University of Twente

Master Thesis

Big Data-Driven Personalized and Preventive Diabetes Care

MSc Business Administration - Digital Business & Analytics
Faculty of Behavioural, Management and Social Sciences
Student: Manolya Nur Kara
Supervisor: dr. ir. A.A.M. Spil
2nd Supervisor: dr. M.R Stienstra

March 18, 2024

Abstract

This study explores the usage of big data in personalized and preventive diabetes care. The increasing number of people affected by diabetes poses a significant global health concern. Technology is rapidly evolving, and big data represents a vital part of digital health technologies, offering significant potential for addressing this issue. Employing both qualitative and quantitative methodologies, we conducted a survey to collect data, followed by interviews with experts in the field. Data analysis was performed to examine insights from the survey, followed by the Gioia methodology for the interviews. Results indicate the importance of lifestyle factors in preventing diabetes, emphasizing the importance of adopting a healthy lifestyle. Furthermore, personalized care seems to be not commonly used in practice, due to the need for its development and the governmental impediments. Moreover, while big data holds potential for personalization and prevention, it is important to acknowledge and overcome the numerous challenges that persist. Our study highlights the importance of personalized interventions, such as coaching, in raising awareness, preventing diabetes and providing tailored interventions with the aid of big data. Additionally, governmental involvement is crucial for using personalized care and fostering awareness of a healthy lifestyle. Moving forward, investment in research and development is necessary to address existing challenges and facilitate the successful integration of big data into diabetes care.

Keywords: big data, healthcare, diabetes, personalization, prevention

Contents

1 Introduction

Nearly 500 million individuals worldwide are living with diabetes, and this number is expected to surge by 25% in 2030 [16]. The increasing prevalence of diabetes and its associated health issues underscores a significant global health problem. Sedentary lifestyles and unhealthy dietary choices are associated with the risk of diabetes [9]. In this context, over the years digital health technologies have been used for more specialized goals, such as diabetes prevention [13]. Furthermore, various digital health technologies play a role in personalizing medicine, such as mobile apps, wearable devices and big data analytics [10]. Ever since the introduction of the electronic medical record (EMR), healthcare professionals and researchers have embraced big data as the solution for handling the vast quantities of data produced by patients in today's healthcare environment [2]. Diabetes in young adults, nominally aged 18–30, is a more aggressive condition than that seen in older age [20], and these advancements in big data can offer new possibilities in the quest to personalize and prevent diabetes. However, while the utilization of big data in this field presents opportunities, it also comes with challenges, such as regarding data privacy and security [4, 11, 12].

While various studies indicate the potential of using big data for diabetes prevention and personalization, as well as the challenges associated with both big data and the connection between lifestyle factors and diabetes, a gap exists in effectively integrating lifestyle factors into personalization and prevention of diabetes while addressing the associated challenges in big data, such as data privacy & security. This study will address the existing gap in effectively using big data in personalized and preventive diabetes care while taking into account the several challenges associated with big data utilization.

This leads to the Research Question:

- How can big data improve healthcare with more personalization and prevention of diabetes?

2 Literature Method

Initially, before commencing the literature review, the research topic and research question were identified. We then conducted searches in online libraries, including ScienceDirect, PubMed, and ResearchGate, to explore relevant academic articles. To identify relevant papers, we used the following search queries: 'big data' and one or more of the following: 'diabetes,' 'personalized', 'prevention' and 'healthcare', mainly with a focus on the inclusion of diabetes in the research paper at least. For instance, we used Boolean operators such as 'AND' and 'OR' to ensure that our search results are accurate. In cases where the keyword 'diabetes' was absent in the title, keywords, or abstract, but was found within the body of the paper, those papers were still included. Also, in case 'diabetes' was missing, a research paper that included at least three keywords and substantially contributed to this research was still included.

We excluded research papers that primarily focused on mathematical aspects and lacked theoretical contributions to this research. These are papers that do not go much into detail regarding the keywords. Additionally, papers highly relevant to other diseases were excluded.

Once the relevant papers were selected, they were further examined to identify one or more recurring themes. Subsequently, we established recurring themes found in one or more papers and grouped the papers accordingly based on these themes.

3 Literature Review

The increasing prevalence of diabetes and its associated health issues indicates a significant global health problem. Identifying effective preventive measures and interventions for individuals at risk of developing diabetes is crucial to mitigate its impact. In recent years, the utilization of big data has emerged for enhancing preventive care in various healthcare domains. This literature review aims to indicate the importance and effect of the usage of big data in personalized and preventive care for diabetes. See Figure 1 for an overview of the literature sources and concepts utilized in this research. See Appendices A and B for a detailed overview.

Big Data-5Vs	Big Data Challenges	Lifestyle factor	Preventive Healthcare	Personalized Healthcare
Rumbold et al. (2020)	Rumbold et al. (2020)	Rumbold et al. (2020)		
Venkatraman et al. (2023)	Venkatraman et al. (2023)			Venkatraman et al. (2023)
		Chawla et al. (2013)		Chawla et al. (2013)
Riddle et al. (2019)			Riddle et al. (2019)	
	Raghupathi & Raghupathi (2014)			
			Fagherazzi en Ravaud (2019)	
Barrett et al. (2013)	Barrett et al. (2013)	Barrett et al. (2013)	Barrett et al. (2013)	
		Glauber and Karnieli (2013)	Glauber and Karnieli (2013)	Glauber and Karnieli (2013)
Razzak et al. (2020)	Razzak et al. (2020)		Razzak et al. (2020)	

Figure 1: An overview of the literature and the concepts.

In the ongoing battle against diabetes, the fusion of healthcare and information technology has emerged as a formidable ally. The term e-health can be described as "the use of information and communications technology in support of health and health-related fields" [1]. It aids in prevention efforts through education and early detection, while also enhancing the management of the condition through remote monitoring, and improved access to healthcare expertise. E-health's role in diabetes care continues to grow, offering hope for better outcomes and a higher quality of life for those affected by the disease. Sources such as electronic medical records have the potential to enhance preventive healthcare, reduce medication mistakes, and streamline the management of population health [14]. The collection, examination, and utilization of big data are integral aspects of an expanding field that has been described using various terms, including e-health [4].

According to Rumbold et al. [15], the routine collection of large amounts of data is the essential bedrock of Big Data. Riddle et al. [14] specify that in diabetes care and research big data come from three main sources, namely electronic medical records (EMRs), surveys and registries, and randomized controlled trials (RCTs). Rumbold et al. [15] and Venkatraman et al. [18] imply that Big Data is characterized by the 5Vs: volume, variety, velocity, value and veracity. Also, Barrett et al. [4] mention that big data is characterized by high-volume, high-variety, and high-velocity information. Rumbold et al. [15] gives the following definitions:

Volume: This refers to the data quantity. Large data sets alone do not present unique analysis challenges. Variety: It encompasses various data types, including numerical, ordinal, and nominal data, and may involve semantic differences. Velocity: This pertains to the speed required for generating usable insights. Veracity: It concerns data accuracy and reliability; large amounts of poor-quality data offer no advantage. Value: This indicates the data's intrinsic worth in providing useful insights; data with low individual value may be valuable when combined with other data.

The main challenges with Big Data include:

- Managing huge amounts of unstructured data
- Integrating different data sets
- Effective exploitation of big data

[1] https://iris.who.int/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1

- Ensuring data privacy and security

As mentioned above, there are a number of challenges associated with big data, such as managing huge amounts of unstructured data and integrating different data sets [12, 15]. About 80 percent of electronic health data is unstructured [12]. In healthcare, real-time big data analytics is a crucial requirement and efforts should be made to reduce the lag between data collection and processing [11]. Moreover, effective exploitation of big data seems to be a challenge [15, 18]. There is a huge amount of healthcare data available, but it is important to exploit them in an effective way [15]. Furthermore, another issue seems to be regarding data privacy [4, 11, 12]. Ensuring patient privacy and information security are essential elements of an efficient healthcare system that contributes to better health outcomes and healthier individuals [12]. Data security manages data access at all points in the data lifecycle, whereas data privacy governs this access according to privacy policies and legal regulations which, for instance, specify who is permitted to access personal data [1].

The integration of extensive secure and individual datasets presents significant logistical and ethical challenges in preserving privacy [4]. Venkatraman et al. [18] add that there are several security and privacy concerns that require attention in pervasive and personalized health to use big data analytics effectively. Data privacy also seems to be a major hurdle in the prevention system [12].

Furthermore, an important aspect of the risk of diabetes is the lifestyle factor [9]. Glauber and Karnieli [8] indicate that although lifestyle changes or medication can help reduce the risk of diabetes, there is uncertainty in predicting who will benefit the most from these lifestyle changes or medication. To add, Rumbold et al. [15] mention that Type 2 diabetes is influenced by genetics, but it's also tied to diet and lifestyle. Diet plays a big role in both preventing and treating diabetes. A low-calorie diet can help people avoid taking medication for diabetes, however, according to Steven et al. [17] it turns out that many people with diabetes struggle to manage their condition effectively. Chawla et al. [5] mention that patients exposed to similar risk, lifestyle and environmental factors may have similar outcomes. Finally, Barrett et al. [4] claim that big data has the potential to be significantly influential in disease prevention. They assert that it can be used for both the discovery of personalized disease risk factors linked to an individual's lifestyle or environment and for facilitating successful behaviour modifications to reduce disease risks [4]. Big data can play a crucial role in personalized preventive care.

Big data plays a key role in preventive healthcare. Big Data-driven analyses will help change the way diabetes and diabetes-related complications and their prevention are being dealt with [6]. As an example, Harvard Medical School and Harvard Pilgrim Health Care used advanced data analysis techniques on electronic health records (EHR) to detect patients with diabetes and categorize them into Type 1 and Type 2 diabetes groups. They examined four years' worth of data from various sources, considering numerous health indicators. This analysis allowed them to identify high-risk patient groups and implement preventive measures, such as introducing new treatment protocols for patients with high cholesterol, to reduce health risks [12].

Furthermore, the utilization of data-driven and network-oriented techniques is in the development of personalized healthcare. Many diseases exhibit identifiable risk factors, some of which can be prevented, or at least, have indicators of potential risk [5]. In personalized healthcare, Chawla et al. [5], created a patient-centred framework using big data, a data-driven computational aid for physicians assessing the disease risks facing their patients. The authors claim that to advance personalized healthcare, there is a need for a computing and analytics framework to gather and integrate big data, uncover patient similarities, and create personalized disease risk profiles. Venkatraman et al. [18] claim that there is a deficiency in studies offering personalized healthcare services to health-conscious IoT users with a preventive approach. They argue that more research is needed to enhance personalized healthcare systems in order to leverage IoT, big data, and analytics. In response to this, they propose an IoT and big data architecture for personalized healthcare [18].

Finally, Glauber and Karnieli [8] call for personalized intervention to prevent diabetes. The authors suggest that a big data approach to personalized medicine will be such a tool, allowing personalized application of proven lifestyle and other primary prevention measures most efficiently and cost-effectively.

Analysis

The literature reveals that lifestyle factors play a pivotal role in the development of diabetes. Authors like Glauber and Karnieli [8] emphasize the uncertainty in predicting who would benefit the most from lifestyle changes or medication. This underscores the need for a personalized approach to diabetes prevention, tailoring interventions to an individual's specific risk factors, such as diet and environmental influences. The theoretical implication here is that a one-size-fits-all approach is not effective in preventing diabetes, and personalized interventions are crucial.

While big data, characterized by the 5Vs, holds promise in addressing the complexity of diabetes and discovering personalized disease risk factors, our research places a specific emphasis on the challenges associated with big data. These challenges include managing unstructured data and ensuring data privacy, highlighting the need for strong data management in diabetes prevention. The theory underlines the necessity for effective data utilization and privacy-preserving techniques to harness the full potential of big data in healthcare.

In the realm of personalized preventive diabetes care, the literature review points to the development of data-driven frameworks. Authors like Chawla et al. [5] advocate for the integration of big data to uncover patient similarities and create personalized disease risk profiles. The theoretical implication is that personalized care should be data-driven, leveraging big data to enhance the accuracy of risk assessment and intervention strategies. This suggests a shift towards a more data-centric approach in diabetic healthcare.

While this literature review has highlighted the significance of big data in the context of personalized preventive care for diabetes, it is crucial to emphasize the existing research gap and potential opportunities for further exploration. Although there is widespread recognition of the importance of big data in this context, a comprehensive understanding of its practical implementation and impact on personalized and preventive care of diabetes is still evolving.

This research gap centres on the limited exploration of solutions for leveraging big data to effectively integrate and utilize lifestyle factors, such as diet and physical activity, within personalized and preventive care for diabetes, all while addressing the main challenges associated with data privacy and security. The focus of our research will be on students/graduates and experts. Lifestyle factors play a pivotal role in diabetes risk and management. However, the practical implementation of big data in personalized and preventive care faces significant challenges, such as concerning data privacy and security. Addressing this research gap is critical to unlocking the full potential of big data in improving outcomes for individuals.

Conclusion

In conclusion, this study has revealed several important findings and implications within the realm of diabetes prevention and personalized care, emphasizing the significance of lifestyle factors and addressing various challenges, including privacy concerns:

Prevention: - The one-size-fits-all approach in diabetes prevention has been found to be ineffective. Personalized interventions are crucial for achieving meaningful results. In practice, this involves the development of personalized lifestyle plans and targeted interventions.

Personalized Care: - Personalized diabetes prevention strategies should be driven by data, harnessing the potential of big data for more accurate risk assessment and intervention strategies. Practical implementation will involve the integration of data-driven strategies in personalized diabetes prevention. In practice, this may include the use of big data analytics to assess individual risk factors and recommend specific interventions. Healthcare providers may adopt data-driven decision-making to enhance patient outcomes.

Lifestyle: - Effective data utilization is essential to unlock the full potential of big data in healthcare, particularly in improving outcomes related to diet and physical activity. Healthcare providers may employ data-driven approaches to guide individuals in making healthier lifestyle choices, with personalized recommendations based on their data profiles.

Challenges: - The challenges outlined in our theoretical framework extend beyond theory and manifest as practical obstacles in the field of diabetes prevention. These challenges, including data privacy and security concerns, persist as ongoing issues and require significant attention to enhance the effectiveness of big data utilization. Although rules and regulations exist for data privacy and security, in practice, these concerns

will require the implementation of robust safeguards and policies to protect patient data and build trust in data-driven healthcare.

In practice, the transition towards personalized and data-driven diabetes prevention will likely lead to more effective and efficient strategies that address individual needs while tackling the challenges associated with data privacy and security.

We anticipate that the effective utilization of big data in personalized and preventive care for diabetes will lead to enhanced outcomes in diabetes. This includes a reduction in the incidence of diabetes and more effective disease management for individuals at risk the expectation is that personalized care, driven by data, will enable more accurate risk assessments and more effective intervention strategies. Also, we expect that lifestyle factors will continue to be a central focus in personalized preventive strategies.

4 Methodology

4.1 Research Design

For this research, both quantitative and qualitative methods will be used. A survey will be used to collect and analyze quantitative data from students, followed by expert interviews to gather qualitative insights. The sequential use of quantitative and qualitative research methods facilitates a thorough exploration of the quantitative findings, providing the foundation for the subsequent expert interviews. Furthermore, the combination of both methods allows for enhancing the validation of the results [19].

4.2 Data Collection

Survey among university students/graduates

The first step involves data collection through a survey. Initially, a pilot study with two students and an expert in the field was done to test the survey's validity. The questions needed to be clear, easily understandable, and indeed aligned with the focus of the study. The survey was designed to gather insights from students and graduates across various educational and cultural backgrounds. The inclusion criteria encompassed university students or graduates. The survey instrument was created using Google Forms and comprised a total of 29 questions, including a question on personal or family history with diabetes. The questionnaire included a mix of open-ended questions, multiple-choice questions, multiple responses and Likert scales to capture both qualitative and quantitative data. The Likert scales were either from 1-5 or 1-10. 1-5 has been chosen for simple questions, whereas 1-10 has been chosen if the question was focused on nuance, such as in questions regarding personal data in which responses can offer more nuance across respondents.

