MSc Thesis Health Science

Input Variable Selection for a Classification Algorithm to Predict Personalised Health Advice in Secondary Prevention

A PILOT STUDY

Enno Kivits
University of Twente | 17-08-2023

Submitted to the faculty at the University of Twente in partial fulfilment of the requirements for the master's degree in Health Science
Enschede

Student:	Enno Kivits
Registration number:	HS-2022066-2-3
Faculty:	Faculty of Science and Technology
Chair group:	Health Technology and Services research (HTSR)
First supervisor:	Prof. Dr. W. van Harten
Secondary supervisor:	Dr. S. Oude Wesselink
Date:	Augustus 17, 2023

Acknowledgements

First and foremost, I would like to express my deepest appreciation to my first supervisor, Prof. Dr. Wim van Harten, who guided me throughout this project. The incisive yet occasionally cryptic feedback triggered me to critically reflect upon my work. I would like to express my deepest gratitude to my secondary supervisor Dr. Sandra Oude Wesselink, who assisted me with the data analyses and for giving confidence during the process. I would like to extend my sincere thanks to my daily supervisor Anne-Jet Jansen. Thank you for being such a kind, enthusiastic and patient supervisor, for always taking time to listen to my struggles and for quickly undertaking action when I needed information. Especially, during the data collection and the process of obtaining approval from the local feasibility committee. I would be remiss in not mentioning dRural team Rijnstate for their support and engagement. I want to thank the staff of the 'Gezondheidsplein' for their cooperation during this research and Rijnstate 'Intern transport & Bevoorrading' for their flexibility which enabled me to combine study and work. Last but not least, I would like to thank Rijnstate 'Innovatie & Zorgtransformatie' for making this research possible.

Abstract

The Dutch healthcare system faces challenges such as an ageing population, staff shortages, fragmentation, and disparities in health literacy. Developing a digital platform that concretely brings together the supply and demand of care and social services can contribute to addressing these challenges. This study examines the relationship between personal characteristics and the domains of personalised health advice that is needed for developing the underlying algorithm of this platform.

To meet this objective, an explanatory pilot study was conducted. The study identified a set of candidate predictors based on existing literature, analysed and evaluated the observations and results of the pilot study, and proposed a revised study design to ensure more robust follow-up research.

Data was collected from the electronic health record and an outpatient clinic in Rijnstate Hospital. Chi-square tests and logistic regression were used to examine the relationships between fourteen personal characteristics (five socio-demographic, two behavioural and seven medical conditions) and twelve domains of health care or social services. Univariate logistic regression analysis indicated a significant association between 'smoking status' and personalised health advice related to nutrition. Another significant association was found for 'bodily functioning', one of the dimensions of positive health, and nutrition. After conducting a multivariate logistic regression analysis with these two variables, only 'smoking status' remained significant (OR = 0.02, 95% CI: 0.00 - 0.26, $p<0.05$). However, the results should be interpreted with caution as this study is not without limitations. The small sample size, a large number of missing data and inappropriate variable selection affect the reliability of the results. A larger sample size is needed to achieve valid results.

As demonstrated in the pilot study, obtaining valid and reliable results regarding the relationship between personal characteristics and the domains of personalised health advice is not feasible due to various limitations. Considering the significant role of the outpatient clinic in recruiting potential patients, it is crucial to ensure that processes, such as the referral process, are established effectively in the first place. The revised study design offers a framework to enhance patient inclusion, data quality, and candidate predictor selection. This revised design can be used in further research to gain a better understanding of the relationship between personal characteristics and the service domains. Considering alternative classification algorithms for selecting the input variables for the final algorithm could be valuable. Future studies should compare techniques to identify optimal input variables for the final algorithm and determine the best-performing underlying algorithm for the platform.

Keywords: Lifestyle coach, Non-communicable diseases, Positive health, Predictors, Regression analysis

Table of Contents

List of abbreviations

AI	Artificial Intelligence
AIC	Akaike information criterion
CVD	Cardiovascular diseases
EHR	Electronic Health Record
ML	Machine Learning
PH	Positive Health
PH_Q	Positive Health Questionnaire
$\chi 2$	Chi-square

1 Introduction

Access to state-of-the-art healthcare services is one of the pillars of the Dutch healthcare system. Although the Dutch healthcare system is considered among the best-performing in the world, it faces major challenges in the future that requires resilience. First, an ageing population, staff shortages, and disparities in health literacy put pressure on the accessibility of the healthcare system. Second, the quality of care is under pressure due to the fragmentation of the system and inefficient and ineffective provision of care since not all medical cases have, by definition, a medical solution. Last, healthcare costs are predicted to triple by 2060 because of an increase in care demand and the use of costly medicines (1). The implementation of new technologies that facilitate the transition to high-quality care and the organizational restructuring of care is needed to anticipate these challenges and contribute to a sustainable healthcare system (2).

