enhanced AI sensitivity comes with many normal images included in the selected frames for review (false positives). AI selected 14640 SB images for review over the 39 studies with 1293 positive findings giving a true-positive rate of just 8.8%. More significantly, however, the mean number of images for re-

view for each study was reduced to 375.4 (SD 329.4, interquartile range [IQR] 139–684).

Missed lesions

A total of 17 images of significant lesions were not noted in the AI mode (▶ Table 4); four images of P1 angiodysplasia in an otherwise positive study for angiodysplasia, nine images of a circumferential ulcer at the ileoeceal calve in an area of poor prep were also missed, and four images of a clip placed during a recent antegrade enteroscopy (▶ Fig. 1). In addition, four images of reds spots and eight images of lymphangiectasia were also not selected for review. Interestingly, tattoo placement during previous enteroscopy was identified in five images.
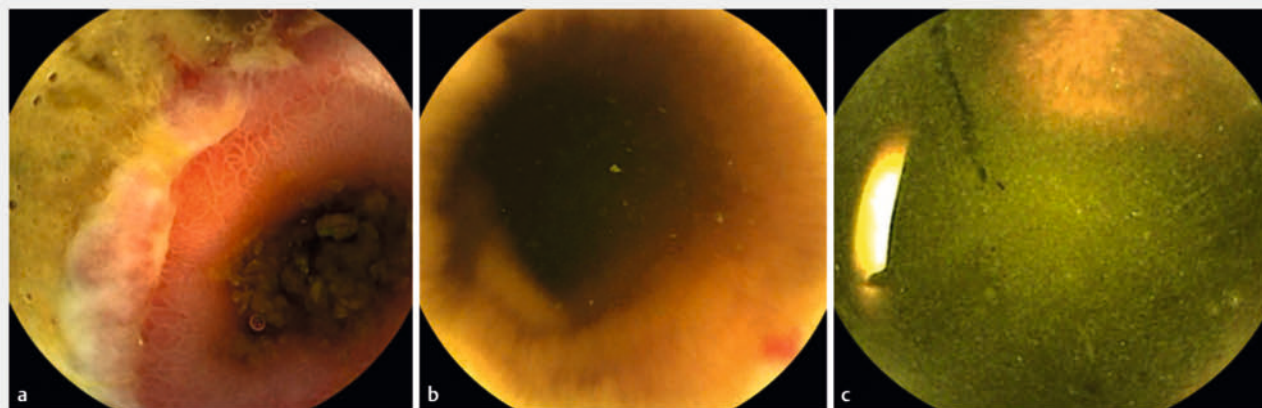
▶ Fig. 1 Examples of missed lesions.

Table 5 Small bowel reading time.

SBCE reading time (minutes)	AI (mins)	SR	Difference	P value
Mean (SD)	2.29 (1.39)	29.69 (11.82)	27.40	< 0.001
Median	2	27	25	< 0.001
Range	1–4.5	15–63		

Non-parametric paired Wilcoxon rank sum test: Z = -5.4424; P < 0.0001. Small bowel capsule endoscopy; SD, standard deviation; AI, artificial intelligence-assisted reading; SR, standard reading.

Table 6 Intestinal landmark identification.

Landmark	SR time	AI time	Mean difference	Range	
First duodenal image (minutes)	45.0	94.2	−49.2	−543 to + 107	(r = 0.222, P = 0.275)
First cecal image (minutes)	304.5	205.3	99.2	−567 to + 337	(r = 0.085, P = 0.679)

SR, standard reading; AI, artificial intelligence.

Reading time

AI assistance significantly reduced reading time. The mean (SD) reading time for SB images went from 29.7 minutes (11.8) with SR to 2.3 minutes (1.4) with AI mode (z = -5.44, P < 0.001). This gives a mean time savings of 92.3% or 27.4 minutes per study (▶ Table 5).

