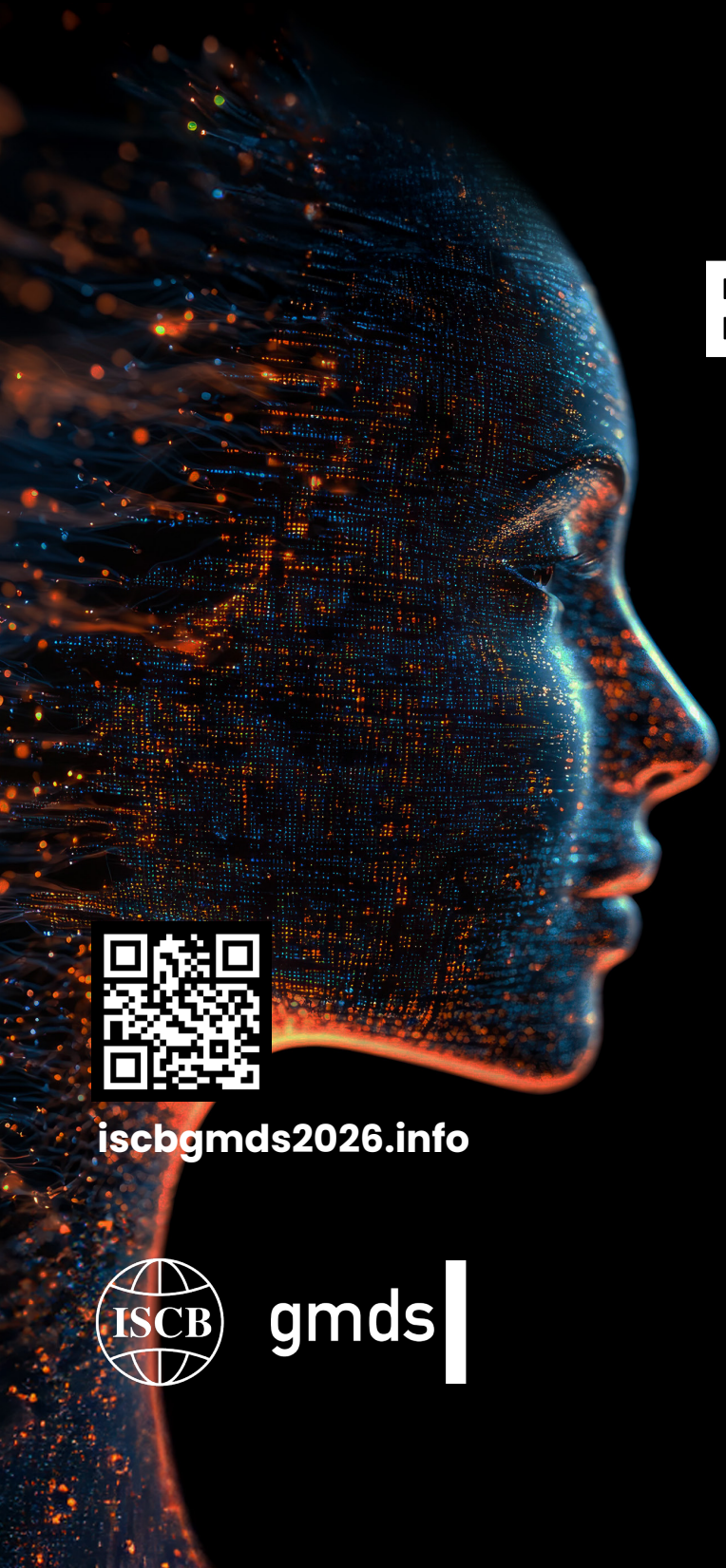
ISCB GMDS 2026

Conference Book



BIG DATA
small data
Your Data

Transitions in the age
of biomedical AI

iscbgmds2026.info

FREIBURG | GERMANY

27.09. - 01.10.26

ISCB gmds

Contents

Welcome Messages **8**
Mayor of the City of Freiburg - Martin W. W. Horn ... 8
Conference President, SPC Chair - Nadine Binder ... 9
Conference President, LOC Chair - Harald Binder ... 10

About the Conference **11**
Conference Overview ... 11
Contact ... 13
Our Sponsors ... 14
Committees ... 15
Scientific program committee ... 15
Local organizing committee ... 17
Awards ... 18
ISCB Awards ... 18
GMDS Awards ... 20
Best Poster Awards ... 22

General On-Site Information **23**
Venue ... 23
Venue Floor Plans ... 25
Venue Access and Logistics ... 28
Practical Information ... 30
Conference App ... 31

Social Program **33**
Overview ... 33
Conference Excursions ... 36
Excursion to Freiburg's Local Mountain: Schauinsland ... 36
Kaiserstuhl Experience - Wine Tradition at site ... 36
Schlossberg Walk - Hike from the Old Town in the Vines ... 37
City Tour with Cathedral Tower Ascent ... 37
"Schwarzwald-Mädel" City Tour ... 38
Freiburg - City Rally ... 38
Traditional City Tour - A Journey Through Time ... 39
Wine Tasting at "Alte Wache" in the Old Town ... 39

Culinary City Tour - Chocolate 40
Cocktail Course - Typical Black Forest Cocktails 41