Participants were recruited through a convenience sampling method, thus the sample is drawn from easily accessible sources to the researcher [3]. Fellow students and individuals from the researcher's academic network were contacted and requested to distribute the survey among their peers. The survey was disseminated online through the Google Forms platform. Participants received a clear description of the research and its objectives on the first page of the survey. Before starting the survey, participants were explicitly asked for their consent to participate. The anonymity of respondents was emphasized, and no personal data, such as email addresses, was collected to ensure confidentiality. Ethical considerations were addressed by obtaining informed consent from participants. The study adhered to ethical standards, emphasizing participant confidentiality and privacy throughout the data collection process. See Appendix C for the survey.

Interview with experts

In conjunction with the survey component of this research, interview data collection served as a crucial method for validation, enriching and complementing the quantitative findings. Especially, since the number of respondents to the surveys is lower than 50. Since the existing literature in the review of this paper had no existing interview questions, the pilot-tested survey questions and results have been used as the foundation. The integration of interviews was undertaken with the specific objective of validating the survey results and ensuring a comprehensive understanding of the research phenomena. For the interview phase of this study, data was sought from experts specializing in the field of diabetes care. The aim was to capture insights from professionals such as healthcare practitioners, researchers, and educators who possess a deep understanding of

current practices, challenges, and advancements in diabetes care. Through semi-structured interviews, these experts were given a platform to share their specialized knowledge, providing valuable qualitative data that supplements the broader quantitative survey findings. Finally, ethical considerations, which include obtaining informed consent and maintaining participant confidentiality, were rigorously adhered to throughout the expert interview data collection process. With participants' consent, interviews were audio-recorded to capture the richness of responses accurately. See Appendix D for the interview questions.

4.3 Data Preparation

Survey among university students/graduates

After the data collection phase, the raw data was processed and transformed into a CSV file to enhance readability and facilitate further analysis. An initial exploration of the dataset was conducted to understand its characteristics. During this exploration, it was identified that some open-ended questions contained missing data.

To address these missing values, a standard practice of imputation was employed. Specifically, the missing data in the open-ended questions were handled by assigning the label "No response" to indicate instances where participants chose not to provide input. This approach was chosen to maintain consistency in the dataset and to differentiate between deliberate non-responses and other types of missing data.

In the analysis of survey responses from a relatively small sample size of 40 respondents, it was observed that a subset of participants provided ratings that significantly deviated from the central tendency of the dataset. Specifically, these outliers manifested in the form of notably low ratings (e.g., a rating of 1) on the scale assessing concerns about anonymity. It is crucial to clarify that these extreme values were not the result of errors or data entry issues but rather reflect intentional and valid responses from a limited number of individuals with divergent viewpoints.

Despite the small sample size, it is important to acknowledge and celebrate the diversity of perspectives within the respondent group. The intentional outliers underscore the richness of opinions within our surveyed population, providing valuable insights into the spectrum of attitudes regarding concerns about anonymity. While the majority of respondents expressed a high level of confidence in the anonymity of the survey, the intentional inclusion of all opinions adds depth to our understanding of the subject matter.

This analysis opts for transparency in reporting by keeping these intentional outliers in the dataset. The decision to keep these extreme values, despite the limited sample size, aligns with the study's commitment to capturing the full spectrum of respondent perspectives. By doing so, we aim to present a more comprehensive and nuanced portrayal of the attitudes and concerns expressed by our relatively small sample of survey participants.

Interview with experts

The data preparation process involved the transcription of interview recordings, focusing on maintaining accuracy and fidelity to participants' expressions. Transcripts were cross-checked with the audio recordings to identify and rectify any discrepancies. Anonymization has been done to protect participants' identities.

4.4 Data Analysis Procedures

Survey among University Students/Graduates

In this study, data analysis was conducted using the programming language R and R Studio to gain insights into the relationships and patterns within the collected dataset. The dataset, sourced from the survey in Google Forms, was loaded into R Studio. The dataset, named rawsurveydata.csv was imported into R Studio. Initial data cleaning steps included checking for missing values, and when necessary, imputing missing data to ensure the dataset's integrity.

Descriptive statistics, such as mean values for numerical variables and frequency tables for categorical variables, were computed for analysis. To provide a visual representation of the data, histograms and bar charts were created using the 'ggplot2' package. These visualizations offered a preliminary understanding of the distribution and characteristics of variables. Due to the nature of the type of questions (variety of Likert scales, multiple-choice and open-ended questions in all themes) and structure of the surveys (4 themes:

lifestyle, prevention, personalization and big data), we refrained from employing further regression analysis due to the limitations posed by the survey structure and sample size, rendering regression analysis unsuitable and potentially misleading.

This data analysis procedure ensured a systematic exploration of the dataset, offering valuable insights.

Interview with experts

A total of 4 interviews have been conducted with the experts in the area. The data analysis process for the interviews adhered to an inductive coding approach using the Gioia method, allowing themes and patterns to naturally emerge from the participants' narratives [7]. To do so, ATLAS.ti [2] tool has been used, a qualitative data analysis software, to systematically analyze interview transcripts. Following data import, we familiarized ourselves with the content and initiated an inductive coding process, creating codes based on emerging themes and patterns. ATLAS.ti's tools facilitated organized data management, allowing for efficient grouping of codes into categories and constant comparison to identify relationships. Using the Gioia methodology a data structure has been obtained, see Figure 9 in section 5.2.

Initially, the three-step coding process started with focusing on informant-centric terms in the transcripts. Data segments in the interview transcripts that are relevant or interesting for the research are highlighted. In this phase around 80 categories have been identified. Then, similarities and differences between these categories have been sought which also helped to reduce the number of categories to a more manageable number of concepts. In this phase, labels have been created to categorise similarities and differences. This resulted in the 1st order concepts, as depicted in Figure 9 in section 5.2. Next, we focus on creating the 2nd order themes by considering the theoretical realm. To do so, the grouped 1st order concepts are given a specific suitable theme that is rather theory-centric and overarches that specific group of 1st order concepts. As an example, the first group of 1st order concepts in Figure 9 are made into the 2nd order theme "Lifestyle and Health". The same is done with all the other 1st order concepts in the figure. Afterwards, 2nd order themes have been brought together and have been made to an aggregate dimension. Using the Gioia method, the aggregate dimension is a more abstract and theoretically overarching the 2nd order themes [7]. As an example, the aggregate dimension of "Holistic Health Engagement" is determined by us by trying to create an abstract and theoretically overarching definition of the 2nd order themes "Lifestyle and Health", "Wellness Engagement" and "Personalization Development". Since the three 2nd order themes cover health, engagement and the broad picture of the individual, the aggregate dimension became "Holistic Health Engagement". While determining these dimensions, we try to make the name abstract yet overarching the corresponding 2nd order themes. The same approach has been used for the other 2nd order themes. So for the aggregate dimension, we try to create a definition that is more abstract and overarches the 2nd order themes that were brought together.

The three-step coding process has been executed iteratively to ensure validity. This implies that, during the analysis phase, the coded data underwent continuous review. At each stage, including the creation of 1st order concepts, 2nd order themes, and aggregate dimensions, revisions and refinements were made as necessary to create a precise and accurate data structure. See the resulting data structure in Figure 9 in section 5.2.

5 Results

In this section, we present the comprehensive findings derived from the analysis of survey responses focusing on four key thematic areas: lifestyle, prevention, personalization, and big data. The survey, conducted to understand perceptions and behaviours of our research focus, gathered data through a combination of open-ended and closed-ended questions. Through rigorous analysis, we aim to unravel nuanced insights within each thematic domain.

[2] https://atlasti.com/atlas-ti-web

5.1 General survey results

5.1.1 Demographic Information

The demographic information of survey respondents that will be used for further analysis is depicted in Figure 13 in Appendix E. The demographic information is the age group, gender and ethnicity of participants. The majority of the respondents are aged between 18-24. Moreover, there are more females than males and ethnicity seems to be mostly white.

Additionally, personal or family history of diabetes has been recorded. The majority of respondents have no diabetes history, either personally or in the family. See Figure 14 in Appendix E.

5.1.2 Lifestyle

The lifestyle category in the survey is about the importance of lifestyle choices, specifically diet and physical activity in preventing diabetes. Around 55% have a balanced diet, with an active lifestyle. Also, on a scale from 1 to 10, The motivation to maintain a healthy lifestyle among respondents is on average 7.4. Furthermore, the awareness of the impact of lifestyle choices on preventing and managing diabetes is 52.5% "somewhat aware". See Figure 15 in Appendix E for the awareness among respondents. It turns out that the majority is somewhat aware of the impact of lifestyle choices.

In response to the open-ended question probing participants on their beliefs regarding the prevention of diabetes through lifestyle choices, a diverse range of responses was received. We categorized these responses into three main groups: 'Diet only,' 'Diet and exercise,' and 'Uncategorized.' The distribution of responses across these categories is presented in the bar chart below, see Figure 2. Out of the total survey participants, a subset of 30 individuals responded to the open-ended question regarding their beliefs on preventing diabetes through lifestyle choices. Given that this question was voluntary and not mandatory, the responses represent a subset of the overall participant pool. While the sample size is relatively small, the insights gathered from these 30 participants provide valuable qualitative perspectives on the topic. Based on the respondents, the majority claim that both diet and exercise are essential in preventing diabetes through lifestyle choices.

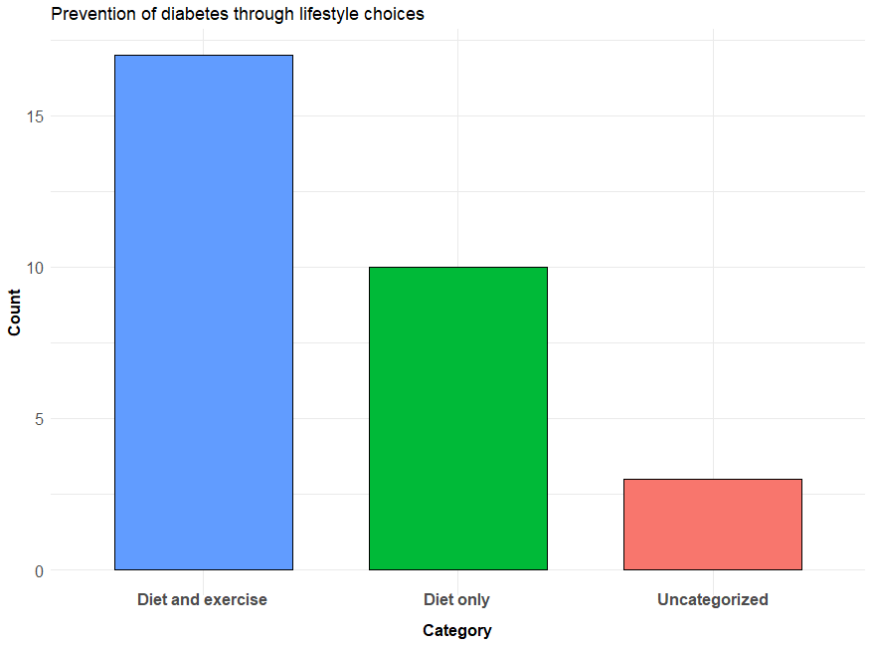


Figure 2: Frequencies of diabetes prevention categories

5.1.3 Prevention

Regarding the prevention of diabetes, the confidence level for knowledge of strategies for preventing diabetes is depicted in Figure 3. The majority of the respondents seem to be "somewhat aware", followed by being "not aware at all".

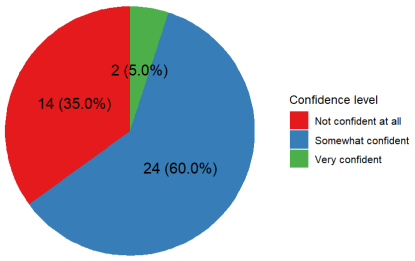


Figure 3: Confidence level for knowledge of strategies

Furthermore, the survey participants were asked to identify barriers to adopting a lifestyle for diabetes prevention, with the option to select one or more predefined responses or provide additional prompts. The majority of participants chose one or more barriers from the predefined options. However, it is noteworthy that some participants also gave other prompts not included in the predefined list. The additional prompts have been mostly categorized into "Motivation". The coded responses were categorized for analysis, with the findings presented in Figure 4. From the results, it emerges that the lack of awareness is the biggest barrier.

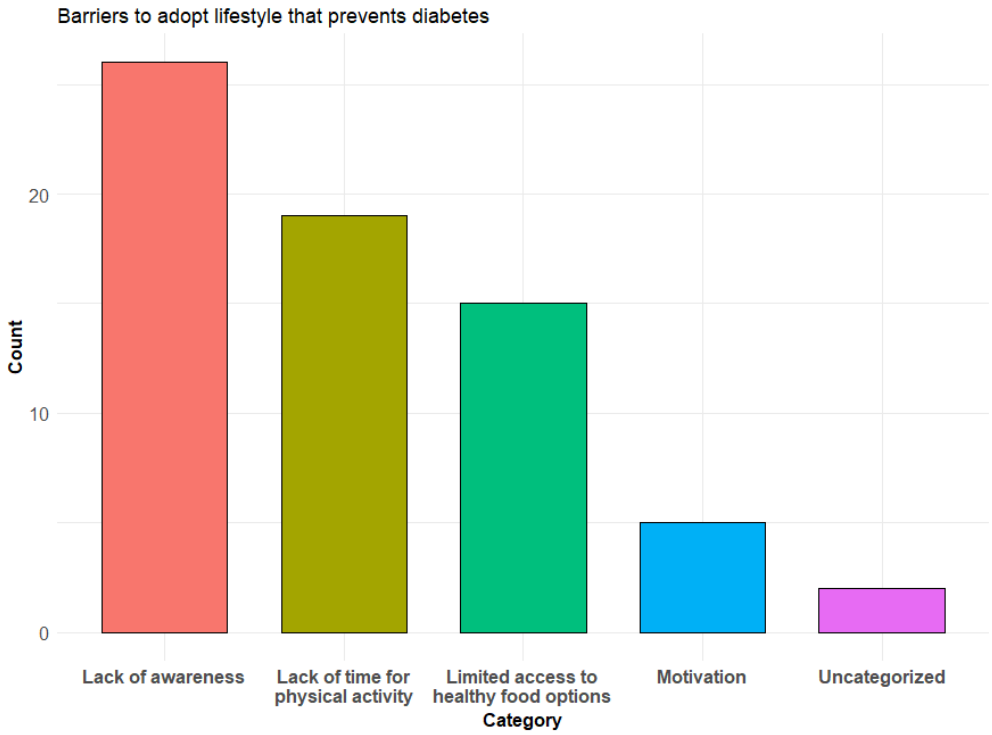


Figure 4: Barriers to adopt a lifestyle that prevents diabetes

5.1.4 Personalization

On the personalization aspect of diabetes care, it was asked how effective personalized approaches can be in diabetes care. On a scale from 1 to 5, the average was 4.1, indicating that personalized approaches are effective. Moreover, regarding the awareness of personalized approaches in diabetes care that cater to individual needs and characteristics, it turns out that 47.5% of the respondents are not aware at all. Potential barriers to adopting personalized approaches have been asked, and 23 out of 40 participants responded. While the data collected provides valuable insights, it is important to note the number of respondents, and the findings should be interpreted with consideration of this limitation. The responses have been categorized, see Figure 5.

