# DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

**Abstract book**

DZD
German Center for Diabetes Research
Health Data Science & AI Certificate Program

# Poster Abstracts

| | | |
|---|---|---|
| **Poster session I**<br>April 28, 15:30-17:00 \| Helmholtz Pioneer Campus, 3rd Floor Creativity Walk | | |
| 1 | Ivona Anastasova | Unlocking the potential of DZD clinical study data for secondary research |
| 2 | Angela Dedie | Why we need a standardized core data set for diabetes research and beyond |
| 3 | Raeesa Yousaf | An AI System for Robust Plain Language Translation of Clinical Trials |
| 4 | Nuha Shugaa Addin | Personalized Prevention of Cardiometabolic Diseases: Proteomic Insights from the KORA Study |
| 5 | Nicolas Wilms | Metabolic Adjustment and Latest Developments in Therapy for Children and Adolescents with Type 1 Diabetes – 30 Years of Real-World-Data from the Multinational DPV Registry |
| **Poster session II**<br>April 29, 10:45-12:15 \| Helmholtz Pioneer Campus, 3rd Floor Creativity Walk | | |
| 6 | Christine El-Khoury | Genetic and Epigenetic Factors Modifying the Association between Diet Quality and Incident Type 2 Diabetes: The EPIC-Potsdam Cohort |
| 7 | Michael Stein | High-Resolution Accelerometry to Quantify Diurnal Physical Activity Timing in Relation to Obesity and Diabetes in the German National Cohort (NAKO) |
| 8 | Sowjanya Batchu | Identifying the Medication Patterns in the KORA FF4 study and its Metabolic Associations |
| 9 | Filippo Michelotti | Imaging signatures of pancreas across different diabetes subtypes |
| 10 | Rupshali Dasgupta | AI-Driven Knowledge Graph for Cross-Study Analysis of Cardiometabolic Transcriptomic Data |

**Abstract book**

# Unlocking the potential of DZD clinical study data for secondary research

**Ivona Anastasova**, DZD Datamanagement

Ivona Anastasova, Angela Dedie, Kirstin Tümler, Lars Oest

German Center for Diabetes Research (DZD), Munich-Neuherberg, Germany.

Data silos remain a major barrier to the large-scale utilization of health data beyond the primary aims of data collection. The exchange of such data can be difficult to navigate and may lead to misinterpretations and data quality issues if not properly prepared. In parallel, initiatives like the European Health Data Space II and the Observational Health Data Sciences and Informatics (OHDSI) network highlight the growing opportunity for secondary research. These movements align with the mission of the German Center for Diabetes Research (DZD) to improve exploration and secondary reuse of its clinical study data, which is collected across heterogeneous electronic data capture systems and data formats.

To support these goals, the DZD has (1) established a Core Data Set to enable the prospective harmonization of essential diabetes parameters in all new DZD clinical studies since 2021 and (2) adopted the OHDSI OMOP Common Data Model (CDM) for retrospective harmonization of its clinical study data. Prospective harmonization ensures consistent data collection so that data from different DZD clinical studies is already structurally and semantically aligned at the point of data capture. In contrast, retrospective harmonization aligns preexisting data from different studies to enable comparability and joint analysis.

The OMOP CDM facilitates harmonization by standardizing both data structure (through a relational data model) and semantics (through integration of international terminologies like SNOMED CT, ICD, LOINC, and ATC via the OHDSI ATHENA vocabulary). This approach offers several key benefits for conducting secondary research: it establishes a common scientific ground that is understood by health data scientists worldwide, increases the statistical power for testing hypotheses through combining datasets and enables the testing of the transportability of findings across studies.

To date, 11 DZD multicenter clinical studies involving over 6,000 volunteers across 10 sites in Germany have been conducted. As a first step toward improved data standardization and harmonization, a subset of two clinical DZD studies are currently retrospectively harmonized using the OMOP CDM. This data will be integrated with proteomics data of a subset of patients to unlock new clinical insights.