Conference Schedule **42**

Pre-Conference Courses **48**
Sun 27.09.2026 48
Performing Comparative Effectiveness Research Using Network Meta-Analysis: An Introduction to Theoretical and Practical Advances 48
Beyond "Randomized": Designing and Evaluating Randomization in Clinical Trials 50
Efficient Bayesian Analysis of Longitudinal and Survival Data using R-INLA 52
Causal inference for time-to-event outcomes - with practical applications in R 54
Prediction Under Interventions 56
Clinical Text Mining under Real-World Constraints 58
Genomic AI Agents for Personal Genomics: Hands-on Analysis with Just-DNA-Lite and LLM agents 60

Opening Session - Welcome to ISCB GMDS 2026 **62**

Keynote Sessions **63**
Cathie Sudlow: Future of Population Health Research: a UK Health Data Perspective 63
Marylyn Ritchie: The End of Data-Driven Medicine: Entering the Era of AI-Native Science ... 64
Annette Kopp-Schneider: Leveraging External Data in Clinical Trials: Principles, Perils & Potential 66
Andre Dekker: AI for Better Health Care 68

Invited Sessions **70**
Mon 28.09.2026 70
11:00–12:30 70
Biomedical Knowledge Integration 70
13:30–15:00 73
Causal inference with truncation by death and missing data: from estimands to estimation 73
15:30–17:00 76
Innovations in Master Protocols Evaluating Multiple Treatments and (Rare) Diseases 76
Tue 29.09.2026 79
11:00–12:30 79
AI/ML-Driven Causal Inference with Real-World Data Complexities 79
Wed 30.09.2026 82
09:00–10:30 82
Accelerating evaluation and adoption of complex digital outcome measures in trials 82
11:00–12:30 85
Biostatistics for personalized medicine 85
14:00–15:30 88

Evaluating Fairness in Clinical Prediction Models and AI: Current and Emerging Topics . 88
Signature Session: AI and Me ... 92

Workshops **93**
Mon 28.09.2026 ... 93
11:00–12:30 ... 93
Designing Next-Generation Respiratory Virus Trials: Estimands, Core Outcomes, and Adaptive Pandemic-Ready Frameworks ... 93
From Research Question to Access: Using Nationwide Claims Data Today - and Electronic Health Record Data Tomorrow ... 97
Developing and Executing Privacy-Preserving Federated Analyses with FLAME ... 99
From Theory to Practice in the EU HTA JCA: Early Insights and Lessons Learned ... 103
13:30–15:00 ... 106
Analysis and Modelling of Infectious Diseases in the Age of AI: Lessons from COVID-19 106
Between Data, Practice, and Evidence: (Inter)Nationally EHR Research ... 111
Mapping metadata schemas to the HealthDCAT-AP standard: Lessons learned and hands on community examples (1/2) ... 113
A Guided Experiment on Clinical Trial Data Sharing Quality and Data Reuse (1/2) ... 117
15:30–17:00 ... 120
Uncertainty in Clinical Predictions ... 120
Large Language Model Informed Real-World Cohort Data ... 123
Mapping metadata schemas to the HealthDCAT-AP standard: Lessons learned and hands on community examples (2/2) ... 126
A Guided Experiment on Clinical Trial Data Sharing Quality and Data Reuse (2/2) ... 128
Tue 29.09.2026 ... 130
11:00–12:30 ... 130
Reliable Subgroup Identification and Analysis - Results of GMDS Biostatistics Competition 2026 ... 130
Introduction to Individual Participant Data Meta-Analysis ... 132
On the Ethical and Technical Aspects for Deployment and Evaluation of AI-Driven Clinical Decision Support Systems ... 133
13:30–15:00 ... 139
Sponsor Workshop: SAS Data Maker + Friedrich Wingert Stiftung ... 139
Wed 30.09.2026 ... 141
09:00–10:30 ... 141
AI-driven Therapeutic Modelling in Biomedical Informatics (1/2) ... 141
Benchmarking AI scientists ... 143
Forum Digital Medicine and Artificial Intelligence (1/4) ... 146
Regulatory Topics for Statisticians: Implementation of Risk-Based Quality Management from Biostatistical Perspective ... 148
Ordinal Outcomes: Friend or Foe? ... 151
11:00–12:30 ... 155
AI-driven Therapeutic Modelling in Biomedical Informatics (2/2) ... 155

Forum Digital Medicine and Artificial Intelligence (2/4) 158
Population-Adjusted Indirect Comparisons in the Framework of Health Technology Assessment 162
AI supported Data Retrieval and Data Management in the Health Domain 167
Use of Bayesian Methods for Clinical Trial Design and Analysis (1/2) 171
13:00–14:00 175
Forum Digital Medicine and Artificial Intelligence (4/4) 175
14:00–15:30 177
Forum Digital Medicine and Artificial Intelligence (3/4) 177
What Makes a Trial Complex? Evidence, Experience, and AI 180
Teaching Statistical Literacy in Times of Political Changes 183
Deep Dive Into eHealth With the Project MINT Meets Medicine 188
Use of Bayesian Methods for Clinical Trial Design and Analysis (2/2) 192

Oral Sessions **193**
Mon 28.09.2026 193
11:00–12:30 193
Frameworks for Clinical Research 193
Longitudinal and Joint Models of Clinical Trajectories 200
Design and Analysis Choices in Clinical Trials 207
Missing Data and Measurement Error 214
Digital Health Tools and Engagement 221
Semantic Health Data and Meta-Research 229
Survival Inference 236
13:30–15:00 242
LLMs for Clinical Workflows 242
Learning Latent Clinical Trajectories 250
Adaptive and Comparative Trial Design Strategies 258
Prediction and Imputation with Missing Data 264
Health Data Infrastructure and Secondary Data Use 271
Diagnostic Evidence Synthesis and Test Evaluation 278
Survival Prediction Models 285
15:30–17:00 293
Clinical Imaging 293
GMDS Awards 300
Adaptive Trial Inference and Error Control 308
Infectious Disease Modelling 315
Digital Health Education and Transformation 322
Network Meta-Analysis and Evidence Reliability 329
Dynamic Survival Modeling and Prediction 336
Tue 29.09.2026 343
11:00–12:30 343
LLMs for Clinical Text and Data 343
Clinical Trajectories and Events 351