The challenge of an ageing population is associated with an increasing prevalence of non-communicable diseases (NCDs) such as cancer, diabetes, cardiovascular diseases (CVD), and Chronic Respiratory Diseases (CRDs). A substantial proportion of this group has multiple chronic conditions. NCDs cause 80% of the disease burden in European countries and leading causes of preventable premature mortality (3). Mortality because of NCDs is higher in people with low socioeconomic status (SES). In high-income countries, SES markers such as income level, educational attainment, employment status, and environmental factors are associated with CVD, CRDs, cancer, and diabetes (4–7). In these studies, patients with a low SES show an increased risk for NCDs.

The needs in healthcare associated with NCDs are complex. Consequently, the need for effective chronic disease prevention and management initiatives grows (8). Broadly, preventive initiatives can be provided at three levels: primary, secondary, and tertiary. Initiatives at the primary level aim to lower or prevent the risk of adverse health conditions. Secondary prevention aims to prevent the progression of a present condition. Tertiary prevention involves clinical interventions such as treatments and rehabilitation (9). In the Dutch healthcare setting preventive initiatives vary from giving advice, making referrals, and counselling to offering lifestyle interventions within the social domain. Generally, lifestyle interventions focus on smoking, unhealthy diets, physical inactivity, and overweight (10). Multiple studies have shown the positive effects of primary and secondary preventive interventions on risk factors of NCDs and the impact of behavioural factors such as smoking, alcohol consumption, unhealthy diet, and physical inactivity, that are associated with NCDs (11–14).

In the last two decades, there has been a growing body of literature that recognizes the importance of exploring innovative approaches to delivering healthcare. One approach is to establish a strong network between the healthcare and social domain. A challenging task because both fields have different approaches. The collaboration between the social domain, primary care, and specialist care is far from being an inherent certainty (15,16). In general, healthcare services provide a medical solution while services in the social domain provide non-medical solutions to meet the patient's needs. Breaking down existing boxes in both sectors is challenging due to the historical multi-sourced funding structures. Furthermore, due to the lack of a clear and comprehensive overview of existing healthcare, social, and welfare services, it is difficult to connect patients to the most suitable care and improve their health. Integrated care is an approach aimed to connect different parts of the healthcare system to improve continuity of care and meet patients' needs in such a way that the right care is delivered at the right place and time and optimal experienced health is achieved (17). Social prescribing (SP) is a strategy to connect healthcare and the social domain. In brief, SP refers patients with health-related needs to non-medical services in the social domain (18).

Rijnstate, a large teaching hospital (Arnhem), recognised the importance of a network in which healthcare and social domain collaborate to deliver optimal care for their patients. In early 2023, Rijnstate opened the 'Gezondheidsplein', an initiative to support patients in improving their lifestyle to prevent the progression of current conditions (i.e., secondary prevention). Patients who are eligible for a consultation with an employee of the 'Gezondheidsplein' are referred by nurses or physicians. After referral, the first step is to identify the patient's needs. The person who identifies the patient's needs and directs them to the appropriate healthcare or social service is referred to as 'lifestyle coach'. In Rijnstate, lifestyle coaches are educated to support patients in maintaining or achieving a healthy lifestyle. The lifestyle coaches include specialized nurses, sports coaches, and an integrative medicine physician. Before the consult, patients complete the 17-item version of the Positive Health Questionnaire (PH_Q) proposed by van Vliet et al. (19). Positive Health (PH) is the underlying concept of the PH_Q, taking a holistic approach to health that encompasses physical, mental, and social well-being and emphasizes the patient's strengths, resources, and abilities. Huber et al. defined health as 'the ability to adapt and self-manage in the face of social, physical, and emotional challenges' (20). PH includes six dimensions: bodily functions, mental functions and perception, spiritual/existential dimension, quality of life, social and societal participation, and daily functioning (21). The PH_Q is used to identify the need for support in these six dimensions. However, the PH_Q tool has not been validated as a measurement tool, so its output cannot be used as such in this study (19). The lifestyle coach uses knowledge and expertise to identify the patient's needs, preferences, and intrinsic motivation and provide a bit of health advice. The health advice encompasses a suggestion for care or social services aimed at enhancing the patients' health through the selection of the most suitable service. In this thesis, the term 'Care or social services', next referred to as services, will be used in its broadest sense to refer to all services, both care and social services, advised by lifestyle coaches aimed to improve health in terms of secondary prevention. After the consultation, the lifestyle coach completes a score table that specifies the domain of the service. The score table consists of 12 domains that cover each service within healthcare and the social domain. From here on the specific domains of the services will be referred to as the service domain 'name', e.g., the service domain 'Lifestyle behaviours'.

