Topic Exploration Report [1]

AI-assisted pathology to triage suspected gastric cancer biopsies

What is a Topic Exploration Report?

Topic Exploration Reports are not health technology assessments. These reports provide a high-level briefing on new topics submitted to Health Technology Wales and are not based on exhaustive or systematic literature searches. Instead, they rely on a focussed scan of key resources.

What evidence is used in a Topic Exploration Report?

Priority is given to summarising the most relevant or useful evidence, rather than covering all possible evidence. Information reported is typically based on abstracts and study authors' own conclusions, rather than detailed scrutiny of full texts.

What are the aims of a Topic Exploration Report?

Topic Exploration Reports offer an overview of the available evidence on a topic and aim to highlight any uncertainties or gaps in the evidence. These reports outline the quantity and type of evidence found, but no critical appraisal or formal evidence synthesis is conducted.

How should a Topic Exploration Report be used?

Topic Exploration Reports can be used to indicate what evidence may be available for a topic, and do not provide definitive guidance on how a technology should be used. The evidence presented within the reports should be interpreted with caution.

---

[1] Cyfieithu dogfennau HTW wedi'u cyhoeddi o'r Saesneg i'r Gymraeg
Translation of published technical HTW documents from English into Welsh

Topic exploration report number	TER587
Topic	AI-assisted pathology to triage suspected gastric cancer biopsies
Summary of findings	AI review of gastric biopsies suspected of cancer or pre-cancerous conditions prior to pathologist review is hoped to improve the accuracy of review and speed up the review of cases in which the AI detects suspicious lesions.

Ibex Gastric is currently being implemented as part of the Small Business Research Initiative (SBRI) in Wales, so HTW searched for evidence of AI tools which can help triage and review gastric biopsies.

Four individual studies and the ongoing SBRI project were identified. All identified studies reviewed a different AI technology, with none being consistently used across multiple publications. No peer reviewed published evidence for the technology being used within the SBRI project (Ibex Gastric) was identified. AI alone review of biopsies did appear to improve accuracy and reduce review time compared to pathologist alone. The AI did appear to be better at identifying negative cases than positive cases, and there was some variability in the technology's sensitivity and specificity. |

Introduction and aims

Gastric biopsies are commonly taken during gastroscopy procedures if there is suspicion of disease. These biopsies are then assessed for cancer and precancerous lesions and precursors including Helicobacter Pylori. Around 50% of the 2.5 million diagnostic endoscopies undertaken per year in the UK result in biopsies being taken, and these all require review by a pathologist- causing a great demand and burden on these services, which are commonly understaffed. The aim of artificial intelligence (AI) assisted review of these biopsies is to triage those which need urgent review by pathologists, and those which do not appear to have cancer or precursors. This aims to ensure that those biopsies with a higher chance of containing cancer or a high-risk precursor can be reviewed sooner, with the intention that patients can start treatment sooner, or have confirmation there is no disease- thus reducing anxiety for them and their families.

This topic was submitted to HTW via the Small Business Research Initiative (SBRI) project. The AI technology, Ibex Gastric, which has a CE mark certificate in this indication as an In Vitro Diagnostic Device, is being rolled out at a few health boards in Wales. The aim of the project is to assess the effect of Ibex Gastric on turnaround times to diagnosis, efficiency gains on pathologist workloads, and acceptability to pathologists.

Health Technology Wales researchers searched for evidence on the clinical and cost effectiveness of AI-assisted review of gastric biopsies in the detection of gastric cancer and its precursors.

Evidence overview

Individual studies

Yoo et al (2024) used an AI technology called SeeDP as a quality control check to review cases for incorrect diagnoses. They re-reviewed discordant cases, where 2 pathologists disagreed. Due to scanner failure or other issues, only 67.7% of the slides submitted (572,254 out of 844,906) were scanned and 0.8% of the scans were not reviewed by AI. AI came to a different conclusion from the pathologist's diagnosis in 7.7% cases (42,760). Following a more detailed review- only 5.5% of the disagreements (25 of 454) were true misdiagnoses. SeeDP detected more misdiagnosis (7 v 14) in shorter timeframe (3.6 v 38.7 days, $p<0.001$) than when conventional error recognition methods were used.

Zhu et al (2022) reviewed semi-manual assessment whereby pathologists used the endoscopic gastric biopsy assistant system (EGBAS) to assist during review and diagnosis. The overall four-tier classification (negative, low-grade dysplasia, high-grade dysplasia and intramucosal invasive neoplasia) accuracy increased from 66.49% ± 7.73% to 73.83% ± 5.73% ($p < 0.05$). Length of time to complete dataset reduced from 461.44 mins ± 117.96 mins to 305.71 mins ± 82.43 ($p < 0.001$).

