Patient Identification for Rare Diseases With Recent ICD-10-CM Designation Using Claims Data

Queeny Ip and Pam Kumparatana — Komodo Health, New York, NY, and San Francisco, CA

ISPOR 2026
SA18

Scan here to download poster or inquire for more info.

Introduction

- Identifying sufficient rare disease patients in claims data is challenging, particularly for conditions with recently assigned ICD-10-CM codes.
- Previous studies uncovered several identification barriers for diseases with either no specific ICD-10-CM code or recently assigned ICD-10-CM code.
  - Code aggregation: multiple rare diseases sharing one "Other" code (Mazzucato et al, 2023)
  - Adoption lag: new specific codes are not used by practitioners for years (Bluematter 2025; Nelson et al, 2024)
  - Mapping errors: ICD-9-CM to ICD-10-CM-10 transitions lose rare disease specificity (Hsu et al, 2024)
  - Diagnostic practice: patients coded for symptoms before code introduction (Zhu et al, 2017)
- Using warm autoimmune hemolytic anemia (wAIHA) as a case study (ICD-10-CM code D59.11, introduced in October 2022), we evaluated the performance of a commonly used claims-based algorithm by calculating key diagnostic metrics: sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Period	ICD-10 -CM Code	Description
Before Oct. 1, 2022	D59.1	Other autoimmune hemolytic anemias (generalized)
After Oct. 1, 2022	D59.10	Autoimmune hemolytic anemia, unspecified
	D59.11	Warm autoimmune hemolytic anemia
	D59.12	Cold autoimmune hemolytic anemia
	D59.13	Mixed type autoimmune hemolytic anemia
	D59.19	Other autoimmune hemolytic anemia

Objective

- To evaluate the performance of a commonly used claims-based algorithm in identifying wAIHA patients

Methods

Data Source

- Komodo Research Dataset (KRD)

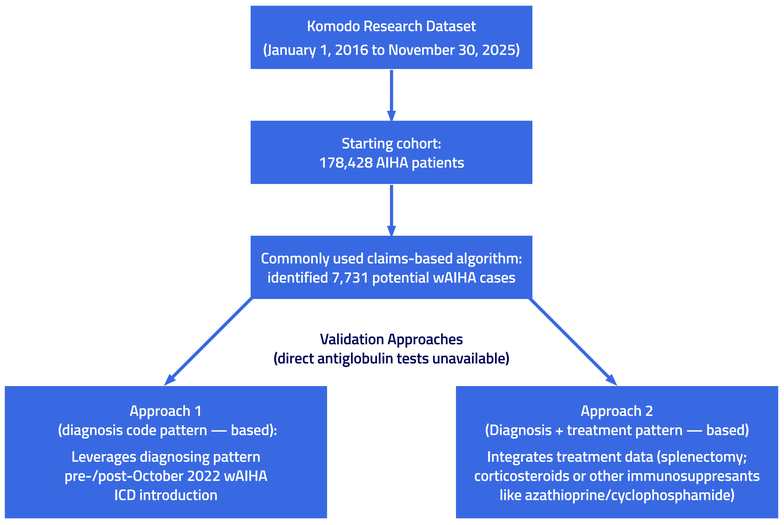
Komodo Research Dataset (KRD)

Composed of administrative data and claims, KRD captures routinely collected health services utilization records for over 330 million de-identified unique individuals in the United States. Native to HIPAA-compliant, privacy-preserving tokens, KRD offers extended patient-level observations of medical encounters and outpatient pharmacy dispensings via linkage across health and pharmacy insurance plans. Data availability is as early as 2016. Specialty datasets such as genomics, laboratory test results, and electronic medical records are readily accessible via additional linkage. KRD is the optimized schema of the underlying Healthcare Map® from Komodo Health® for real-world evidence generation.