Currently, clinical physicians do not use any tool in the referral decision, while lifestyle coaches use the PH_Q as input for health advice. However, technological advancements in data science are expected to automate this process. Big data analytics, which involves analysing vast amounts of data, can provide personalized health services and support decision-making through automated algorithms (22,23). In healthcare, algorithms such as Artificial Intelligence (AI) and Machine Learning (ML) are used for various purposes, including prevention, diagnosis, and treatment prediction (13,24–27). Machine learning algorithms can train prediction models based on sample data, and these models can support clinical decision-making. Recommendation systems, which are algorithms developed to suggest the most suitable intervention for a specific clinical condition, can be used to predict suitable services and connect patients to the right service. To achieve this, a clear and comprehensive overview of the patient's behaviour, lifestyle, genetics, and living conditions is necessary (28). Therefore, it is expected that, in addition to the outcomes of the PH_Q, personal characteristics will enhance such algorithms to predict a suitable service.

1.1 Objective

The objective of this study is to determine which personal characteristics are predictor variables of the domain of a 'Positive Health' based recommendation of care or social services aimed at secondary prevention in non-communicable disease patients. To achieve this objective, the following steps will be taken:

- Identifying a set of candidate predictors that have been highlighted in existing literature as potential predictors for the service domain of the health advice;
- Analysing and evaluating the initial observations and results of the pilot study;
- Revising the initial study design to ensure more robust follow-up research.

1.2 Significance of the study

This study is part of the first phase of the development of an algorithm that will be used for a digital platform. The platform provides an environment where supply and demand for care and social services meet. In this way, the platform will facilitate collaboration between primary care, secondary care, and the social domain. The aim is to improve the current recommendation process by identifying possible predictor variables of secondary care patients to the right secondary preventive service across all levels of healthcare and the social domain. Automating the recommendation process based on personal characteristics will improve the quality of care, in terms of increased efficiency, effectiveness and patient-centredness care. An automatised recommendation system could make a quick, consistent, and accurate recommendation based on large datasets, provided that the model is validated. A validated model can recommend the right healthcare, social- or welfare service with a low error rate. This reduces waste of time, care, capacity and the provision of services to those not likely to benefit. Moreover, the personalized recommendation is attentive to the patient's specific needs, values, and preferences.

2 Theoretical framework

As briefly described in the introduction, the Dutch healthcare system is under increasing pressure. A shift from care towards health has been initiated to maintain the quality, accessibility, and affordability of care in the Netherlands. Prevention and collaboration between healthcare and the social domain are the central topics of two Dutch agreements: the Integral care agreement and the Healthy and Active Life Agreement.

A strong network between all levels of care and the social domain is essential to support people in their lives, health, participation and coping with their health and vulnerability. Promoting a healthy lifestyle and mental health starts with a healthy living environment. A healthy living environment is safe and attractive, invites everyone to engage with one another, promotes healthy behaviours, and is pleasant to live in. Support from both domains has a fundamental role in promoting healthy behaviours.

2.1 Health

Since the WHO introduced its ground-breaking definition of health in 1948, there has been a growing interest in how health should be defined. At that time, WHO defined health as a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity (29). By including mental and social dimensions the WHO widened the view on health. Saracci (30) questioned this definition and stated that a state of complete physical, mental, and social well-being relates considerably more to happiness than to health. Therefore, Sarracci proposed to define health as a condition of well-being, free of disease or infirmity and as a basic and universal human right. Bircher (31) argues that the definition of health should have a dynamic component as demands of life change with each stage of life, depend on culture, and must be fulfilled with personal responsibility. In the following years, the need for a dynamic approach increased. Thus, a few years later, a new concept of health was introduced: 'Health is the ability to adapt and to self-manage, in the face of social, physical and emotional challenges' (20). Between 2011 and 2012 Huber et al. (21) conduct a study to operationalise this concept. The study identified health indicators which were categorised into six dimensions: bodily functions, mental functions and perception, spiritual/existential dimension, quality of life, social and societal participation, and daily functioning, with 32 underlying aspects. For this broad perception of health, Huber et al. proposed the term 'Positive Health' (20,21). The six dimensions and 32 underlying aspects are incorporated into the PH_Q. As mentioned before, the PH_Q has not been validated as a measurement tool; therefore, the output cannot be used as such in this. Vliet et al. (19) aimed to develop a measurement tool for PH and proposed a 17-item model with a six-factor structure, comprising the factors of physical fitness, mental functions, future perspective, contentment, social relations, and daily life management.

