Aus dem Deutschen Diabetes-Zentrum (DDZ)
Leibniz-Zentrum für Diabetes-Forschung
an der Heinrich-Heine-Universität Düsseldorf
Institut für Biometrie und Epidemiologie
Direktor: Univ.-Prof. Dr. Oliver Kuß

Diabetes subtyping for precision medicine – From practical implementation to methodological challenges

Dissertation

zur Erlangung des Grades eines Doktors in Public Health
der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Tim Tomiyuki Mori
(2025)

Als Inauguraldissertation gedruckt mit Genehmigung der
Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez.:
Dekan: Prof. Dr. med. Nikolaj Klöcker
Erstgutachter: Prof. Dr. sc. hum. Oliver Kuß
Zweitgutachterin: Prof. Dr. rer. nat. Annika Hoyer

Widmung

Heike, Yukio, Christa, Willi & Tobi

Teile dieser Arbeit wurden veröffentlicht:

Mori, T., Prystupa, K., Straßburger, K., Bonn, M., Zaharia, O.P., Spörkel, O., Kuß, O., Roden, M., and Wagner, R., 2025. A web-based application for diabetes subtyping: The DDZ Diabetes-Cluster-Tool. *Acta Diabetologica*, (62) 281-283. https://doi.org/10.1007/s00592-024-02436-5

Mori T, Zaharia OP, Straßburger K, Dennis JM, Mai K, Kabisch S, Bornstein S, Szendroedi J, Blüher M, Meyhöfer S, Seissler J, Birkenfeld A, Stefan N, Roden M, Wagner R, Kuß O, and the GDS group 2025. Recognising, quantifying and accounting for classification uncertainty in type 2 diabetes subtype. *Diabetologia.* Online ahead of print. https://doi.org/10.1007/s00125-025-06486-4

Zusammenfassung

Diabetes-Subtypen für die Präzisionsmedizin – von der praktischen Umsetzung zu methodischen Herausforderungen

Es besteht ein wachsendes Interesse an der Präzisionsmedizin, um die Behandlung von Menschen mit Diabetes zu optimieren. Derzeitige Forschungsarbeiten zielen darauf ab, die Einteilung in Typ 1 und Typ 2 Diabetes durch eine differenziertere Klassifikation zu ergänzen. Insbesondere wurde vorgeschlagen, Typ 2 Diabetes in milde (altersbedingte, adipositasbedingte) und schwere (insulindefiziente, insulinresistente) Subtypen zu unterteilen.

Derzeit existieren jedoch keine Softwarelösungen, die diese Neuklassifizierung des Diabetes im klinischen Alltag ermöglichen. Zudem lassen sich manche Personen möglicherweise nicht eindeutig einem der vorgeschlagenen Subtypen zuordnen. Bisher wurde diese Problematik jedoch nicht systematisch untersucht. Ziel dieser Arbeit war es daher, (i) eine benutzerfreundliche Softwareimplementierung zu entwickeln, die eine Klassifizierung der Diabetes-Subtypen in der klinischen Praxis ermöglicht, und (ii) eine neue Methode zur Quantifizierung der Klassifikationsunsicherheit bei Typ 2 Diabetes-Subtypen vorzuschlagen.

Es konnte gezeigt werden, dass (i) das neu entwickelte *DDZ Diabetes-Cluster-Tool* Ärzt:innen eine vereinfachte Zuordnung von Personen zu den Diabetes-Subtypen erlaubt und (ii) die *normalisierte relative Entropie* ein Maß ist, mit dem sich die Sicherheit von Klassifizierungen auf einer Skala von 0 (vollständige Unsicherheit) bis 1 (vollständige Sicherheit) quantifizieren lässt. Bei Personen aus der Deutschen Diabetes-Studie mit neu diagnostiziertem Typ 2 Diabetes betrug die mediane normalisierte relative Entropie 0.127 (95% Konfidenzintervall: 0.119, 0.135), was auf eine erhebliche Klassifikationsunsicherheit hindeutet.

Insgesamt ist das DDZ Diabetes-Cluster-Tool die erste benutzerfreundliche Softwarelösung, mit der sich Diabetes im klinischen Alltag in Subtypen einteilen lässt. Gleichzeitig deuten die Ergebnisse darauf hin, dass die vorgeschlagene Neuklassifizierung aufgrund unsicherer Zuordnungen nur begrenzten Nutzen für die Präzisionsmedizin bietet.

Summary

Diabetes subtyping for precision medicine – From practical implementation to methodological challenges

There is growing interest in precision medicine to optimise the treatment of people living with diabetes. Recent research has focused on complementing the classification into type 1 and type 2 diabetes with a more nuanced subtyping approach. In particular, it has been proposed to subdivide type 2 diabetes into mild (age-related, obesity-related) and severe (insulin-deficient, insulin-resistant) subtypes.

However, no software tools are currently available to support this reclassification of diabetes in routine clinical care. Moreover, some individuals may not be clearly assignable to any of the proposed subtypes, an issue that has not yet been systematically investigated. The aim of this work was therefore to (i) develop a user-friendly software implementation to enable diabetes subtyping in clinical practice, and (ii) propose a novel method to quantify classification uncertainty in type 2 diabetes subtypes.

It was shown that (i) the newly developed *DDZ Diabetes-Cluster-Tool* provides clinicians with a practical means of assigning individuals to diabetes subtypes, and (ii) the *normalised relative entropy* provides a measure to quantify classification certainty on a scale from 0 (complete uncertainty) to 1 (complete certainty). Among individuals with newly diagnosed type 2 diabetes from the German Diabetes Study, the median normalised relative entropy was 0.127 (95% confidence interval: 0.119, 0.135), indicating substantial classification uncertainty.

In conclusion, the DDZ Diabetes-Cluster-Tool is the first user-friendly software solution to enable diabetes subtyping in routine clinical care. At the same time, the results suggest that the proposed reclassification may offer only limited value for precision medicine due to uncertainty in subtype assignments.

List of abbreviations

ADA	American Diabetes Association
ANDIS	All New Diabetics in Scania
BMI	Body mass index
C-index	Concordance index
CPRD	Clinical Practice Research Datalink
CVD	Cardiovascular disease
DDZ	*Deutsches Diabetes-Zentrum*
DPP-4i	Dipeptidyl peptidase-4 inhibitors
EASD	European Association for the Study of Diabetes
eGFR	Estimated glomerular filtration rate
GADA	Glutamic acid decarboxylase antibodies
GDS	German Diabetes Study
GLP1-RA	Glucagon-like peptide-1 receptor agonists
HbA1c	Glycated haemoglobin
HOMA-B	Homoeostasis model assessment of beta-cell function
HOMA-IR	Homoeostasis model assessment of insulin resistance
KDIGO	Kidney Disease: Improving Global Outcomes
MARD	Mild age-related diabetes
MOD	Mild obesity-related diabetes
NRE	Normalised relative entropy
PMDI	Precision Medicine in Diabetes Initiative
SAID	Severe autoimmune diabetes
SCORE	Systematic Coronary Risk Evaluation
SGLT2i	Sodium-glucose cotransporter 2 inhibitors
SIDD	Severe insulin-deficient diabetes
SIRD	Severe insulin-resistant diabetes

Acknowledgements: ChatGTP (version 4) was used to check grammar and spelling of the manuscript.

Table of contents

1 Introduction

Diabetes mellitus is a collection of metabolic diseases characterised by elevated blood glucose (hyperglycaemia) [1], affecting about one out of ten people in Germany [2]. Diagnosis is commonly based on glycated haemoglobin (HbA1c) levels ≥ 6.5%, which reflect increased blood glucose over the past 2-3 months [1]. Type 1 diabetes is an autoimmune disease that accounts for 5-10% of diabetes cases and usually requires insulin treatment due to an absolute insulin deficiency [1]. Type 2 diabetes, which accounts for 90-95% of diabetes cases, is characterised by a progressive, relative loss of insulin secretion not caused by an autoimmune process [1]. People with type 2 diabetes typically have a reduced response to insulin, which is referred to as insulin resistance. While the progression rates can differ depending on the type, there are common complications for which people with diabetes are at risk [1, 3]. In particular, diabetes is associated with chronic complications caused by damage to small blood vessels (microvascular complications) and large blood vessels (macrovascular complications) [3]. For example, people with diabetes have an increased risk for kidney disease and nerve damage, as well as potentially fatal cardiovascular events such as stroke and myocardial infarction. In 2010, the annual healthcare costs in Germany related to type 1 and type 2 diabetes were about €1 billion and 28€ billion and have been projected to further increase in the coming decades [4].

1.1 Heterogeneity of diabetes mellitus

While the distinction between type 1 and type 2 diabetes is well-established and guides treatment decisions in clinical practice, not every person can be clearly classified at the time of their diagnosis [1]. Both diseases are heterogeneous and the clinical manifestations may vary substantially between individuals. For example, not every person with type 1 diabetes is autoantibody positive or develops the disease at a younger age. Similarly, not every person with type 2 diabetes is overweight or develops the disease at an older age. Such ambiguities have challenged the established classification system and have led to increasing efforts to understand the variability in the clinical presentation of diabetes [1, 5, 6]. Notably, the heterogeneity of type 2 diabetes has received considerable attention in recent years [7-9].