Small Sample and Sequential Trial Methods ... 359
Robust Causal and Statistical Methods for Complex Data ... 366
Data Foundations for Clinical AI ... 374
Health Technology Assessment ... 382
Strategies for Prediction Models ... 389
13:30–15:00 ... 395
Data Quality, Selection, and Measurement ... 395
Statistical Learning under Uncertainty ... 403
Tailoring Statistical Models ... 410
Strategies for Evidence Synthesis ... 418
Adaptive Monitoring and Borrowing in Trials ... 425
Target Trial Emulation ... 432
Research-Ready Data Integration and Software Workflows ... 439
Prediction and Treatment ... 447
Treatment Effect Heterogeneity and Clinical Interpretation ... 454
Building and Updating Prediction Models ... 461
Wed 30.09.2026 ... 468
09:00–10:30 ... 468
Dose Finding and Adaptive Trial Design ... 468
Causal Inference with Survival and Competing Events ... 475
Interoperable Data and Digital Evaluation ... 483
Omics Discovery ... 491
Strategies in Clinical Trials ... 498
Recurrent Events Modelling ... 505
11:00–12:30 ... 512
Clinical Prediction ... 512
Basket Trials and Adaptive Borrowing ... 520
Causal Estimands and Biases ... 528
Federated Health Data Systems ... 535
Diagnostic Estimands and Modeling ... 543
Advances in Multi-State Modeling ... 550
14:00–15:30 ... 557
Platform and Cluster Trial Designs ... 557
Omics in Modeling ... 564
Causal Estimands and External Controls ... 572
Real-World Evidence, Registries, and Surveillance ... 579
Cancer Screening ... 587
Time-to-Event Methods and Meta-Analysis ... 595

Poster Sessions **603**
Mon 28.09.2026 ... 603
P1 - Digital Health and Data Systems ... 603
P2 - Prediction and Personalized Medicine ... 639
P3 - Clinical Trials ... 665

P4 - Causal Methods and Epidemiology 682
P5 - Time-to-event and Longitudinal Methods 699
P6 - Evidence and Decision Sciences 721
P7 - AI and Biomedical Data 739
Tue 29.09.2026 750
P1 - Digital Health and Data Systems 750
P2 - Prediction and Personalized Medicine 783
P3 - Clinical Trials 810
P4 - Causal Methods and Epidemiology 828
P5 - Time-to-event and Longitudinal Methods 847
P6 - Evidence and Decision Sciences 865
P7 - AI and Biomedical Data 882
Wed 30.09.2026 893
P1 - Digital Health and Data Systems 893
P2 - Prediction and Personalized Medicine 925
P3 - Clinical Trials 952
P4 - Causal Methods and Epidemiology 970
P5 - Time-to-event and Longitudinal Methods 988
P6 - Evidence and Decision Sciences 1009
P7 - AI and Biomedical Data 1027

Closing Session **1037**

Mini Symposia **1038**
SmallData Symposium 1038
SmallData Symposium Session 1: Similarity 1040
SmallData Symposium Session 2: Transfer 1042
SmallData Symposium Session 3: Uncertainty 1053
Poster sessions 1056
Poster Session: P1 - SmallData Similarity 1056
Poster Session: P2 - SmallData Transfer 1067
Poster Session: P3 - SmallData Uncertainty 1079
STRATOS Symposium 1087
STRATOS Symposium Session 1A: Methodological Aspects 1089
STRATOS Symposium Session 1B: Methodological Aspects 1091
STRATOS Symposium Session 2A: Practical Aspects of Medical Research 1094
STRATOS Symposium Session 2B: Practical Aspects of Medical Research 1096
Early Career Researcher Day 1100
Early Career Researcher Day Session 1 1102
Early Career Researcher Day Session 2 1103
Early Career Researcher Day Session 3 1104

SC/AG Meetings **1106**

Author Index 1111

Welcome Messages

Mayor of the City of Freiburg - Martin W. W. Horn

Dear conference participants,



Source: Stadt Freiburg

I am really delighted that the 47th Annual Conference of the International Society for Clinical Biostatistics together with the 71st Annual Conference of the German Association for Medical Informatics, Biometry and Epidemiology is being held in Freiburg. A warm welcome to all of you! Welcome to Freiburg!

This conference and the City of Freiburg: they're a perfect match. Firstly, because Freiburg is a leading centre for science and research. Freiburg is home to one of Germany's oldest and most prestigious universities, as well as several other institutions of higher education. With more than 15,000 employees, our university hospital is one of the largest in all of Germany. With five institutes and approximately 2,500 employees, Freiburg is the largest Fraunhofer location in Germany. Our city is a major hub for medical technology and pharmaceuticals - with nearly 40,000 people employed in this industry. And, Freiburg is one of the top 10 startup hubs in all of Germany - not least because startups here benefit from a very special environment of cutting-edge research and companies open to innovation.