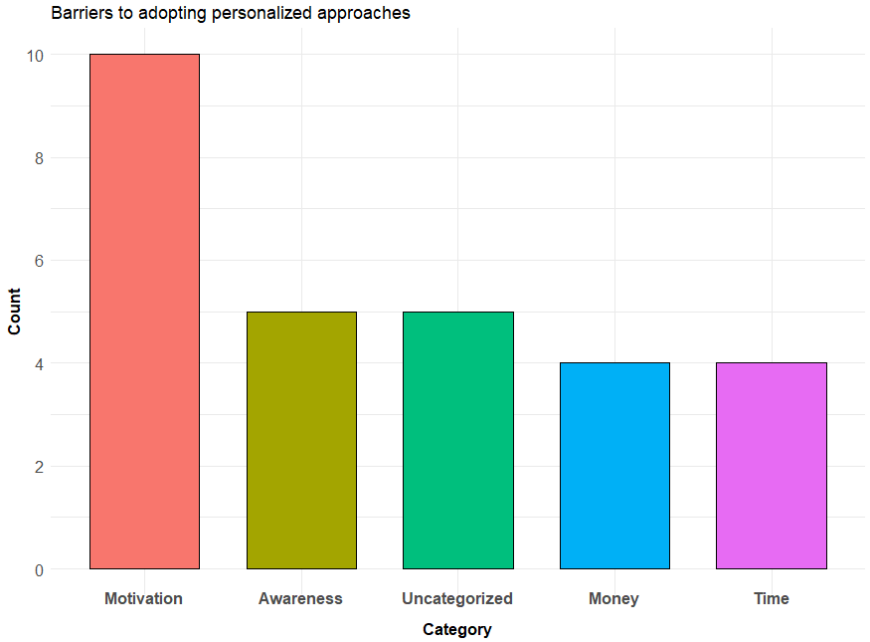


Figure 5: Barriers to adopting personalized approaches in diabetes

Additionally, it was asked what aspects of personalization would be beneficial in diabetes care. The findings are depicted in Figure 16 in Appendix E. "Tailored diet plans" seem to be most beneficial, followed by "Personalized exercise regimes".

5.1.5 Big Data

Finally, the findings on big data indicate that 65% of the respondents are not aware at all of the utilization of big data in preventing and personalizing diabetes care. Furthermore, the positive impact of big data on personalization and prevention of diabetes is reflected in an average score of 3.9 on a scale of 1 to 5. The survey result indicates that there is a lack of awareness of big data usage among students.

In the context of personalizing diabetes care with big data, participants were asked about lifestyle factors they believe should be prioritized for data collection. The frequencies of preferences are summarized in Figure 6. Mostly, dietary habits have been chosen for data collection, followed by physical activity level.

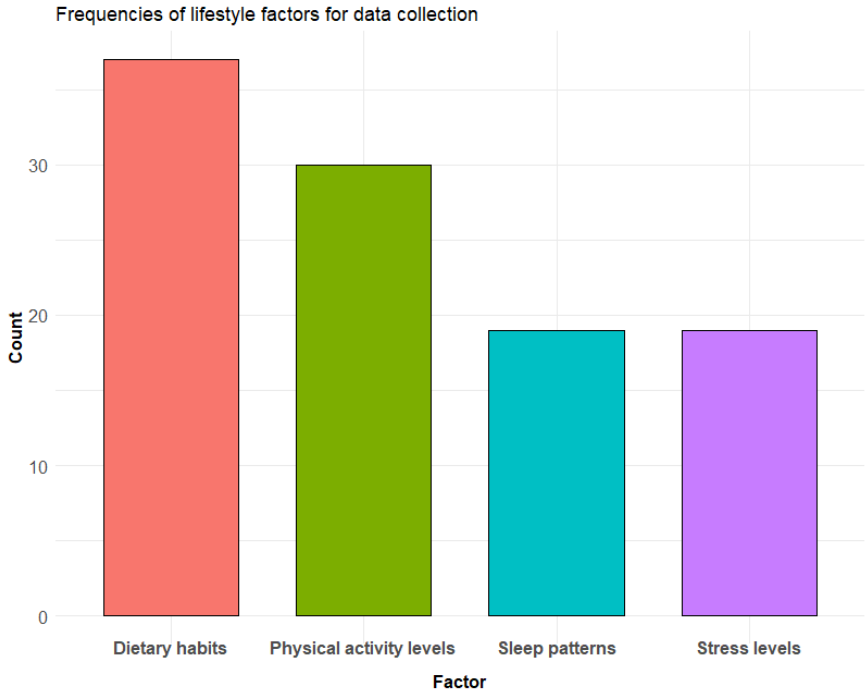


Figure 6: Frequencies of lifestyle factors for personalization with big data

Furthermore, the willingness to share lifestyle data for the improvement of big data-driven personalization was rated on a scale of 1-10, with an average score of 6.4.
The concern regarding privacy and security of health and lifestyle data in big data-driven diabetes care is on average 6.3, on a scale of 1-10. Moreover, the distribution for the preferred level of anonymity regarding health and lifestyle data can be found in Figure 17 in Appendix E. Respondents prefer to have partial anonymity, followed by being fully anonymous.

Also, the following results are obtained from how it can be best made sure to keep health and lifestyle information private and secure, see Figure 7. The results indicate using strong protection, closely followed by ensuring authorized people to see the information is one of the most important.

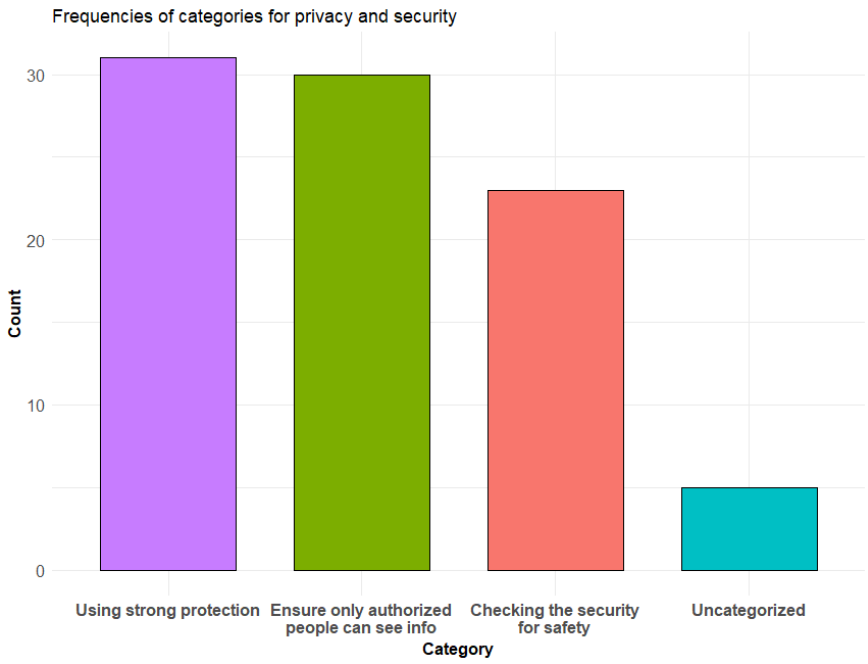


Figure 7: Frequencies of categories for ensuring privacy and security

Finally, see Figure 8 regarding combining different types of health information. Using tools to put all data together has the lead, closely followed by making different systems work together.

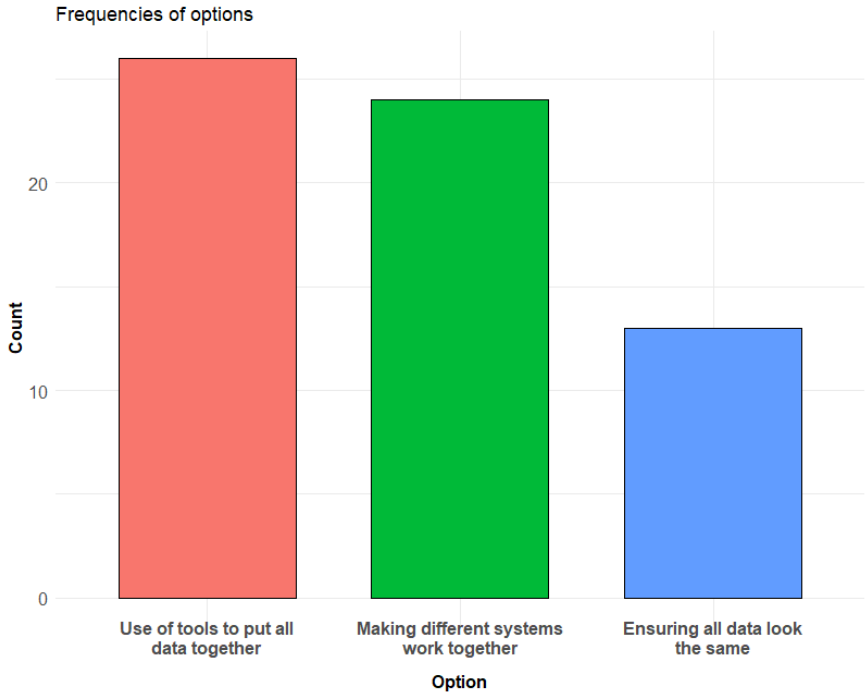


Figure 8: Frequencies of options for combining health data

5.1.6 Analysis based on gender

This section contains the key insights derived from the analysis made based on the gender of the survey respondents. The percentages mentioned in this section and the figures are proportionally calculated. 75% of the respondents are females, and 25% are males.

To start with, a comparison is made between the lifestyles across gender. It turns out that overall females are more moderately active with a balanced diet than males. On average, males have a more active lifestyle with a healthy diet. See Figure 18 in Appendix E.

Regarding motivation to maintain a healthy lifestyle, females score an average of 7.3 while males score an average of 7.6. The confidence level for knowledge of strategies for preventing diabetes is depicted in Figure 19 in Appendix E.

Figure 20 in Appendix E shows the barriers to adopting a lifestyle that prevents diabetes, by gender. The majority of females found that "Lack of awareness" is the biggest barrier, while for males it is the "lack of time for physical activity".

Furthermore, the average grade on a scale of 1-5 for the effectiveness of personalized approaches in diabetes care is 4.0 for females and 4.2 for males, indicating on average a high score across genders. The beneficial personalized aspects mentioned mostly are depicted in Figure 21 in Appendix E. For both males and females, tailored diet plans, followed by personalized exercise regimes are most beneficial.

Another key insight indicates the awareness level by gender regarding big data usage in diabetes care, see Figure 22 in Appendix E. The results show that the awareness levels across the genders are quite low.

On a scale of 1 to 10, the willingness among males to share lifestyle data for the improvement of big data-driven personalized diabetes care is rated at 8, whereas this figure is 5.9 among females. Additionally, the average rating for privacy and security concerns associated with sharing lifestyle data is 6.4 among females and 6.1 among males. The preferred anonymity level for health and lifestyle data regarding big data-driven diabetes care is depicted in Figure 23 in Appendix E. The preferred anonymity level among females is "fully anonymous" and among males is "partially anonymous".

5.1.7 Analysis based on diabetes history

This section contains the key insights derived from the analysis made based on the personal or family history of the survey respondents. The percentages mentioned in this section and figures are proportionally calculated. 62.5% respondents have no personal or family history of diabetes, and 37.5% have personal or family history of diabetes. Initially, Figure 24 in Appendix E shows the lifestyle habits based on diabetes history. In both groups, respondents are "Moderately active with a balanced diet". However, following up, people with diabetes history tend to be "inactive with poor dietary habits". Also, respondents with a diabetes history are a minority in being "active with healthy dietary habits".

Furthermore, on the prevention through lifestyle choices, the categories as identified earlier are "Diet only", "Diet and exercise" and "Uncategorized. Figure 25 in Appendix E shows the differences based on the diabetes history of the respondents. The percentages in the bar plot are based on diabetes history from the respondents who answered the question. Roughly, both groups indicate similarities in answers. The most popular category is "Diet and exercise", followed by "Diet only".

In terms of confidence in knowledge regarding prevention strategies, 40% of individuals with a history of diabetes express "no confidence at all," while the corresponding percentage for those with no history of diabetes is 32%. The numbers are proportional. Also, barriers to adopting a lifestyle that prevents diabetes based on diabetes history is depicted in Figure 26 in Appendix E.

Moreover, regarding awareness of personalized approaches in diabetes care for individual needs and characteristics, it turns out that 53% of people with a history of diabetes are "not aware at all," while people with no history of diabetes have a score of 44%. Furthermore, the belief in the effectiveness of personalized approaches in diabetes care, rated on a scale of 1 to 5, averages 4.0 for individuals with no history of diabetes, while those with a history of diabetes give an average rating of 4.1.

Figure 27 in Appendix E shows the beneficial personalized aspects by diabetes history. In both groups "Tailored diet plans" is the most popular category.

Regarding awareness of using big data in diabetes care based on diabetes history, see Figure 28 in Appendix E. Both groups indicate a low awareness level. The gap seems to be greater among people with no diabetes history.

The willingness to share lifestyle data for improvement of big data-driven personalized and preventive diabetes care is graded with an average of 6.7 for respondents with a diabetes history, while respondents with no diabetes history have graded with 6.3. The concern regarding privacy and security of health and lifestyle data in the context of big data-driven diabetes care is graded with 6.1 for respondents with a diabetes history, while respondents with no diabetes history have graded with 6.4. Finally, the preferred anonymity level regarding sharing health and lifestyle data by diabetes history is depicted in Figure 29 in Appendix E. The "Not concerned" level is low across both groups.

5.2 Interview results

From the coding process, the following data structure has been obtained, see Figure 9:

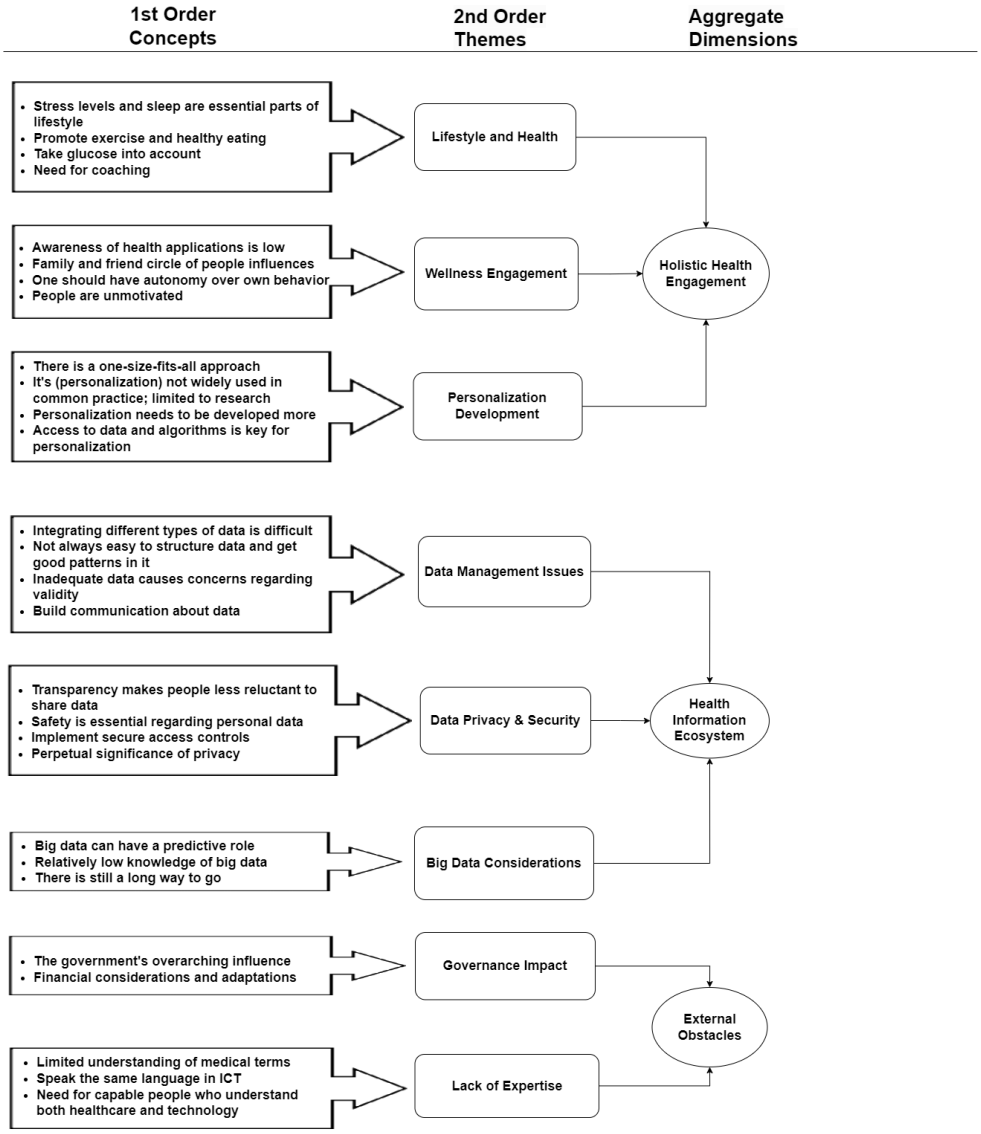


Figure 9: Data structure

Several themes and dimensions have been created. To start with, the aggregate dimension of "Holistic Health Engagement" consists of the themes "Lifestyle and Health", "Wellness Engagement" and "Personalization Development". The "Lifestyle and Health" theme consists of essential lifestyle and health aspects that are important for maintaining a healthy lifestyle and thereby preventing diabetes. Besides diet and exercise, it is important to focus on a broader picture of an individual and include various aspects such as stress, sleep and glucose values. Expert C says: *"We know from research that in a very large amount of patients diabetes can be prevented by a focus on a healthier lifestyle, yeah, and it's not solely diet and nutrition it's also stress and sleep, everything to be in balance, actually."*
Furthermore, regarding patients who already have diabetes, Expert A says: *...but if it were possible, it would also be very nice if they have stress and sleep, also have a lot of influence on glucose levels, so it may also be the case that someone who, for example, has his exercise and nutrition quite well under control, that he still has very fluctuating glucose levels."*.

