

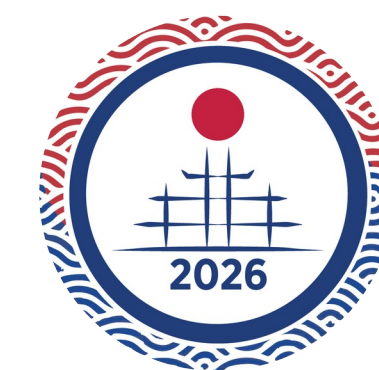


P3.380: Establishing a Taiwan Multi-Institutional Thymic Tumor Registry: Web-Based Platform Architecture and Data Quality Assessment

Ying-An Chou¹, Ping-Chung Tsai², Hsiao-Hung Lu³, Wei-Li Ma⁴, Yen-Shou Kuo⁵, Yi-Ting Yen^{6*}

¹Chi-Mei Medical Center, Tainan, Taiwan; ²Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan; ³Shin Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan; ⁴National Taiwan University Hospital, Taipei, Taiwan; ⁵Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan; ⁶National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

*Correspondence: Ying-An Chou, MD, agnh330284@gmail.com



IASLC 2026 World Conference on Lung Cancer

SEPTEMBER 12 - 15, 2026
SEOUL, REPUBLIC OF KOREA

INTRODUCTION

Thymic epithelial tumors (TETs) are rare mediastinal malignancies for which large-scale multi-institutional data are critical to inform staging, treatment strategy, and long-term outcomes.

No Taiwan-specific data are represented in any major international thymic tumor registry, despite Taiwan's high surgical case-ascertainment rate.

Taiwan's National Health Insurance cancer registry lacks WHO histological subtype, Masaoka-Koga/TNM staging, tumor size, and resection-margin data — precluding outcome research.

We describe the establishment of a Taiwan-wide integrated thymic tumor registry, emphasizing technical platform design, systematic data-quality monitoring, and preliminary survival analysis.

METHODS

1. Multi-Site Data Upload

Site-specific Excel / SPSS exports with heterogeneous variable names and coding conventions



2. AI-Assisted Standardization

LLM-guided mapping of site variables to a unified coding schema



3. Automated Quality Control

Per-site, per-variable missing-rate and logical-consistency checks (date errors, survival conflicts, staging conflicts)



4. Interactive QC Dashboard & Feedback

Real-time visual reports returned to site coordinators for correction before pooled analysis

Figure 0. AI-assisted data-integration and quality-control pipeline, from heterogeneous site exports to reconciled, pooled data.

As of submission, 8 institutions contributed surgically managed thymic tumor cases (1989–2025); a 9th site's cases are pending pathological confirmation.

A custom web-based data entry platform was developed using the FastAPI framework with AI-assisted programming, incorporating a 97-variable case report form (CRF) spanning pre-operative assessment (CT morphology, myasthenia gravis workup), intra-operative findings (surgical approach, resection status), and pathological/oncological outcomes (WHO classification, Masaoka-Koga and TNM 8th-edition staging, adjuvant therapy, overall survival, disease-free survival, recurrence) (Figure 1).

The platform enables real-time multi-institutional data synchronization with role-based institutional access control.

Data quality was systematically evaluated for per-variable completeness and logical consistency — temporal sequence violations, staging contradictions, survival-status conflicts, and numeric outliers (Figure 2; ongoing site-level QC monitoring in Figure 3).

Kaplan-Meier survival analysis with log-rank testing and post-hoc pairwise comparisons (Bonferroni correction) evaluated overall survival by Masaoka-Koga stage.

DISCUSSION AND CONCLUSION

Actionable Take Away Message

- A custom FastAPI-based platform with AI-assisted development enables real-time, multi-institutional thymic tumor data collection with built-in automated quality monitoring — a scalable model for rare-disease registries. Preliminary survival analysis confirms significant OS stratification by Masaoka-Koga stage.
- Given the demanding clinical workload of practicing thoracic surgeons, AI assistance with coding, debugging, and data-field mapping meaningfully lowered the extra non-clinical burden of building and maintaining this registry, making a physician-led, multi-institutional data infrastructure feasible without dedicated technical staff.

Future Direction for Research

- Expand enrollment (9th-site validation), link with Taiwan's National Death Registry for complete survival ascertainment, and validate AI-assisted QC against manual chart-review audits. This validated infrastructure supports subsequent prognostic modeling and cross-institutional outcome benchmarking.