And the second reason why this conference and the City of Freiburg are such a good fit: Because Freiburg is not only one of the sunniest cities in Germany, but also has a lot of sunshine in its heart. Our city stands for openness, diversity and hospitality.

Freiburg is a city of science and research, but it has much more to offer. I sincerely hope that you will find some time to discover our city in all its wonderful facets: It wasn't long ago that Freiburg even made it into Lonely Planet as a top travel destination. Why? Because of Freiburg's "enviable quality of life." Because bicycles are the number one mode of transportation in Freiburg. Because of our wonderful old town with its many cafes. And because we have the "Bachle" and the "Munster" Cathedral. Or in short: because Freiburg is one of the "most youthful, relaxed, and sustainable cities in the country".

Dear conference participants, I wish you many new insights and encounters, as well as unforgettable days in our city. Welcome to Freiburg!

Kind regards

Martin W. W. Horn

Mayor of the City of Freiburg

Conference President, SPC Chair - Nadine Binder



Dear colleagues,

A good scientific conference should do more than confirm what we already know. It should surprise us with an idea we had not considered, a question we had not thought to ask, or a conversation with someone we would otherwise never have met. With that hope, it is a great pleasure to welcome you to Freiburg for ISCB GMDS 2026.

In shaping the scientific program, our ambition was to create more than a meeting of the two scientific societies, ISCB and GMDS. We wanted to bring together communities that approach biomedical data from different perspectives, traditions and areas of expertise, because some of the most interesting things happen when different ways of thinking meet.

This feels particularly timely. Artificial intelligence is rapidly changing biomedical research, yet many fundamental questions remain: How do we learn reliably from complex or limited data? How do we quantify uncertainty? How do we design studies that answer the questions that matter? And how do we ensure that methodological innovation ultimately serves patients, researchers and society?

Our theme, "BIG DATA, small data, Your Data: Transitions in the age of biomedical AI," captures these contrasts: big data and small data, established methods and emerging ones, excitement about what is becoming possible and healthy skepticism about what it actually means. You will find these conversations throughout the program, in keynotes and invited sessions, contributed talks and posters, workshops, courses, mini-symposia and interactive formats.

A program like this is a collective achievement. I sincerely thank the Scientific Program Committee, the organizers of workshops and courses, and the editors and reviewers of the GMDS full papers. Most importantly, I thank everyone who submitted their research; your contributions are the foundation of this conference. My sincere thanks also go to the Local Organizing Committee and everyone working behind the scenes to make this meeting possible.

But the program in this book is only the framework. The conference itself begins when we start talking to one another. So be curious. Ask questions. Attend a session outside your usual field. Talk to someone you do not yet know. Share an unfinished idea. Challenge an assumption, including your own.

On behalf of the Scientific Program Committee, I wish you an inspiring conference, stimulating discussions, unexpected connections, and an enjoyable stay in Freiburg.

Welcome to ISCB GMDS 2026!

Nadine Binder

Chair, Scientific Program Committee

Conference President, LOC Chair - Harald Binder



Welcome to ISCB GMDS 2026!

Welcome to interdisciplinarity: to biostatistics, medical informatics, bioinformatics, epidemiology, and the many areas where these disciplines meet, overlap, and inspire each other.

Welcome to experiencing what our disciplines can contribute in a rapidly changing world, as biomedical AI, new forms of data, and new methodological possibilities challenge us to rethink established approaches. Welcome to Big Data, Small Data, Your Data, and to exploring the transitions that lie ahead.

Welcome to a showcase of the Collaborative Research Center Small Data, illustrating how interdisciplinary collaboration can make AI happen in setting that seemed too challenging. And welcome to Freiburg, a city shaped by science and scholarship and home to one of Germany's oldest universities.

Most importantly, welcome to discussions with peers: to exchanging ideas, questioning assumptions, making unexpected connections, and sparking change that matters. Welcome to five intense days of talks, posters, workshops, courses, encounters, and conversations. Welcome to food for thought, and to thoughts that may travel well beyond these five days.

This conference program and abstract book of offers a first glimpse of what awaits you. I hope that discovering its contents will raise your curiosity, lead you towards topics beyond your usual field, and provide starting points for many fruitful conversations.

Welcome also to the shaping of what will hopefully become some great memories.

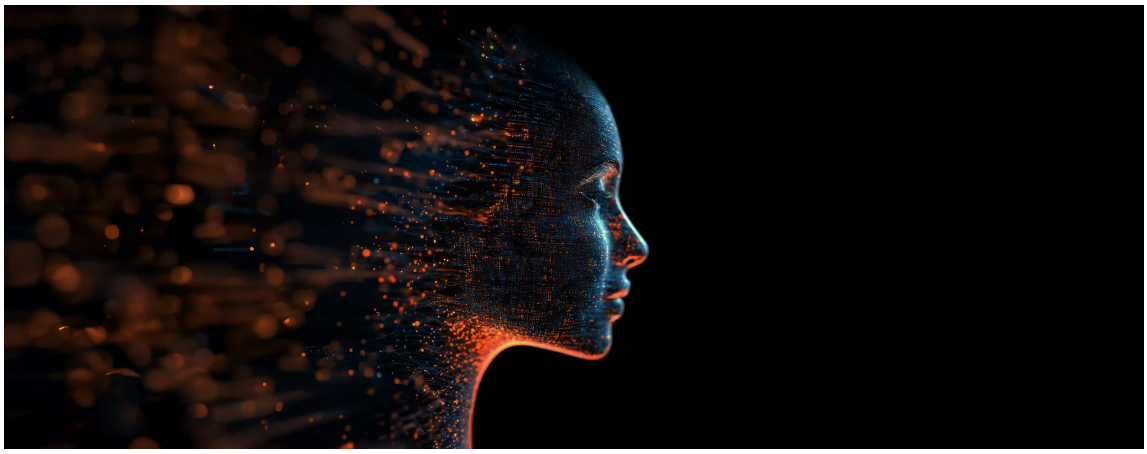
Welcome to the future.