Moreover, there is a need for coaching to guide people to a healthier lifestyle. Expert A mentions: *"So coaching is also a part of it (diabetes prevention) because otherwise, they (people) aren't aware of healthy and unhealthy foods..".*

The "Wellness Engagement" theme contains the engagement of people with their overall well-being. It turns out that motivation is a hurdle in engaging with well-being. Expert A says: *"Especially motivation. I think most people I see don't have a lack of time, but they most are unmotivated.".* Expert D mentions regarding motivation *"motivation is super important [..]also for the caregiver.* Moreover, awareness is low for the usage of health applications. Expert B says: *think there's very little awareness left that such apps exist....I think just a lot of people don't really know where to look and also yes if you want to use an app like that, yes, which app do you choose, right?".* On the other hand, Expert C mentions: *"It's (Apps) not their natural way of being educated or supported by an app for health..".*

The "Personalization Development" theme is about the use of personalized approaches in diabetes care. The current landscape indicates that personalization is not a lot used in healthcare. Expert C on topic: *"Yeah, I think that it's insufficient personalised. Definitely, because we do one-size-fits-all.".* Expert A on topic mentions *"It's (personalization) not really used in common practice right now, only in, only in research."* To make it more personalized Expert A says: *"..it (personalization) has to be developed more...".* Also, regarding key aspects of personalization, Expert C says *"...So I think that having access to data and having access to algorithms to make it more personalised is one of the key aspects.".*
Finally, the aspects of tailored diet plans, personalized exercise regimens and remote monitoring and feedback are useful. Expert B says *Yes, I think all of that (aspects) is already very useful.*, while Expert D says: *I think the first two (tailored diet plans and personalized exercise regimes) is all well and good, but I don't think those are the most important ones..."..*

The "Health Information Ecosystem" dimension consists of the themes: "Data Management Issues", "Data Privacy & Security" and "Big Data Considerations". The "Data Management Issues" contains the challenges regarding managing data. Various challenges arise, such as data structuring, integrating data, and concerns regarding validity. Expert B says the following: *"...you can put a lot of different types of data there, but it's not always easy to structure it all and get good patterns in it..."* and *"I can indeed imagine that integrating different type data can be a real challenge.".* Another issue seems to be concerns regarding the validity of data. Expert C mentions *"...we ask people to keep a diary and we see that you do have a lot of inadequate information..".*

In the second theme "Data Privacy Security", experts mention the importance of safeguarding data and how to do it. Expert D says: *Yes, you just have to have a good feeling (on personal data) [..] it has to be safe.".* Regarding solution, for example, Expert A mentions: *"You have to have passwords in your computer and behind these files etc...".* Furthermore, people become less reluctant to share their personal data when there is transparency, as Expert B says: *"So if you make clear where the data is used for who is using the data and I think that people are less reluctant to sharing data".* Expert C mentions that *...who gets the data. For what is it used? Make it transparent, that's also because if you make it transparent, people give trust..".*

Regarding the "Big Data Considerations" theme, it becomes obvious that big data is a relevantly new concept in healthcare settings. Expert B says: *"I know, in healthcare big data is not really applied yet, in the care of diabetes patients.* and Expert C: *"I think it's limitedly used".* Big data could be potentially used for its predictive role as per Expert B:*"I think big data can play a role in that, maybe also a predictive role of "well, we're not just going to try something, but we can do it based on the data". We say perhaps this is the best treatment option for this patient.".* However, there is still a long way to go in big data in healthcare for diabetes care. Expert B says: *..."I think it's just a very long way that we have to go before you can do this properly.".* Expert A says: *"Big data. I think it really is the future...".*

Finally, the "External Obstacles" dimension, consists of the themes "Governance Impact" and "Lack of Expertise". The role of the government is influential, in various areas. As an example, Expert B mentions: *"If you want to raise awareness among people, then the government just has to well, there has to be an approach*

that also encourages a healthy lifestyle, so to speak.". Also, Expert C mentions *"It's (Dutch guidelines) really prohibit personalized care."*. Furthermore, "Lack of Expertise" seems to be an issue. Expert C claims *"...lots of people in healthcare, in general, have very low health literacy."*. Additionally, Expert D mentions *"...well, before that, that you all speak the same language in the field of ICT. That remains a very big stumbling block."*.

6 Discussion

Based on the Gioia Methodology [7], we use the 2nd order themes and aggregate dimensions from the data structure in Figure 9 in section 5.2 to create this emerging model. This model gives an overview and relations between the 2nd order themes and aggregate dimensions. See Figure 10.

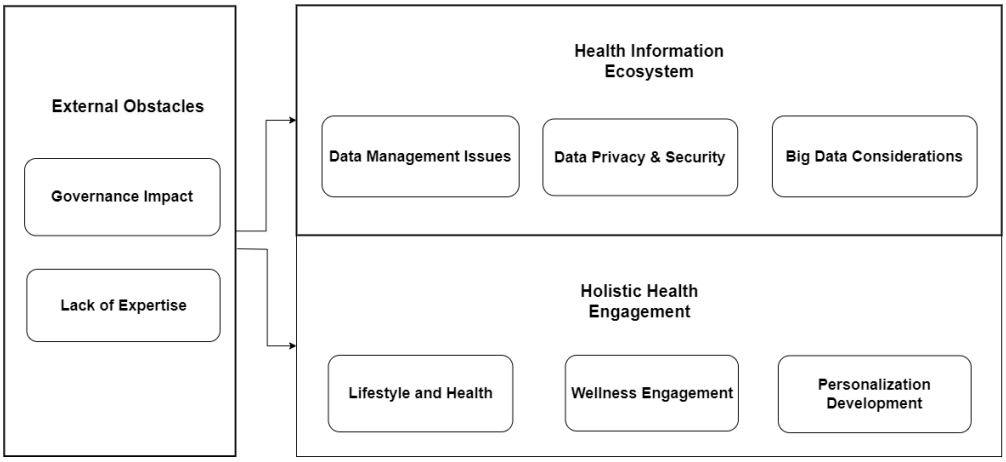


Figure 10: Emerging Model

In Figure 10, each aggregate dimension and the corresponding 2nd order themes are shown in rectangles.

Besides the overview that this figure provides, we see the relationship between the dimension "External Obstacles" and the dimensions of "Health Information Ecosystem" and "Holistic Health Engagement" in the form of pointing out arrows. These arrows indicate that the dimension "External Obstacles" affects both dimensions "Health Information Ecosystem" and "Holistic Health Engagement". To be more specific, as described in section 5.2, "Governance Impact" seems to play a role in encouraging a healthy lifestyle and personalization of healthcare. Furthermore, "Lack of Expertise" seems to be a problem both in healthcare and ICT. Hence, we depict that in Figure 10 as outgoing arrows from "External Obstacles".

Regarding lifestyle, survey respondents indicate a low awareness of the impact of lifestyle choices on preventing and managing diabetes. Also, survey results indicate "diet and exercise" to have the highest frequency in the prevention of diabetes through lifestyle choices. From the interview results, it turns out that the role of lifestyle is very essential for the prevention of diabetes. In the model in Figure 10, the 2nd order theme "Lifestyle and Health" consists of lifestyle factors including diet and exercise, as described earlier. The interview results, as previously discussed, also indicated that lifestyle encompasses more than just diet and exercise; factors such as stress and sleep were also essential components of lifestyle, indicating the broader picture of the individual is important. The findings align with previous research on the importance of lifestyle factors in the prevention of diabetes. Namely, existing research [6, 8, 12] associates a healthy lifestyle including various aspects such as diet and physical activity, with a highly reduced risk of developing diabetes. This means that with regards to diabetes prevention, adopting a healthy lifestyle, and focusing on its various aspects remains an essential method for diabetes prevention.

Additionally, some experts indicated the need for coaching people due to a lack of awareness of healthy and unhealthy food options. The lack of awareness is in line with the lack of awareness that emerged from the survey results on lifestyle and prevention. Furthermore, regarding barriers to adopting a lifestyle that prevents diabetes, motivation was among the survey responses. The experts validate it by mentioning that motivation is very important and that it seems to be a challenging subject for people. Although motivation is less mentioned than "Lack of time for physical activity", the expert view indicates that mostly "lack of time" is not the issue, but motivation is. In line with existing research [8] coaching programs can help with accessing effective interventions for individuals at risk. Additionally, dieticians play a critical role in the change of lifestyle and thereby prevention of diabetes [12]. On another note, as described earlier in the interview results about "External Obstacles", the government plays an overarching role in various areas, specifically in terms of awareness of people and encouraging a healthy lifestyle. See Figure 10 for the relation of the dimension "External Obstacles". The survey results did not reveal any external obstacles.

Our findings indicate that adopting a healthy lifestyle focusing on various aspects including diet, exercise, sleep and stress remains very important for diabetes prevention. It is essential to look at the broader picture of the individual. Additionally, there is a need for coaching people, especially in terms of following a healthy lifestyle and motivation. This implies that a personalized approach can be useful for prevention. Furthermore, the government has great potential in terms of increasing awareness and encouraging a healthy lifestyle.

From the survey results, personalized approaches are promising, with the most frequent barriers being "Motivation" and "Awareness". To adopt personalized care by people it is important to raise awareness of personalized care and motivate people to use such care when needed. The interview results indicate that personalization has to be developed more and that government and guidelines limit personalized care. Furthermore, there is still a one-size-fits-all approach according to experts. In line with the survey results, there could be a link to a lack of awareness and motivation of people because personalized care is not a lot employed. Overall, some experts agreed that "Tailored diet plans", "Personalized exercise regimens" and "Remote monitoring and feedback" are data that are very useful for personalized care, but not necessarily the most beneficial.

The existing literature [8] implies that a big data approach to personalization can be a tool for the application of proven lifestyle and other prevention measures in an efficient way. The research [8] adds that until the use of such a tool, encouraging a healthy lifestyle remains key. This means that personalized care can be effective if applied, however until then preventive measures need to be encouraged.

The findings on personalization shed light on the current status of personalized care and offer insights into the types of data that could be beneficial for personalized care. Furthermore, there could be a potential link between the lack of awareness and motivation among individuals and the limited utilization of personalized care. With regards to implementing personalized care, development and government and guidelines form a challenge to personalized care.

The use of big data in preventing and personalizing diabetes care has a low awareness among survey respondents. The interview results validate by indicating that big data is a relatively new concept and that it is currently not really used in healthcare, but it is promising for the future. The lack of awareness in the survey results could therefore be linked to big data being a new concept. It seems that there is still a long way to go to implement big data properly for diabetes. An integrated and improved management of big data holds promise for diabetes care and research [14].

Findings, as described earlier, indicate that personalized care needs to be developed first. On top of this, research [18] indicates that more research is required to enhance the personalized healthcare system for the general population, enabling them to take advantage of recent advancements in the industry including big data, proactively. Our findings suggest that big data holds potential for personalization and prevention, such as coaching individuals in diabetes prevention by facilitating the adoption of a healthy lifestyle and offering tailored advice. Previous research emphasizes that big data can help for prevention and personalization in healthcare and diabetes [4, 5, 6], but for now, it remains for the future [14, 15].

Among the survey respondents, the willingness to share data for the improvement of personalization with big data-driven was rated at 6.4 in general. The concerns regarding privacy and security of health and lifestyle data with 6.3 and the preferred anonymity level was 47.5% "partially anonymous". In the interview results it turns out that indeed data privacy and security are very common issues and important. The low rates of people willing to share data are likely due to concerns related to data privacy and security, such as a lack of trust. Experts highlight that to make people less reluctant to share data, transparency is key. Also, experts indicate secure access controls need to be implemented, such as passwords, which is in line with the highest frequency categories in the survey results regarding ensuring privacy and security. The interview results emphasize that privacy is essential and data needs to be safe. The findings align with previous research on the presence and importance of ensuring data privacy & security, being also one of the main challenges in big data as addressed by various research [4, 11, 12, 18].

Aside from data privacy and security, several data management issues arise from the interview results. Various challenges arise, such as the ability to structure data, integrate data, and concerns regarding the validity of data. In the survey results regarding combining data, it is mentioned that mostly it is essential to use tools to put data together and make different systems work together. The findings in data management issues align with previous research that indicates that the main challenges in big data include managing unstructured data and integrating different datasets [12, 15]. Thus, the data management issues consist of mostly main challenges of big data as indicated in previous research [4, 11, 12, 18], however additionally another concern is the validity of data. Data is not always adequate, this could be potentially linked to concerns of people regarding their data privacy and security.

Our findings highlight the current main challenges to big data and provide insights into how these challenges could be overcome. Investment in research and development of the challenges in big data is needed to overcome these challenges. These challenges form an impediment to properly implementing big data in diabetes healthcare.

Our findings lead to several practical implications. Firstly, adopting a healthy lifestyle is still the way to go to prevent diabetes. Promoting a healthy lifestyle will likely lead to increased awareness among individuals and improved public health by reducing the number of individuals with diabetes. The government can play an important role in promoting this awareness, for example by promoting healthy food options and stimulating physical activity. This could for example be done by using educational programs. The government easing limitations on personalized care could encourage healthcare organizations to adopt and implement personalized care for diabetes. Moreover, investing in research for the development of big data and personalized and preventive care is essential. When it comes to big data, it's essential to establish transparent policies and secure access controls for personal data. This builds trust and encourages individuals to share their data for research and healthcare purposes. To further develop big data, investments could be made that focus on the main challenges that are associated with big data. These include developing tools to effectively structure and integrate data.

For the surveys, analysis based on gender and diabetes history has been made, see sections 5.1.6 and 5.1.7. The anticipation was to uncover significant results that could offer insights for this study. However, our findings did not correspond with our expectations. The analysis did not produce significant results that could be validated through the interview data. Hence, this data has not been used for further exploration.

7 Limitations and Future Work

It is important to note that there are limitations to this study. To start with, it is important to acknowledge the limited sample size of the survey, which comprised only 40 respondents. Although interview data has been used to ensure the validity of this research, a larger sample size would enhance the validity and generalizability of the findings. Moreover, data analysis techniques, such as regression analysis, were not feasible within the scope of this research, such as due to the small sample size. Using these methods for future work would provide deeper insights into the variable dependencies of the survey. We made an analysis based on gender and diabetes history for the surveys, however, no significant results have been obtained for further exploration. This is likely due to the small sample size of the survey. For future research, the analysis based on gender and diabetes history could give different results, providing valuable insights. Also, for future research surveying and interviewing diabetes patients or people with a risk for diabetes only could provide interesting and nuanced insights on the topic. Furthermore, the survey questions and interview questions could not be taken from existent literature, hence the questions may not cover all relevant aspects of the topic. Additionally, despite the pilot study, some questions could still be unclear or vague to respondents which unintentionally could lead to bias. For future research, it is important to use survey and interview questions from existing sources to decrease potential bias and increase the validity of the research. The data analysis of the interviews is done by using the Gioia methodology [7], allowing for continuous review of the coded data for increased accuracy and validity. However, researcher subjectivity and interpretation could potentially impact the results of the interviews. For future research, coding the interviews with more than one researcher could lead to higher validity and accuracy.