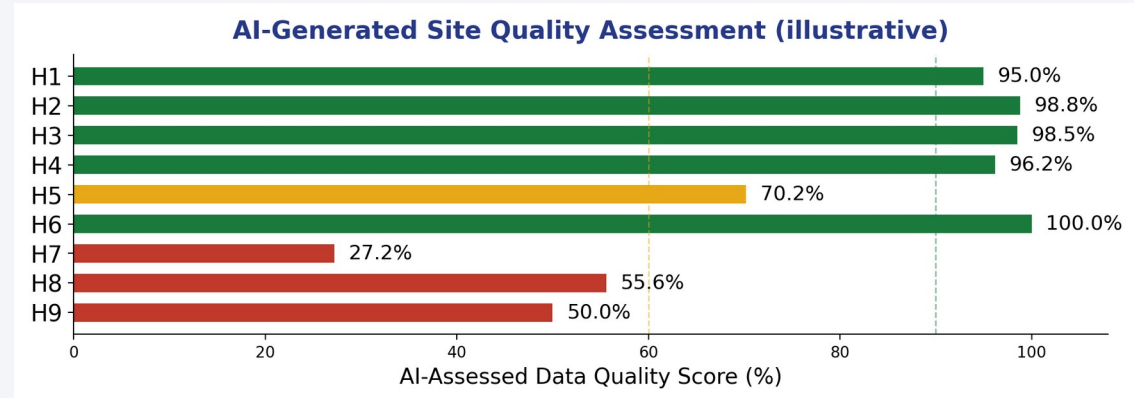


Figure 4. AI-generated site quality assessment (Methods, Step 4): automated scoring flagged H7-H9 for data-completeness follow-up and produced targeted QC recommendations for site coordinators. Illustrative; independent of the 229-issue count in Figure 2.

REFERENCES & ACKNOWLEDGEMENTS

Data standards: WHO Classification of Thoracic Tumours (5th ed.); Masaoka-Koga staging system; AJCC/UICC TNM staging (8th ed., thymic tumors).

- Acknowledgments: We thank the thoracic surgery teams at all nine participating centers for their contribution to this registry.

Artificial Intelligence (AI) Use Disclaimer

In accordance with the standards and guidance set forth by the Accreditation Council for Continuing Medical Education (ACCME), individuals in control of content are required to disclose the use of artificial intelligence (AI) in the development of educational materials. This includes identification of the specific AI tool(s) utilized and a description of how AI contributed to content creation as a slide reference.

IASLC maintains oversight of all educational content and conducts a thorough review process to ensure that materials are evidence-based, fair-balanced, and free from commercial bias, consistent with ACCME Standards for Integrity and Independence.

While AI may be used as a supportive tool in content development, it does not replace clinical expertise or independent medical judgment. Learners are advised to critically evaluate all information presented and apply their own professional judgment when incorporating knowledge into clinical practice.

*All individuals responsible for content must disclose within the slides if AI was used in the development of the material and specify which AI tool was utilized.
Specific disclosure: Claude (Anthropic) was used as an AI coding assistant to help develop the FastAPI-based web platform and to assist in LLM-guided mapping of site-specific data fields to the registry's standardized schema. AI was not used to analyze or interpret the clinical/statistical results on this poster; all analysis, interpretation, and conclusions were performed and verified by the study authors.

RESULTS

1,675

Validated Cases
(8 institutions)

97

CRF Variables

229

QC Issues Flagged
for Reconciliation

1,862

Total Cases
(All Sites)

9

Participating
Hospitals

729

Complete Cases

76.8%

Overall Quality
Score*

*Overall Quality Score = mean per-site completeness across core CRF fields (gender, age, operation date, staging), from the live platform dashboard; a system-level snapshot distinct from the 229 reconciled QC issues above.

8 institutions contributed cases (range 40–518 per site; median follow-up 57 months, range 0–363); a 9th site contributes 103 additional cases pending validation. Continuous site-level QC monitoring spans all 9 sites (Figure 3); the 229 issues above are those reconciled within the locked, validated 8-institution cohort (Figure 2).

Core pathological variables (Masaoka-Koga stage, WHO classification, R-status) achieved high completeness at most institutions; TNM 8th-edition completeness was lower and more variable (28.8% overall, 483/1,675), reflecting heterogeneous adoption of the newer system.

Kaplan-Meier analysis: overall survival differed significantly across Masaoka-Koga stages (log-rank $p < 0.001$); 5-year OS ranged from 92.0% (Stage I) to 39.8% (Stage IVb); median OS declined from 209 months (Stage I) to 44 months (Stage IVb).