Study Design

- Potential wAIHA cases were identified by implementing a commonly used algorithm: ≥2 diagnostic codes (i.e., ≥2 [D59.1 or D59.11] in this specific case) with at least 2 claims ≥30 days apart during the study period of January 1, 2016 to November 30, 2025.
- Without wAIHA confirmatory test information (direct antiglobulin test), true wAIHA cases were estimated using 2 claims-based approaches. Approach 1 (APP1) focuses on diagnosis patterns. Approach 2 (APP2) includes treatments specific for wAIHA.
- Performance metrics were derived from the resulting true/false positives and negatives.
- AI tools used were validated by the quality check in the research process.
  - Gemini 3 Flash was used as a writing assistant (accessed 04/01/2026)
  - Kpmodo's Marmot™ was used to assist in study design and perform preliminary analyses (accessed 12/10/2025)
- True positive qualification criteria
  - APP1:
    - ≥2 D59.11 (at least 30 days apart) post October 2022 + no (D59.1, D59.10, D59.12, D59.13, D59.19) post-October 2022
    - ≥2 D59.1 (at least 30 days apart) pre-October 2022 + ≥1 D59.11 and no (D59.1, D59.10, D59.12, D59.13, D59.19) post-October 2022
  - APP2:
    - ≥2 D59.11 (at least 30 days apart) post October 2022 and no (D59.1, D59.10, D59.12, D59.13, D59.19) post-October 2022
    - ≥2 D59.1 (at least 30 days apart) pre-October 2022 + ≥1 D59.11 and no (D59.1, D59.10, D59.12, D59.13, D59.19) post-October 2022
    - ≥1 D59.11 and no (D59.1, D59.10, D59.12, D59.13, D59.19) + corticosteroids or other immunosuppressants (lazathioprine or cyclophosphamide) or splenectomy post-October 2022
- True negative qualification criteria
  - APP1:
    - ≥2 (D59.12, D59.13) (at least 30 days apart) + no (D59.1, D59.10, D59.11, D59.19) post-October 2022
  - APP2:
    - ≥2 (D59.12, D59.13) (at least 30 days apart) + no (D59.1, D59.10, D59.11, D59.19) post-October 2022
    - ≥1 (D59.12, D59.13) + no (D59.1, D59.10, D59.11, D59.19) post-October 2022 + rituximab + no + corticosteroids or other immunosuppressants (lazathioprine or cyclophosphamide) or splenectomy
- False positive qualification criteria
  - APP1:
    - ≥1 D59.11 pre-October 2022 + no D59.11 post-October 2022
    - ≥2 D59.1 (at least 30 days apart) pre-October 2022 + ≥1 D59.11 + ≥1 (D59.1, D59.10, D59.12, D59.13, D59.19) post-October 2022
    - ≥2 D59.11(at least 30 days apart) + ≥1 no (D59.1, D59.10, D59.12, D59.13, D59.19) post-October 2022
  - APP2:
    - ≥1 D59.11 pre-October 2022 + no D59.11 post-October 2022
    - ≥2 D59.1 (at least 30 days apart) pre-October 2022 + ≥1 D59.11 + ≥1 (D59.1, D59.10, D59.12, D59.13, D59.19)
    - ≥2 D59.11 (at least 30 days apart) + ≥1 (D59.1, D59.10, D59.12, D59.13, D59.19) post-October 2022
    - ≥1 D59.11 + ≥1 (D59.1, D59.10, D59.12, D59.13, D59.19) + corticosteroids or other immunosuppressants (lazathioprine or cyclophosphamide) or splenectomy post-October 2022
- False negative qualification criteria
  - APP1:
    - ≥2 (D59.12, D59.13) (at least 30 days apart) + ≥1 (other (D59.1, D59.10, D59.11, D59.19) post-October 2022
  - APP2:
    - ≥2 (D59.12, D59.13) (at least 30 days apart) + ≥1 (D59.1, D59.10, D59.11, D59.19) post-October 2022
    - ≥1 (D59.12, D59.13) (at least 30 days apart) + ≥1 (D59.1, D59.10, D59.11, D59.19) + rituximab + no corticosteroids, splenectomy, or other immunosuppressants (azathioprine or cyclophosphamide)