A key takeaway from this project is that early consideration of structural and semantic interoperability during study design is critical to reduce downstream harmonization efforts. This includes designing variables according to international standards, minimizing highly

study-specific variables, avoiding free text entries where possible, implementing centralized data collection, and providing data with rich metadata.

Further reading

1. Cheng C, Messerschmidt L, Bravo I, Waldbauer M, Bhavikatti R, Schenk C, Grujic V, Model T, Kubinec R, Barceló J. A General Primer for Data Harmonization. Sci Data. 2024 Jan 31;11(1):152. doi: 10.1038/s41597-024-02956-3.

2. Observational Health Data Sciences and Informatics (OHDSI). Standardized data: the OMOP common data model [Internet]. [cited 2026 Apr 10]. Available from: https://www.ohdsi.org/data-standardization/

3. Henke E, Zoch M, Peng Y, Reinecke I, Sedlmayr M, Bathelt F. Conceptual design of a generic data harmonization process for OMOP common data model. BMC Med Inform Decis Mak. 2024 Feb 26;24(1):58. doi: 10.1186/s12911-024-02458-7.

# DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

DZD German Center for Diabetes Research
Health Data Science & AI Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

## Why we need a standardized core data set for diabetes research and beyond

**Angela Dedié**, DZD Datamanagement

Angela Dedié, Ivona Anastasova, Kirstin Tümler, Lars Oest

German Center for Diabetes Research (DZD), Munich-Neuherberg, Germany.

The German Center for Diabetes Research (DZD) conducts large clinical multicenter studies in the field of diabetes and metabolic research. It is part of the German Centers for Health Research (DZG) which focus on novel therapies for diabetes, infections, lung diseases, cancer, mental disorders, cardiovascular, pediatric and neurodegenerative diseases. In this vein, a core data set (CDS) provides the descriptions of variables and definitions that are relevant for clinical research in data dictionary for purposes of consistency, data validity, reliability and interoperability. The design of a CDS enables harmonization and standardization in the collection, measurement and reporting of minimal information necessary for collaborative research.

In 2021, the DZD partner institutions agreed on a common DZD Core Data Set (DZD CDS). Its parameters were developed and harmonized in a multi-stage consensus process in close coordination with all DZD partner sites. The DZD CDS was drafted according to the FAIR criteria which emphasize the importance of data that is findable, accessible, interoperable and reusable. Rich metadata including standardized vocabularies (SNOMED CT and LOINC) was attached and made publicly available together with standard operating procedures (SOPs). An open licence was chosen for this data set, metadata and SOP (Creative Commons BY-SA 4.0). Therefore, the material is free to share and adapt for non-commercial use.

The resulting DZD CDS consists of 8 modules (master data (biodata), anthropometry, vital signs, laboratory, diabetes data, medical history, comorbidities, questionnaires) and a total of 153 parameters that are defined interdisciplinary. There are also optional parameters that can be included in data analyses depending on the research question. The focus of the DZD CDS is on prediabetes, type 1 and type 2 diabetes, metabolic issues and pathophysiology of metabolism. In 2024, the core data set of the DZG was integrated within the DZD CDS to support researchers in identifying patient cohorts across the various DZG and create new opportunities for collaboration. The DZD CDS establishes definitions for the collection of diabetes-relevant clinical parameters and thus makes a valuable contribution to diabetes research and forms the basis for future progress in the treatment and prevention of this and other widespread diseases. Download flyer on the DZD Core Data Set Link to the DZD Core Data Set Link to DZD CDS related Metadata and SOP

# DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

DZD
German Center for Diabetes Research
Health Data Science & AI Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

Further reading

1. Inau, E.T., Dedié, A., Anastasova, I. et al. The Journey to a FAIR CORE DATA SET for Diabetes Research in Germany. Sci Data 11, 1159 (2024). https://doi.org/10.1038/s41597-024-03882-0

2. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

# DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

DZD German Center for Diabetes Research
Health Data Science & AI Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

## An AI System for Robust Plain Language Translation of Clinical Trials

**Raeesa Yousaf**, Helmholtz Munich

Raeesa Yousaf, Stephanie Stapfer, Sebastian Lobentanzer

Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany

German Center for Diabetes Research (DZD), Munich-Neuherberg, Germany.