Harald Binder

Chair, Local Organizing Committee

About the Conference

Conference Overview



ISCB GMDS 2026

ISCB

GMDS

Come with us and change the world

We are looking forward to jointly hosting the 47th Annual Conference of the International Society for Clinical Biostatistics (ISCB) together with the 71st Annual Conference of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) from September 27 to October 1, 2026 in Freiburg im Breisgau, Germany.

The rise of AI techniques and the increased public awareness of the value of data present great opportunities to data science disciplines. In particular, medical data sciences, as represented by the GMDS, and application fields such as novel treatment paradigms in medicine, which are hotly debated at ISCB, promise great impact of methodological developments, including transitions in research paradigms. This can optimally be achieved jointly with other disciplines, in particular mathematics, computer science, and also close collaboration with biomedical researchers, as exemplary reflected in the Freiburg CRC Small Data.

The ISCB GMDS 2026 Conference, together with the CRC Small Data Symposium, will provide a scientific platform for exchange on these exciting novel developments and discussing how our disciplines master the corresponding transitions.

Conference presidents



Harald Binder
LOC Chair

"I am grateful for the opportunity to host the conference as an interdisciplinary methods platform that will propel advancements in modeling and AI."



Nadine Binder
SPC Chair

"It is exciting to design a highly interdisciplinary program, reflecting the joint contributions of all methods disciplines, and I am looking forward to many contributions from the scientific communities."

About Freiburg

Freiburg is a scenic, vibrant university city and home to one of the oldest universities in Germany. Founded in 1120, picturesque Freiburg is located in southwest Germany's black forest. Locally rooted yet well connected at the European and global levels. Freiburg enjoys an excellent reputation as a science location with a focus on medicine, sustainability and innovation as well as education, science and research. The conference venue is situated in the heart of the city. Freiburg's Konzerthaus can look back on a history of over 10.000 hosted events and enjoys a high reputation far beyond the region.

Contact

1. Emergency & Medical Help

112: Emergency - ambulance / fire brigade

110: Police

116 117: Medical on-call service for urgent, non-life-threatening medical problems

0761 19240: Poison Control Centre Freiburg

0800 00 22 8 33: Emergency pharmacy service

116 116: Central service for blocking cards and some electronic banking services

2. Essential Local Information

International country code: +49

Freiburg area code: 0761; from abroad: +49 761

(When calling Germany from abroad, omit the initial 0 of the area/mobile code.)

3. ISCB GMDS 2026 Contacts

Conference Office & Organiser

Registration & logistics: iscbgmds2026@intercongress.de

Scientific content & formats: bemb.iscbgmds2026@uniklinik-freiburg.de

Submission-system technical support: helpdesk@webtek.at

During the Conference

Info Desk: General enquiries, lost & found and on-site guidance

Conference App: Personal schedule, programme updates and notifications

On-site emergency: Call 112 first in a life-threatening emergency; inform the conference team when safe

Our Sponsors

We sincerely thank all sponsors and partners for their valuable support of the ISCB GMDS 2026 conference.

Institutional & Platinum Sponsors







Silver Sponsors









Bronze Sponsors & Media Partners

















Committees

Scientific program committee


Nadine Binder (SPC Chair)


Harald Binder


Adriani Nikolakopoulou


Andre Karch


Andre Scherag


Antonia Zapf


Cecile Proust-Lima


Dagmar Waltemath


Diane Uschner


Dominic Magirr


Fabian Prasser


Frank Ueckert


James Carpenter


Jan Gorodkin


Klaus Jung


Laure Wynants


Liangyuan Hu


Margarita Moreno-Betancur


Maja Pohar Perme


Maria Grazia Valsecchi


Maria Kalweit


Martin Posch


Nancy Leung


Nicole Radde


Spiros Denaxas


Suzanne Cadarette


Stefan Michiels


Stijn Vansteelandt


Ulrich Sax


Valeria Vitelli

Local organizing committee


Harald Binder (LOC Chair)


Nadine Binder


Sania Simran (Conference Manager)


Marie Weigold


Lea Grober


Olivia Kuhnrich


Claudia Giuliani


Derek Hazard


Georg Koch


Jean-Pierre Gnimatin


Jorg Sahlmann


Ksenija Fahrlander


Lukas Maximilian Horstmeier


Max Behrens


Moritz Hess


Olga Sieber


Sophia Nolde

13:30–15:00

LLMs for Clinical Workflows

Mon, 28.09.26, 13:30–15:00, Room: K2

Chair: Fabian Kabus

Rule-Based Prompting for Extraction of TNM Stage and Tumor Grade from Population-Level Lithuanian Pathology Reports Using a Medical Large Language Model

13:30-13:45

Presenter: Tadas Žvirblis

Authors: Tadas Žvirblis [1]; Tomas Griguola [2]; Justas Mileika [1]; Paulius Milmantas [1, 3]; Miglė Gervytė [1, 4]; Agnė Irkinaitė [1]; Valdas Pečeliūnas [1]

Affiliations: [1] Cancer Registry, Laboratory of Cancer Epidemiology, National Cancer Institute, Vilnius; [2] National Cancer Institute, Vilnius; [3] Institute of Informatics, Faculty of Mathematics and Informatics, Vilnius University, Vilnius; [4] Institute of Data Science and Digital Technologies, Faculty of Mathematics and Informatics, Vilnius University, Vilnius

Introduction

Pathology reports contain key oncological variables, including tumor grade and TNM stage, but these are typically embedded in unstructured free text. This limits secondary use for cancer surveillance, outcomes research, and clinical data integration, particularly in low-resource languages such as Lithuanian. Large language models (LLMs) may provide a scalable approach for structured information extraction from routine clinical text.

