Algorithmic Case Finding Approaches for Type 1 Gaucher Disease in Primary Care Records

Patrick Deegan[*], Calum Grant[†], Elizabeth Morris[*], Rand Dubis[†], Amanda Worker[†], Hadley Mahon[†], Peter Fish[†].

[*]*Addenbrooke's Hospital, Cambridge, United Kingdom,* [†]*Mendelian, London, United Kingdom*

Diagnostic Delay in Type 1 Gaucher Disease

Early treatment improves outcomes: provides symptomatic relief, slows disease progression, and prevents severe complications[1]

Significant diagnostic delays persist:

- 5 year delay between symptom onset and diagnosis in UK settings[2]
- 1 in 6 patients diagnosed >7 years after first consulting a physician[3]
- Patients see up to 8 specialists before diagnosis[4]
- Avoidable severe or irreversible complications by diagnosis[4]

Reasons for delays: rarity / limited physician awareness, varied clinical presentations, non-specific early symptoms (e.g. fatigue, pain, nosebleeds)[2]

Case Finding Approaches

Opportunity for big data / AI approaches to address diagnostic delays.

A rule-based case-finding approach was attempted prior to this project:

- Not successful due to broad, non-specific, and common clinical features
- Optimising for sensitivity lowered specificity ➔ impractical flagging rates
  - Exacerbated in rare disease when account for prevalence

A machine learning approach was then attempted:

- Low no. of available cases + low feature density ➔ poorly trained models
- Approaches to mitigate failings again leads to impractical flagging rates when accounting for real world GD prevalence

Alternate Case Finding Approach

Focused on **surfacing cohorts warranting further investigation**

Informed by literature and multi-disciplinary team:

- Gaucher disease experts; Primary Care coding experts; Data scientists

Tested and refined in research data:

- De-identified, structured, primary care Electronic Health Record (EHR)
- 28 million, geographically distributed, UK
- Generally representative of UK Population

Target performance

> 3 in 4 of flagged cases 'valuable' to review

> 1 in 5 flagged cases appropriate for testing

> 3% Positive Predictive Value (prevalence-adjusted)

Seven distinct case-finding algorithms were created, each concerned with specific clinical presentation that is known to experience delay:

Algorithm	Reason
Splenomegaly	59% of presentations have splenomegaly
Splenectomy and subsequent bone pathology	13.3% of patients undergo splenectomy –> then have bone pathology
Fragility Fractures at a young age	Can be main feature in minority of patients
Concurrent Anaemia and High Ferritin	High Ferritin is hallmark of Gaucher, exploration of instances without alternative explanation
Immunoglobulin Abnormalities	Unexplained Occurrence with other features of GD
Pregnancy Presentations	GD can be incidentally identified during pregnancy
Paediatric Presentations	Give higher weighting to symptom clusters in younger patients

Table 2: The seven distinct case-finding algorithms approaches, with respective clinical rationale.

Algorithm Methodology

Overarching algorithm logic:

Clinical features constructed

➔ EHR flagged based on features

➔ Diagnostic exclusions applied

➔ Clinical features assigned 'points', with clinical weightings assigned

➔ Output EHRs get a total score, score thresholds can be considered

Benefits:

- Score thresholds manage flag rate and prioritise high-interest patients
- No machine learning 'black box', approach is interpretable and transparent
- Easily adaptable to different datasets and geographies