2.2 Healthcare

Healthcare services include services being delivered by physicians or other licenced healthcare professionals and provide a medical solution to meet the patient's needs. Health care is delivered on different levels. In 1920, Dawson's Report introduced three hierarchal levels of care (primary, secondary, and tertiary). More than a century later these three levels of care are still relevant to structured healthcare systems. Medical specialisation, fragmentation of health services, and the introduction of publicly funded healthcare all contributed to the necessity for structuring the healthcare system. Primary care, also known as the first level of care, is generalist care, consisting of general medical, paramedical, and pharmaceutical care, nursing and supportive care, physiotherapy and occupational therapy care, and non-specialised mental and social healthcare, together with preventive and health educational activities linked to these forms of care. Care is delivered close to the patient's home and outside the walls of the hospital (32). In contrast to primary care, secondary care is delivered by specialists within the walls of the hospital. This specialist care is provided on an ambulatory or outpatient basis, usually after a referral from primary care. The third level of care, tertiary care, which includes highly specialized care, is beyond the scope of this thesis.

2.3 Social domain

Support in the social and public domain is non-care related and organised by the municipality. Social and welfare services are public or private services providing support and assistance to ensure the patient's well-being. Examples are walking clubs to improve physical activity, social dining to reduce loneliness and nutrition or activities in community centres.

Social determinants of health (SDH) can broadly be defined as non-medical factors that influence health outcomes (33). It encompasses "the conditions in the environments where people are born, live, learn, work, play, worship, and age that affects a wide range of health, functioning, and quality-of-life outcomes and risks". SDH is categorised into 5 key areas: 'economic stability', 'education access and quality', 'health care access and quality', 'neighbourhood and built environment', and 'social and community context'. These categories can be sub-categorised in dimensions. Examples of dimensions are income, employment, education, ethnicity, marital status, living environment and behavioural and mental health. SDH is associated with increased healthcare utilisation and poorer health outcomes (34,35). SDH may impact blood sugar levels, cholesterol and blood pressure resulting in an increased risk of diabetes and CVD (36). In addition, SDH has been associated with Type 2 diabetes, cancers and kidney diseases (37).

According to Gottlieb et al. (38) collecting data on SDH is effective in improving health outcomes. Data on SDH can be used in clinical decision-making and referral to social services. Vest et al. (39) studied the association between adult patient characteristics and SDH service needs to identify patients with SDH needs. In their study, SDH services included: nutritional counselling, behavioural health, respiratory therapy services, patient navigation, and financial counselling. Finding the relation between patient characteristics and SDH services was not part of the study.

2.4 The intersection of Healthcare and the Social domain

Collaboration among diverse stakeholders in the health and social sectors is imperative for addressing impending challenges for a sustainable healthcare system in the future. Nonetheless, attaining such collaboration in practical terms proves to be challenging. This can be elucidated by the existence of distinct barriers that impede collaboration between the various tiers of healthcare and the social domain. In this study patients from secondary care were referred to the 'Gezondheidsplein'. After the consultation, patients received health advice to improve their health. The 'Gezondheidsplein' refers patients to healthcare services in the first level of care or social and welfare services in the social domain. Integrated care and social prescribing are two concepts related to the intersection of healthcare and the social domain.

2.4.1 Integrated care

Previous paragraphs described healthcare and the social domain. It is important to emphasize that both affect health outcomes. Integration of both systems might lead to better health outcomes. Multiple studies show evidence of improved quality of care, increased patient satisfaction and improved access to care because of integrated care. Moreover, integrated care may reduce hospital admission rates and lengths of hospital stay (40). Valentijn et al. (41) define integrated care as ambulatory care settings in which a network of multiple professionals and organisations across the health and social care system provide accessible, comprehensive, and coordinated services to a population in a community. In integrated care, effective collaboration between all levels of care, the social domain and other community care providers is crucial to meet the population's diverse behavioural, social, and physical needs (42). Therefore, a strong network of healthcare institutes, municipalities, and non-medical actors in the social domain is required (43).