Methods

We evaluated MedGemma-27B for extraction of tumor grade (G) and clinical (c) / pathological (p) TNM variables (cT, cN, cM, pT, pN, pM) from Lithuanian pathology and related clinical reports collected since 2020 according to TNM 8th edition definitions. The dataset represented routine population-level real-world data, with no preselection by tumor site, diagnosis, or other clinical criteria. Two prompting strategies were compared: basic JSON prompt and iteratively refined rule-based prompt with explicit instructions for distinguishing clinical from pathological staging and assigning null values when evidence was absent or ambiguous. Prompts were tested in Lithuanian and English. Performance was evaluated on 100,000 manually annotated reports using accuracy and extraction coverage.

Results

Rule-based prompting improved extraction accuracy for most variables, including tumor grade (G: 0.83 vs 0.70) and clinical staging components (cN: 0.43 vs 0.32 cM: 0.49 vs 0.42). Accuracy also improved for pathological variables (pT: 0.79 vs 0.73 pN: 0.70 vs 0.66). Accuracy for pM declined under structured prompting (0.44 vs 0.55). Clinical T-stage extraction remained challenging under both prompting strategies (cT: 0.38 vs 0.27). Prompt language had no meaningful effect on performance. However, coverage varied across variables, with increases observed for some (e.g., G, pT) and decreases for others, particularly for rare variables such as pM.

Discussion

In heterogeneous, non-selected real-world pathology data, prompt design materially affected both accuracy and completeness of LLM-based extraction. Rule-based prompting improved accuracy across most variables without model retraining, supporting the feasibility of structured oncology data extraction in a low-resource language setting. However, these improvements were accompanied by variable changes in extraction coverage, with reduced completeness observed for some variables. Persistently low accuracy for cT and reduced performance for rare variables such as pM highlight limitations of prompt-based approaches. Task-specific validation remains essential before downstream clinical or epidemiological use.

LLM-Based Semi-Automated Information Extraction from Histopathological Reports in T-Cell Lymphoma Patients

13:45-14:00

Presenter: Christoph Demus

Authors: Christoph Demus [1]; Filip Kramarzyk [2]; Annemarie Weise [1]; Christian Jäger [1]; Thomas Weber [2]; Jan Christoph [1]

Affiliations: [1] Junior Research Group (Bio-)Medical Data Science, Faculty of Medicine, Martin-Luther-University Halle-Wittenberg, Halle (Saale); [2] Department of Internal Medicine IV, Hematology and Oncology, University Hospital Halle (Saale), Halle (Saale)

Introduction

Clinical registries are essential for research on rare and heterogeneous diseases such as T-cell lymphomas. However, manual extraction of relevant information from unstructured clinical documents remains time-consuming and limits data completeness and quality. Large language models (LLMs) offer the opportunity to support information extraction from clinical routine data. The DQ-LIR project aims to improve data density and quality in entity-specific clinical registries by implementing an LLM-based pipeline for semi-automated extraction of structured information from histopathology reports.

Methods

We developed a locally hosted LLM-based processing pipeline to extract predefined clinical variables from pseudonymized histopathology reports of patients with T-cell lymphomas. The training dataset contained 218 reports from 68 patients recorded in the University Hospital Halle. After text extraction and preprocessing of PDF reports, relevant variables (including diagnosis, tissue type, ICD-10 code, and multiple immunohistochemical markers) were automatically extracted using LLMs with specific prompts for each marker. The system combines prompt-based LLM extraction with rule-based preprocessing, including regular-expression filtering of relevant text passages and synonym mappings. Performance was evaluated against a manually annotated gold standard created by medical experts. To assess the reliability of the annotators, the inter-annotator agreement was also determined. Three open-source LLMs were compared, and iterative optimization was performed through repeated evaluation cycles.

Results

Among the tested models, Llama-3.1-8B-Instruct achieved the best overall performance with a mean accuracy of 0.90. Extraction accuracy varied across variables: several immunohistochemical markers achieved accuracies above 0.95, while more complex tasks such as diagnosis classification showed lower performance (accuracy of 0.84). The integration of rule-based preprocessing significantly improved results, primarily by reducing hallucinations when information was absent in the reports. Extracted data were exported into registry-compatible CSV files for semi-automated integration into the clinical registry.

Discussion

Our results demonstrate that LLM-based approaches can substantially support the extraction of structured registry data from unstructured pathology reports. A hybrid architecture combining LLMs with rule-based preprocessing proved particularly effective in this specialized medical domain. The locally deployed system enables privacy preserving use in a clinical environment and has the potential to reduce manual workload while improving registry data completeness.