8 Conclusion

Our research sheds light on the usage of big data in personalized and preventive diabetes care. The research underlines the importance of lifestyle factors in preventing diabetes. Adopting a healthy lifestyle remains essential for preventing diabetes. Moreover, personalized care is currently not commonly used in practice, due to the need for its development and the governmental impediments. Furthermore, while big data holds potential for personalization and prevention, it is important to acknowledge and overcome the numerous challenges that persist. These challenges underscore the necessity for further development in the field of big data. It seems that data privacy and security remain an important issue in the use of big data. Regarding prevention, there is a need for increasing awareness among individuals to adopt a healthy lifestyle. Government can provide support to increase awareness about adopting a healthy lifestyle for the prevention of diabetes. Personalized care, such as coaching individuals, can increase awareness of adopting a healthy lifestyle, preventing diabetes and providing tailored interventions with the help of big data. However, further development is needed for personalization. Additionally, the government should facilitate the adoption of personalized approaches by removing these barriers. Investments in research and development of big data are necessary to overcome the challenges and implement big data properly in diabetes care.

References

[1] Karim Abouelmehdi, Abderrahim Beni-Hssane, Hayat Khaloufi, and Mostafa Saadi. Big data security and privacy in healthcare: A review. *Procedia Computer Science*, 113:73–80, 2017.

[2] CA Alexander and L Wang. Big data in healthcare: a new frontier in personalized medicine. *Am J Hypertens Res*, 1(1):15–18, 2017.

[3] Chittaranjan Andrade. The inconvenient truth about convenience and purposive samples. *Indian Journal of Psychological Medicine*, 43(1):86–88, 2021.

[4] Meredith A Barrett, Olivier Humblet, Robert A Hiatt, and Nancy E Adler. Big data and disease prevention: from quantified self to quantified communities. *Big data*, 1(3):168–175, 2013.

[5] Nitesh V Chawla and Darcy A Davis. Bringing big data to personalized healthcare: a patient-centered framework. *Journal of general internal medicine*, 28:660–665, 2013.

[6] Guy Fagherazzi and Philippe Ravaud. Digital diabetes: Perspectives for diabetes prevention, management and research. *Diabetes & metabolism*, 45(4):322–329, 2019.

[7] Dennis A Gioia, Kevin G Corley, and Aimee L Hamilton. Seeking qualitative rigor in inductive research: Notes on the gioia methodology. *Organizational research methods*, 16(1):15–31, 2013.

[8] Harry Glauber and Eddy Karnieli. Preventing type 2 diabetes mellitus: a call for personalized intervention. *The Permanente Journal*, 17(3):74, 2013.

[9] Kalliopi Karatzi and Yannis Manios. The role of lifestyle, eating habits and social environment in the prevention and treatment of type 2 diabetes and hypertension, 2021.

[10] Naresh Kasoju, NS Remya, Renjith Sasi, S Sujesh, Biju Soman, C Kesavadas, CV Muraleedharan, PR Harikrishna Varma, and Sanjay Behari. Digital health: trends, opportunities and challenges in medical devices, pharma and bio-technology. *CSI Transactions on ICT*, pages 1–20, 2023.

[11] Wullianallur Raghupathi and Viju Raghupathi. Big data analytics in healthcare: promise and potential. *Health information science and systems*, 2:1–10, 2014.

[12] Muhammad Imran Razzak, Muhammad Imran, and Guandong Xu. Big data analytics for preventive medicine. *Neural Computing and Applications*, 32:4417–4451, 2020.

[13] Sang Youl Rhee, Chiweon Kim, Dong Wook Shin, and Steven R Steinhubl. Present and future of digital health in diabetes and metabolic disease. *Diabetes & Metabolism Journal*, 44(6):819–827, 2020.

[14] Matthew C Riddle, Lawrence Blonde, Hertzel C Gerstein, Edward W Gregg, Rury R Holman, John M Lachin, Gregory A Nichols, Alexander Turchin, and William T Cefalu. Diabetes care editors' expert forum 2018: managing big data for diabetes research and care, 2019.

[15] John MM Rumbold, M O'Kane, Nada Philip, and Barbara K Pierscionek. Big data and diabetes: the applications of big data for diabetes care now and in the future. *Diabetic Medicine*, 37(2):187–193, 2020.

[16] Pouya Saeedi, Inga Petersohn, Paraskevi Salpea, Belma Malanda, Suvi Karuranga, Nigel Unwin, Stephen Colagiuri, Leonor Guariguata, Ayesha A Motala, Katherine Ogurtsova, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas. *Diabetes research and clinical practice*, 157:107843, 2019.

[17] Sarah Steven, Kieren G Hollingsworth, Ahmad Al-Mrabeh, Leah Avery, Benjamin Aribisala, Muriel Caslake, and Roy Taylor. Very low-calorie diet and 6 months of weight stability in type 2 diabetes: pathophysiological changes in responders and nonresponders. *Diabetes care*, 39(5):808–815, 2016.

[18] Sitalakshmi Venkatraman, Sazia Parvin, Wathiq Mansoor, and Amjad Gawanmeh. Big data analytics and internet of things for personalised healthcare: opportunities and challenges. *International Journal of Electrical & Computer Engineering (2088-8708)*, 13(4), 2023.

[19] Marja J Verhoef, Ann L Casebeer, et al. Broadening horizons: integrating quantitative and qualitative research, 1997.

[20] Jencia Wong, Glynis P Ross, Sophia Zoungas, Maria E Craig, Elizabeth A Davis, Kim C Donaghue, Louise J Maple-Brown, Margaret J McGill, Jonathan E Shaw, Jane Speight, et al. Management of type 2 diabetes in young adults aged 18–30 years: Ads/adea/apeg consensus statement. *Medical Journal of Australia*, 216(8):422–429, 2022.

9 Appendices

9.1 Appendix A

	Paper	Main points	Relevance		Concepts/Focus				
					Big data	Diabetes	Preventive	Personalized	Healthcare
1	Rumbold, J., O'Kane, M., Philip, N., & Pierscionek, B. K. (2020). Big Data and diabetes: the applications of Big Data for diabetes care now and in the future. Diabetic Medicine, 37 (2), 187–193. https://doi.org/10.1111/dme.14044	Healthcare data are being produced at ever-increasing rates, and this information has the potential to transform the provision of diabetes care. Big Data is beginning to have an impact on diabetes care through data research. The use of Big Data for routine clinical care is still a future application. Vast amounts of healthcare data are already being produced, and the key is harnessing these to produce actionable insights. Considerable development work is required to	Big Data has great potential but it is important to understand the limitations of the approach.	Problem/limitations	√	√	√		√
2	Venkatraman, S., Parvin, S., Mansoor, W., & Gawanmeh, A. (2023). Big data analytics and internet of things for personalised healthcare: opportunities and challenges. International Journal of Power Electronics and Drive Systems, 13 (4), 4306. https://doi.org/10.11591/ijece.v13i4.pp4306-4316	In this paper, we identify the current personalized healthcare trends and challenges. Then, propose an architecture to support big data analytics using POC test results of an individual. The proposed architecture can facilitate an integrated and self-managed healthcare as well as remote patient care by adapting three popular machine learning algorithms to leverage the current trends in IoT, big data infrastructures and data analytics for advancing personalized healthcare of the future.	This paper presented the current trends in personalized healthcare and proposed a big data analytics framework for healthcare in adapting popular machine learning algorithms such as DT, k-NN and NB for enhancing the inferences on health risks of an individual.	Solution	√			√	√
3	Chawla, N. V., & Davis, D. A. (2013). Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework. Journal of General Internal Medicine, 28 (S3), 660–665. https://doi.org/10.1007/s11606-013-2455-8	We present the foundations of work that takes a Big Data driven approach towards personalized healthcare, and demonstrate its applicability to patient-centered outcomes, meaningful use, and reducing re-admission rates.	We present the foundations of work that takes a Big Data driven approach towards personalized healthcare, and demonstrate its applicability to patient-centered outcomes, meaningful use, and reducing re-admission rates.	Solution, Framework	√			√	√
4	Riddle, M. C., Blonde, L., Gerstein, H. C., Gregg, E. W., Holman, R. R., Lachin, J. M., Nichols, G. A., Turchin, A., & Cefalu, W. T. (2019). Diabetes Care Editors' Expert Forum 2018: Managing Big Data for Diabetes Research and Practice. Diabetes Care, 42(6), 1136–1146. https://doi.org/10.2337/dci19-0020	In medicine—including diabetes care and research—big data come from three main sources: electronic medical records (EMRs), surveys and registries, and randomized controlled trials (RCTs). EMRs continuously accumulate information about patients and make it readily accessible but are limited by missing data or data that are not quality assured. Because EMRs vary in structure and management, comparisons of data between health systems may be difficult. Registries and surveys provide data that are consistently collected and representative of broad populations but are limited in scope and may be updated only intermittently. RCT databases excel in the specificity, completeness, and accuracy of their data, but rarely include a fully representative sample of the general population. Also, they are costly to build and seldom maintained after a trial's end. To consider these issues, and the challenges and opportunities they present, the editors of Diabetes Care convened a group of experts in management of diabetes-related data on 21 June 2018, in conjunction with the American Diabetes Association's 78th Scientific Sessions in Orlando, FL.	This article summarizes the discussion and conclusions of that forum, offering a vision of benefits that might be realized from prospectively designed and unified data-management systems to support the collective needs of clinical, surveillance, and research activities related to diabetes.	Experts view/solution	√	√	√		√
5	Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health Information Science and Systems, 2(1). https://doi.org/10.1186/2047-2501-2-3	Big data analytics has the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make informed decisions. In the future we'll see the rapid, widespread implementation and use of big data analytics across the healthcare organization and the healthcare industry. Several challenges must be addressed. As big data analytics becomes more mainstream, issues such as guaranteeing privacy and safeguarding security.	There are challenges regarding big data in healthcare such as daha privacy and security	Problems	√		√		√
6	Faghenazzi, G., & Ravaud, P. (2019). Digital diabetes: Perspectives for diabetes prevention, management and research. Diabetes & Metabolism, 45(4), 322–329. https://doi.org/10.1016/j.diabet.2018.08.012	The present review looks at how the digitization of diabetes can impact all fields of diabetes – its prevention, management, technology and research – and how it can complement, but not replace, what is usually done in traditional clinical settings. S	Big Data-driven analyses and AI applied to diabetes data will change the way diabetes and diabetes-related complications, and their prevention and management, are dealt with. They should complement, but not replace, what is usually done in traditional clinical settings. Indeed, they truly constitute a game changer that should be embraced by all, as they can provide solid research results transferable to patients, improve general health literacy, and offer tools to facilitate everyday decision-making processes for both HCPs and people living with diabetes.	Solution, Conclusion	√	√	√	√	√
7	Barrett, M. A., Humblet, O., Hiatt, R. A., & Adler, N. E. (2013). Big Data and Disease Prevention: From Quantified Self to Quantified Communities. Big Data, 1(3), 168–175. https://doi.org/10.1089/big.2013.0027	Big data has a potentially critical role to play in preventing disease. It can both allow the discovery of new, personalized disease risk factors related to lifestyle or the environment, and also help people to successfully modify their risk behaviors.	Big data has a potentially critical role to play in preventing disease. It can both allow the discovery of new, personalized disease risk factors related to lifestyle or the environment, and also help people to successfully modify their risk behaviors.	Evidence	√		√	√	√
8	Glauber, H., & Kemiel, E. (2013). Preventing Type 2 diabetes mellitus: a call for personalized intervention. The Permanente Journal, 17(3), 74–79. https://doi.org/10.7812/tpp/12-143	A large and growing number of people worldwide already have DM2 or are likely to develop DM2. Nevertheless, available evidence shows very low rates of identification and management of diabetes risk. Given the large population at risk for DM2 and given the cost and limitations of preventive interventions, there is increasing recognition that identification of those at greatest risk for progression to DM2 will maximize the cost-effectiveness of an intervention program. It is likely that a bigdata approach to personalized medicine will be such a tool, allowing personalized application of proven lifestyle and other primary prevention measures in the most efficient and cost-effective manner.	A big data approach to personalized medicine is likely to be a valuable tool for efficiently and cost-effectively applying proven lifestyle and primary prevention measures to those at risk.	Problem/Potential solutions	√	√	√	√	√
9	Razzak, M. I., Imran, M., & Xu, G. (2020). Big data analytics for preventive medicine. Neural Computing and Applications, 32(9), 4417–4451. https://doi.org/10.1007/s00521-019-04095-y	The aim of this study is to provide a comprehensive and structured overview of extensive research on the advancement of data analytics methods for disease prevention. This review first introduces disease prevention and its challenges followed by traditional prevention methodologies.	Harvard Medical School and Harvard Pilgrim Health Care applied analytical methods to EHR data to identify patient with diabetes and classify them into groups (Type 1 and Type II diabetes). Four years worth of data based on numerous indicators from multiple sources have been analyzed. Patient could be grouped into high-risk disease groups and risk could be minimized by preventive care, i.e., new preventive treatment protocols could be introduced among patient groups with high cholesterol.	Evidence/Problems	√	√	√		√

Figure 11: An overview of the used research papers and their relevance

9.2 Appendix B

	Paper	Big Data-5Vs	Big Data Challenges	Lifestyle factor	Preventive Healthcare	Personalized Healthcare
1	Rumbold, J., O'Kane, M., Philip, N., & Pierscionek, B. K. (2020). Big Data and diabetes: the applications of Big Data for diabetes care now and in the future. Diabetic Medicine, 37(2), 187–193. https://doi.org/10.1111/dme.14044	Description of 5Vs	Effective exploiting healthcare data. Difficulty in combining datasets not least because of concerns over privacy with such sensitive data	Type 2 diabetes is influenced by genetics, but it's also tied to diet and lifestyle.		
2	Venkatraman, S., Parvin, S., Mansoor, W., & Gawanmeh, A. (2023). Big data analytics and internet of things for personalised healthcare: opportunities and challenges. International Journal of Power Electronics and Drive Systems, 13(4), 4306. https://doi.org/10.11591/ijece.v13i4.pp4306-4316	Big data is characterized by 5Vs	1. Several security and privacy concerns require attention in pervasive and personalised health to use big data analytics effectively 2. The use of big data in several contexts is not exploited to the full potential			There is a deficiency in studies offering personalized healthcare services to health-conscious IoT users with a preventive approach.
3	Chawla, N. V., & Davis, D. A. (2013). Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework. Journal of General Internal Medicine, 28(S3), 660–665. https://doi.org/10.1007/s11606-013-2455-8			Patients exposed to similar risk, lifestyle and environmental factors may have similar outcomes		To advance personalized healthcare, there is a need for a computing and analytics framework to gather and integrate big data, uncover patient similarities, and create personalized disease risk profiles.
4	Riddle, M. C., Blonde, L., Gerstein, H. C., Gregg, E. W., Holman, R. R., Lachin, J. M., Nichols, G. A., Turchin, A., & Cefalu, W. T. (2019). Diabetes Care Editors' Expert Forum 2018: Managing Big Data for Diabetes Research and Care. Diabetes Care, 42(6), 1136–1146. https://doi.org/10.2337/dci19-0020	Specify further that in diabetes care and research big data come from three main sources, namely electronic medical records (EMRs), surveys and registries, and randomized controlled trials (RCTs).			Sources such as electronic medical records have the potential to enhance preventive healthcare, reduce medication mistakes, and streamline the management of population health	
5	Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health Information Science and Systems, 2(1). https://doi.org/10.1186/2047-2501-2-3		1. In healthcare, real-time big data analytics is a crucial requirement and efforts should be made to reduce the lag between data collection and processing. 2. In Big Data privacy will garner attention.			
6	Fagherazzi, G., & Ravaud, P. (2019). Digital diabetes: Perspectives for diabetes prevention, management and research. Diabetes & Metabolism, 45(4), 322–329. https://doi.org/10.1016/j.diabet.2018.08.012				Big Data-driven analyses will help change the way diabetes and diabetes-related complications and their prevention are being dealt with	
7	Barrett, M. A., Humblet, O., Hiatt, R. A., & Adler, N. E. (2013). Big Data and Disease Prevention: From Quantified Self to Quantified Communities. Big Data, 1(3), 168–175. https://doi.org/10.1089/big.2013.0027	Characterized by high-volume, high-variety, and high-velocity information	The integration of extensive secure and individual datasets presents significant logistical and ethical challenges in preserving privacy	They assert that it can be used for both the discovery of personalized disease risk factors linked to an individual's lifestyle or environment and for facilitating successful behavior modifications to reduce disease risks	Big data has the potential to be significantly influential in disease prevention.	
8	Glauber, H., & Karnieli, E. (2013). Preventing Type 2 diabetes mellitus: a call for personalized intervention. The Permanente Journal, 17(3), 74–79. https://doi.org/10.7812/tpp/12-143			Lifestyle changes or medication can help reduce the risk of diabetes, there is uncertainty	Call for call for personalized intervention to prevent diabetes	Need for personalized intervention for diabetes
9	Razzak, M. I., Imran, M., & Xu, G. (2020). Big data analytics for preventive medicine. Neural Computing and Applications, 32(9), 4417–4451. https://doi.org/10.1007/s00521-019-04095-y	About 80 percent of electronic health data is unstructured	Data privacy is another major hurdle in development of prevention system		Harvard Medical School and Harvard Pilgrim Health Care used advanced data analysis techniques on electronic health records (EHR) to detect patients with diabetes and categorize them This analysis allowed them to identify high-risk patient groups and implement preventive measures, such as introducing new treatment protocols for patients with high cholesterol, to reduce health risks	

Figure 12: An overview of the literature and the concepts.