Clinical trial registry entries often contain complex medical terminology, creating barriers for patients who search online for relevant studies. Limited accessibility of trial information contributes to challenges in patient recruitment and retention, with up to 80% of trials failing to meet enrollment timelines. Large Language Models (LLMs) are increasingly used to generate plain language summaries (PLS) to improve accessibility. In the high-stakes domain of health information, where registries provide patients access to novel therapeutics and contribute to scientific advancement, summaries must preserve accuracy and clarity. Although current approaches aim to improve readability, they rarely incorporate structured validation to detect distortions or omissions of crucial details.

We investigate whether multi-agent critique systems can reduce distortions and improve alignment between AI-generated PLS and their source clinical trial registry entries. Specifically, we assess whether iterative revision preserves clinically relevant details. We compare two pipelines: (1) a baseline single-pass LLM summarization system and (2) an agentic architecture with a generator agent followed by critique agents targeting eligibility preservation and explicit uncertainty representation. Critique agents provide structured feedback that triggers iterative revision until alignment criteria are met.

We construct a dataset of ClinicalTrials.gov registry entries and implement a multi-dimensional evaluation framework. (1) Content preservation is measured with embedding-based similarity scores, such as BERTScore. (2) Consistency checks ensure key clinical details remain aligned with the source, combining rule-based extraction (e.g., regex for numeric values or age ranges) with a ground-truth–driven LLM-as-a-judge. (3) Readability is assessed using standard metrics, such as the Flesch Reading Ease Score.

This study quantifies whether agentic validation layers reduce clinically relevant distortions while maintaining accessibility and proposes a reproducible benchmarking methodology for evaluating generative models in high-stakes health communication, where fidelity to source material is essential.

# DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

DZD German Center for Diabetes Research
Health Data Science & AI Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

## Personalized Prevention of Cardiometabolic Diseases: Proteomic Insights from the KORA Study

**Nuha Shugaa Addin**, Helmholtz Munich

Nuha Shugaa Addin [1,6,] Agnese Petrera [2], Christian Herder [3,4,5], Steffi Hauck [3,6], Michael Roden [3,4,5], Annette Peters [1,6,7], Barbara Thorand [1,6]

[1] Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Epidemiology, Neuherberg, Germany

[2] Metabolomics and Proteomics Core, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Neuherberg, Germany

[3] German Center for Diabetes Research (DZD), Partner Düsseldorf, Neuherberg, Germany.

[4] Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany

[5] Department of Endocrinology and Diabetology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[6] German Center for Diabetes Research (DZD), Partner München-Neuherberg, Neuherberg, Germany

[7] Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Faculty of Medicine, LMU Munich, Pettenkofer School of Public Health, Munich, Germany.

Introduction: The KORA study (Cooperative Health Research in the Region of Augsburg) is a population-based prospective cohort that has contributed to the identification of risk factors and biomarkers, including proteomic signatures, for the prediction and prevention of cardiometabolic diseases such as excess/dysfunctional adiposity, type 2 diabetes, and chronic kidney disease (CKD).

Methods: The KORA cohort, initiated within the WHO MONICA project, comprises 17,602 participants (aged 25–74 years) recruited between 1984 and 2001, with comprehensive clinical, lifestyle, and biomarker data collected at baseline and follow-up. Proteomic profiling in the KORA S4 study (1999–2001) was conducted in participants aged 55–74 years, quantifying 233 proteins using a proximity extension assay (PEA). Prospective analyses utilized follow-up examinations (KORA F4 and KORA FF4) to assess outcomes including incident type 2 diabetes and changes in creatinine-based estimated glomerular filtration rate (eGFR). Protein–outcome associations were analyzed using linear, logistic, and Cox proportional hazards regression models, as appropriate. Least Absolute Shrinkage and Selection Operator (LASSO) regression was further applied for the identification of robust proteins, as well as for prediction modeling.