A diagnostic challenge of type 2 diabetes is that it is, in principle, a diagnosis by exclusion of other causes [1, 8]. Specifically, a person is diagnosed with type 2 diabetes

if their increased blood glucose cannot be attributed to a specific cause such as an autoimmune process (type 1 diabetes) or a single gene defect (monogenic diabetes) [1, 8]. In other words, while autoantibody tests are used to diagnose type 1 diabetes, there is no diagnostic test for type 2 diabetes [8]. At the same time, type 2 diabetes has some hallmark features such as insulin resistance and some degree of remaining insulin secretion, often combined with obesity or overweight [1]. In terms of pathophysiology, it has been linked to several pathways, including beta cell function, adipose tissue distribution, and glucagon and incretin function [8]. However, the relative contribution of the different pathophysiological processes varies between individuals, and the aetiology of type 2 diabetes is not yet fully understood [1, 8, 10]. As a consequence, people diagnosed with type 2 diabetes may differ significantly from one another in terms of their clinical profile. For example, it is well established that some individuals with type 2 diabetes are relatively obese and insulin-resistant, whereas other are rather lean and insulin-deficient [7]. In general, there are large differences between individuals across defining clinical features such as blood glucose levels, age at diagnosis, blood pressure and lipid levels [11]. In spite of these individual differences, current guidelines for the management of type 2 diabetes typically employ a one-size-fits-all approach with little personalised recommendations [12].

1.2 Precision medicine in diabetes research

In response to the clinical heterogeneity, and given increasing availability of large datasets and advanced analytical methods, there has been growing interest in the concept of precision medicine [1, 6, 13, 14]. In 2018, the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) co-founded the Precision Medicine in Diabetes Initiative (PMDI) with the aim of promoting precision medicine research, assessing the current evidence base and developing guidelines for clinical implementation [6]. In a joint consensus report, ADA and EASD defined precision medicine as "*an approach to optimise the diagnosis, prediction, prevention, or treatment of diabetes by integrating multidimensional data, accounting for individual differences*." [6] (p. 1620). They define the individual components of precision medicine as follows:

- "*Precision diagnosis involves refining the characterisation of the diabetes diagnosis for therapeutic optimisation or to improve prognostic clarity using*

information about a person's unique biology, environment and/or context." [6] (p. 1620)

- "*Precision prognostics focuses on improving the precision and accuracy with which a patient's disease-related outcomes are predicted* (…)." [6] (p. 1620)
- "*Precision prevention includes using information (...) to determine their likely responses to health interventions and risk factors and/or to monitor progression toward disease*." [6] (p. 1620)
- "*Precision treatment involves using information (...) to guide the choice of an efficacious therapy to achieve the desired therapeutic goal or outcome, while reducing unnecessary side effects*." [6] (p. 1620)

While these individual components provide a helpful framework for precision medicine, they can overlap and need not be considered strictly separated from one another [15]. The overarching aim is to improve the management of diabetes by tailoring clinical decisions to subgroups of individuals with similar characteristics [6]. As such, it can be seen as an attempt to reduce errors in medical decision-making.

It should be noted that the terms 'precision', 'stratified', 'personalised', and 'individualised' medicine are sometimes used interchangeably. The ADA/EASD consensus report [6] considers 'precision medicine' to be synonymous with 'stratified medicine' and pertaining to tailored recommendations based on findings from a reference population [15]. In other words, it involves analysing datasets of a heterogeneous population in order to derive recommendations for individuals sharing similar clinical features [15]. Of note, this can be achieved in two ways. One option is to explicitly assign individuals into different subgroups sharing similar clinical features. The other option is to directly use those features to predict complications or treatment response [15, 16]. For example, a prediction model developed in a reference population can recommend an optimal glucose-lowering medication based on clinical features such as disease duration and body weight [17].

In comparison to precision medicine, 'personalised' (or 'individualised') medicine goes one step further by making recommendations based on the analysis of data from a single person [14, 15]. For example, data from a person's continuous glucose monitoring device can be used to optimise that person's dose and timing of insulin administration in real-time [15]. Franks et al [15] further distinguish between 'objective' and 'subjective' personalisation. 'Objective' personalisation entails using quantitative, patient-specific

data to inform individual recommendations. For example, following an N-of-1 trial of glucose-lowering medications, a person might choose the drug which resulted in the greatest glycaemic improvement [18]. By contrast, 'subjective' personalisation entails tailoring recommendations based on qualitative features such as patient preferences, capabilities and circumstances [15]. For example, following the same N-of-1 trial, a shared decision-making process might result in a person choosing the drug on which they personally felt best [19]. At present, research efforts appear to focus more on precision medicine than on personalised medicine [6, 14], with the former considered more achievable in the near future [15].

1.3 Subtyping of adult-onset diabetes

Given the growing interest in precision medicine, much of recent research has focused on classifying diabetes into more granular 'subtypes' (or 'clusters', 'phenotypes', 'endotypes') than the current type 1 and type 2 distinction [1, 6, 9, 20, 21]. In the PMDI's systematic review on precision subclassification of type 2 diabetes, Misra et al [9] distinguish between so-called simple and complex subtyping approaches for precision diagnosis. Simple approaches use routinely available clinical features and categorise individuals based on cut-off values or other straightforward techniques [9]. For example, type 2 diabetes has previously been subclassified into an obese and non-obese type based on a 30 kg/m^2 body mass index (BMI) cut-off [22]. By contrast, complex subtyping approaches categorise individuals using advanced statistical or machine-learning approaches that combine several clinical features [9]. For example, machine-learning methods have been used to identify type 2 diabetes subtypes defined by age, BMI, waist circumference and HbA1c [23]. Besides incorporating routine clinical features, complex subtyping approaches may also make use of genetic data [9, 24].

In their literature review, Misra et al [9] identified 51 studies that employed simple subtyping approaches and 62 studies that employed complex subtyping approaches to study the heterogeneity of type 2 diabetes. The simple approaches covered a variety of clinical features, focusing on measures of beta-cell function, lipid profiles, BMI and pancreatic autoantibodies, among others [9]. Given the different clinical features and methods used, these simple subtyping approaches were rather fragmented and there was a consistent lack of replications [9]. By contrast, complex subtyping approaches have

increasingly converged towards a framework for adult-onset diabetes first introduced by Ahlqvist et al [20].

In their seminal paper published in 2018, Ahlqvist et al [20] performed a data-driven cluster analysis, which suggested that newly diagnosed adult-onset diabetes (including both type 1 and type 2 diabetes) can be classified into five different subtypes. Their subtyping approach relied on six clinical features: Presence of glutamic acid decarboxylase antibodies (GADA), HbA1c, BMI, age at onset of diabetes, homoeostasis model assessment of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR) [20]. GADA is the primary autoantibody used to differentiate type 1 diabetes from type 2 diabetes in newly diagnosed adults [1]. HOMA-B and HOMA-IR are proxy measures of insulin secretion and insulin resistance, which are computed from fasting glucose and fasting insulin or C-peptide concentrations [25]. The five clusters derived from their analysis were labelled and described by Ahlqvist et al [20] as follows:

- **Cluster 1 – Severe autoimmune diabetes (SAID)**: GADA-positive, early disease onset, low BMI, insulin-deficient, high HbA1c
- **Cluster 2 – Severe insulin-deficient diabetes (SIDD)**: GADA-negative, otherwise similar to SAID
- **Cluster 3 – Severe insulin-resistant diabetes (SIRD)**: GADA-negative, late disease onset, high BMI and high insulin resistance
- **Cluster 4 – Mild obesity-related diabetes (MOD)**: GADA-negative, early disease onset, high BMI but low insulin resistance
- **Cluster 5 – Mild age-related diabetes (MARD)**: GADA-negative, late disease onset, low BMI, low insulin resistance

The clusters were derived by applying k-means clustering to data from 8,980 adults in the Swedish All New Diabetics in Scania (ANDIS) cohort [20]. The k-means algorithm assigns individuals to a pre-specified number of clusters such that the squared Euclidean distances of individuals to their cluster means (i.e., the average dissimilarity) are minimised [26]. The optimal number of five clusters was determined beforehand in a data-driven manner, which also grouped together all individuals who were GADA-positive (i.e., the SAID subtype) [20]. The k-means clustering was then performed only in GADA-negative individuals, using the continuous clustering variables (HbA1c, BMI, age, HOMA-B, HOMA-IR) to derive the four non-autoimmune clusters (i.e., the SIDD,

SIRD, MOD, and MARD subtypes). In other words, within the subtype framework proposed by Ahlqvist et al [20], adult-onset type 1 diabetes (typically GADA-positive [1]) has been relabelled as SAID, whereas adult-onset type 2 diabetes (GADA-negative [1]) has been subdivided into SIDD, SIRD, MOD, and MARD. That is, the innovation of their approach lies in the more granular subclassification of type 2 diabetes [6, 9, 15]. A graphical overview of the clustering approach is shown in Figure 1.

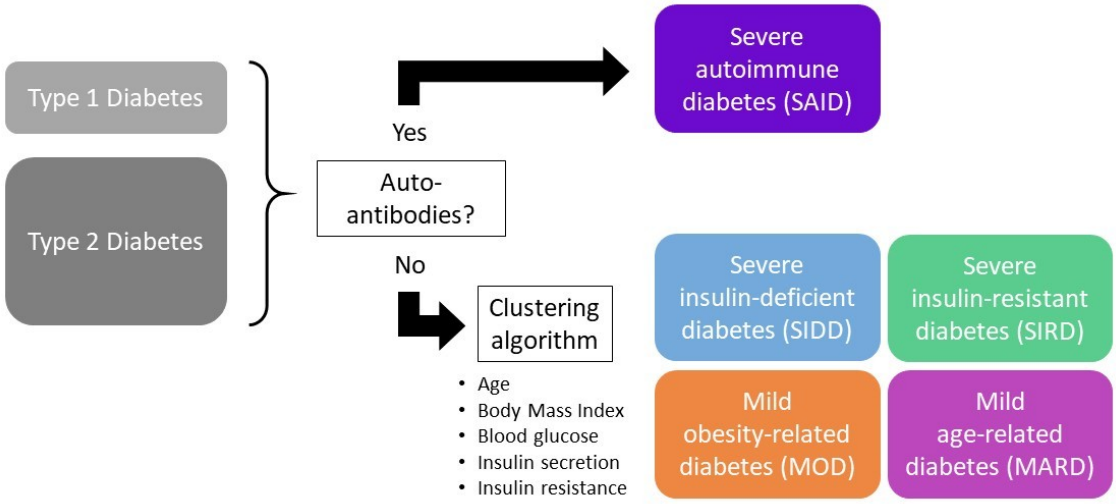

Fig. 1 – Graphical overview of the clustering approach of Ahlqvist et al [20]. SAID corresponds to adult-onset type 1 diabetes (due to the presence of autoantibodies), whereas SIDD, SIRD, MOD, and MARD correspond to subtypes of adult-onset type 2 diabetes derived using k-means clustering.