Figure 1: Study Design



Key Terms

- New diagnosis code; algorithm; patient identification
- Sensitivity; specificity; positive predictive value; negative predictive value

Results

- Sensitivity: Critically low across both approaches (APP1: 1.72%; APP2: 1.77%)
- True positives: The algorithm identified only 2,858 true cases
- False negatives: Failed to capture the vast majority of cases (APP1: 158,531; APP2: 162,829)
- Specificity: Moderate performance, with APP2 showing a slight improvement (APP1: 61.75%; APP2: 71.40%)
- PPV: 36.97% for both approaches, meaning only about 37% of flagged patients were true cases
- NPV: Remained low (APP1: 4.75%; APP2: 7.54%)

SENSITIVITY
APP1: 1.72% | APP2: 1.77%
Critically Low: missed the vast majority of cases (false negatives: 158K-162K) Identified only 2,858 true positives.

POSITIVE PREDICTIVE VALUE (PPV)
APP1: 36.97% | APP2: 36.97%
Insight: only ~37% of flagged patients were likely true wAHIA cases.

SPECIFICITY
APP1: 61.75% | APP2: 71.40%
Insight: moderate performance in excluding true negatives.

NEGATIVE PREDICTIVE VALUE (NPV)
APP1: 4.75% | APP2: 7.54%
Insight: extremely low probability that a negative algorithm result indicates a true non-wAIHA patient.

Conclusion

- Requiring ≥ 2 D59.11 codes ≥ 30 days apart is insufficient as a stand-alone rule for identifying wAIHA patients.
- The low sensitivity suggests a mismatch with clinical coding practices, where many true cases are recorded with single or clustered codes.
- To achieve an acceptable capture rate, researchers may consider:
  - Developing temporal requirements or augmenting definitions with broader clinical context.
  - Developing and testing an identification algorithm using claims data only or claims plus lab data.
  - Developing a prediction model by training and testing machine learning algorithms using claims data only or claims plus lab data.

References

1. Angin C, Mazzucato M, Weber S, et al. Coding undiagnosed rare disease patients in health information systems: recommendations from the RD-CODE project. *Orphanet J Rare Dis.* 2024;19(1):28. doi:10.1186/s13023-024-03030-2
2. Bhat N. Case study: Unlocking the value of patient identification in rare disease markets. Blue Matter Consulting website. Published October 15, 2025. Accessed March 31, 2026. https://bluematterconsulting.com/insights/blog/ai-patient-identification-rare-diseases/
3. Cutillo CM, et al. Estimating rare disease prevalence and costs in the USA: a cohort study approach using the Healthcare Cost Institute claims data. *BMJ Public Health.* 2024;2(1):e000346. doi:10.1136/bmjph-2023-000346
4. Mazzucato M, Visonà Dalla Pozza L, Facchin P, et al. ORPHAcodes use for the coding of rare diseases: comparison of the accuracy and cross country comparability. *Orphanet J Rare Dis.* 2023;18(1):267. doi:10.1186/s13023-023-02864-6
5. Nelson SJ, Yin Y, Trujillo Rivera EA, et al. Are ICD codes reliable for observational studies? Assessing coding consistency for data quality. medRxiv. Published online January 17, 2024. doi:10.1101/2024.01.17.24301268
6. Roll M, Gladbach A, Selinger I, et al. Dilemma of missing specific disease codes: an approach to assess the incidence and prevalence of a rare nephrology disease. *BMC Nephrol.* 2025;26(1):346. doi:10.1186/s12882-025-04271-4
7. Zhu Z, Yan C, Sun J, Chen J. Identification of sufferers of rare diseases using medical claims data. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2017:372-377. doi:10.1109/BIBM.2017.8217641

komodo®