Results: We identified several proteins associated with markers of excess/dysfunctional adiposity, including established markers such as leptin and adiponectin, as well as less well-

characterized proteins, including α-L-Iduronidase (IDUA) and CD8 Alpha chain. Similarly, we identified 14, 24, and 4 proteins associated with prevalent prediabetes, prevalent type 2 diabetes, and incident type 2 diabetes, respectively. Of these, IL-17D, IL-18 receptor 1, carbonic anhydrase-5A, IL-1 receptor type 2 (IL-1RT2) and matrix extracellular phosphoglycoprotein were novel candidates. The addition of 12 priority LASSO-selected biomarkers significantly improved the prediction of type 2 diabetes (ΔAUC = 0.0219; 95% CI: 0.0052–0.0624). In CKD analyses, 66 proteins were inversely associated with the annual rate of change in eGFRcr. The most prominent associations were observed for Kidney Injury Molecule-1 (KIM1), N-terminal pro-B-type natriuretic peptide (NT-proBNP), and EPH receptor B4 (EPHB4). Furthermore, 21 of the 66 proteins were associated with incident CKD. The strongest associations were observed for TRAIL receptor 2 (TRAIL-R2), Tumor necrosis factor receptor superfamily member 9 (TNFRSF9), and Tumor necrosis factor receptor superfamily member 11A (TNFRSF11A).

Conclusion: Long-term cohort studies such as the KORA study are crucial for advancing personalized prevention, including the identification of proteomic biomarkers of cardiometabolic diseases.

# DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

DZD German Center for Diabetes Research
Health Data Science & AI Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

## Metabolic Adjustment and Latest Developments in Therapy for Children and Adolescents with Type 1 Diabetes – 30 Years of Real-World-Data from the Multinational DPV Registry

**Nicolas Wilms**, University of Ulm

Nicolas Wilms[1,2], Alexander Eckert [1,2], Julia Grimsmann [1,2], Beate Karges[3,4], Torben Biester[5], Claudia Böttcher[6], Katharina Warncke[7,8], Christina Reinauer[9], Clemens Kamrath[10,11], Sabine Hofer[12], Ulrike Schierloh[13], Stefanie Lanzinger[1,2], for the DPV Initiative

[1]Institute of Epidemiology and Medical Biometry, ZIBMT, Ulm University, Ulm, Germany.

[2]German Center for Diabetes Research (DZD), Munich-Neuherberg, Germany.

[3]Department of Pediatrics and Adolescent Medicine, Bethlehem Hospital Stolberg, Stolberg, Germany.

[4]Department of Endocrinology and Diabetology, University Hospital Aachen, Faculty of Medicine, RWTH Aachen University, Aachen, Germany.

[5]Diabetes Center for Children and Adolescents, Hannover, Germany.

[6]Paediatric Endocrinology and Diabetology, University Children's Hospital, University of Bern, Bern, Switzerland.

[7]Department of Pediatrics, Kinderklinik München Schwabing, Technical University of Munich School of Medicine, Munich, Germany.

[8]Institute of Diabetes Research, Helmholtz Munich, German Center for Environmental Health, Munich, Germany.

[9]Department of General Pediatrics, Neonatology and Pediatric Cardiology, Medical Faculty, University Hospital Düsseldorf, Düsseldorf, Germany.

[10]Centre for Paediatrics and Adolescent Medicine, University of Freiburg, Freiburg, Germany.

[11]Faculty of Medicine, Justus-Liebig-University Giessen, Giessen, Germany.

[12]Department of Paediatrics 1, Medical University of Innsbruck, Innsbruck, Austria.

[13]Department of Pediatric Diabetes and Endocrinology, Clinique Pédiatrique, Centre Hospitalier, Luxembourg, Luxembourg.

Introduction: To examine 30-year trends in metabolic control, diabetes technology and acute complications across children and adolescents with type 1 diabetes (T1D) from Germany, Switzerland, Austria, and Luxembourg.