9.3 Appendix C

Big Data-Driven Personalized and Preventive Diabetes Care.

This research endeavors to explore the relationship between big data utilization and its impact on preventing and personalizing diabetic care. This survey aims to understand your perspectives on the intersection of big data, lifestyle choices, and diabetic care. By gathering your insights, we seek to enhance our understanding of how big data can contribute to the prevention and personalization of diabetes care. This survey is designed for university students and graduates, representing diverse backgrounds and experiences.

This survey should take approximately 3 to 5 minutes to complete.

Your responses will be kept confidential. The data collected will be aggregated and anonymized, ensuring that individual responses cannot be traced back to specific participants. The information gathered from this survey will be used solely for research purposes. Please be informed that your participation in this research project is entirely voluntary, and you have the right to discontinue your involvement at any time without providing a reason. If you choose to withdraw, your decision will not result in any negative consequences. Your comfort and well-being are our priorities, and we encourage you to reach out to us if you have any concerns during the course of the study. Your individual responses will not be disclosed, and the results will contribute to advancing our understanding of diabetes care.

If you have any questions or concerns regarding the survey, please feel free to contact us at m.n.kara@student.utwente.nl
Thank you for your valuable participation in this survey.

Sign in to Google to save your progress. Learn more

* Indicates required question

Please check the box below to indicate your consent to participate in this survey: *

- [ ] I consent to participate in the survey "Big Data-Driven Personalized and Preventive Diabetes Care."

Big Data-Driven Personalized and Preventive Diabetes Care.

* Indicates required question

1. What is your current age group? *

   Mark only one oval.

   - ◯ 18-24
   - ◯ 25-34
   - ◯ 35-44
   - ◯ 45-54
   - ◯ 55-64
   - ◯ 65+

2. What is your gender? *

   Mark only one oval.

   - ◯ Male
   - ◯ Female
   - ◯ Prefer not to say
   - ◯ Other: _____

3. What do you identify as your ethnicity or race? *

   Mark only one oval.

   - ◯ White
   - ◯ Black or African American
   - ◯ Asian
   - ◯ Hispanic or Latino
   - ◯ Native American or Alaska Native
   - ◯ Other: _____

4. What is your current field of study? Please indicate in the text box below. *

   _____

5. Do you have any personal or family history of diabetes? *

Mark only one oval.

- ◯ Yes
- ◯ No

Questions on Lifestyle

6. How would you describe your current health and wellness habits, considering both physical activity and diet? *

Mark only one oval.

- ◯ Inactive with Poor Dietary Habits
- ◯ Moderately Active with Balanced Diet
- ◯ Active with Healthy Dietary Habits

7. On a scale from 1 to 5, how important do you believe healthcare is in maintaining a healthy lifestyle? *

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Extremely important

8. How aware are you of the impact lifestyle choices can have on preventing and managing diabetes? *

Mark only one oval.

- ◯ Not aware at all
- ◯ Somewhat aware
- ◯ Very aware

9. To what extent are you motivated to maintain a healthy lifestyle? *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Highly motivating

10. In what ways do you personally believe you can contribute to the prevention of diabetes through your lifestyle choices?

Questions on prevention of diabetes

11. How confident are you in your knowledge of strategies for preventing diabetes? *

Mark only one oval.

- ◯ Not confident at all
- ◯ Somewhat confident
- ◯ Very confident

12. To what extent do you believe that lifestyle choices, such as diet and physical activity, can effectively prevent diabetes? *

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Highly effective

13. What do you perceive as the most significant barrier to adopting a lifestyle that prevents diabetes? Please select all that apply. If you think there are other barriers, please specify in "Other". *

Check all that apply.

- ☐ Lack of awareness
- ☐ Limited access to healthy food options
- ☐ Lack of time for physical activity
- ☐ Other: _____

14. How motivated are you to actively engage in behaviors that can prevent the onset of chronic diseases, such as diabetes? *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Highly motivated

Questions on personalization in diabetes care

15. How aware are you of personalized approaches in diabetes care that cater to individual needs and characteristics? *

Mark only one oval.

- ◯ Not aware at all
- ◯ Somewhat aware
- ◯ Very aware

16. To what extent do you believe that personalized approaches in diabetes care can be effective? *

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Very effective

17. If given the option, which personalized aspects in diabetes care would you find beneficial? *

Check all that apply.

- ☐ Tailored diet plans
- ☐ Personalized exercise regimens
- ☐ Remote monitoring and feedback
- ☐ Individualized medication plans
- ☐ Other: _____

18. How willing are you to actively participate in personalized approaches for managing diabetes or supporting others in diabetes management? *

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Highly willing

19. What do you perceive as potential barriers to adopting personalized approaches in diabetes?

The usage of Big Data in personalized preventive diabetes care.

Note: Big data refers to large and diverse collections of information that continuously expand, originating from various sources and arriving rapidly.

20. How aware are you of the utilization of big data in personalizing and preventing diabetes care? *

Mark only one oval.

- ◯ Not aware at all
- ◯ Somewhat aware
- ◯ Highly aware

21. In the context of personalizing diabetes care with big data, which lifestyle factors do you think should be prioritized for data collection? (Select one or more) *

Check all that apply.

- ☐ Physical activity levels
- ☐ Dietary habits
- ☐ Stress levels
- ☐ Sleep patterns
- ☐ Other:

22. To what extent do you believe big data can positively impact the personalization and prevention of diabetes care? *

Mark only one oval.

	1	2	3	4	5	
No i	◯	◯	◯	◯	◯	Highly impactful

23. If given the option, which personalized aspects in diabetes care driven by big data would you find most valuable? (Select all that apply) *

Check all that apply.

- [ ] Tailored lifestyle recommendations
- [ ] Individualized treatment plans
- [ ] Predictive risk assessments
- [ ] Continuous monitoring and feedback
- [ ] Other: _____

24. To what extent would you be willing to share your lifestyle data for the improvement of big data-driven personalized diabetes care? *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Highly willing

25. How accessible do you think big data-driven personalized/preventive diabetes care is to the general population? *

Mark only one oval.

- ◯ Not accessible
- ◯ Somewhat accessible
- ◯ Highly accessible

Concerns regarding the use of Big Data

26. To what extent would you be concerned about the privacy and security of your health and lifestyle data in the context of big data-driven diabetes care? *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extremely concerned

27. When it comes to big data-driven diabetes care, what level of anonymity would you prefer for your health and lifestyle data? *

Mark only one oval.

- ◯ Fully anonymous
- ◯ Partially anonymous
- ◯ Not concerned about anonymity
- ◯ Other: _____

28. How can we best make sure that personal health information is kept private and secure in using technology for diabetes care? Please select one or more. *

Check all that apply.

- ☐ Making sure only authorized people can see the information
- ☐ Using strong protection for the information
- ☐ Checking the security often to keep information safe
- ☐ Other: _____

29. What do you think is most important for combining different types of health information for diabetes care? Please select one or more. *

Check all that apply.

- ☐ Making sure all data looks the same
- ☐ Making different systems able to work together
- ☐ Using tools to put all the data together
- ☐ Other: _____

9.4 Appendix D

Interview

Short introduction
Hello, I am Manolya, currently doing my master's thesis on the usage of big data in personalized and preventive diabetes care at the University of Twente. Initially, I collected data through surveys about the perspectives/opinions of students/graduates on the usage of big data in personalized and preventive diabetes care. As a second step, I wanted to gather expert views to validate the results of the survey. The interview contains several parts. Firstly, I will ask questions on lifestyle choices and prevention then personalization and finally the usage of big data.

1. To start, can you tell me a little bit about yourself and your professional background?

Lifestyle choices and prevention

2. Firstly, what do you think of the role of lifestyle, such as diet and physical activity in preventing diabetes?

Students seem to be somewhat aware of the impact of lifestyle choices to prevent diabetes and manage it, around 53% seem to be somewhat aware.

3. From your experience, what key factors contribute to awareness of the impact of lifestyle choices in preventing and managing diabetes, and what strategies would promote this awareness?

One question was on how to prevent diabetes through lifestyle choices. The most common answers were: eating healthy, reducing carbs/sugar intake, and exercising.

4. Would you agree with this?

5. What, in your opinion, are the most effective strategies or recommendations for diabetes prevention through lifestyle choices?

6. Interestingly, no respondent mentions the usage of personalized approaches or usage of social applications to this question. Why would this be the case you think, could it be linked to a lack of awareness?

7. To students, I asked what barriers there are to adopting a lifestyle that prevents diabetes. The answers are in order mostly lack of awareness, followed by a lack of time for physical activity and finally limited access to healthy food. A part of the students also mentioned motivation.

   Would you agree with the results? And can motivation play a key role?

The next questions are on the personalization aspect.

8. How would you describe the current landscape of personalized diabetes care, and what advancements have been made in recent years?

9. What are the key challenges and opportunities in personalizing diabetes care?

10. What are the barriers to adopting personalized approaches in diabetes?

11. On the question of what personalized aspects would be found beneficial, respondents mentioned in the most common order: tailored diet plans, personalized exercise regimes, remote monitoring and feedback. Would you agree or do you think there are other personalized aspects to be beneficial?

The next questions are on the big data aspect.
Big data refers to vast and complex datasets marked by their size, speed, and diversity, which pose challenges to traditional methods of data processing.

12. Do you believe big data can positively impact the personalization and prevention of diabetes care?

13. Can you provide an overview of how big data is currently being used in the healthcare sector, specifically in the context of diabetes prevention and personalization?

14. What challenges or obstacles are commonly faced in the effective utilization of big data for diabetes prevention and personalization, and how are these challenges addressed?

15. In the survey responses, 65% of respondents seem to be not aware at all of the usage of big data in diabetes care. What do you think is the reason and how could we increase awareness?

16. Respondents mention that mostly dietary habits and physical activity should be prioritized for data collection to personalizing diabetes care with big data. Do you agree that these should be prioritized?

17. In your opinion, what specific types of data do you believe are crucial to collect in the process of preventing diabetes?

18. The main challenges in big data include managing huge amounts of unstructured data, integrating different data sets and ensuring data privacy and security. Do you agree, or do you think there are other challenges as well?

19. How can we overcome these challenges, you think?

20. The majority of respondents are willing to have partial anonymity when sharing data followed by a willingness to be fully anonymous. What are your thoughts on this?

21. How can we best make sure that personal health information is kept private and secure in using technology for diabetes care?

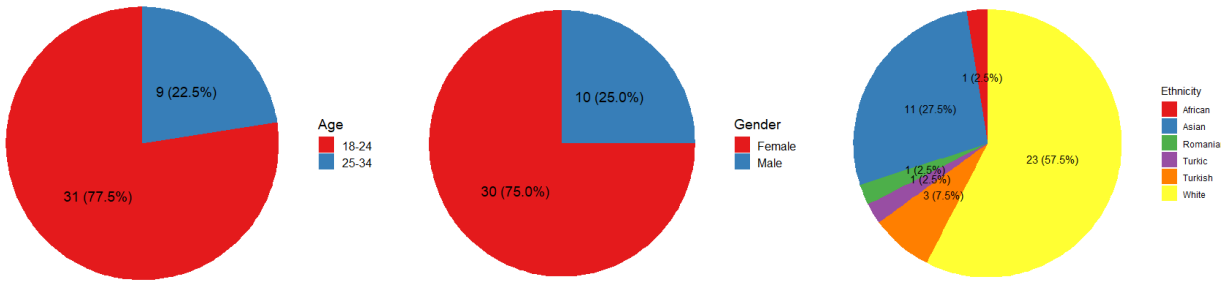
22. What do you think is most important for combining different types of health information for diabetes care?

    Finally, these were the questions from my side. Is there anything else that you would like to share or tell?

Thank you for your participation!