The clustering procedure was carried out separately for men and women, but the same five clusters with similar characteristics were identified in both sexes [20]. While Ahlqvist et al [20] initially replicated the diabetes subtypes in three Scandinavian cohorts, they have now been replicated in at least 22 studies, which included around 88,000 individuals from diverse ethnic and ancestral origins [9]. For example, replication studies have identified the diabetes subtypes in German [27], North American [28], Chinese [28], and Japanese [29] cohorts.

Beyond their contribution to precision diagnosis, the diabetes subtypes have received considerable attention for their potential regarding precision prognostics [9]. In their initial study, Ahlqvist et al [20] found that the subtypes differed in their disease progression, both in terms of complications and medication usage. For instance, individuals with SIRD had the highest risk of chronic kidney disease, whereas individuals with SIDD had the highest risk of retinopathy. In terms of treatment, individuals with SIDD were the earliest to require an additional glucose-lowering medication and, along with those with SAID, were most likely to receive insulin treatment. Subsequent studies

confirmed that the subtypes differed in their risks of both microvascular conditions (such as nephropathy) and macrovascular conditions (such as major adverse cardiovascular events) [9]. However, while the diabetes subtypes and their associations with clinical outcomes have been replicated in several studies [9], there remain fundamental challenges regarding the procedure of assigning individuals to these subtypes.

1.4 Challenges in the assignment of diabetes subtypes

1.4.1 Lack of practical tools for subtype assignment

The first challenge of the subtyping approach is that, so far, there exist no software tools to enable diabetes subtyping in clinical practice [6]. While assignment to the SAID subtype is based simply on the presence or absence of GADA, the procedure for assignment to the type 2 diabetes subtypes (SIDD, SIRD, MOD, and MARD) is less straightforward. Even if the required continuous clinical variables (HbA1c, BMI, age, HOMA-B, and HOMA-IR) are available, a clinician cannot readily determine an individual's subtype due to the complexity of the clustering algorithm [9].

To assign individuals to the type 2 diabetes subtypes, Ahlqvist et al [20] proposed the so-called 'nearest centroid approach' [27]. In this approach, one first computes the Euclidean distances of a person to the sex-specific average clinical profile of each subtype ('centroid') as reported in the ANDIS cohort [20]. These distances provide a measure of dissimilarity between the person and the respective subtype. The person is then assigned to the subtype to which the distance was the smallest ('nearest centroid'). In other words, each person is assigned to the subtype to which they were comparatively the most similar. Note that, in line with the original analysis in the ANDIS cohort [20], the five continuous clinical features must be standardised (i.e., centred and scaled) prior to carrying out the nearest centroid algorithm.

Overall, the subtype assignment procedure requires several computational steps, which researchers typically perform in statistical software programs such as R [30]. However, this requires programming knowledge and is often not feasible for clinicians working in clinical practice. Therefore, a user-friendly software tool, which requires only an individual's clinical information as input and performs all the necessary calculations in the background, would be highly valuable to enable diabetes subtyping in clinical practice.

1.4.2 Classification uncertainty in type 2 diabetes subtypes

The second challenge of the subtyping approach is that, while each individual with type 2 diabetes can technically be assigned to a subtype using the nearest centroid approach, not every individual will match with that subtype equally well [8, 31]. Already in the original publication of Ahlqvist et al [20], a close examination of the distribution of clinical features reveals large variability within each subtype. For example, not every individual with SIRD shows elevated insulin resistance and not every individual with SIDD exhibits reduced insulin secretion. In line with this, Pearson [8] argues that the phenotypic variability of type 2 diabetes might be difficult to capture with just four categories and that individual classification probabilities might be low. In particular, an individual's clinical profile might be in between two subtypes (e.g. sharing similarities with both MOD and MARD) [32] or not match with any of the four subtypes [8].

In response to this problem, the second PMDI consensus report [14] insists that diabetes subtype assignments should always be accompanied by a measure of classification uncertainty. So far, however, the quantification of classification uncertainty in type 2 diabetes subtypes has received limited attention [33-35]. Kahkoska et al [33] compared the fit of the assigned subtype to that of the next best subtype by computing the ratio of the two Euclidean distances obtained from the nearest centroid approach. In another study, Tanabe et al [34] developed a machine-learning model that predicted classification probabilities for each subtype based on an extensive set of clinical features. The latter approach is comparable to the soft-clustering approach of Wesolowska-Andersen et al [35], which directly provides classification probabilities as part of the clustering model. In both approaches the authors addressed classification uncertainty by excluding individuals with low classification probabilities. Individuals with classification probabilities lower than 0.60 for their assigned subtype were referred to as 'undecidable' [34] or 'mixed ethology' [35], respectively. In contrast, Kahkoska et al [33] chose a cut-off value of 0.80 for their distance ratio measure of classification uncertainty.

There are several limitations regarding the above approaches for quantifying classification uncertainty. First, the use of different cut-off values illustrates the problem with determining hard thresholds beyond which a cluster assignment is deemed uncertain. Frequently, the decisions for specific cut-off values are data-driven [35], which can make comparisons between studies challenging [36]. Second, by focusing solely on the largest classification probability (i.e., the probability of the assigned cluster), these approaches ignore relevant information about similarity to the other three clusters. As a consequence,

they cannot differentiate between an individual who is at the border between two clusters and one who does not fit well into any cluster. While the distance ratio measure of Kahkoska et al [33] considers the two best fitting clusters, it still fails to account for the potential overlap with all four type 2 diabetes subtypes. Finally, none of the studies investigated whether classification uncertainty varies across type 2 diabetes subtypes or what factors might influence it. Overall, this highlights a significant gap in our understanding of classification uncertainty in the type 2 diabetes subtypes and how to properly account for it.

1.5 Aims of this doctoral thesis

At present, no software tools are available to support diabetes subtyping in routine clinical care. Moreover, some individuals may not be clearly assignable to any of the proposed subtypes. To address these challenges in the assignment of diabetes subtypes, the aims of this doctoral thesis were to:

1) develop a user-friendly software tool that enables clinicians to assign individuals to the diabetes subtypes [37], and
2) propose a novel, statistically sound method to quantify classification uncertainty in type 2 diabetes subtypes [38].

The remainder of the dissertation is structured as follows: Chapter 2 introduces the DDZ Diabetes-Cluster-Tool, which was launched in May 2023 and provides a web-based application for diabetes subtyping [37]. Chapter 3 describes how the normalised relative entropy (NRE) statistic can be used to recognise, quantify and account for classification uncertainty in type 2 diabetes subtypes [38]. It is then applied to examine the classification uncertainty of recently diagnosed individuals from the German Diabetes Study (GDS). The GDS is approved by the ethics board of the Medical Faculty of Heinrich Heine University Düsseldorf (reference no. 4508). Finally, Chapter 4 discusses limitations of the diabetes subtyping approach, considers alternative strategies, and examines broader challenges to precision medicine.

2 **A web-based application for diabetes subtyping: The DDZ Diabetes-Cluster-Tool**, Mori T, Prystupa K, Straßburger K, Bonn M, Zaharia OP, Spörkel O, Kuß O, Roden M, and Wagner R, *Acta Diabetologica* 62(2): 281-283, (2025)

DOI: https://doi.org/10.1007/s00592-024-02436-5

3 **Recognising, quantifying and accounting for classification uncertainty in type 2 diabetes subtypes**, Mori T, Zaharia OP, Straßburger K, Dennis JM, Mai K, Kabisch S, Bornstein S, Szendroedi J, Blüher M, Meyhöfer S, Seissler J, Birkenfeld A, Stefan N, Roden M, Wagner R, Kuß O, and the GDS group, *Diabetologia*, online ahead of print, (2025)

DOI: https://doi.org/10.1007/s00125-025-06486-4

4 Discussion

The publications presented here covered a tool for the practical implementation of diabetes subtyping, but also highlighted the methodological challenge of classification uncertainty in type 2 diabetes subtypes. In brief, the DDZ Diabetes-Cluster-Tool provides a user-friendly online tool for clinicians to perform diabetes subtyping [37]. To assess uncertainty in subtype assignments, the NRE statistic has been proposed and examined in the prospective, observational GDS cohort [38]. A graphical overview of the proposed NRE method is shown in Figure 2.

4.1 Challenges of diabetes subtyping in clinical practice

While the DDZ Diabetes-Cluster tool [37] provides a useful software interface, there are additional obstacles to implementing diabetes subtyping in clinical practice. The extensive set of clinical features used in the clustering algorithm might be available in well-characterised cohort studies [9], but availability in real-world clinical practice is limited [6]. In particular, to compute the HOMA indices of insulin secretion and insulin resistance, fasting insulin or C-peptide concentrations are needed [25]. At present, however, these are not routinely measured in clinical practice [6, 33, 39, 40]. In consequence, the diabetes subtyping approach may have limited utility for improving diabetes care in real-world clinical practice. Even for individuals in the original ANDIS cohort, GADA and C-peptide were not part of the standard measurements in clinical practice, but were measured additionally as part of the study protocol [41]. In addition, access to the necessary measurements may be even more restricted in low-income and middle-income countries with resource-constrained healthcare systems [42]. Consequently, there is a risk that such a precision medicine approach, when implemented in clinical practice, could further exacerbate global health disparities [14, 43].