However, the fragmentation of the Dutch healthcare system makes it difficult to manage and coordinate such initiatives. Furthermore, in practice cooperation and coordination are limited due to barriers on the macro-, meso- and micro-level (37). Barriers on the macro-level are administrative/regulatory and funding related. In practice, organisations do not have or are not able to acquire sufficient resources to develop integrated care. Furthermore, regulations on different levels hinder collaboration by making the process complex, time-consuming, and costly (44). On a meso-level, inter-organisational and organisation barriers caused by a lack of leadership and coordination impede cooperation and coordination (44,45). In the healthcare system, various stakeholders with divergent interests are involved. Conflicting interests between different organisations hamper collaboration. Barriers on the micro-level cover the domains of service delivery and clinical practice. Lack of commitment, communication, mutual understanding, and technological standards are described as a barrier to corporate (44,46).

2.4.2 Social prescribing

Social prescribing entails the referral of healthcare patients to services in the social domain. In literature, no consensus about the definition of SP, also known as social referral (SR), exists. A rapid review by the University of York defined SP as a way of linking patients in primary care with sources of support within the community (47). The National Health Service has a similar definition and defines SP as an approach that connects people to activities, groups, and services in their community to meet the practical, social and emotional needs that affect their health and well-being (48). In general, referrals from healthcare to the social domain are from primary care, but cases from outpatient services (e.g. oncology and gynaecology), community-based nursing, mental health teams, rehabilitation and intermediate care and acute care or emergency departments have been described (49). Rempel et al. (50) prefer to use the term social referral to describe the links between healthcare and third-sector organisations as it encompasses a broader scope. In this broader scope, they do not limit social referrals to primary care and include referrals by other healthcare professionals. In general, both definitions describe the links between healthcare and the social domain. However, they do not incorporate health and well-being. Vidovic et al. (51) use a more extensive definition: ’Social prescribing is a community-based, person-centred, holistic health coaching scheme, which supports individuals to better understand their needs and take action to improve their health and well-being to reduce demand on health and social care’. This definition is used in this study since it is coherent with the concept of PH and addresses the personal aspect by incorporating patients’ needs.

2.5 Lifestyle coaching

In the last years, lifestyle coaching has been recognised as a potential supporting tool in disease management. Health coaching is described as the process of health promotion and education to improve lifestyle and prevent the progression of chronic conditions (52). To manage their disease, patients receive support in identifying needs, evaluating interventions and choosing the right strategy towards an improved lifestyle (53). A meta-analysis conducted by Long et. al. showed that lifestyle coaching has a significant positive effect on the quality of life and leads to a significant reduction in COPD-related hospital admissions (54). Moreover, a study conducted by Racey et al (55) shows a clinically significant reduction in blood sugar levels (HbA1c) and small benefits for BMI, waist circumference, body weight, and depression/distress post-treatment. A major limitation in the evidence of lifestyle coaching is the lack of a consensual definition of what coaching entails. In addition, it is primarily the interventions or services to which the patient is referred that contribute to the enhancement of health outcomes.

2.6 Automating the process

This relationship between personal characteristics and the provided health advice will be used to automatise the referral process. Currently, clinical physicians and lifestyle coaches provide health advice based on their knowledge and expertise. Clinical physicians do not use any tool in the referral decision. Recent developments in the field of data science have led to an increased interest in automating healthcare processes. Recommendation systems and digital social prescribing are tools described in literature used to provide automatically generated health advice.

2.6.1 Recommendation systems

Studies describing personalised recommendation systems are limited available in the literature. Few studies have demonstrated promising results in personalised recommendations of lifestyle interventions. Chi et al. (56) used an algorithm to learn patterns in personal characteristics and to connect patients to the most suitable intervention that prevents cardiovascular diseases. To determine the relationship between the patient characteristics and an intervention a predictive model was used. Nam and Kim (57) proposed a personalised recommendation system that provides systematic recommendations for lifestyle interventions such as exercise and diet to obese patients. However, these studies focused on interventions in the domain of primary prevention and the target populations of these studies were limited to CVD and obesity.

2.6.2 Digital social prescribing

Patel et al. (58) propose to define digital social prescribing (DSP) as: "social prescription that has been facilitated through the use of technology, such as mobile phone apps and online platforms intended to benefit users.". DSP uses social data combined with health data (i.e., patient characteristics) from the electronic health record (EHR) to enhance clinical care. Patient data from the EHR can be analysed to predict social needs using AI and algorithms (49). Digital social prescribing tools (DSPTs), such as 'Joy' and 'Access elemental' use algorithms to connect patients to the right service (59,60). These DSPTs use patient preferences, comorbidities, and locally offered services to provide personalised health advice. As far as known the algorithm did not include demographic, lifestyle, and disease-specific characteristics. Furthermore, the concept of positive health is not incorporated.