Methods: Data from 104,518 children and adolescents aged <18 years with T1D and a diabetes duration >3 months were included from the diabetes prospective follow-up (DPV) registry between 1995 and 2024. We applied regression models to analyze temporal trends in continuous subcutaneous insulin infusion (CSII), continuous glucose monitoring (CGM),

automated insulin delivery (AID), HbA1c, event rates of severe hypoglycemia (SH), and diabetic ketoacidosis (DKA). Age was categorized as <6, 6-<12, and 12-<18 years. Regression models were adjusted for age, sex, diabetes duration, and migration background.

Results: CSII use increased from 3.4% (95% CI: 3.1-3.8) in 2000 to 75.9% (75.4-76.4) in 2024, with the highest proportion among children <6 years of age (93.6%; 92.6-94.5), followed by children aged 6-<12 years (81.3%; 80.6-82.1) and ≥12 years (68.5%; 67.8-69.2). CGM use increased from 24.7% (24.2-25.3; 2016) to 97.4% (97.2-97.6; 2024), reaching 98.1% (97.6-98.6), 97.8% (97.5-98.0), and 96.9% (96.6-97.1) within the respective age groups. AID use increased from 5.2% (4.9-5.4; 2020) to 58.3% (57.7-58.8; 2024), with 69.9% (68.1-71.6), 64.6% (63.7-65.6), and 52.8% (52.0-53.5) for each age group. Between 1995 and 2024, HbA1c (%) decreased from 8.1% (8.1-8.2) to 7.3% (7.3-7.3), and the proportion of persons achieving an HbA1c <7% increased from 25.8% (24.4-27.2) to 46.6% (46.0-47.1). Rates of SH decreased from 18.6 (15.7-21.9) to 4.9 (4.6-5.3) per 100 person-years (PY), and DKA from 2.1 (1.6-2.8) to 1.3 (1.1-1.4) PY.

Conclusion: The use of modern diabetes technologies has increased substantially across all age groups, accompanied by improved metabolic control and a significant reduction in acute complications. However, further advances are needed in order to support children and adolescents with the achievement of current glycemic targets.

# DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

DZD
German Center for Diabetes Research
Health Data Science & AI Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

## Genetic and Epigenetic Factors Modifying the Association between Diet Quality and Incident Type 2 Diabetes: The EPIC-Potsdam Cohort

**Christine El-Khoury**, German Insitute of Human Nutrition Potsdam-Rehbrücke (DIfE)

Christine El-Khoury[1,2], Fabian Eichelmann[1,2], Franziska Jannasch[1], Matthias B. Schulze[1,2,3]

[1]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany

[2]German Center for Diabetes Research (DZD), München-Neuherberg, Germany

[3]Institute of Nutritional Science, University of Potsdam, Nuthetal, Germany

Objective: To assess whether the effect of diet quality on type 2 diabetes risk differs across subgroups with different genetic or epigenetic susceptibility.

Research Design and Methods: We used data from a nested case-cohort in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam cohort (random subsample of 2,204 participants, 750 verified incident type 2 diabetes cases, median follow-up: 6.22 years). The Alternative Healthy Eating Index (AHEI-2010), Dietary Approaches to Stop Hypertension (DASH), Dietary Inflammatory Index (DII) and the Mediterranean Diet Pyramid score (MedPyr) were used to assess diet quality. Genetic risk was characterized using global polygenic risk scores (PRS) and pathway-specific polygenic risk scores (pPRS). A blood DNA methylation risk score (MRS) was used to reflect epigenetic risk. Cox proportional hazards models were used to assess the modification of the diet quality-diabetes risk association by the risk scores.

Results: Higher adherence to the AHEI-2010 was associated with a lower risk of type 2 diabetes, while higher PRS, pPRS and MRS predictably indicated higher risk in this population. We did not find convincing evidence of modification of the diet quality-diabetes risk association by genetic risk, but detected significant modification by the MRS for AHEI-2010 ($p=0.002$) and MedPyr ($p=0.001$). In particular, the AHEI-2010 association with diabetes was more pronounced in the low and medium MRS strata.

