Landscape of natural language processing (NLP) capabilities at clinical sites: insights from a real-world (RW) gastric cancer (GC) study

S. Jones[1], L. Turner[1], A. Rashid[1], J. Gallinaro[2], M. Borges[3], K. Vitanova[2], E. Eldridge[4], M. Conill[5], V. Saglimbene[6], I. Guerra[2]

[1]AstraZeneca, Baar, Switzerland, [2]IQVIA, London, United Kingdom, [3]IQVIA, Oeiras, Portugal, [4]IQVIA, Durham, NC, USA, [5]IQVIA, Barcelona, Spain, [6]IQVIA, Milan, Italy

RWD113

Background and objective

- Medical notes contain valuable clinical information, yet they are often underutilized in real-world evidence generation due to cost and complexity of manual curation.
- NLP offers solutions to mitigate this issue by providing efficient ways to extract information from notes in an unstructured format.
- However, the adoption of NLP across clinical sites is still unclear.
- This study aimed to assess the current landscape of NLP capabilities for research purposes across sites participating in a real-world gastric cancer study.

Methods

- A feasibility questionnaire (FQ) was developed in collaboration with NLP experts to capture information on sites' NLP capabilities, including technical details of NLP, regulatory compliance and quality assurance processes in place (Table 1).
- The FQ was sent to 27 sites, across six countries: five in Canada, four in UK, three in France, six in Germany, five in Italy, and four in Switzerland.
- Sites were chosen for their expertise in GC treatment and RWE capabilities, with NLP availability considered as a secondary factor. Several of selected sites belonged to IQVIA's Oncology Evidence Network (OEN).
- Follow-up interviews were conducted with sites to answer questions and clarify responses when needed.

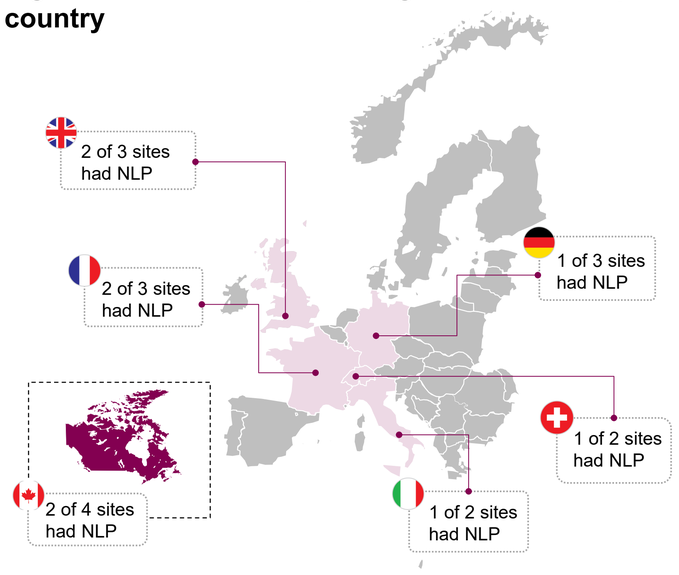
Table 1 Topics included in the feasibility questionnaire

Topic	Description
NLP solution	Overview, elements of the pipeline, type of NLP in use, preprocessing steps, data handling, transfer and formatting.
Variable availability	Assessment of which study relevant variables have been extracted or could be extracted using NLP, level of flexibility in how extracted variables are defined.
Regulatory compliance	Steps taken to satisfy regulatory compliance, including relevant guidelines and regulations followed.
Quality assurance	Procedures in place to ensure completeness of the data and the approach to ongoing monitoring and quality assurance of data.
Performance evaluation	Performance metrics used, level at which metrics are calculated (hit level, assertion level, patient level), frequency of evaluation.
Publication	Any publicly available information about the NLP algorithm implemented, for example any published study that used NLP for variable extraction.

Results

Of the 17 responding sites, nine reported having NLP capabilities (Figure 1).