9.5 Appendix E

Figures for general survey results:



(a) Age groups (b) Gender (c) Ethnicity

Figure 13: Pie charts for Age, Gender, Ethnicity

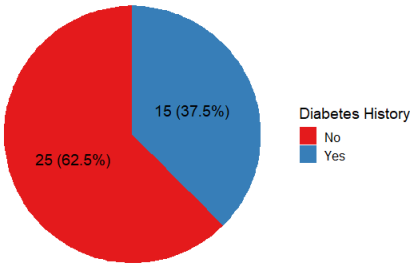


Figure 14: Diabetes History

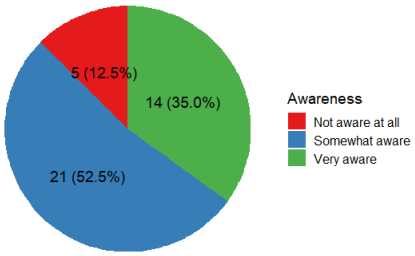


Figure 15: Awareness of the impact of lifestyle choices on preventing and managing diabetes

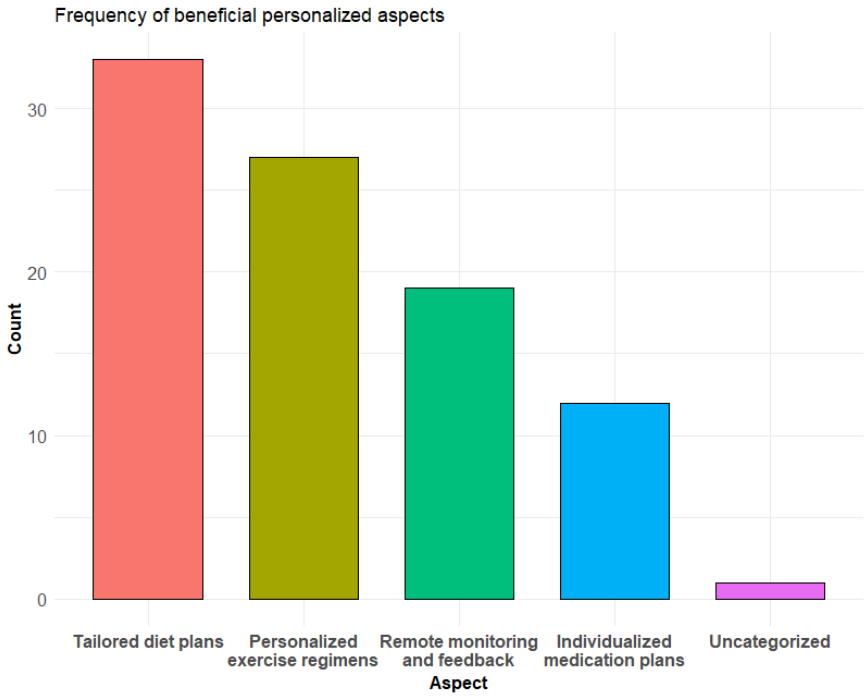


Figure 16: Frequency of beneficial personalized aspects

Figures for analysis based on gender:

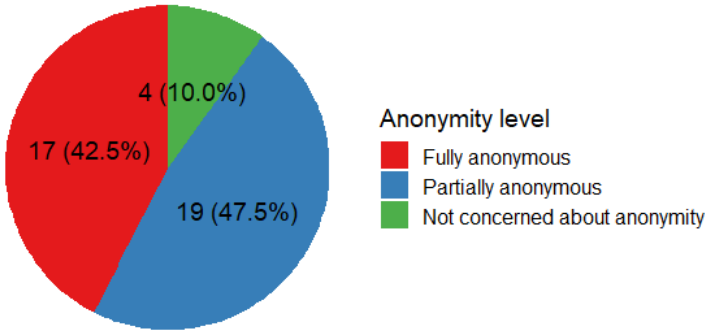


Figure 17: Distribution for the level of anonymity for health and lifestyle data

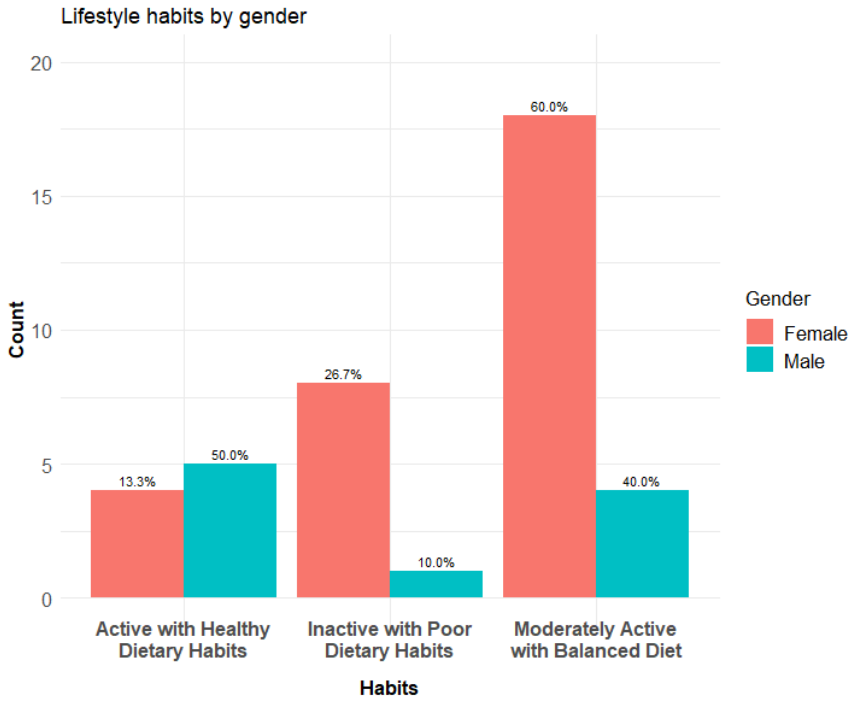


Figure 18: Lifestyle habits based on gender

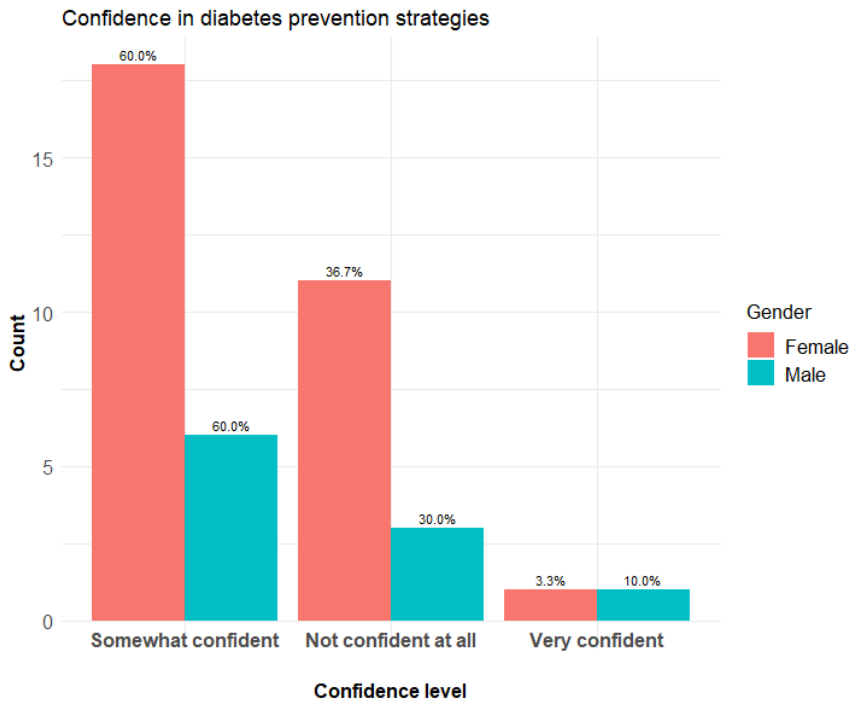


Figure 19: Confidence in diabetes prevention strategies by gender

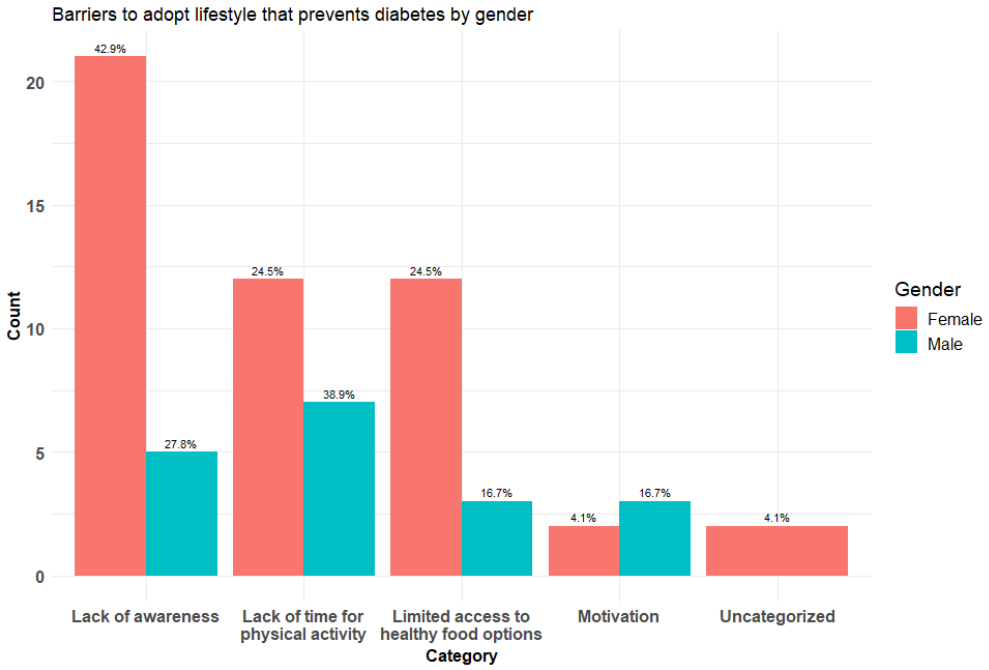


Figure 20: Barriers mentioned to adopt a lifestyle that prevents diabetes

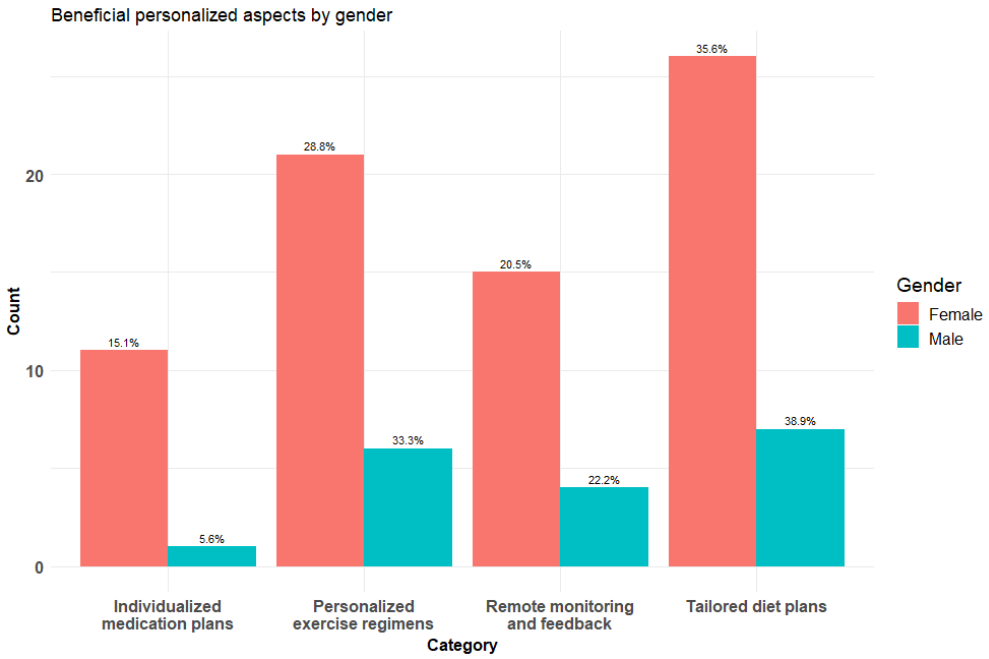


Figure 21: Beneficial personalization aspects by gender

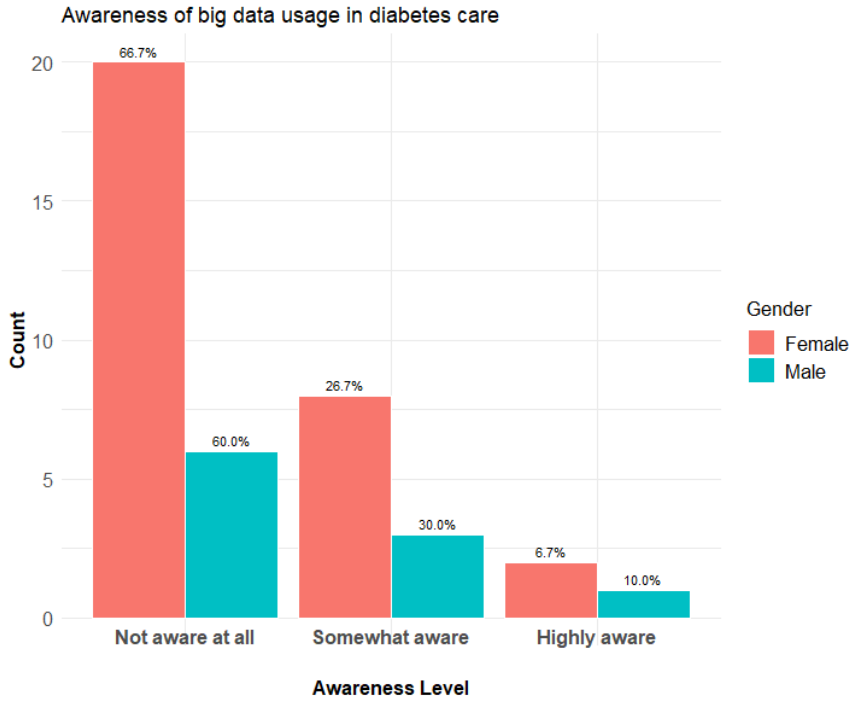


Figure 22: Awareness of big data usage in diabetes care

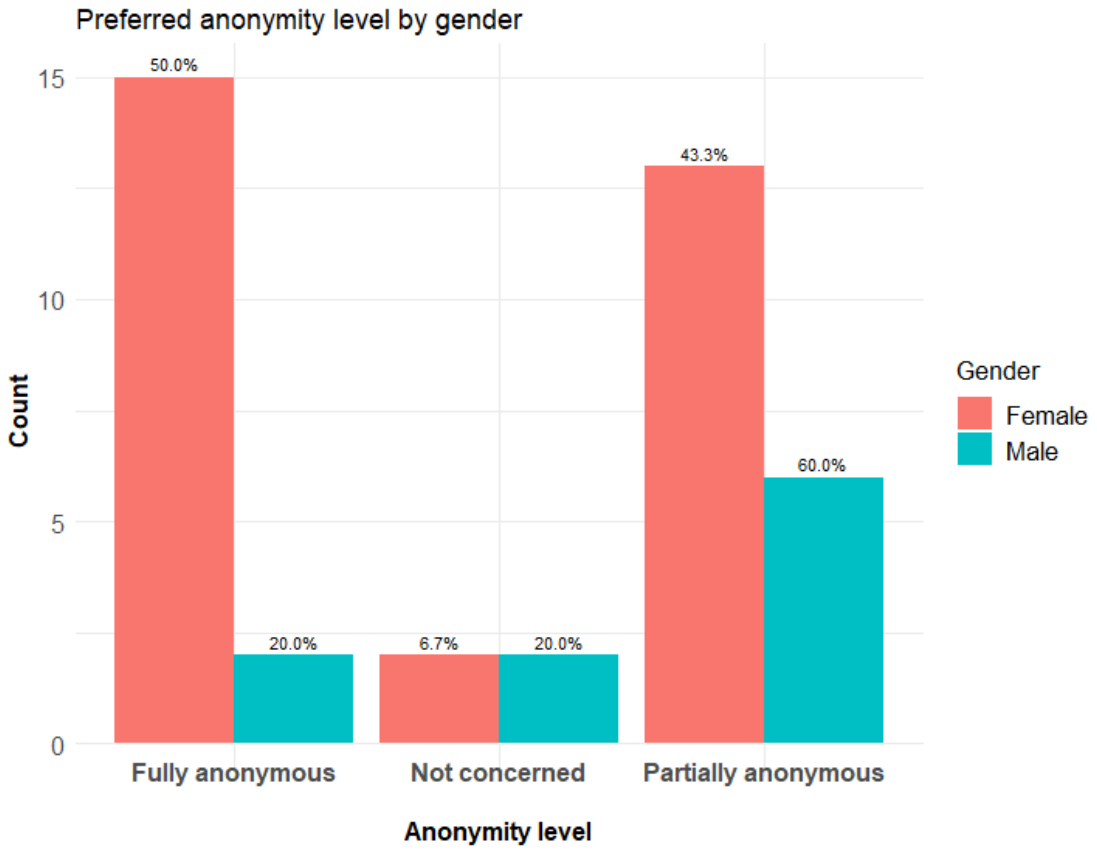


Figure 23: Preferred level of anonymity in sharing data by gender

Figures for analysis based on diabetes history:

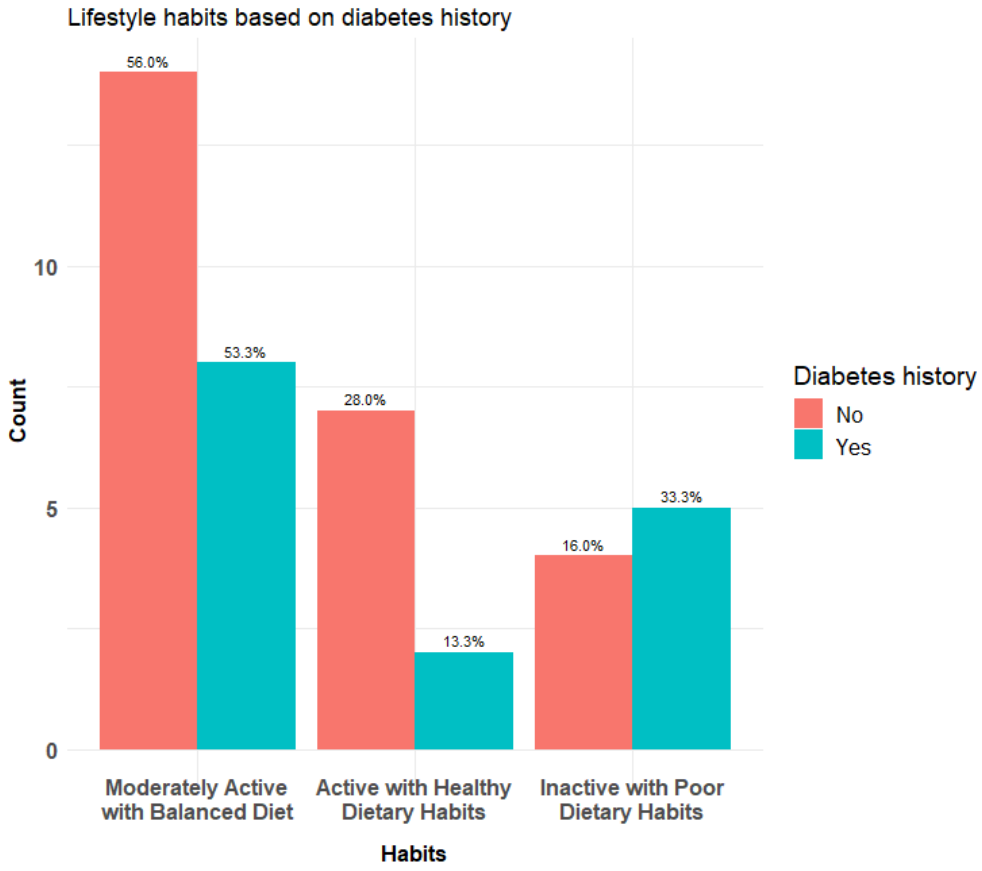


Figure 24: Lifestyle habits based on diabetes history

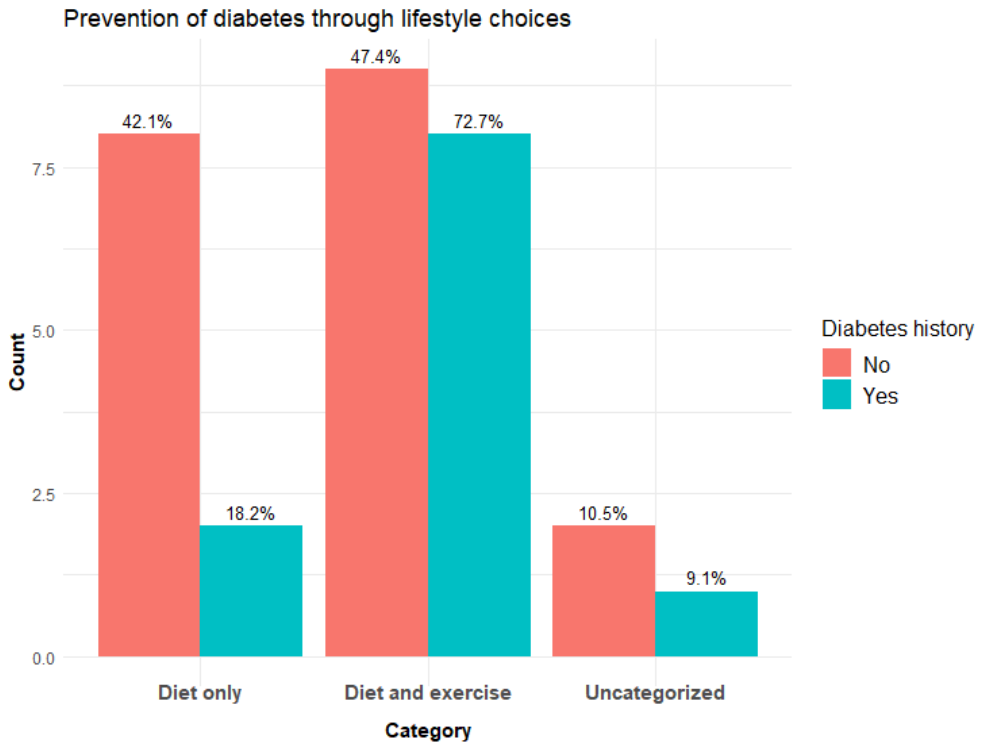


Figure 25: Prevention through lifestyle choices based on diabetes history

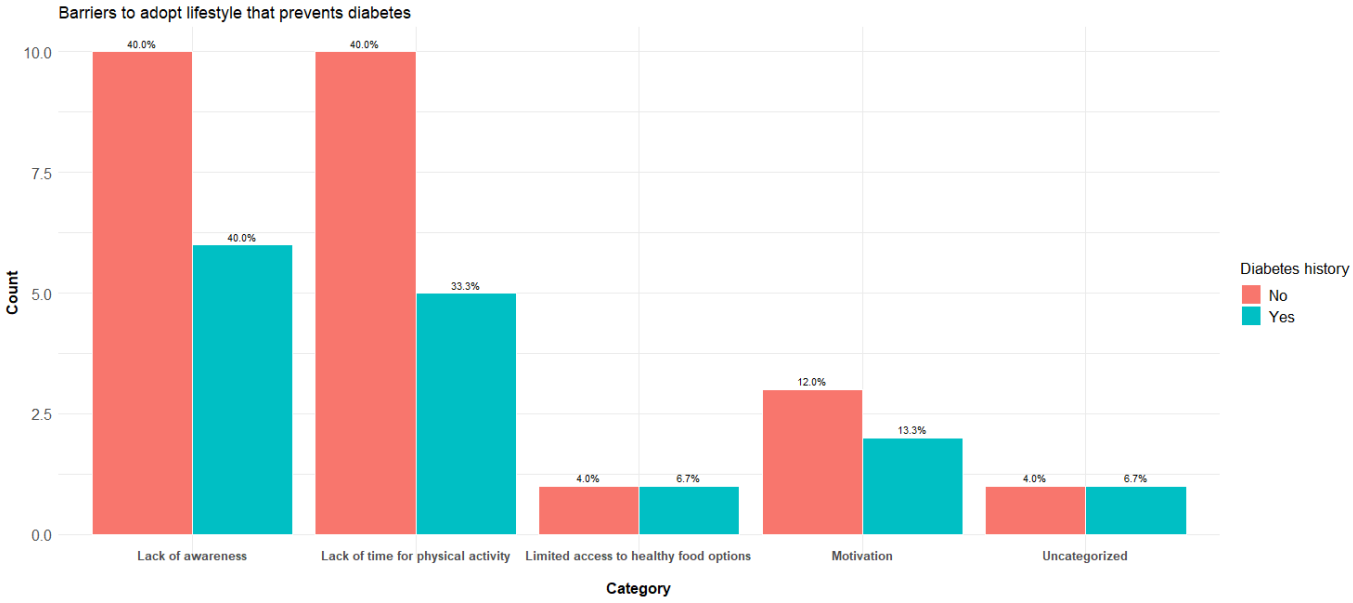


Figure 26: Barriers to adopting a lifestyle that prevents diabetes by diabetes history

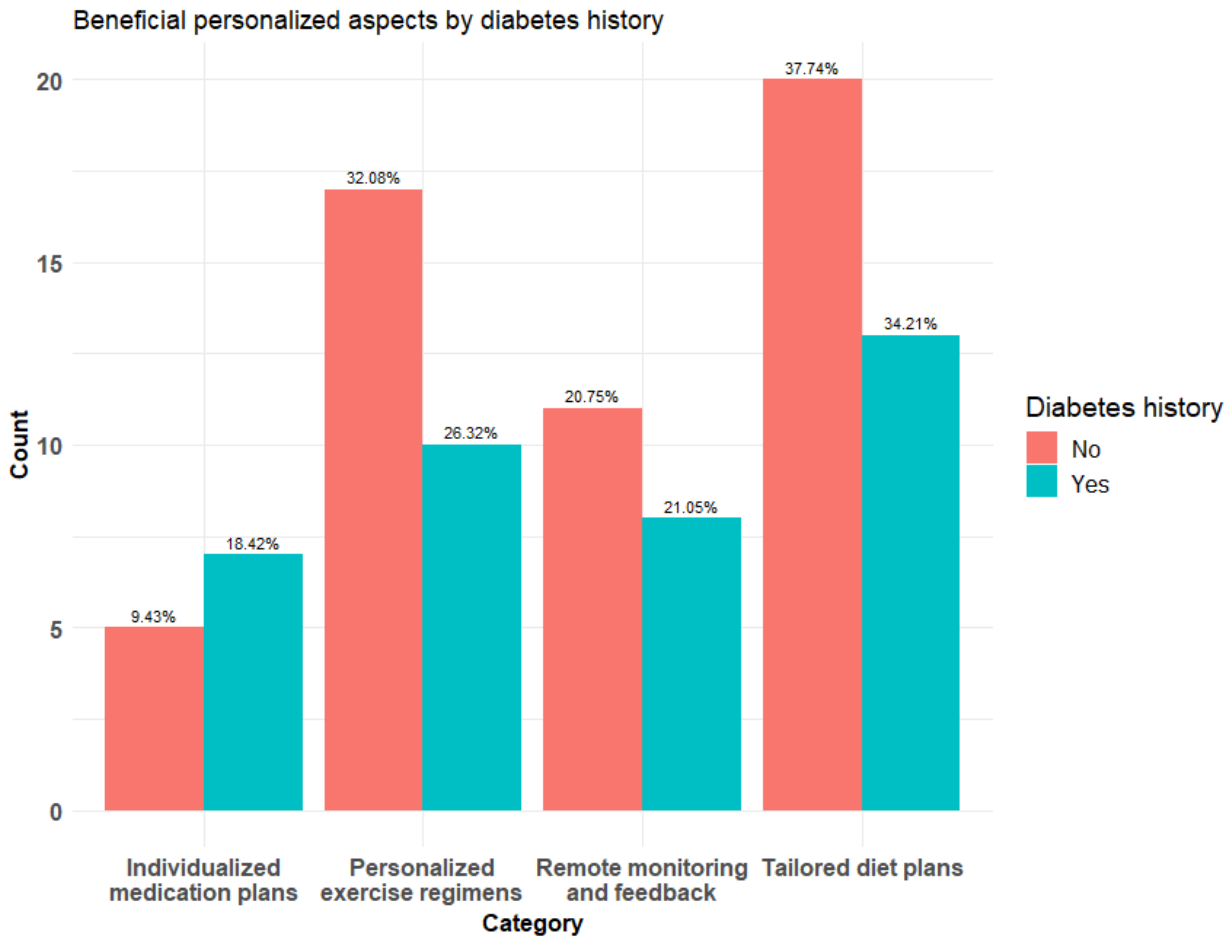


Figure 27: Beneficial personalized aspects by diabetes history

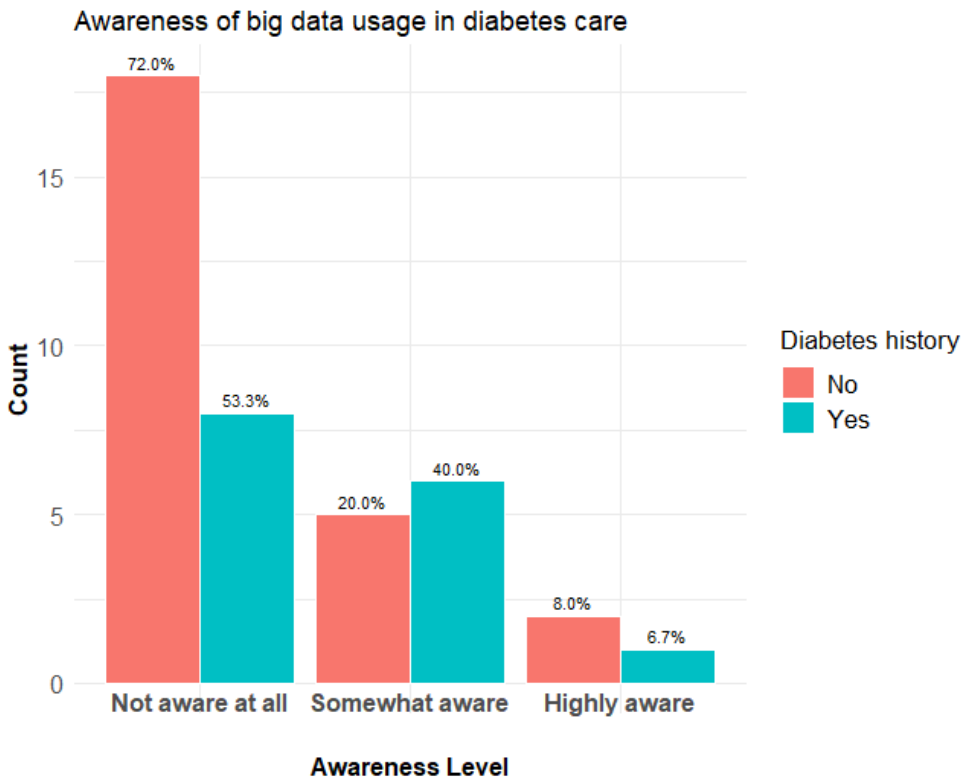


Figure 28: Awareness of big data usage in diabetes care by diabetes history

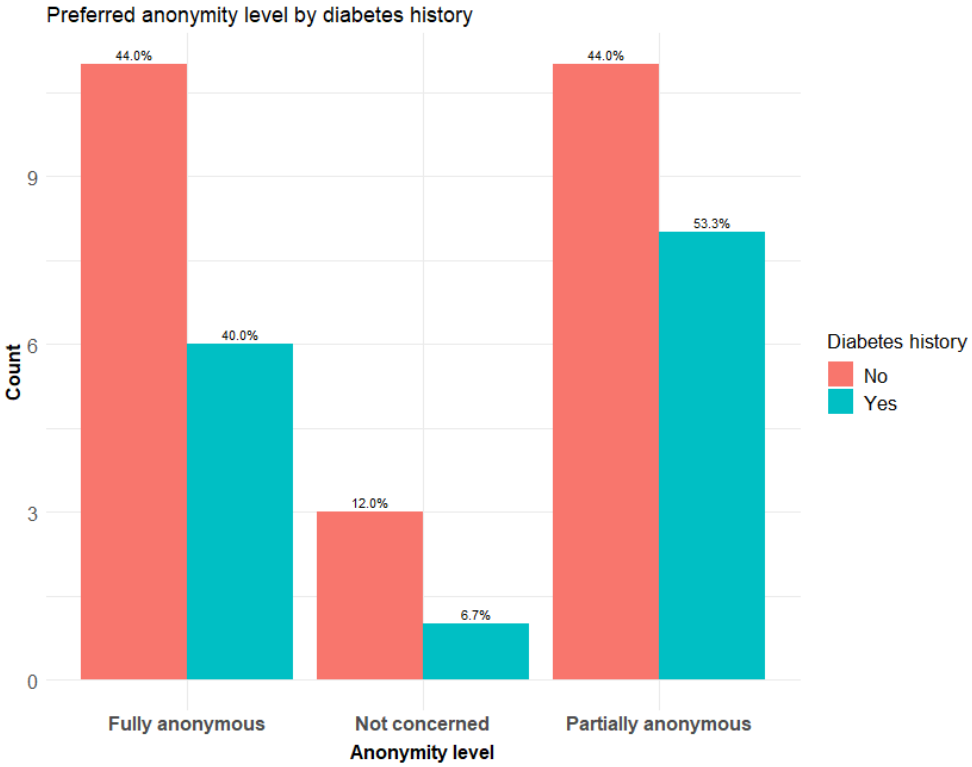


Figure 29: Preferred level of anonymity in sharing data by diabetes history