2.7 Predictors

Algorithms for recommendation systems and DSPTs that predict the patients' needs for services use predictor variables, next referred to as predictors. Predictive models are used to determine the relationship between patient characteristics (predictor variables) and lifestyle interventions (outcomes). Predictive models provide estimates of a patient's probability of receiving certain health advice based on personal characteristics. It is expected that the patient's condition is associated with the provided health advice. Correspondingly, certain personal characteristics are associated with the patient's condition. In this thesis, risk factors of NCDs are used as protentional predictors in predicting health advice. Budreviciute et al. (61) proposed a model to classify the risk factors of NCDs. In their model risk factors are classified into five classes: genetic, environmental, socio-demographic, behavioural and medical conditions. Table 1 provides an overview of personal characteristics associated with the NCDs CVD, diabetes, CRDs and cancer. Genetic and environmental risk factors are beyond the scope of this study and therefore not included. In addition, positive health scores are included as predictors. However, the use of these scores in such a manner has not yet been validated in the literature.

Table 1: Risk factors associated with various non-communicable diseases (NCDs). Studies supporting the associations are referenced in parentheses.

Class	Risk factors	NCD	References
Socio-demographics	Age	CVD, Diabetes, CRDs, Cancer	(62–66)
	Gender	CVD, Diabetes, CRDs	(62–71)
	Degree of urbanisation	CVD	(68,72,73)
	Educational level	CVD, Diabetes, CRDs, Cancer	(73)
	Employment	CVD, Diabetes, CRDs, Cancer	(73)
Behavioural	Alcohol consumption	CVD, Diabetes, CRDs, Cancer	(62,64,71,73,74)
	Smoking	CVD, CRDs, Cancer	(62–74)
Medical condition	Diastolic blood pressure	CVD, Diabetes, Cancer	(61–63,70)
	Systolic blood pressure	CVD, Diabetes, Cancer	(61–63,67,68,70,72)
	BMI	CVD, Diabetes, Cancer	(61–63,65,68,70,71,74)
	Total cholesterol	CVD, Diabetes	(61–63,68,72)
	LDL cholesterol	CVD, Diabetes	(61,62)
	HDL cholesterol	CVD, Diabetes	(61–63,68,72)

3 Method Pilot study

3.1 Setting

This pilot study was conducted at Rijnstate Hospital, which comprises four locations situated in the Arnhem region of the Netherlands. Rijnstate Hospital is recognized as a top clinical hospital, serving a service area encompassing approximately 433,000 inhabitants. Annually, the hospital attends to 500,000 outpatients and 60,000 inpatients (75). With a strong emphasis on prevention, the hospital strives to establish itself as a frontrunner in this domain and is one of the co-founders of the prevention network Community of Care Gelderland Midden (76) . This collaborative network aims to foster cooperation among stakeholders dedicated to the promotion of health and well-being among the region's residents.

This study is part of the dRural project, a European project with over 30 participants from 8 countries (77). The project part in the Netherlands focuses on the so-called “Care region of the future” in which digitalisation plays an important role. A study to develop a digital marketplace to connect patients to the right service is part of this and contributes to overcoming barriers in prevention as described in Chapter 2 (Theoretical framework). The digital marketplace will be a platform where healthcare and social care organisations offer their services. Inhabitants and caregivers can use the platform to search for suitable services and meet demands. The platform aims to connect the right healthcare or social service suitable to the patient's needs. To support this process, Rijnstate initiated the development of the underlying algorithm.

3.2 Study design

This explanatory quantitative observational study includes data collection, regression analyses and a research report. Patients’ positive health scores, patients’ characteristics, and recommendations from the employees of the ‘Gezondheidsplein’ will be used in the regression analysis. The report will provide an overview of what personal characteristics should be included in the algorithm that will be developed to identify patients’ personal needs for services. The steps are described in more detail below.

3.3 Study population

This study exclusively focused on patients (≥ 18 years old) who were referred to the 'Gezondheidsplein' between January 12, 2023, to June 30, 2023, and received health advice aimed at secondary prevention. The population consisted of patients who visited the outpatient clinics and departments of Cardiology, DIVAN (Diabetes, Vascular, and Nephrology), Gastrointestinal Oncology, Gynaecology, Pulmonology, and Oncology at Rijnstate and were referred based on expert opinion by nurses or physicians. No participation from other hospitals occurred within the scope of this study.