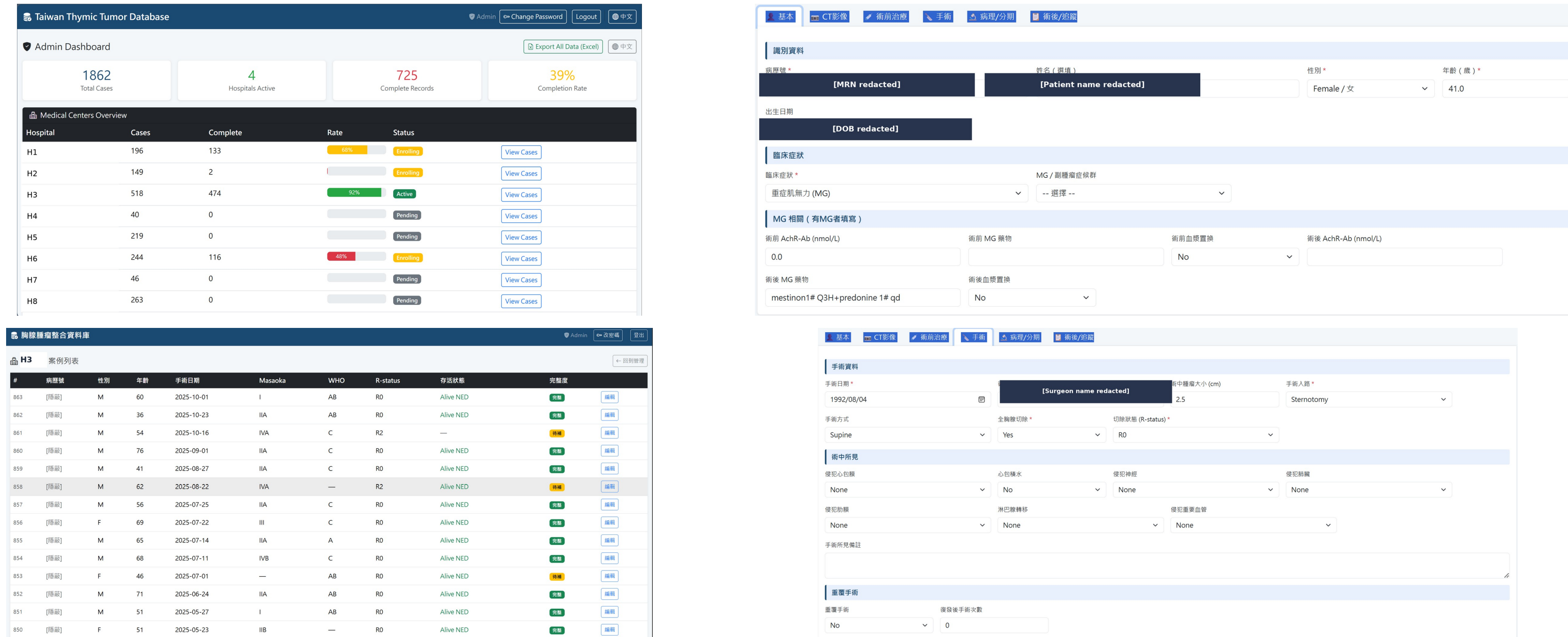


Figure 1. Web-based registry platform. (A) Real-time admin dashboard tracking enrollment across all 9 sites. (B) Structured case-entry form capturing the 97-variable CRF (myasthenia gravis workup shown). (C) Per-site case list (medical record numbers hidden). (D) Operative-detail entry (surgeon redacted). De-identified throughout: institution identities, patient identifiers, and clinician names redacted or coded (H1-H9).