Conclusions: The associations of healthy diets do not seem to depend on the genetic predisposition for type 2 diabetes, however, they could depend on the diabetes risk captured by blood DNA methylation profiles.

DZD Health Data Science & AI
Kick-Off Workshop

April 27-29, 2026

DZD
German Center for
Diabetes Research
Health Data Science & AI
Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

# High-Resolution Accelerometry to Quantify Diurnal Physical Activity Timing in Relation to Obesity and Diabetes in the German National Cohort (NAKO)

**Michael Stein**, Helmholtz Munich

Michael J. Stein, on behalf of the NAKO collaborators

Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Epidemiology, Neuherberg, Germany

Background: Physical activity supports weight regulation and metabolic health, but its timing in relation to obesity and diabetes remains unclear. We aimed to assess the diurnal timing of physical activity and its association with obesity and diabetes.

Methods: We derived interpretable physical activity metrics from high-resolution accelerometer data collected over seven consecutive days. In a cross-sectional analysis of 61,116 participants aged 20–75 years from the German National Cohort (2015–2019), we classified activity into sex- and age-standardized quartiles within four time-of-day windows: morning (06:00–11:59), afternoon (12:00–17:59), evening (18:00–23:59), and night (00:00–06:00). Using multivariable logistic regression, we estimated associations of physical activity timing with obesity (BMI ≥ 30.0 kg/m2) and diabetes (self-reported or HbA1c ≥ 6.5%). We accounted for sex, age, study region, education, employment, risky alcohol use, smoking, night shift work, and sleep duration.

Results: High afternoon (top vs. bottom quartile, OR: 0.36, 95% CI: 0.33–0.38) and evening physical activity (OR: 0.45, 95% CI: 0.42–0.48) showed lower obesity odds than high morning activity (OR: 0.71, 95% CI: 0.66–0.76), whereas nighttime activity increased obesity odds (OR: 1.58, 95% CI: 1.48–1.68). Associations were similar for diabetes, with the lowest odds for afternoon (OR: 0.47, 95% CI: 0.42–0.53), followed by evening (OR: 0.56, 95% CI: 0.50–0.62) and morning activity (OR: 0.80, 95% CI: 0.71–0.89), and higher odds for nighttime activity (OR: 1.43, 95% CI: 1.29–1.58). Findings were not modified by employment status, night shift work, and sleep duration.

Conclusions: Our cross-sectional findings require longitudinal corroboration but suggest afternoon and evening activity provide greater metabolic health benefits than morning activity, while nighttime activity is discouraged.

DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

DZD German Center for Diabetes Research
Health Data Science & AI Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

# Identifying the Medication Patterns in the KORA FF4 study and its Metabolic Associations

**Sowjanya Batchu**, Helmholtz Munich

Lakshmi Sowjanya Batchu[1,2], Homa Bazireh[1,2], Annette Peters[1,2], , Wolfgang Rathmann[3], Lars Schwettmann[4], Schlesinger Sebastian[5], Harald Grallert[1], Sapna Sharma[1]

[1]Institute of Epidemiology, Research Unit of Molecular Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, 85764 Neuherberg, Germany.

[2]Chair of Epidemiology, Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Faculty of Medicine, Ludwig Maximilian University of Munich, Munich, Germany.

[3]Institute for Biometrics and Epidemiology, German Diabetes Center, Auf'm Hennekamp 65, 40225 Düsseldorf, Germany; German Centre for Diabetes Research, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany.

[4]Institute of Health Economics and Health Care Management, Helmholtz Zentrum München – German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany; Department of Economics, Martin Luther University Halle Wittenberg, 06099 Halle (Saale), Germany.

[5]Institute of Epidemiology, Kiel University, Kiel, Germany.