4.2 Methodological limitations of type 2 diabetes subtyping

The analysis of the NRE presented above [38] highlights that classification uncertainty is a fundamental challenge of type 2 diabetes subtyping. In the GDS cohort, the median NRE (on a scale from 0 to 1) was 0.127 (95% CI 0.119, 0.135) and the median

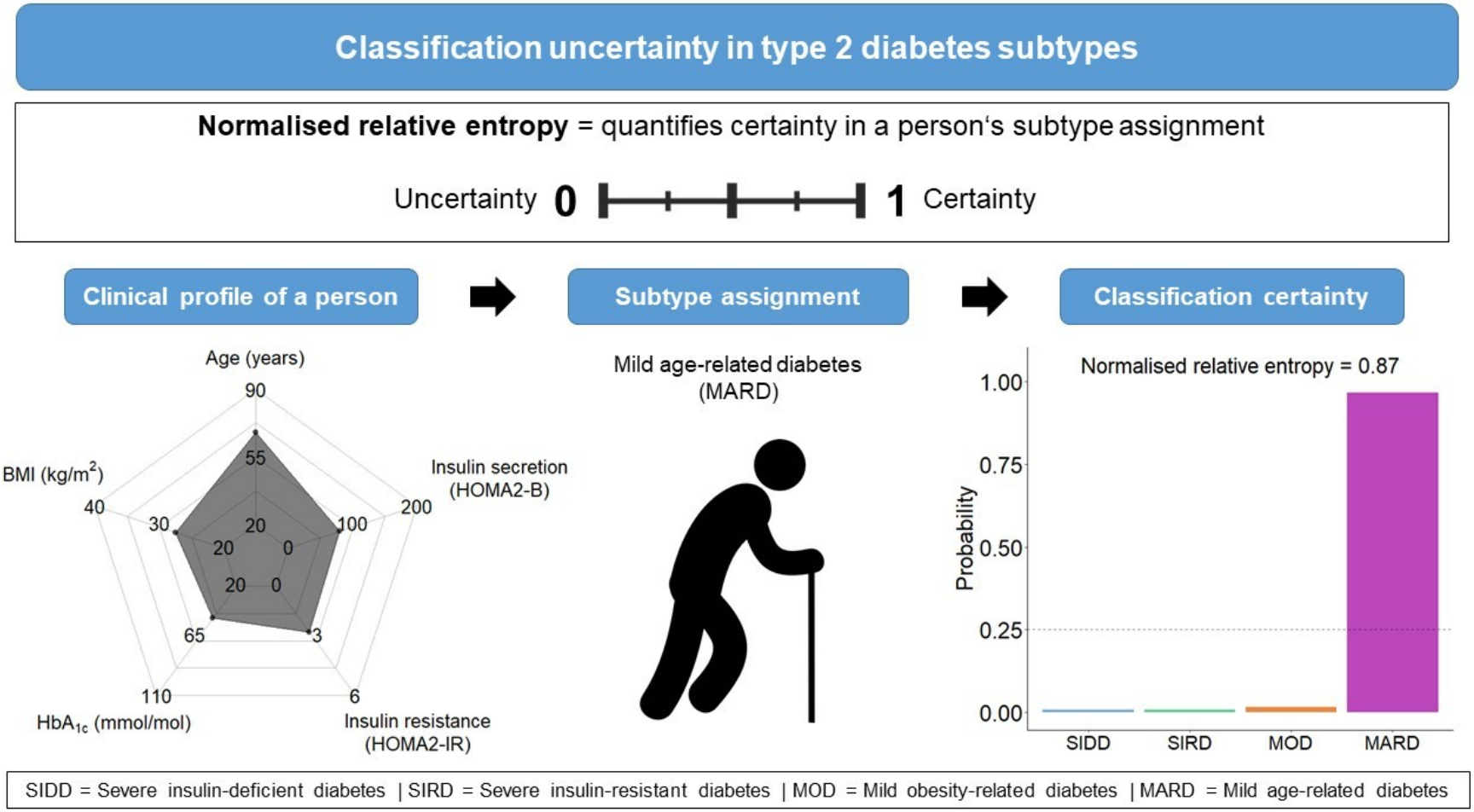

Fig. 2 – Graphical overview of the proposed normalised relative entropy (NRE) method to assess classification uncertainty in type 2 diabetes subtype assignments. Reprinted from Mori et al [38], licensed under CC BY 4.0. https://creativecommons.org/licenses/by/4.0/

This figure has been designed using resources from Flaticon.com. The icon of the elderly person has been designed by Freepik from Flaticon. The scale icon has been made by egorpolyakov from Flaticon.

classification probability (on a scale from 25% to 100%) was 48.8% (95% CI 47.3, 50.2) [38]. These findings indicate that the average classification certainty for subtype assignments was low. In other words, many individuals in the GDS cohort could not be mapped well onto the adult-onset type 2 diabetes subtypes proposed by Ahlqvist et al [20]. These results provide, for the first time, clear empirical evidence supporting concerns about classification uncertainty that several authors had previously raised [8, 14, 31, 32]. For example, Franks and Sargent [31] recently warned that "*many individuals cannot be assigned definitely to a given cluster, meaning that such labelling will likely (...) be incorrect in many patients*" (p. 6). In line with this, our analysis showed that individuals in the GDS cohort with a low NRE often had clinical features that did not match well with their assigned subtype [38]. For example, some were assigned to the MARD subtype despite being younger than 40 years. For these individuals, the label 'age-related' diabetes is clearly misleading.

More broadly, our study adds to the mounting evidence that challenges the utility of the diabetes subtyping approach for precision medicine. Shortly after the original publication by Ahlqvist et al [20], two important methodological limitations of their cluster analysis were highlighted in a response by van Smeden et al [44]. The first problem with this data-driven (or 'unsupervised') cluster analysis is that there is no 'ground truth' available. That is, there is no way of knowing whether the clusters correspond to true, biological subgroups of diabetes. Alternatively, it is also possible that they simply appear because a set of interdependent variables (e.g. BMI, insulin resistance, HbA1c) is entered into the clustering algorithm. In that sense, clustering can be seen as just another way of capturing dependencies among clinical features. However, unlike traditional statistical models, clustering methods like k-means do not come with inferential tools like confidence intervals [26, 45]. In other words, the results of a cluster analysis may suggest certainty where none exists.

The second problem [44] is that clustering inevitably results in a loss of information, as it reduces an individual's precise combination of clinical features to a few broad disease categories. This is disadvantageous for predicting clinical outcomes (i.e., precision prognostics), and it would be more informative to use statistical models that directly incorporate the clinical features as predictors. This conceptual criticism of the type 2 diabetes subtypes has since been substantiated by several empirical studies [16, 40, 46], which are outlined below.

4.2.1 Empirical and simulated evidence against discrete subtypes

Regarding the lack of 'ground truth', Lugner et al [40] examined whether the subclassification into exactly four subtypes could be replicated using data from 114,231 individuals in the National Diabetes Register of Sweden. Following the same k-means clustering procedure as Ahlqvist et al [20], they assessed the optimal number of clusters using statistical criteria (silhouette statistic, gap statistic) and an elbow plot. The latter is a visual method to assess whether adding more clusters leads to more distinct, homogenous subtypes [26]. A 'bend' in the plot suggests that the optimal number of clusters has been reached, as further clusters no longer meaningfully improve how well individuals are grouped. In the analysis of Lugner et al [40], the elbow plot did not show a pronounced bend, making it difficult to justify the choice of any specific number of clusters. Additionally, none of the statistical criteria supported the presence of exactly four clusters. The silhouette statistic suggested two clusters, while the gap statistic indicated only one (i.e., no clustering at all). The latter finding is particularly noteworthy, given that the gap statistic is known to correctly identify the absence of meaningful clusters if, indeed, none are present in the data [26].

The findings above were corroborated in a separate analysis by Aoki et al [46], who assessed the replicability of the subtypes using data from 4,694 individuals in the multinational SAVOR-TIMI 53 cardiovascular outcome trial [47]. In particular, their elbow plot showed a similar, ambiguous shape and the silhouette statistic again suggested two rather than four clusters to be optimal. Given these ambiguities in diabetes classification, it is not surprising that many individuals in the GDS cohort showed low classification certainty and did not align well with their assigned subtype [38]. In particular, the nearest centroid algorithm forcibly assigned them to one of the four subtypes, even though none provided a good fit. Taken together, these findings challenge the proposal of Ahlqvist et al [41] that four clusters provide an optimal precision subclassification of type 2 diabetes.

Setting aside the issue of the optimal number of clusters, the more fundamental question is whether distinct biological subtypes of type 2 diabetes truly exist. Both Lugner et al [40] and Aoki et al [46] carried out simulation studies to assess whether data-driven clustering would identify diabetes subtypes even in the absence of an underlying subgroup structure. Lugner et al [40] simulated data without any subgroups from a multivariate normal distribution based on the empirical covariance matrix from the Swedish national registry data. When they applied k-means clustering to their simulated

data, the resulting elbow plot closely resembled that of their original analysis. In both cases, the elbow plot had an ambiguous shape and did not support a distinct number of clusters.

Aoki et al [46] expanded on this by assessing whether the four subtypes could be replicated solely based on the known correlations between the clustering variables (HbA1c, BMI, age, HOMA-B, and HOMA-IR). Based on the SAVOR-TIMI 53 trial, they simulated data without any underlying subgroups and incrementally introduced the empirical correlations between the clustering variables (e.g. between age of onset and HbA1c). Their results showed that once the correlations were fully introduced, k-means clustering reproduced the four subtypes, even though the data were generated from a single distribution with no inherent groupings. In other words, they demonstrated that the subtypes could have emerged solely from known interdependencies among the clinical features, rather than reflecting distinct biological entities.

4.2.2 Limited prognostic and therapeutic value of subtypes

The second methodological issue, loss of information, has been investigated in several studies comparing the performance of type 2 diabetes subtypes with conventional prediction models [16, 40, 46]. Using Swedish national registry data, Lugner et al [40] compared the prognostic performance of the diabetes subtypes to a Cox model that included the clustering variables as individual predictors. For cardiovascular disease (CVD) events, the Cox model showed superior predictive discrimination, with a concordance index (c-index) [48] of 0.77 (95% CI: 0.76, 0.77) compared to 0.66 (95% CI: 0.65, 0.66) for the subtype approach.