The inclusion criteria were:

- Patients completed the PH_Q (Appendix 1)
- Patients received health advice from a lifestyle coach of the ‘Gezondheidsplein’.
- The patients’ scoring table (Appendix 2) is filled out by a lifestyle coach.

3.4 Patient inclusion

To access patient data, approval was requested by Rijnstate’s local feasibility committee (LFC. As this study uses patient data, the patient’s informed consent was required. A patient information form was used to inform the patient (Appendix 3). After the LFC’s approval was obtained, the inclusion process started. The researcher did not have a treatment relationship with the patients. Therefore, the lifestyle coaches, who did have a treatment relationship with the patients, were tasked with asking them if they were willing to participate in this study. This was done by sending an e-mail to the patients with a patient information form (PIF). At least 7 days after receiving the PIF, patients who did not object to being contacted were called by the researcher and asked if there were any questions regarding the study. During the phone call, the patients were given the opportunity to indicate their understanding of the provided information and declare (verbally) that their patient data could be used for this study. Patients who had a consultation with the 'Gezondheidsplein' after the data collection started, received the PIF during the consultation. After reading the PIF, the lifestyle coach asked whether the patient wanted to participate or not. Patients who met the inclusion criteria declared to fully understand the PIF and provided verbal consent were enrolled in this study.

To ensure patient privacy, study IDs were used. Before anonymisation, a key list with identifiable data was made. In this list, patient name, date of birth, phone number, HiX numbers (patient EHR number) and study IDs were documented. The key list enabled the researchers to contact patients in case of problems such as a data leak. The list was not used to collect any data. Subsequently, a screening list with study IDs was created to log which patients were approached by the researchers and whether they provided consent.

3.5 Data collection

The study data were collected from May 19, 2023, to June 30, 2023. During this period, eligible participants were recruited and enrolled in the study. Before the consultation, the patient filled out the PH_Q. The patient gains access to the questionnaire through a link that is sent to them after scheduling the appointment. The patient completes this questionnaire independently. In case a patient did not fill out the PH_Q, this was done at the start of the consultation. Filling out the PH_Q was part of the regular care process and was done to identify the patient's needs for services. Needs were identified by a score within the six dimensions of PH on a scale from 1 – 10. Scoring was done using a digitalised 17-item version of the PH_Q proposed by van Vliet et al. (19) (Appendix 1). In this study only items 10, 17, 22, 29, 36, and 43 were used to determine the PH-score's per dimension. This digital version was made before the start of this study. The researcher was not involved in the development of the questionnaire.

After the consult, the lifestyle coach filled out the scoring table to specify the health advice given to the patients. Lifestyle coaches noted the service domain and the specific advice. Both the PH_Q and the scoring table were filled out using Microsoft Forms.

Personal characteristics were extracted from HiX using the data extraction tool CTcue. The software has two workflows: patient finder and clinical data collector. 'Patient finder' was used to compile the study cohort. Patients were added to the study manually by searching the patient's HiX number. CTcue automatically generates pseudo-IDs to ensure patient privacy. After selecting the study cohort, data were extracted from the EHR using the clinical data collector. In CTcue queries are used to collect data in structured and unstructured form. In general, structured data comprise numeric values such as lab and vital sign measurements. Unstructured data is any type of data in free text fields and consists of documentation reported by health professionals. Appendix 4 shows the query used to collect data.

3.6 Study variables

3.6.1 Candidate predictors

Patient characteristics were selected as candidate predictors predicting the service domain. These predictors were categorised into socio-demographic, behavioural, medical conditions, and positive health characteristics. The candidate predictors were selected by conducted literature research (Table 1). Candidate predictors were both continuous and categorial.