Background: Multimorbidity is raising a global health concern, and it often requires multiple treatment regimens. The regular use of five or more medications is defined as polypharmacy. While preclinical studies suggest that polypharmacy may have synergistic or complex effects on metabolism, evidence from real-world population studies remains limited. Linking medication patterns to metabolomic profiles may help better understand the biological consequences of polypharmacy.

Aim: This study aims to identify medication patterns in the KORA (Cooperative Health Research in the Region of Augsburg) FF4 cohort using archetypoidal clustering and to explore how these patterns differ across clinical characteristics, including polypharmacy prevalence. We also examine the associations between medication patterns and circulating metabolites, as well as with individual drugs. Through this, we aim to better understand the combined effects of multiple medications and comedication on metabolism.

Methods: Data from the population-based KORA FF4 cohort (n ≈ 1,401 participants with medication use; n ≈ 520 with metabolomics) were analyzed. The prevalence of polypharmacy was found to be 26.5% among participants with documented medication use. Archetypoidal clustering, an unsupervised learning technique, was used to identify pure or extreme medication patterns in the study population. The archetypoidal clusters were compared to clinical characteristics, and their associations with circulating metabolites were assessed using multivariate regression in both basic and full models. In the basic model,

analyses were adjusted for age, sex, and BMI. In the full model, additional adjustment for comedications was performed along with age, sex, and BMI.

Results: We identified four clusters Angiotensin Converting Enzyme dominant cardiovascular cluster (ACE cluster), Angiotensin Antagonist Dominant Cardiometabolic cluster (ARB cluster), Thyroid cluster, and Painkiller cluster. The ACE and ARB clusters mainly included older individuals with higher cardiometabolic risk and more polypharmacy, while the painkiller cluster appeared healthier with the lowest medication burden.

The metabolic changes were observed in the ACE and ARB clusters. In the ACE cluster, there was a clear decrease across several lipid classes, including ceramides, sphingomyelins, cholesteryl esters, and phosphatidylcholines, along with increases in amino acid–related metabolites and acylcarnitine's. The ARB cluster showed a similar pattern overall, but with altered variation in phosphatidylcholines and Lys phosphatidylcholines, and an increase in triglycerides. No significant metabolite for Thyroid and painkillers clusters.

When adjusting for comedications, many of the individual drug associations weakened, suggesting that the unadjusted results may reflect combined medication effects. In polypharmacy, we observed lower lipid levels and Dehydroepiandrosterone sulfate (DHEAS), along with higher levels of amino acid–related metabolites and indoxyl sulfate (Ind SO4).

Conclusion: These results suggest that multiple medications have a greater metabolic influence on the body than individual drugs. The decrease in DHEAS suggests the overall medication burden in the polypharmacy individual, and the increase in indoxyl sulfate suggests its influence on the systemic and metabolic effects associated with polypharmacy. It highlights the importance of considering interactions when studying the biological impact of medications.

# DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

DZD German Center for Diabetes Research
Health Data Science & AI Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

## Imaging signatures of pancreas across different diabetes subtypes

**Filippo Michelotti**, German Diabetes Center (DDZ)

F.C. Michelotti[1,3], K. Bódis[1,2,3], C. Binsch[1,2,3], K.S. Massold[1,3], S. Trenkamp[1,3], M. Roden[1,2,3], R. Wagner[1,2,3], V.B. Schrauwen-Hinderling[1,3]

[1]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[2]Department of Endocrinology and Diabetology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University, Düsseldorf, Germany

[3]German Center for Diabetes Research (DZD), Partner Düsseldorf, München-Neuherberg, Germany

Background and aims: Previous studies showed that pancreas features such as volume, border irregularity or lipid content assessed by magnetic resonance imaging (MRI) differ between people with type 1 (T1D) or type 2 diabetes (T2D) and normal glucose tolerance (NGT). Some of these features appear to be related to alterations in insulin secretion. However, available data are inconclusive and largely based on small cohorts. In this study, we aimed to 1) evaluate the reliability of MRI-derived pancreas features quantified using an automated deep learning (DL) approach, 2) compare pancreas MRI features across different diabetes subtypes, and 3) assess their relationship with gold-standard measures of insulin secretion.