Building on this, Aoki et al [46] showed that it is not even necessary to fit new models using the clustering variables, since established clinical risk scores already provide robust prognostic information. In their analysis of the SAVOR-TIMI 53 trial, they employed the widely used Coronary Risk Evaluation (SCORE) [49] and the Kidney Disease: Improving Global Outcomes (KDIGO) classification [50] to predict CVD and kidney outcomes, respectively. They then compared the prognostic performance of these risk scores to that of the diabetes subtypes. Even though the CVD SCORE was treated as a broad ordinal variable (10-year CVD risk of ≤ 5%, 5-10%, 10-15%), its prognostic performance was similar to that of the subtypes, with c-indices of 0.57 and 0.56, respectively. For kidney events, prediction using the KDIGO classification clearly outperformed the subtype approach with c-indices of 0.62 and 0.54, respectively.

Perhaps the most influential criticism of the subtyping approach comes from Dennis et al [16], who evaluated its clinical utility using data from the ADOPT [51] and RECORD [52] trials. Regarding precision prognostics, they found that age at diagnosis alone was as informative as the subtype classification for predicting glycaemic disease progression in the ADOPT trial (R^2 of 0.09 and 0.08, respectively). This finding is striking, given that the subtype assignment requires five additional clinical features beyond age at diagnosis. That is, the information loss due to the discrete clustering was comparable to omitting five prognostic factors. For kidney disease progression, baseline estimated glomerular filtration rate (eGFR), a routine measure of kidney function, clearly outperformed prediction based on diabetes subtypes (R^2 of 0.18 vs 0.01). While accounting for classification uncertainty, as described above [38], might improve the predictive performance of the subtypes, it still discards valuable information compared to a model that directly includes the clinical features as predictors. Therefore, it is unlikely that accounting for uncertainty in subtype assignment provides a sufficient remedy for the information loss inherent to diabetes subtyping.

Beyond precision prognostics, the key contribution of Dennis et al [16] was their critical assessment of the subtypes' potential for precision treatment. Using the available clinical trial data, they examined whether the subtypes differed in their treatment response and could potentially inform glucose-lowering treatment selection. Although the subtypes differed modestly in their drug response in the ADOPT trial, a model that directly used the clustering variables (without the HOMA measures) as predictors better explained individual treatment response. For example, when predicting response to metformin, the prediction model had an R^2 of 0.35 compared to 0.15 for the subtype approach. In an external validation using the RECORD trial, treatment recommendations based on the prediction model yielded greater benefits than subtype-based recommendations, with an additional HbA1c reduction of 0.36 percentage points compared to 0.16 percentage points. In other words, for precision treatment, the diabetes subtypes were outperformed by a standard prediction model that only required routinely collected clinical features.

4.3 Individualised prediction models as an alternative to type 2 diabetes subtyping

Following their seminal work [16], Dennis and colleagues continued to develop and apply the so-called 'individualised prediction' framework for precision treatment of type 2

diabetes [17, 53-56]. In this framework [53], the aim is to use simple clinical features to directly model treatment response (i.e., precision treatment) rather than to identify distinct aetiologies of diabetes (i.e., precision diagnosis). In other words, the goal is to identify the optimal treatment for an individual based on their continuous clinical features, without requiring an explicit subtype assignment. Note that, according to the PMDI nomenclature [6], this constitutes a precision medicine rather than a personalised (or 'individualised') medicine approach. Unlike personalised medicine approaches, this framework does not rely on the analysis of data from a single person to generate recommendations. Rather, the prediction model is developed in a reference population and subsequently applied to generate individual treatment recommendations [53].

Conceptually, the individualised prediction framework requires the identification of features that are associated with differential treatment effects of glucose-lowering drugs (i.e. drug-by-feature interactions). To ensure applicability in clinical practice, Dennis [53] emphasises the use of routinely available clinical features, such as age, BMI, and eGFR. In terms of data analysis, the framework proposes to utilise both observational and clinical trial data on glucose-lowering medications. Specifically, electronic health record databases can be used in an initial discovery step to identify potential drug-by-feature interactions in a heterogeneous, real-world population. Subsequently, data from existing trials can be used to replicate these interactions in a well-controlled setting with randomised treatment assignments. Dennis [53] refers to this two-step approach as triangulation of observational and clinical trial data. The choice of the specific modelling approach is flexible and may include multiple regression models [16, 17, 56], flexible Bayesian models [54] or machine-learning models [55].

So far, applications of the individualised prediction approach have focused on treatment selection models (or 'treatment algorithms') for second- and third-line glucose-lowering drugs for type 2 diabetes [17, 54, 56]. For example, Dennis and colleagues developed models to optimise prescribing of sodium-glucose cotransporter 2 inhibitors (SGLT2i) versus dipeptidyl peptidase-4 inhibitors (DPP-4i) [17] and glucagon-like peptide-1 receptor agonists (GLP1-RA) [54]. The most recent advance is the development of a 'five-drug class model' [56]. This model allows simultaneous prediction of an individual's 12-month HbA1c for any of the five oral glucose-lowering drugs commonly prescribed in primary care (SGLT2i, DPP-4i, GLP1-RA, sulfonylurea, and thiazolidinedione). It was developed using observational primary care data from the UK Clinical Practice Research Datalink (CPRD), comprising more than 100,000 drug

initiations. Nine routine clinical features (age, duration of diabetes, sex, HbA1c, BMI, eGFR, HDL cholesterol, total cholesterol, and alanine aminotransferase) were included as predictors of differential treatment response. The modelling approach was a flexible linear regression model using five-knot restricted cubic splines. Model validation was performed in an independent CPRD validation cohort as well as in three clinical trials, one of which was a three-way cross-over trial [18].

In the CPRD cohort, the five-drug class model predicted most individuals to have an optimal 12-month HbA1c response on GLP1-RA (33.4%), SGLT2i (28.9%), or sulfonylureas (27.6%) [56]. Treatment in line with these recommendations was associated with an additional HbA1c reduction of 0.50 percentage points compared to non-optimal treatment. The model-predicted HbA1c benefits showed good calibration in external validation, although they were smaller in the more selective clinical trial populations (average additional HbA1c reductions of 0.22 to 0.24 percentage points). Overall, the five-drug class model performed well at predicting optimal glucose-lowering treatment, with benefits also extending to cardiorenal outcomes during the 5-year follow-up in the observational CPRD data.

The key difference between the individualised prediction approach and the diabetes subtyping approach lies in how individual information is leveraged for precision medicine. A graphical comparison of the two approaches, including their advantages and disadvantages, is shown in Figure 3 (reprinted with permission from Dennis [53]). In the subtyping approach, an individual's subtype is usually determined at a given time point (e.g. at the time of diagnosis) and this categorisation then serves as the basis for precision prognostics or precision treatment [20, 53]. Consequently, this approach makes the assumption that treatment response is homogeneous for all individuals assigned to a given subtype [53]. However, based on our analysis of classification uncertainty in the GDS cohort [38], and given the observed lack of fit for some individuals, this assumption appears difficult to justify. The individualised prediction approach avoids this problem by directly using an individual's clinical information at the time of treatment decision to predict the optimal treatment [53]. Since the individualised prediction approach relies only on routinely collected clinical features (i.e. not requiring HOMA or C-peptide) it can be implemented at low costs in most healthcare systems worldwide [56].