From Free-Text to ICD-10 Codes: Comparison of Embedding Models and Retrieval-Augmented Generation for Automated Coding of German Tumor Diagnoses

14:00-14:15

Presenter: Fatma Alickovic

Authors: Fatma Alickovic [1]; Stefan Lenz [1]; Arsenij Ustjanzew [1]; Lakisha Ortiz Rosario [1]; Georg Vollmar [2]; Thomas Kindler [2]; Torsten Panholzer [1]

Affiliations: [1] Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the Johannes Gutenberg University Mainz, Mainz; [2] University Cancer Center of the Johannes Gutenberg University Mainz, Mainz

Introduction

Accurate ICD-10 coding of free-text tumor diagnoses is essential for cancer registries and research. However, the heterogeneity of diagnostic text makes automated coding difficult. While retrieval-augmented generation (RAG) [1] and embedding models [2] have been explored for this task, the key objective in this study is to assess the added value of LLMs and RAG over purely embedding-based approaches.

Methods

Our evaluation is based on a real-world test dataset from the local tumor documentation system that contains short diagnosis descriptions and corresponding codes (2,024 entries 238 distinct codes). The Alpha-ID (12,021 entries 776 codes) and ICD-10-GM classification (859 entries 859 codes) served as a knowledge base. Nine different embedding models were used to predict codes via semantic search and alternatively via an added neural network classifier trained on the knowledge base embeddings. For RAG, the top-3 retrieved entries were incorporated into prompts for Llama-3.1-8B and Llama-3.3-70B to generate ICD-10 codes. The performance was evaluated using accuracy of exact-match (full codes) and partial-match (three-characters only) of predicted codes in the test dataset. A data quality analysis indicated a maximum of approximately 70% and 89% for attainable exact and partial-match accuracy.

Results

Among the embedding models, Qwen3-Embedding-8B achieved the best performance in both semantic search (44.3% exact-match, 67.7% partial-match accuracy) and combined with the classification network (47.8% exact-match, 72.1% partial-match accuracy). RAG substantially improved the performance of the base LLMs. RAG with Llama-3.3-70B achieved the highest partial-match accuracy (81%) but did not surpass Qwen3-Embedding-8B in exact-match performance, while Llama-3.1-8B did not outperform embedding approaches in either metric.

Discussion

Our results suggest that while RAG can enhance the coding performance of LLMs without requiring additional model training, LLMs with and without RAG are not necessarily better than pure embedding-based approaches. Fine-tuning embedding models could be a promising direction for future work.

[1] Krumscheid et al. (2025), Retrieval-Augmented Generation for ICD-10 Coding in German Clinical Texts - A Technical Case Report. https://doi.org/10.3233/SHTI251397

[2] Kreuzthaler et al. (2025), Secondary Use of Clinical Problem List Descriptions for Bi-Encoder Based ICD-10 Classification. https://pmc.ncbi.nlm.nih.gov/articles/PMC12099355/

Benchmarking AI Vibe Coding for Clinical Statistical Analysis: A Structured Evaluation in Pulmonary Hypertension Research

14:15-14:30

Presenter: Jeeva Sam

Authors: Jeeva Sam [1, 2]; Til Karsten Spreuer [1, 2]; Max Breger [1, 2]; Athiththan yogeswaran [1, 2]; Khodr Tello [1, 2]; Patrick Janetzko [1, 2]; Werner Seeger [1, 2]; Jonas Bienzeisler [3]; Raphael W. Majeed [1, 2, 3]

Affiliations: [1] Department of Internal Medicine, Universities of Giessen and Marburg Lung Center (UGMLC), Member of the German Center for Lung Research (DZL), Gießen; [2] Cardio-Pulmonary Institute (CPI), Institute for Lung Health (ILH), Gießen; [3] Medizinische Faculty, RWTH University, Aachen, Aachen

Introduction

Statistical analysis of clinical data requires expertise in medical statistics. Large language models (LLMs) are increasingly used for code generation and may support both descriptive and advanced analyses, but their reliability remains uncertain. Errors in grouping, missing-value handling, or model specification may produce outputs that appear plausible while remaining analytically incorrect. This study benchmarked the reliability of current LLMs for predefined clinical statistical analyses against a verified expert reference.

Methods

Five LLMs (GPT 5.3, Claude Sonnet 4.6, Gemini 2.5 Flash, Perplexity, and Grok) were evaluated across three statistical tasks derived from a pulmonary hypertension registry workflow: descriptive table generation, Kaplan-Meier survival analysis, and Cox proportional hazards modelling. All models received identical datasets and standardised prompts. Outputs were compared with analyses performed by two experts trained in mathematical statistics and assessed for grouping correctness, numerical accuracy, missing-value handling, completeness of requested output, and quality of generated R code. Benchmark datasets were derived from the PVRI GoDeep meta-registry and reflected the structure of a published pulmonary hypertension analysis workflow.

Results

All five models produced correct descriptive statistics once dataset variables were specified explicitly. In the initial descriptive benchmark, two models failed because of variable-name ambiguity, but both recovered after prompt clarification. All models reproduced the correct Kaplan-Meier log-rank p-value, although figure completeness differed. In the Cox benchmark, four models reproduced all required hazard-ratio terms, while one omitted the interaction terms required for the full analysis. Even when statistical outputs were correct, substantial differences remained in code quality, formatting, and direct usability.

Discussion

Current LLMs can support routine clinical statistical tasks, but their reliability depends strongly on prompt precision and careful verification of generated code and results. Numerical correctness alone is not sufficient to assess practical usefulness, as code quality and completeness of output varied substantially between models. LLMs may therefore be useful as assistive coding tools in clinical research workflows, but they should not be regarded as autonomous analytical systems. Independent validation remains necessary before generated outputs are used in research practice.