Material and methods: Participants of the German Diabetes Study (GDS), including people with T1D, T2D, and NGT underwent abdominal MRI (Achieva 3T, Philips Healthcare, The Netherlands). Fat fraction (FF) maps were obtained from the acquisition of transversal 2-echo 3D-gradient-echo sequence (time of repetition (TR)=3.76 ms, time of echo (TE)1/2=1.32/2.40 ms, reconstructed voxel dimension=0.78×0.78×2 mm3). Following automatic deep learning (DL)-driven segmentation of the pancreas using a dedicated nnU-Net model, several features including volume, lipid content, fractal dimension (FD) reflecting border irregularity, as well as inverse difference (ID), reflecting FF texture homogeneity, were calculated. The segmentation model was tested in a hold-out set (n=15). Measurement variability was assessed in a scan-rescan repeatability study (n=8). Differences in pancreatic features between T1D (n=126), T2D (n=123), phenotype-based diabetes subtypes and NGT (n=93) were analysed using ANOVA with post-hoc correction. Associations with insulin secretion were examined using multiple linear regression adjusted for sex, age, BMI and visceral adipose tissue (VAT) volume. Additionally, associations of pancreas features with insulin secretion assessed by mix-meal test (AUCC-peptide(0-120 min)/AUCglucose(0-120 min)) were assessed.

Results: Comparison of DL-driven segmentation of the pancreas with manually annotated regions yielded dice similarity coefficient of 0.83±0.10. Repeated measurements analysis showed CV<10% for all pancreas features. People with severe autoimmune diabetes (SAID or T1D) had lower FD (mean, CI: -0.074, [-0.114; -0.035] a. u.) and lower pancreatic volume (-8,561, CI: [-10,649; -6,474] cm3) compared with the NGT. Increased pancreatic lipid content (2.5, [0.6; 4.4] FF%) and lower FF texture homogeneity (-0.020, [-0.031; -0.009] a. u.) were found in T2D. Lower homogeneity was observed in mild age-related diabetes (MARD) (-0.004, CI: [-0.021; -0.002] a. u.) and mild obesity-related diabetes (MOD) (-0.024, [-0.042; -0.005] a. u.). Pancreatic lipid content (β, CI: -0.28, [-0.46; -0.09]) inversely, volume (β=0.39, CI: [0.23; 0.55]) and FF texture homogeneity (0.24, [0.05; 0.42]) positively associated with postprandial insulin secretion.

Conclusion: Overall, DL-driven pancreas segmentation led to reproducible MRI signatures, which differ across diabetes subtypes. Pancreas features independently relate to insulin secretion, supporting MRI-based pancreas phenotyping as a scalable marker of disease heterogeneity.

# DZD Health Data Science & AI Kick-Off Workshop

April 27-29, 2026

DZD German Center for Diabetes Research
Health Data Science & AI Certificate Program

**Helmholtz Pioneer Campus**, Ingolstädter Landstraße 1, 85764 Neuherberg

## AI-Driven Knowledge Graph for Cross-Study Analysis of Cardiometabolic Transcriptomic Data

**Rupshali Dasgupta**, Helmholtz Munich

Rupshali Dasgupta, Sebastian Lobentanzer

Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany

Biomedical research generates large volumes of heterogeneous data across multi-omics and clinical studies. Integrating this data is essential for understanding biological relationships, yet querying knowledge graphs (KGs) typically requires specialist expertise, limiting accessibility for domain scientists. Existing KG and AI tools for biomedical data exploration lack support for integrating heterogeneous, multi-study transcriptomic datasets within a unified, schema-driven framework accessible to non-technical researchers. We present the first implementation of a schema-driven KG platform that integrates multi-study transcriptomic data by representing differential expression as structured evidence. A natural language interface enables clinical researchers to query this biological evidence without requiring query language expertise, supporting intuitive cross-study exploration of cardiometabolic disease data.