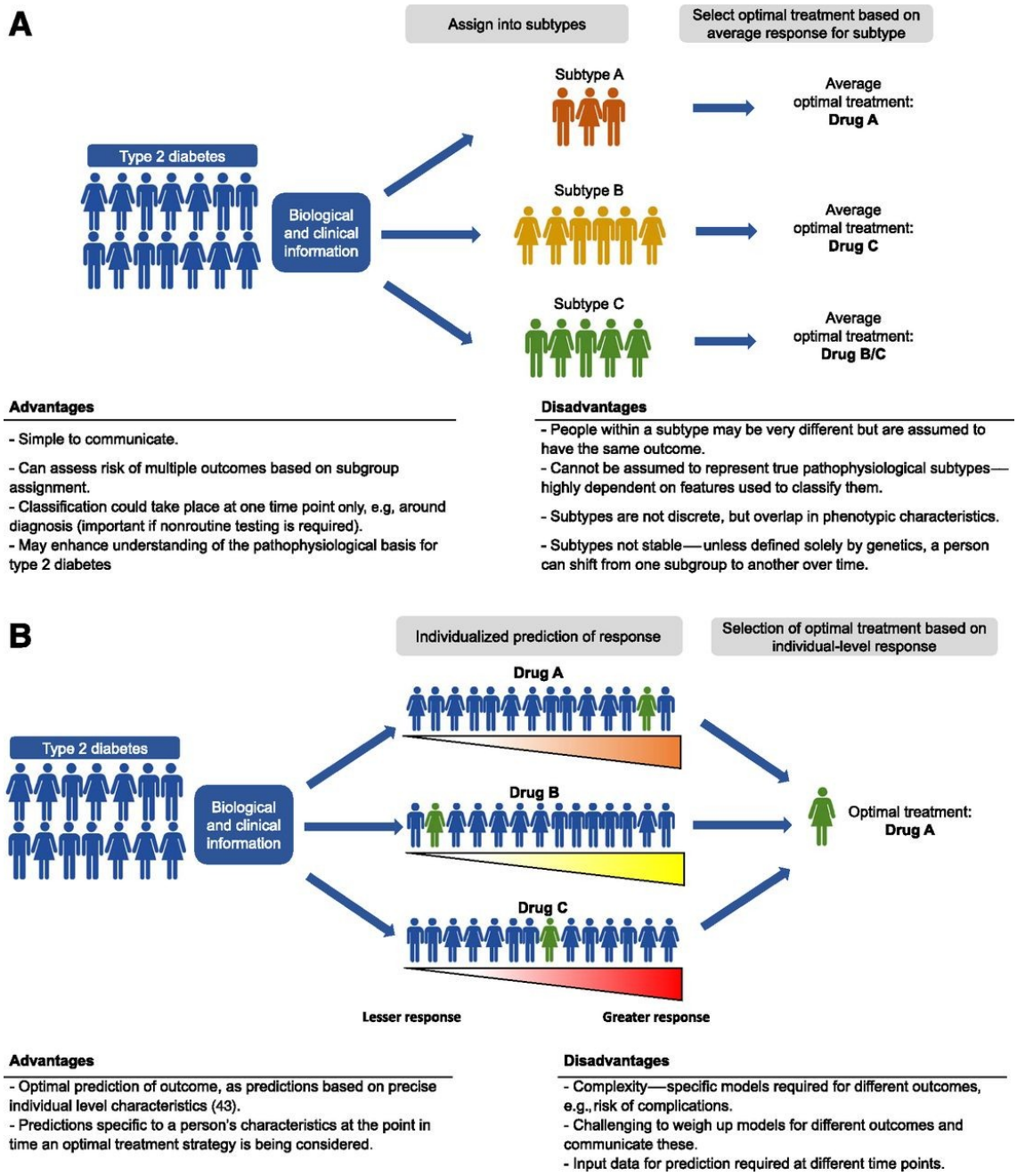

Fig. 3 – Individualised prediction compared with classification into subtypes: advantages and disadvantages of two strategies to apply a precision medicine approach in type 2 diabetes. A) Classification into subtypes. B) Individualised prediction. Reprinted with permission from Dennis [53]. © 2020 American Diabetes Association.

In terms of clinical utility, the five-drug class model [56] offers the advantage of simultaneously considering five drug classes, as opposed to previous models that only compared two at a time [17, 54]. This makes it particularly relevant as a decision aid for primary care physicians, who must navigate an increasingly complex landscape of glucose-lowering medications [57]. Indeed, the second PMDI consensus report highlights that such individualised prediction models are a key future research priority [14]. In

contrast, a recent clinical trial of individuals classified as SIDD and SIRD found no differential treatment response to GLP1-RA and SGLT2i [58]. In other words, the subtype classification did not translate into clinical utility for selecting glucose-lowering treatment. Taken together, these developments position the individualised prediction approach as a promising alternative to subtype-based stratification, particularly for guiding treatment decisions in clinical practice.

4.4 Challenges to the precision medicine paradigm

While the second PMDI consensus report [14] acknowledges the potential of individualised prediction models, it also emphasises that the overall evidence base for precision medicine is, as yet, weak. In particular, the report highlights the poor reporting quality of many precision medicine studies, which precludes conclusions about the validity and clinical utility of proposed methods. For example, many studies fail to report statistical tests for heterogeneity, statistical power to detect interaction effects, or effect estimates on a clinically relevant scale [14]. In response, an expert consortium recently published the BePRECISE guidelines for precision medicine research [59] to support standardised reporting with an emphasis on clinical relevance.

Beyond reporting practices, the second PMDI consensus report [14] raises concerns about the lack of adequate validation studies. Specifically, it emphasises the limited availability of prospectively designed clinical trials validating proposed precision medicine approaches. A notable exception is the TriMaster study [18], a three-way cross-over trial that demonstrated differential responses to glucose-lowering medications depending on individuals' kidney function and BMI. However, such precision medicine trials remain scarce, and the PMDI consensus report concludes that the clinical translation of precision medicine in diabetes care is, at present, "*largely aspirational*" (p. 2450) [14].

In the absence of precision medicine trials, an alternative approach is to perform meta-analyses of existing clinical trials to quantify the potential of precision medicine [60-62]. In particular, this can be examined by comparing variability in the clinical outcome (e.g., the standard deviation of HbA1c) between treatment and placebo arms following the intervention period [60]. The underlying idea is that if heterogeneous treatment effects (i.e., drug-by-person interactions) exist, this should induce increased variability in the treatment group relative to the placebo group. Kuß et al [60] applied this approach in a meta-regression of 174 clinical trials of glucose-lowering medications, comparing HbA1c

variability between treatment and placebo arms in individuals with type 2 diabetes. They found that HbA1c variability was only minimally increased in the treatment arm, suggesting the absence of clinically relevant treatment effect heterogeneity. In subsequent studies, Kuß and colleagues applied the same approach to analyse variability in body weight reduction [61] and all-cause mortality [62] in clinical trials of glucose-lowering treatment. Again, they found no consistent evidence of clinically relevant treatment effect heterogeneity. Taken together, these findings based on summary-level data from numerous clinical trials provide indirect evidence against the premise of precision treatment in type 2 diabetes.

Besides these empirical findings, the precision medicine paradigm for type 2 diabetes has also been challenged from a clinical perspective. In a critical review, Griffin [63] questions the focus on HbA1c as an outcome, noting that it is only an intermediate marker and that clinical burden arises from complications such as a cardiovascular disease. He argues that instead of aiming for marginal improvements in HbA1c, greater impact could be achieved by improving on the lacking implementation of existing cardiovascular risk guidelines. He also highlights that medication non-adherence is a major barrier in clinical practice, which is often overlooked. However, individualised prediction models have so far been unsuccessful at predicting short-term treatment discontinuation [64]. Instead, Griffin [63] argues that clinicians should prioritise factors such as dosing regimens, individual beliefs, and reminder systems to support adherence. Lastly, given the societal dimension of the diabetes pandemic, he suggests that the greatest gains may come not from precision medicine but from population-level policies that promote healthy diets and physical activity.

4.5 Limitations and recommendations for future research

The work presented in this dissertation has a number of limitations. First, the focus was on disease heterogeneity only after diabetes had been diagnosed. However, recent studies indicate that individuals already differ during the earlier stage of impaired glucose tolerance, suggesting a potential avenue for precision prevention [65]. Indeed, it would be interesting to apply the NRE measure of classification uncertainty to the recently proposed subtypes for individuals with an elevated risk of type 2 diabetes [65]. Since this approach also uses hard clustering, it is reasonable to expect results similar to those reported here for type 2 diabetes subtypes.

Second, the current work concentrated on challenges in assigning type 2 diabetes subtypes. However, classification of adult-onset type 1 diabetes presents its own difficulties. For example, different antibodies besides GADA may be measured [1, 27], and the determination of autoimmunity may depend on the selected cut-off level for the antibody assay [5].

Third, the discussion of disease heterogeneity was restricted to the phenotypic level, focusing on clinical variables rather than genetic information [8]. The potential of genetic testing for precision medicine has been demonstrated in rare monogenic forms of diabetes, where insulin treatment may be ceased in favour of oral treatment with sulfonylurea [14]. However, type 2 diabetes is a complex polygenic disease [8], and genetic testing has thus far shown limited clinical utility [53, 63].

A fourth limitation is that 'subjective' personalisation approaches [15], which emphasise individuals' needs and preferences, were beyond the scope of this work. Nonetheless, recent findings from the TriMaster study suggest that individual preferences for glucose-lowering medications vary and can successfully inform treatment selection [19]. Finally, it is important to acknowledge that the analysis of classification uncertainty presented here was performed in a European population, though the global burden of diabetes is growing most rapidly in other regions [14].

As for future research, an important question is how, and indeed whether, the field should proceed with diabetes subtyping. The findings on classification uncertainty in type 2 diabetes subtypes [38], alongside existing evidence on clinical [8, 16, 58] and methodological [40, 44, 46] limitations, raise doubts about whether subtyping is the most promising path forward. Indeed, as early as 2006, Gale [5] cautioned against an excessive focus on disease classification, arguing that "*Our routine attempt to assign patients to one or other type of diabetes is (…) neither entirely rational, nor measurable, nor useful.*" (p. 1994). In line with this perspective, current evidence suggests that greater gains may be achieved by further developing and validating individualised prediction models [56], as also emphasised by the second PMDI consensus report [14]. Nonetheless, if researchers choose to continue working with diabetes subtypes, it is important to remain mindful of their limitations and to report measures of classification uncertainty, such as the NRE measure proposed here [38].

Given the currently weak evidence base for precision medicine [14], the primary objective for future research should be to generate high-quality evidence from appropriately designed clinical trials [60, 66, 67]. In particular, repeated cross-over trials

(or ‘N-of-1-trials’), in which each individual receives at least one of the treatments twice, are needed [68, 69]. Only these designs can unequivocally demonstrate whether, and to what extent, individuals truly differ in their responses to glucose-lowering medications [60, 62]. While no such trials have yet been conducted for glucose-lowering medications, a successful example has recently been published for blood-pressure-lowering treatment [70]. In the absence of such evidence for diabetes treatment, the value of ongoing efforts to identify phenotypic, genetic, or omics-based strategies for precision medicine remains uncertain [66, 67]. Consequently, established standards of care [12] remain central to improving outcomes for people living with diabetes.

As for the individualised prediction approach, a pragmatic evaluation strategy would be a cluster-randomised trial, where some healthcare providers receive the treatment algorithm and others do not [53]. Such a design would enable a comparison of this precision medicine approach with current standards of care in a routine clinical practice setting [53, 63]. Regardless of the study design, future research should adhere to the BePRECISE guidelines [59] to ensure standardised reporting and facilitate evaluation of the clinical utility of proposed precision medicine strategies.

4.6 Conclusions

In summary, this work provides two important advancements in the field of diabetes subtyping. The DDZ Diabetes-Cluster-Tool is the first user-friendly software solution to enable diabetes subtyping in clinical practice. Moreover, using the normalised relative entropy measure, this is the first study to systematically evaluate classification uncertainty in type 2 diabetes subtypes.