The full paper of this abstract will be published in the GMDS series "Studies in Health Technology and Informatics" with doi: 10.3233/SHTI261007, URL:https://dx.doi.org/10.3233/SHTI261007

LLM-Assisted Generation of Off-Label Medication Requests in Oncology: A QUEST-Based Evaluation Framework

14:30-14:45

Presenter: Valeriya Vishnevskaya

Authors: Valeriya Vishnevskaya [1]; Georg Mathes [1]; Stephanie Berger [2]; Stefan Sigle [1]

Affiliations: [1] MOLIT Institute gGmbH, Heilbronn; [2] SLK Kliniken Heilbronn, Heilbronn

Introduction

Off-label medication requests in precision oncology require extensive documentation of a patient´s treatment history, molecular diagnostics and exhaustion of guideline-based therapies. Generative approaches like Large language models (LLMs) bear potential supporting this workflow but their evaluation remains challenging due to scarcity of domain experts and the nature of generative approaches [1].

Methods

In literature the QUEST framework has been identified as a suitable framework evaluating principles like Quality, Understanding/Reasoning, Expression, Safety and Trust of LLMs in healthcare [2]. Based on QUEST we propose a layered evaluation scheme: Layer 1 (a), applicable without clinical expertise, scores on four dimensions rated on a 1-5 Likert scale: factual faithfulness to input data, patient case completeness, appropriate wording, and clinical readiness. A binary red-flag checklist covers hallucinations, incorrect cancer staging, omitted molecular findings or treatment history, and unsupported guideline citations. Layer 2 (b) extends this process by open-ended questions for domain experts, e.g. did the given output correctly identify which standard options had been used? A joint evaluation of (a) and (b) should give an indication on how a LLM based system for Off-Label requests performs.

Results

Preliminary testing of the proposed methodology identified three failure categories: (1) inference beyond source data: e.g. interpreting lung biopsy as evidence of operability in stage IV patients (2) mismatch between generated medication recommendations and regulatory approval status and (3) hallucination combined with unjustified certainty under data incompleteness.

Discussion

Primary risk factors are hallucination combined with unjustified arguments for personalized off label generation. The proposed framework aims to provide pragmatic evaluation focused on safety and usability. It should uncover safety-relevant failure modes prior to deployment and offers a transferable model for LLM evaluation in clinical documentation.

[1] J. Zhao, H. Liu, Y. Chen, and F. Song, "Application of artificial intelligence tools and clinical documentation burden: a systematic review and meta-analysis," BMC Med. Inform. Decis. Mak., vol. 26, no. 1, p. 29, Dec. 2025, doi: 10.1186/s12911-025-03324-w.

[2] T. Y. C. Tam et al., "A framework for human evaluation of large language models in healthcare derived from literature review," Npj Digit. Med., vol. 7, no. 1, p. 258, Sep. 2024, doi: 10.1038/s41746-024-01258-7.

IRIS: A Multi-Agent Large Language Model Framework for Conversational Clinical Trial Navigation and Protocol Design

14:45-15:00

Presenter: Paolo Messina

Authors: Amirmahdi Abootalebi; Paolo Messina; Mario Torchia; Luca Emili

Affiliations: InSilicoTrials Technologies Spa, Trieste

Introduction

Recent advances in large language models (LLMs) have demonstrated strong capabilities in biomedical natural language understanding, structured reasoning, and multi-turn dialogue, opening new possibilities for intelligent, context-aware clinical research support. However, the integration of LLMs into end-to-end clinical trial workflows remains an open area of investigation.

Objectives

We present the clinical trial reasoning component of IRIS, a multi-agent AI assistant designed to enable conversational, context-aware interaction with clinical trial registries. The primary objective is to demonstrate how an orchestrated pipeline of LLM-based agents can bridge the gap between unstructured user intent and structured clinical trial knowledge, supporting tasks such as eligibility criteria design, trial landscape exploration, and protocol comparison through natural language dialogue.

Methods

IRIS employs a pipeline of specialized, cooperating LLM-based agents. A query interpretation agent converts natural language prompts into structured search parameters, which a retrieval agent uses to access the ClinicalTrials.gov API. To ensure only the most semantically relevant records are passed downstream, a ColBERT-based reranking model scores and re-orders the retrieved candidates. A reasoning agent processes this set using a pre-trained LLM to synthesize design recommendations. Finally, a conversational agent maintains session context across multi-turn exchanges, enabling users to iteratively refine their queries, request comparisons, or explore alternative trial designs. In this work we present some results from the reranking model.

Results

Among the 10 documents retrieved per query, we evaluated the top 5 reranked, ColBERT outperformed the BGE (BAAI General Embedding) baseline across both metrics (mean Normalized Discounted Cumulative Gain @5: 0.970 vs. 0.927 Recall@5: 90% vs. 80%). The performance gap was most pronounced on semantically complex queries, where ColBERT scored 0.954 versus BGE's 0.866, confirming its advantage precisely in the contextually rich query types characteristic of clinical trial design tasks.

Discussion

IRIS demonstrates that a multi-agent LLM architecture, grounded in real-time clinical trial registry data, can meaningfully augment the early phases of trial design and literature navigation. This approach accelerates protocol ideation by surfacing design precedent through natural language. Future work will extend IRIS to additional registry sources, integrate structured outcome data for comparative effectiveness queries, and evaluate system-generated criteria against expert-reviewed gold standards.