Yoshida et al (2017) used an AI technology called ePathologist. Two pathologists reviewed 3062 gastric biopsy specimens, and their findings were compared to e-Pathologist. Reports could be 3 tier (positive, caution/ suspicious, or negative) or 2 tier (positive or negative). From all the slides reviewed, 33.4% showed an abnormal finding. Overall concordance for 3-tier reporting was 55.6%, with a kappa coefficient of 0.28 (interpreted as fair agreement). For negative slides, concordance was reported to be 90.6%, and for positive slides this was <50%. For 2 tier reporting, sensitivity was reported to be 89.5%, specificity was 50.7%, positive predictive value was 47.7%, and negative predictive value was 90.6%.

Iwaya et al (2023) assessed an AI technology called ResNet50, a general AI model for image processing. They used it to assess slides for intestinal metaplasia (IM) which can indicate

Evidence overview

higher risk of gastric cancer. IM can also be graded, to indicate its severity from zero to three, where 0 = no IM, 1 = mild IM, 2 = moderate IM and 3= severe IM. 5753 slides were assessed and classified as having IM (or not) with a sensitivity of 97.7% and specificity of 94.6%. The sensitivity of classifying IM correctly (grade 0 to 3) was 98.5, with a specificity of 94.9%. The IM score allocated by ResNet50 differed to the pathologists in 7.6% of cases, and although ResNet50 could miss small foci of IM, it was reported to successfully identify minimal IM that pathologists missed.

Ongoing studies

Ibex Gastric has CE marking and is being used within the Small Business Research Initiative (SBRI) across Wales. The project aim is to demonstrate Ibex Gastric can be adopted at scale across Wales and can accurately classify and triage gastric samples. Primary objectives are turnaround time for cancer and other clinically significant diagnoses with AI compared to standard of care, efficiency gains (pathologist time saved), user satisfaction via pathologist survey. Due to complete in March 2025 and be analysed/ reported soon after.

Areas of uncertainty

Very limited evidence for AI assisted review of gastric biopsies. Use of AI in evidence identified varied between using the AI prior to pathologist review, concurrent use, or as a quality control check post pathologist review. Evidence is therefore of limited applicability to the suggested use of Ibex Gastric within SBRI context, which is using it prior to pathologist review as a triaging tool.

Some, but limited, evidence for AI assisted review versus pathologist alone, most evidence was AI vs pathologist.

No peer reviewed published evidence for the technology being used within SBRI project, which was the only technology identified which clearly had a CE mark (Ibex Gastric AI). Only conference abstracts were identified for this abstract, and they were not included.

Literature search results

Health technology assessments and guidance
No evidence identified

Evidence reviews and economic evaluations
No evidence identified

Individual studies
Iwaya M., Hayashi Y. et al (2023) Artificial intelligence for evaluating the risk of gastric cancer: reliable detection and scoring of intestinal metaplasia with deep learning algorithms. Gastrointestinal Endoscopy. 98 (6): 925-933. https://doi.org/10.1016/j.gie.2023.06.056   Yoo SY., Hwang Y., et al. (2024). Artificial Intelligence–Assisted Daily Quality Control System for the Histologic Diagnosis of Gastrointestinal Endoscopic Biopsies. Archives of Pathology and Laboratory Medicine. Online ahead of print. https://doi.org/10.5858/arpa.2024-0173-OA   Yoshida H., Shimazu T., Kiyuna T. *et al.* Automated histological classification of whole-slide images of gastric biopsy specimens. *Gastric Cancer* **21**, 249–257 (2018). https://doi.org/10.1007/s10120-017-0731-8   Zhu Y., Yuan W., et al (2022). Two-step artificial intelligence system for endoscopic gastric biopsy improves the diagnostic accuracy of pathologists. Frontiers in Oncology. 12: 1008537. https://doi.org/10.3389/fonc.2022.1008537

Date of search	06 Feb 2025
Concepts used	Artificial Intelligence, gastric/ stomach cancer, biopsy, Ibex

Proposed research question and evidence selection criteria (if selected)

Proposed Research question	What is the clinical and cost effectiveness of Ibex Gastric in the review of gastric biopsies.

	Inclusion criteria	Exclusion criteria
Population	People who have undergone a gastric biopsy for possible gastric cancer	
Intervention	AI assisted biopsy review	AI alone biopsy review
Comparison/ Comparators	Pathologist alone biopsy review	
Outcome measures	Sensitivity for detection of gastric cancer or precursors Specificity for detection of gastric cancer or precursors Slide/ case review time Turnaround time to diagnosis of gastric cancer or identification of precursors Turnaround time to treatment for gastric cancer or precursors User acceptability (pathologists) Health related QoL Resource use Economic outcomes	

Proposed specialities	Cancer, Digestive system