The analysis of recently diagnosed adults from the GDS revealed substantial uncertainty in subtype assignments, particularly for the severe subtypes. Beyond this methodological challenge, the practical implementation of subtyping in routine clinical care may be limited by the requirement for C-peptide or insulin measurements. Taken together, these findings suggest that the proposed reclassification of diabetes may offer only limited value for precision medicine.

As an alternative approach, individualised prediction models have recently shown promise in guiding precision treatment. However, additional evidence from precision medicine trials (e.g. repeated cross-over or cluster-randomised trials) is needed before implementation in clinical practice can be considered. Until further evidence emerges,

clinical practice should continue to prioritise established standards of care to improve the lives of people living with diabetes.

5 References

[1] American Diabetes Association Professional Practice Committee (2025) 2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2025. Diabetes Care 48(Supplement_1): S27-S49. https://doi.org/10.2337/dc25-S002

[2] Tamayo T, Brinks R, Hoyer A, Kuß O, Rathmann W (2016) The prevalence and incidence of diabetes in Germany: an analysis of statutory health insurance data on 65 million individuals from the years 2009 and 2010. Deutsches Ärzteblatt International 113(11): 177-182. https://doi.org/10.3238/arztebl.2016.0177

[3] Forbes JM, Cooper ME (2013) Mechanisms of diabetic complications. Physiol Rev 93(1): 137-188. https://doi.org/10.1152/physrev.00045.2011

[4] Voeltz D, Vetterer M, Seidel-Jacobs E, Brinks R, Tönnies T, Hoyer A (2024) Projecting the economic burden of type 1 and type 2 diabetes mellitus in Germany from 2010 until 2040. Population Health Metrics 22(1): 17. https://doi.org/10.1186/s12963-024-00337-x

[5] Gale EAM (2006) Declassifying diabetes. Diabetologia 49(9): 1989-1995. https://doi.org/10.1007/s00125-006-0348-7

[6] Chung WK, Erion K, Florez JC, et al. (2020) Precision medicine in diabetes: a consensus report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes care 43(7): 1617-1635. https://doi.org/10.2337/dci20-0022

[7] Gale EAM (2013) Is type 2 diabetes a category error? The Lancet 381(9881): 1956-1957. https://doi.org/10.1016/S0140-6736(12)62207-7

[8] Pearson ER (2019) Type 2 diabetes: a multifaceted disease. Diabetologia 62(7): 1107-1112. https://doi.org/10.1007/s00125-019-4909-y

[9] Misra S, Wagner R, Ozkan B, et al. (2023) Precision subclassification of type 2 diabetes: a systematic review. Commun Med 3(1): 138. https://doi.org/10.1038/s43856-023-00360-3

[10] McCarthy MI (2017) Painting a new picture of personalised medicine for diabetes. Diabetologia 60(5): 793-799. https://doi.org/10.1007/s00125-017-4210-x

[11] Schön M, Prystupa K, Mori T, et al. (2024) Analysis of type 2 diabetes heterogeneity with a tree-like representation: insights from the prospective German Diabetes Study and the LURIC cohort. Lancet Diabetes Endocrinol 12(2): 119-131. https://doi.org/10.1016/S2213-8587(23)00329-7

[12] Davies MJ, Aroda VR, Collins BS, et al. (2022) Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes care 45(11): 2753-2786. https://doi.org/10.2337/dci22-0034

[13] Krook A, Mulder H (2022) Pinpointing precision medicine for diabetes mellitus. Diabetologia 65(11): 1755-1757. https://doi.org/10.1007/s00125-022-05777-4

[14] Tobias DK, Merino J, Ahmad A, et al. (2023) Second international consensus report on gaps and opportunities for the clinical translation of precision diabetes medicine. Nat Med 29(10): 2438-2457. https://doi.org/10.1038/s41591-023-02502-5

[15] Franks PW, Cefalu WT, Dennis J, et al. (2023) Precision medicine for cardiometabolic disease: a framework for clinical translation. Lancet Diabetes Endocrinol 11(11): 822-835. https://doi.org/10.1016/S2213-8587(23)00165-1

[16] Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT (2019) Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial

data. Lancet Diabetes Endocrinol 7(6): 442-451. https://doi.org/10.1016/S2213-8587(19)30087-7
[17] Dennis JM, Young KG, McGovern AP, et al. (2022) Development of a treatment selection algorithm for SGLT2 and DPP-4 inhibitor therapies in people with type 2 diabetes: a retrospective cohort study. The Lancet Digital Health 4(12): e873-e883. https://doi.org/10.1016/S2589-7500(22)00174-1
[18] Shields BM, Dennis JM, Angwin CD, et al. (2023) Patient stratification for determining optimal second-line and third-line therapy for type 2 diabetes: the TriMaster study. Nat Med 29(2): 376-383. https://doi.org/10.1038/s41591-022-02120-7
[19] Shields BM, Angwin CD, Shepherd MH, et al. (2023) Patient preference for second-and third-line therapies in type 2 diabetes: a prespecified secondary endpoint of the TriMaster study. Nat Med 29(2): 384-391. https://doi.org/10.1038/s41591-022-02121-6
[20] Ahlqvist E, Storm P, Käräjämäki A, et al. (2018) Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol 6(5): 361-369. https://doi.org/10.1016/s2213-8587(18)30051-2
[21] Herder C, Roden M (2022) A novel diabetes typology: towards precision diabetology from pathogenesis to treatment. Diabetologia 65(11): 1770-1781. https://doi.org/10.1007/s00125-021-05625-x
[22] Prando R, Cheli V, Melga P, Giusti R, Ciuchi E, Odetti P (1998) Is Type 2 Diabetes a Different Disease in Obese and Nenobese Patients? Diabetes Care 21(10): 1680-1685. https://doi.org/10.2337/diacare.21.10.1680
[23] Bancks MP, Chen H, Balasubramanyam A, et al. (2021) Type 2 Diabetes Subgroups, Risk for Complications, and Differential Effects Due to an Intensive Lifestyle Intervention. Diabetes Care 44(5): 1203-1210. https://doi.org/10.2337/dc20-2372
[24] Udler MS, Kim J, von Grotthuss M, et al. (2018) Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS Med 15(9): e1002654. https://doi.org/10.1371/journal.pmed.1002654
[25] Wallace TM, Levy JC, Matthews DR (2004) Use and abuse of HOMA modeling. Diabetes care 27(6): 1487-1495. http://doi.org/10.2337/diacare.27.6.1487
[26] Hastie T, Tibshirani R, Friedman J (2009) Cluster Analysis. In: The elements of statistical learning: data mining, inference, and prediction. Vol 2. Springer, New York, pp 501-527
[27] Zaharia OP, Strassburger K, Strom A, et al. (2019) Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. Lancet Diabetes Endocrinol 7(9): 684-694. https://doi.org/10.1016/S2213-8587(19)30187-1
[28] Zou X, Zhou X, Zhu Z, Ji L (2019) Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. Lancet Diabetes Endocrinol 7(1): 9-11. https://doi.org/10.1016/S2213-8587(18)30316-4
[29] Tanabe H, Saito H, Kudo A, et al. (2020) Factors Associated with Risk of Diabetic Complications in Novel Cluster-Based Diabetes Subgroups: A Japanese Retrospective Cohort Study. Journal of Clinical Medicine 9(7): 2083. https://doi.org/10.3390/jcm9072083
[30] R Core Team (2010) R: A Language and Envrionment for Statistical Computing. R foundation for Statistical Computing, Vienna, Austria

[31] Franks PW, Sargent JL (2024) Diabetes and obesity: leveraging heterogeneity for precision medicine. Eur Heart J 45(48): 5146-5155. https://doi.org/10.1093/eurheartj/ehae746
[32] Leslie RD, Ma RCW, Franks PW, Nadeau KJ, Pearson ER, Redondo MJ (2023) Understanding diabetes heterogeneity: key steps towards precision medicine in diabetes. Lancet Diabetes Endocrinol 11(11): 848-860. https://doi.org/10.1016/S2213-8587(23)00159-6
[33] Kahkoska AR, Geybels MS, Klein KR, et al. (2020) Validation of distinct type 2 diabetes clusters and their association with diabetes complications in the DEVOTE, LEADER and SUSTAIN-6 cardiovascular outcomes trials. Diabetes Obes Metab 22(9): 1537-1547. http://doi.org/10.1111/dom.14063
[34] Tanabe H, Sato M, Miyake A, et al. (2024) Machine learning-based reproducible prediction of type 2 diabetes subtypes. Diabetologia 67(11): 2446-2458. https://doi.org/10.1007/s00125-024-06248-8
[35] Wesolowska-Andersen A, Brorsson CA, Bizzotto R, et al. (2022) Four groups of type 2 diabetes contribute to the etiological and clinical heterogeneity in newly diagnosed individuals: An IMI DIRECT study. Cell Rep Med 3(1): 100477. https://doi.org/10.1016/j.xcrm.2021.100477
[36] Naggara O, Raymond J, Guilbert F, Roy D, Weill A, Altman DG (2011) Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms. Am J Neuroradiol 32(3): 437-440. http://doi.org/10.3174/ajnr.A2425
[37] Mori T, Prystupa K, Straßburger K, et al. (2025) A web-based application for diabetes subtyping: The DDZ Diabetes-Cluster-Tool. Acta Diabetol 62(2): 281-283. https://doi.org/10.1007/s00592-024-02436-5
[38] Mori T, Zaharia OP, Straßburger K, et al. (2025) Recognising, quantifying and accounting for classification uncertainty in type 2 diabetes subtypes. Diabetologia: Online ahead of print. https://doi.org/10.1007/s00125-025-06486-4
[39] Slieker RC, Donnelly LA, Fitipaldi H, et al. (2021) Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study. Diabetologia 64(9): 1982-1989. https://doi.org/10.1007/s00125-021-05490-8
[40] Lugner M, Gudbjörnsdottir S, Sattar N, et al. (2021) Comparison between data-driven clusters and models based on clinical features to predict outcomes in type 2 diabetes: nationwide observational study. Diabetologia 64(9): 1973-1981. http://doi.org/10.1007/s00125-021-05485-5
[41] Ahlqvist E, Prasad RB, Groop L (2022) 100 YEARS OF INSULIN: Towards improved precision and a new classification of diabetes mellitus. J Endocrinol 252(3): R59-R70. https://doi.org/10.1530/joe-20-0596
[42] Anjana RM, Baskar V, Nair ATN, et al. (2020) Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: a data-driven cluster analysis: the INSPIRED study. BMJ Open Diabetes Research and Care 8(1): e001506. https://doi.org/10.1136/bmjdrc-2020-001506
[43] Misra S, Aguilar-Salinas CA, Chikowore T, et al. (2023) The case for precision medicine in the prevention, diagnosis, and treatment of cardiometabolic diseases in low-income and middle-income countries. Lancet Diabetes Endocrinol 11(11): 836-847. https://doi.org/10.1016/S2213-8587(23)00164-X
[44] Van Smeden M, Harrell FE, Dahly DL (2018) Novel diabetes subgroups. Lancet Diabetes Endocrinol 6(6): 439-440. https://doi.org/10.1016/S2213-8587(18)30124-4