Quality of bowel cleansing

Bowel preparation assessment was performed using a validated scoring system described by Brotz [15]. Comparing overall cleansing evaluation scores, which is the standard set by ESGE [10], agreement was seen in 97.4%. Thirty-six of 39 studies (92.3%) in SR mode vs 37 of 39 (94.9%) in AI mode were deemed to have adequate cleansing, giving a fairly strong correlation (r = 0.805, P < 0.001). Using the Brotz qualitative evaluation of prep as poor, fair, good, and excellent, results correlated moderately between the two reading methods with agreement in 54% (r = 0.554, P < 0.001).

Landmark identification/completeness of procedure

Intestinal landmark identification is important for assessing for complete evaluation of the SB as well in localizing lesions for planning further investigations such as direction of enteroscopy. The AI software marks what it believes to be the position of the start and end of the SB during processing by its CNN. These timestamps were compared to those recorded during SR.

Recognition of the intestinal landmarks by the software was poor (▶ Table 6). Time of first cecal image showed a wide discrepancy between AI and SR of 99.2 minutes (range -567 minutes to + 337 minutes, r = 0.085, P = 0.68). Time of first duodenal image also showed a similar level of discrepancy between AI and SR with a mean difference of 49.2 minutes (range -543 to + 165 minutes, r = 0.22, P = 0.27). As such, localization of significant abnormalities as well as calculations of Lewis score [16] will likely be unreliable using the AI landmarks alone.

Discussion

Currently, reporting SBCE is a task for single reviewers. It can be limited by human error and human reading performance has been shown to be disappointing [17]. The reporting accuracy in SBCE declines after reading a single capsule study [18]. In its current iteration, AI in CE needs to recognize all lesions because it remains at this stage a filter of normal from the abnormal prior to human review.

This is the first real-world study of the use of the recently validated OMOM AI system.

Correlation between AI-assisted reading and SR for detection of pathology is excellent. Overall, in a per-procedure analysis, the diagnostic accuracy was equivalent for the two modalities. Indeed, in these 39 cases, reports generated by the two reading methods agreed 100% on overall diagnosis.

When detection of each clinically significant SB lesions was examined, there was a higher rate of detection for the AI-assisted method (98.1% vs 86.2%). This met ESGE expected value for AI in CE of > 95% agreement with an experienced endoscopist for lesion detection. SR missed 179 lesions (13.8%) while AI-assisted reading missed only 25 (1.9%).

When lesions missed by AI-assisted reading were reviewed, bowel preparation was seen to be a factor in all cases. Four were small P1 lesions in areas of suboptimal prep, nine images were circumferential ulcers at the ileocaecal valve also in an area of unprepped colon, while the prep around the endoscopic clip was particularly poor. While in colonoscopy, the endoscopist can focus on an area of interest, clean any debris, and steady the endoscope in CE, the software can only evaluate the images recorded with poor prep and partial or blurred images as a fact of life when reading CE.

AI-assisted reading showed superior performance compared to SR with a 92.3% reduction in mean reading times down to 2.29 minutes from 29.69 minutes per study.

Landmark identification by the AI software showed a poor correlation with expert reader with 49.2 minutes average dif-

ference for first cecal image (r = 0.222, *P* = 0.275), and thus, it does not meet the ESGE expected value of 100% agreement. This will lead to issues with localization of lesions when planning further investigations and therapies as well as assessment of Lewis scores, which are based on identification of tertiles of the SB.

AI in CE will not remove the known limitations of the test, such as the use and timing of bowel preparation and delayed capsule transit [2]. Indeed, standards for bowel preparation may become a more significant issue because missed lesions in this study were associated with areas of suboptimal prep. Reduction in images reviewed to an average of only 375.4 per study will also limit the reader assessment for rapid transit through parts of the SB as well as areas of poor prep. The evaluation of the clinical relevance of any findings remains the role of the clinician.

Overall, the software meets the ESGE expected values of > 95% agreement with a human reader for identification of lesions. Whether this is sufficient for adoption is unclear because liability for missed lesions remains with the clinician and abnormalities may only be seen on a single frame. Also, AI-assisted reading did not meet the ESGE expected values for assessment of completeness of the study or quality of bowel preparation. Given this, AI-assisted reading software, in its current form, would seem more appropriately used as a preread or as a training aid for those beginning capsule reading.