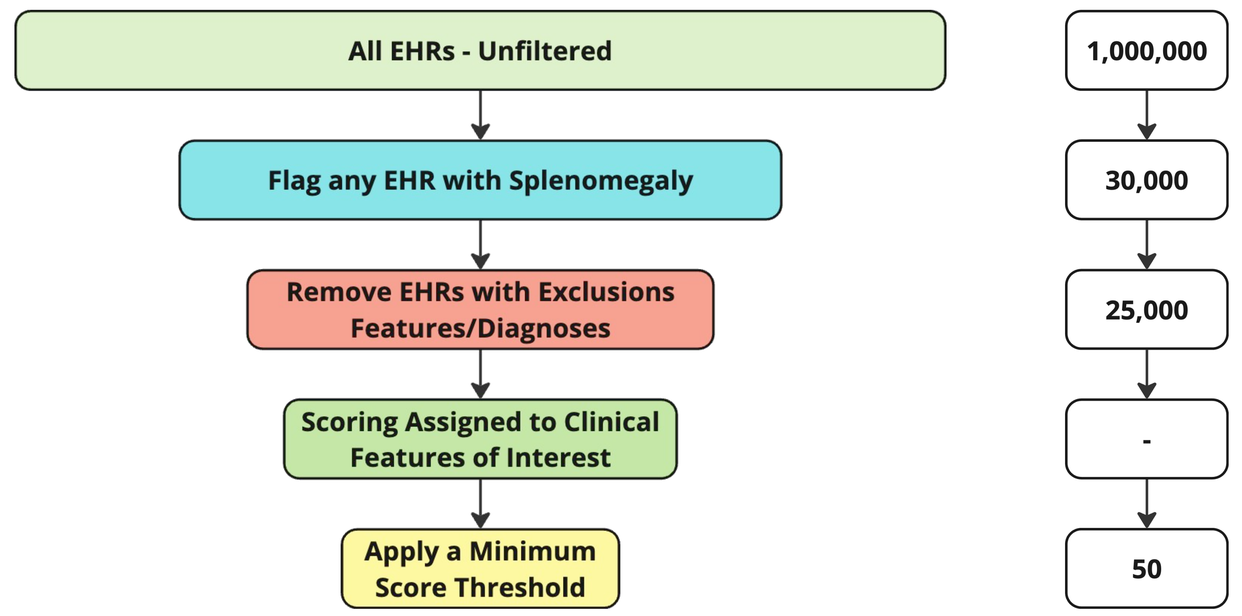
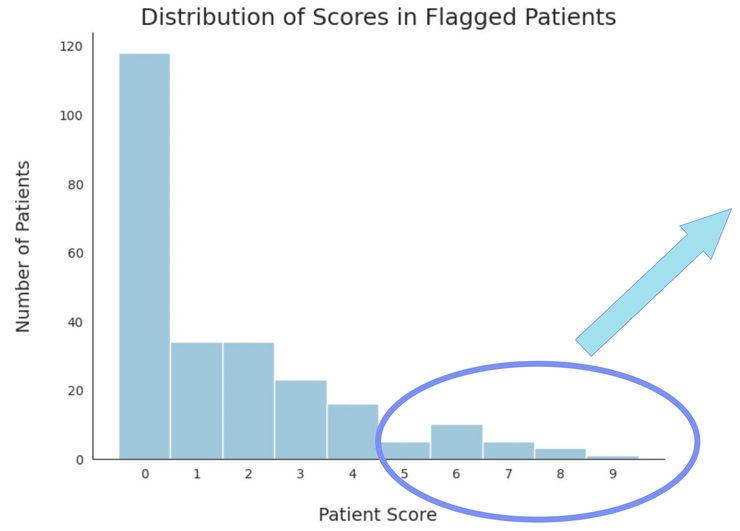


Figure 1: Example Splenomegaly Algorithm Logic with idealised EHRs numbers at each stage

Algorithm Development

Iterative approach with multiple improvement rounds:

- Developed using a test subset of 600,000 random UK primary care EHRs
- Flagging rate and clinical appropriateness of flagged cases were evaluated
- Learnings at each iteration hard-coded into subsequent algorithm versions
- Time constraints limited full development to 4 of the 7 presentations
- The final versions were tested on a second independent 600,000 EHR subset, showing consistent flagging rates and appropriate flagged cases



Score Threshold	No. of Flags	% Appropriate for full EHR review
>8	1	0.0%
>7	4	50.0%
>6	9	55.6%
>5	19	58.0%
>4	24	54.8%

Figure 2: Example output of the splenomegaly-focused algorithm applied to 620,759 EHRs, flagging 249. The histogram shows the distribution of scores, prioritising EHRs with the highest scores for review, as these are most likely associated with Gaucher Disease.

Ongoing limited UK Deployment Study

- The final 4 algorithms were deployed to live UK NHS primary care EHRs
- Deployment capacity was 30 full EHR reviews, and outcomes are pending
- **Key Results of 29 sent forward to date:**

  Identified as "Reasonable Diagnosis": 45% (13)

  1 previously tested for GD (-ve)

  5 referred for testing

  2 had potential GD highlighted to managing team

  5 outcome pending

Conclusions

- Prioritising clinical utility is **crucial** for real-world deployment
- **Rule-based approaches** are able to capture **diverse Gaucher disease presentations** in primary care
- Simple scoring allows flexible thresholds to **prioritise high-suspicion patients**
  - Clinically grounded
  - Interpretable, and transparent—avoiding black-box ML
  - Adaptable to various datasets and geographies
  - Early UK deployment shows promising results
- Next steps include further model refinement & validation in other datasets

References

[1] Mistry, P.K., *et al*. (2007), Consequences of diagnostic delays in type 1 Gaucher disease: The need for greater awareness among Hematologists–Oncologists and an opportunity for early diagnosis and intervention. Am. J. Hematol., 82: 697-701

[2] D'Amore et al; MRC GAUCHERITE Consortium. In-depth phenotyping for clinical stratification of Gaucher disease. Orphanet J Rare Dis. 2021 Oct 14;16(1):431.

[3] Mehta et al journey to diagnosis of Gaucher Mol Genet Metab 2017;122(3):122e9

[4] Mistry PK et al (2011) 86(1):110-5. doi: 10.1002/ajh.21888. PMID: 21080341; PMCID: PMC3058841

Funding:

Sanofi commissioned and paid for Mendelian to develop and integrate the algorithms presented with the MendelScan platform. Additionally, Patrick Deegan and Elizabeth Morris (Addenbrooke's Hospital) provided expert advice in relation to the project and were each paid an honorarium by Sanofi for their services.

Contact:

For any enquiries or further discussion please email corresponding author Dr Calum Grant calum@mendelian.co

Please scan the following QR code for a digital copy of the poster