[45] Toffalini E, Gambarota F, Perugini A, et al. (2024) Clusters that are not there: An R tutorial and a Shiny app to quantify a priori inferential risks when using clustering methods. Int J Psychol 59(6): 1183-1198. https://doi.org/10.1002/ijop.13246
[46] Aoki Y, Hamrén B, Clegg LE, et al. (2021) Assessing reproducibility and utility of clustering of patients with type 2 diabetes and established CV disease (SAVOR-TIMI 53 trial). PLoS One 16(11): e0259372. https://doi.org/10.1371/journal.pone.0259372
[47] Scirica BM, Bhatt DL, Braunwald E, et al. (2013) Saxagliptin and Cardiovascular Outcomes in Patients with Type 2 Diabetes Mellitus. N Engl J Med 369(14): 1317-1326. https://doi.org/10.1056/NEJMoa1307684
[48] Harrell FE, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine 15(4): 361-387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
[49] Conroy RM, Pyörälä K, Fitzgerald AP, et al. (2003) Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 24(11): 987-1003. https://doi.org/10.1016/S0195-668X(03)00114-3
[50] Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney inter., Suppl. 2012; 2:1–138. https://doi.org/10.1159/000339789
[51] Kahn SE, Haffner SM, Heise MA, et al. (2006) Glycemic Durability of Rosiglitazone, Metformin, or Glyburide Monotherapy. N Engl J Med 355(23): 2427-2443. https://doi.org/10.1056/NEJMoa066224
[52] Home PD, Pocock SJ, Beck-Nielsen H, et al. (2009) Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. The Lancet 373(9681): 2125-2135. https://doi.org/10.1016/S0140-6736(09)60953-3
[53] Dennis JM (2020) Precision medicine in type 2 diabetes: using individualized prediction models to optimize selection of treatment. Diabetes 69(10): 2075-2085. https://doi.org/10.2337/dbi20-0002
[54] Cardoso P, Young KG, Nair AT, et al. (2024) Phenotype-based targeted treatment of SGLT2 inhibitors and GLP-1 receptor agonists in type 2 diabetes. Diabetologia 67(5): 822-836. https://doi.org/10.1007/s00125-024-06099-3
[55] Venkatasubramaniam A, Mateen BA, Shields BM, et al. (2023) Comparison of causal forest and regression-based approaches to evaluate treatment effect heterogeneity: an application for type 2 diabetes precision medicine. BMC Med Inform Decis Mak 23(1): 110. https://doi.org/10.1186/s12911-023-02207-2
[56] Dennis JM, Young KG, Cardoso P, et al. (2025) A five-drug class model using routinely available clinical features to optimise prescribing in type 2 diabetes: a prediction model development and validation study. The Lancet 405(10480): 701-714. https://doi.org/10.1016/S0140-6736(24)02617-5
[57] Seidu S, Ikhile D (2025) Personalised type 2 diabetes therapy in primary care. The Lancet 405(10480): 672-674. https://doi.org/10.1016/S0140-6736(24)02844-7
[58] Dwibedi C, Ekström O, Brandt J, et al. (2024) Randomized open-label trial of semaglutide and dapagliflozin in patients with type 2 diabetes of different pathophysiology. Nat Metab 6(1): 50-60. https://doi.org/10.1038/s42255-023-00943-3
[59] Lim SS, Semnani-Azad Z, Morieri ML, et al. (2024) Reporting guidelines for precision medicine research of clinical relevance: the BePRECISE checklist. Nat Med 30(7): 1874-1881. https://doi.org/10.1038/s41591-024-03033-3
[60] Kuss O, Opitz ME, Brandstetter LV, Schlesinger S, Roden M, Hoyer A (2023) How amenable is type 2 diabetes treatment for precision diabetology? A meta-

regression of glycaemic control data from 174 randomised trials. Diabetologia 66(9): 1622-1632. https://doi.org/10.1007/s00125-023-05951-2
[61] Vargas KG, Rütten T, Siemes B, et al. (2024) Assessing the potential for precision medicine in body weight reduction with regard to type 2 diabetes mellitus therapies: A meta-regression analysis of 120 randomized controlled trials. Diabetes Obes Metab 26(6): 2139-2146. https://doi.org/10.1111/dom.15519
[62] Kuss O, Roden M, Schlesinger S, Hoyer A (2024) The potential of precision diabetology for type 2 diabetes treatment—evidence from a meta-regression for all-cause mortality from large cardiovascular outcome trials. Acta Diabetol 62(7): 1073-1080. https://doi.org/10.1007/s00592-024-02425-8
[63] Griffin S (2022) Diabetes precision medicine: plenty of potential, pitfalls and perils but not yet ready for prime time. Diabetologia 65(11): 1913-1921. https://doi.org/10.1007/s00125-022-05782-7
[64] Cardoso P, Young KG, Hopkins R, et al. (2025) Evaluating prediction of short-term tolerability of five type 2 diabetes drug classes using routine clinical features: UK population-based study. Diabetes Obes Metab 27(8): 4320-4329. https://doi.org/10.1111/dom.16470
[65] Wagner R, Heni M, Tabak AG, et al. (2021) Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. Nat Med 27(1): 49-57. https://doi.org/10.1038/s41591-020-1116-9
[66] Gewandter JS, McDermott MP, He H, et al. (2019) Demonstrating heterogeneity of treatment effects among patients: an overlooked but important step toward precision medicine. Clin Pharmacol Ther 106(1): 204-210. https://doi.org/10.1002/cpt.1372
[67] Dworkin RH, McDermott MP, Farrar JT, O'Connor AB, Senn S (2014) Interpreting patient treatment response in analgesic clinical trials: implications for genotyping, phenotyping, and personalized pain treatment. Pain 155(3): 457-460. https://doi.org/10.1016/j.pain.2013.09.019
[68] Senn S (2016) Mastering variation: variance components and personalised medicine. Statistics in medicine 35(7): 966-977. https://doi.org/10.1002/sim.6739
[69] Senn S (2018) Statistical pitfalls of personalized medicine. Nature 563(7733): 619-621. https://doi.org/10.1038/d41586-018-07535-2
[70] Sundström J, Lind L, Nowrouzi S, et al. (2023) Heterogeneity in Blood Pressure Response to 4 Antihypertensive Drugs: A Randomized Clinical Trial. JAMA 329(14): 1160-1169. https://doi.org/10.1001/jama.2023.3322

6 Supplement

Online supplement from:

Mori T, Zaharia OP, Straßburger K, Dennis JM, Mai K, Kabisch S, Bornstein S, Szendroedi J, Blüher M, Meyhöfer S, Seissler J, Birkenfeld A, Stefan N, Roden M, Wagner R, Kuß O, and the GDS group 2025. Recognising, quantifying and accounting for classification uncertainty in type 2 diabetes subtype. *Diabetologia.* Online ahead of print. https://doi.org/10.1007/s00125-025-06486-4

Danksagung

Zuallererst möchte ich Prof. Oliver Kuß für die Betreuung meiner Doktorarbeit danken. Ich weiß das Vertrauen, das Sie mir stets entgegengebracht haben, sehr zu schätzen und bin jedes Mal bestärkt aus unseren ‚Update-Meetings' gegangen. Ich bin sehr froh, dass ich von Ihnen als Vorgesetzten in den letzten drei Jahren nicht nur statistisch, sondern auch menschlich vieles lernen konnte.

Als Nächstes möchte ich mich herzlich bei Prof. Annika Hoyer für die Übernahme der Zweitbetreuung dieser Doktorarbeit bedanken. Es hat mich sehr gefreut, dass wir im Rahmen der Doktorarbeit sowie durch die GMDS in fachlichem Kontakt standen.

Dr. John Dennis möchte ich dafür danken, dass er mich im vergangenen Jahr als Mentor begleitet und mir einen Forschungsaufenthalt bei seinem hervorragenden Team in Exeter ermöglicht hat.

Darüber hinaus möchte ich allen Ko-Autor:innen danken, die mit mir gemeinsam an diesen (und anderen) Publikationen gearbeitet haben. Stellvertretend möchte ich an dieser Stelle Prof. Robert Wagner vom Klinischen Studienzentrum nennen, der mir die Mitarbeit an vielen spannenden Projekten ermöglichte.

Edyta Schaefer und Dr. Nina Trinks danke ich vielmals für die kritische Durchsicht dieser Arbeit.

Prof. Jill Jacobson möchte ich für ihre Statistikvorlesung danken, die mich damals dazu inspiriert hat, eine Karriere auf diesem Fachgebiet anzustreben.

Zuletzt möchte ich meiner Familie und vor allem meiner Mutter dafür danken, dass sie mich auf meinem (Bildungs-)Weg immer unterstützt hat.