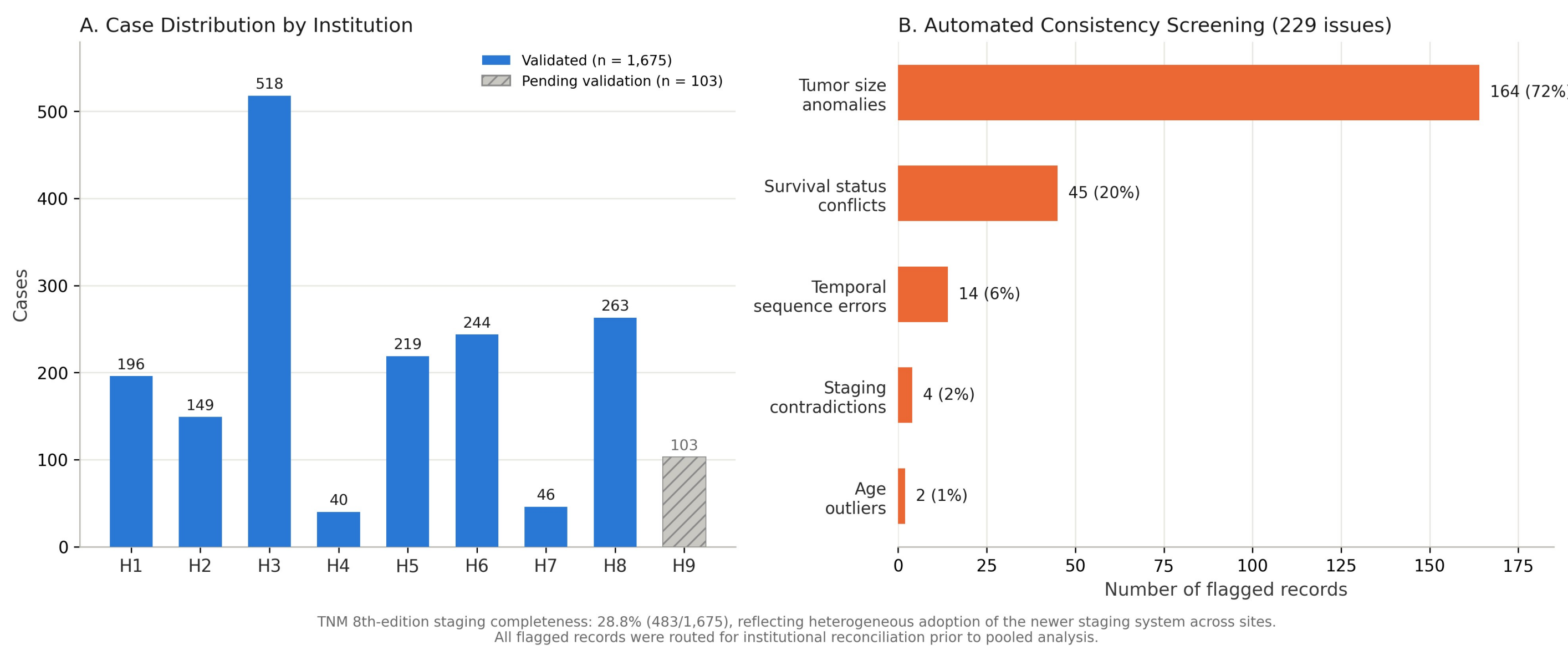


Figure 2. Case distribution and automated data-quality screening across the validated 8-institution cohort ($n = 1,675$).

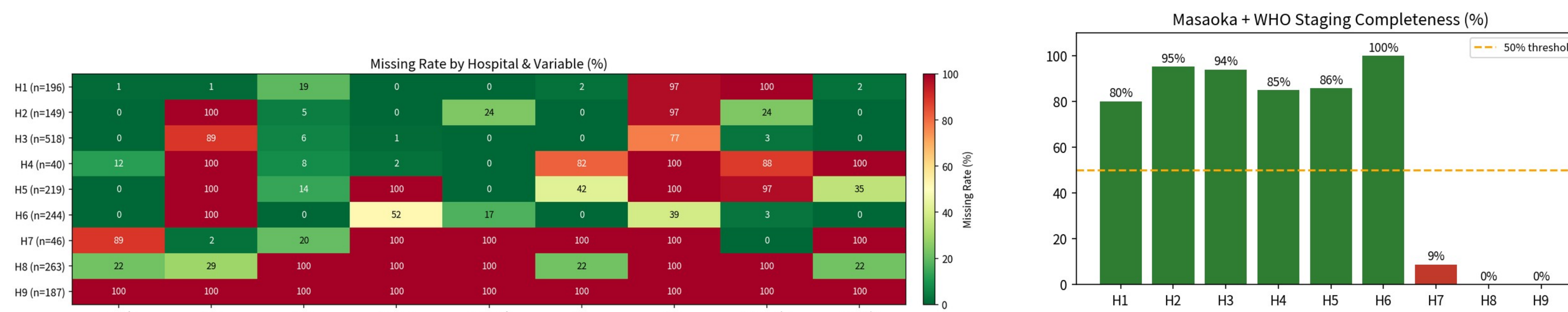


Figure 3. Site-level data-quality monitoring across all 9 sites: per-field missingness (left) and Masaoka-Koga + WHO staging completeness by site (right). Supplementary; independent of the 229 reconciled QC issues in Figure 2.