This study does have a few limitations. This was a single-center study of a small number of capsules. Use of the SR mode as the reference standard is limited by the abilities of that reader. Ideally each study would be reviewed by a panel of readers; however, the effort involved was deemed to be prohibitive.

Conclusions

In conclusion, with recognition of its limitations, AI software has the potential to significantly reduce reading time in CE without negatively affecting pathology recognition and diagnostic yield. Further studies to increase the dataset are required to evaluate the clinical utility of the software in its current form.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Iddan G, Meron G, Glukhovsky A et al. Wireless capsule endoscopy. Nature 2000; 405: 417 doi:10.1038/35013140

[2] Rondonotti E, Spada C, Adler S et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endos-copy (ESGE) Technical Review. Endoscopy 2018; 50: 423–446 doi:10.1055/a-0576-0566

[3] Rondonotti E, Pennazio M, Toth E et al. How to read small bowel capsule endoscopy: a practical guide for everyday use. Endosc Int Open 2020; 8: E1220–E1224 doi:10.1055/a-1210-4830

[4] Dray X, Iakovidis D, Houdeville C et al. Artificial intelligence in small bowel capsule endoscopy - current status, challenges and future promise. J Gastroenterol Hepatol 2021; 36: 12–19 doi:10.1111/jgh.15341

[5] Soffer S, Klang E, Shimon O et al. Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. Gastrointestinal Endoscopy 2020; 92: 831–839 doi:10.1016/j.gie.2020.04.039

[6] Tsuboi A, Oka S, Aoyama K et al. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. Digestive Endoscopy 2020; 32: 382–390 doi:10.1111/den.13507

[7] Trasolini R, Byrne MF. Artificial intelligence and deep learning for small bowel capsule endoscopy. Dig Endosc 2021; 33: 290–297 doi:10.1111/den.13896

[8] Vieira PM, Freitas NR, Lima VB et al. Multi-pathology detection and lesion localization in WCE videos by using the instance segmentation approach. Artif Intell Med 2021; 119: 102141 doi:10.1016/j.artmed.2021.102141

[9] Beg S, Card T, Sidhu R et al. The impact of reader fatigue on the accuracy of capsule endoscopy interpretation. Dig Liver Dis 2021; 53: 1028–1033 doi:10.1016/j.dld.2021.04.024

[10] Messmann H, Bisschops R, Antonelli G et al. Expected value of artificial intelligence in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy 2022; 54: 1211–1231 doi:10.1055/a-1950-5694

[11] Xie X, Xiao Y-F, Zhao X-Y et al. Development and validation of an artificial intelligence model for small bowel capsule endoscopy. Video review. JAMA Network Open 2022; 5: e2221992 doi:10.1001/jama-networkopen.2022.21992

[12] Korman LY. Capsule Endoscopy Structured Terminology (CEST): Proposal of a standardized and structured terminology for reporting capsule endoscopy procedures. Endoscopy 2005; 37: 951–959 doi:10.1055/s-2005-870329

[13] Pennazio M, Spada C, Eliakim R et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. Endoscopy 2015; 47: 352–376 doi:10.1055/s-0034-1391855

[14] Shim KN, Moon JS, Chang DK et al. Guideline for capsule endoscopy: obscure gastrointestinal bleeding. Clin Endosc 2013; 46: 45–53 doi:10.5946/ce.2013.46.1.45

[15] Brotz C, Nandi N, Conn M et al. A validation study of 3 grading systems to evaluate small-bowel cleansing for wireless capsule endoscopy: a quantitative index, a qualitative evaluation, and an overall adequacy assessment. Gastrointest Endosc 2009; 69: 262–270

[16] Cotter J, Dias de Castro F, Magalhães J et al. Validation of the Lewis score for the evaluation of small-bowel Crohn's disease activity. Endoscopy 2015; 47: 330–335

[17] Zheng Y, Hawkins L, Wolff J et al. Detection of lesions during capsule endoscopy: physician performance is disappointing. Am J Gastroenterol 2012; 107: 554–560

[18] Beg S, Card T, Sidhu R et al. ADWE-07 How many capsule endoscopy cases can be read before accuracy is affected? Gut 2018; 67: A164–A164