Figure 1 Number of responding sites with NLP per country



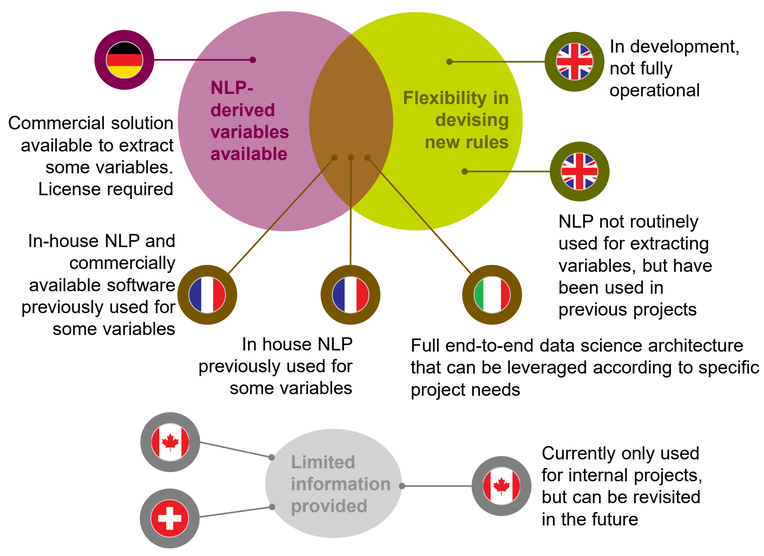
Variable availability

Among sites reporting to have NLP capabilities, four sites had already extracted study-relevant variables using NLP. Three of these four also indicated capacity to extract additional variables (Figure 2).

Two sites from the United Kingdom declared to have prior NLP experience but did not have existing algorithms available that could be easily adapted for variable extraction in new studies (Figure 2).

One Canadian site declared to be piloting NLP internally but not for external use; the other Canadian site and the Swiss site did not provide further details beyond having NLP capabilities available (Figure 2).

Figure 2 Variable availability across sites



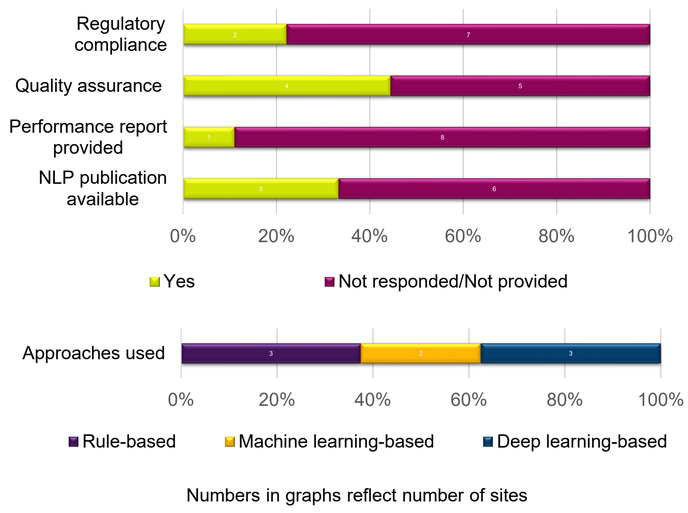
NLP capabilities

The NLP approaches used varied, including rule-, machine learning- and deep learning-based algorithms (including LLMs) developed or fine-tuned in-house or commercially available software (Figures 2 and 3).

Of the four sites that declared to have study-relevant NLP-extracted variables available:

- Two sites confirmed regulatory compliance to different regulations such as GDPR, EU AI Act, DPIA or local ethics committee (Figure 3).
- All four sites declared to have quality assurance processes in place, for example through comparison of NLP output with in-house developed gold standards established by pathology, clinicians and technical experts (Figure 3).
- Although only one site shared a performance report (Figure 3), all four sites reported evaluating performance on an individual variable level, and three sites reported using precision, recall and F1 score as evaluation metrics.
- Three sites reported evaluating performance continuously, with one site reporting a weekly frequency evaluation.

Figure 3 Overview of NLP capabilities reported by sites



Numbers in graphs reflect number of sites

As the use of NLP becomes more widespread, it is very important that sites disclose AI use in a transparent way, including informing if variables available in structured format have been previously extracted using NLP.

To ensure transparency, sites intending to use previously validated NLP algorithms for variable extraction in new studies should ideally provide performance reports or publications with results of validation.

Conclusions

- NLP adoption for variable extraction in clinical settings remains limited. While half of responding sites have explored NLP internally or in past studies, only a subset have validated, reusable algorithms readily available for research purposes.
- More mature NLP pipelines at sites and standardized validation could enable model reuse across projects. Combined with clearer guidance from regulators and HTA bodies, these improvements could support broader adoption of NLP for more efficient variable extraction in RWE studies.

Abbreviations

NLP: Natural language processing; RW: Real-world; RWE: RW evidence; GC: Gastric cancer; FQ: Feasibility questionnaire; OEN: Oncology evidence network; GDPR: General Data Protection Regulation; DPIA: Data protection impact assessment; LLM: Large language model

Supported by IQVIA

Presented at the International Society for Pharmacoeconomics and Outcomes Research Europe 2025, Glasgow, Scotland, UK, 9-12 November 2025