- **Age:** Captured from the EHR at the time of inclusion.
- **Gender:** Captured from the EHR. Reported as:
    - male
    - female
- **Degree of urbanisation:** Derived from postal code captured from the document. Reported as (inhabitants per km2 land):
    - not urbanised (<250)
    - hardly urbanised (250 < 500)
    - moderately urbanised (500 < 1000)
    - strongly urbanised (1000 < 2000
    - extremely urbanised (> 2000)
- **Educational level:** Captured from the PH_Q form. Reported as:
    - low (Basisonderwijs/VMBO/MAVO)
    - medium (MBO/HAVO/VWO)
    - high (HBO/Universiteit)
- **Employment:** Captured from the scoring table filled out by lifestyle coaches. Reported as:
    - employed
    - unemployed
    - retired
    - volunteering
- **Alcohol consumption:** Patient's alcohol consumption. Captured from the EHR. Reported as:
    - no
    - yes
- **Smoking status:** Patient's current or past smoking status. Captured from the EHR. Reported as:
    - never
    - current
    - stopped
- **Department of referral:** The department or outpatient clinic within the hospital referred the patient to the 'Gezondheidsplein'. Captured from the documentation of the 'Gezondheidsplein', reported as:
    - cardiology
    - DIVAN
    - gastrointestinal oncology
    - gynaecology
    - lung
    - oncology
- **Body Mass Index (BMI) (km/m^2)**: A measurement used to assess the patient's body weight in relation to their height. Captured from the EHR as the last available measurement.
- **Blood pressure systolic (mmHG):** Pressure in the blood vessels when the heart contracts (systolic). Captured from the EHR as the last available measurement.
- **Blood pressure diastolic (mmHG):** Pressure in the blood vessels in between contractions (diastolic). Captured from the EHR as the last available measurement.
- **Cholesterol (mmol/L)**: Captured from the EHR as the last available measurement.
- **Low-density lipoprotein (mmol/L)**: Captured from the EHR as the last available measurement.
- **High-density lipoprotein (mmol/L)**: Captured from the EHR as the last available measurement.

- **Bodily function:** Relates to the patient's body. The extent to which the patient feels healthy, feels fit, and can exercise. Self-reported by the patients in the PH_Q. Measured on a scale from 1 to 10, where 1 represents the worst score and 10 represents the best score.
- **Mental well-being:** Relates to the patient's feelings and thoughts. The extent to which the patient can remember things, concentrate, and knows what to do when things are not going well. Self-reported by the patients in the PH_Q. Measured on a scale from 1 to 10, where 1 represents the worst score and 10 represents the best score.
- **Meaningfulness:** Relates to a meaningful life. The extent to which the patient worries about the future and is eager to do things in life. Self-reported by the patients in the PH_Q. Measured on a scale from 1 to 10, where 1 represents the worst score and 10 represents the best score.
- **Quality of life:** Relates to the patient's quality of life. The extent to which the patient is happy, feels good and can cope with life. Self-reported by the patients in the PH_Q. Measured on a scale from 1 to 10, where 1 represents the worst score and 10 represents the best score.
- **Participation:** Relates to the patient's participation in society. The extent to which the patient interacts with others, has a sense of belongingness, and can ask people around them for help. Self-reported by the patients in the PH_Q. Measured on a scale from 1 to 10, where 1 represents the worst score and 10 represents the best score.
- **Daily functioning:** Relates to the patient's daily life. The extent to which the patient knows their own capabilities, how to live healthy and can plan their day. Self-reported by the patients in the PH_Q. Measured on a scale from 1 to 10, where 1 represents the worst score and 10 represents the best score.

Measurements older than 1 year were reported as missing.

3.6.2 Outcome variables

The outcome variables were the twelve service domains as specified by an expert group consisting of, among others: a PhD student, a strategic policy advisor social domain (Municipality Lingewaard), and employees of an organization that is engaged in the implementation of broad welfare work in an adjacent municipality (Lingewaard).

The following twelve service domains were specified:

- **Lifestyle behaviours:** aimed to support with bad lifestyle behaviours such as addictions. Example services: smoke cessation, alcohol or drugs counselling, and gambling therapy
- **Mental well-being**: aimed to support emotional stability, resilience, positive self-perception, and the ability to cope with the stresses of life. Example services: Adult day care, social worker, psychologist, and psychiatrist.
- **Nutrition**: aimed to support with dietary habits. Example services: social dining, meal provision, dietitian, and combined lifestyle interventions (CLIs).
- **Physical activity**: aimed to support with exercise. Example services: walking groups, personal training, physiotherapist, and CLIs.
- **Sleep/Relaxation**: aimed to reduce stress and improve night's sleep. Example services: relaxation techniques, meditation, music therapy, and acupuncture.
- **Physical well-being:** aimed to support with physical fitness and overall functional well-being. Example services: general practitioner, pharmacist, pain relief techniques, acupuncture, and osteopathy.
- **Residence/Household**: aimed to support with the living situation and conditions. Example services: supported living, home adaptations, home care, and informal care.

- **Social environment/contacts**: aimed to support with the social setting in which the patient lives and contact with others. Example services: clubs, groups and events in local communities, community centres, adult day care, and informal care.
- **Financial/work**: aimed to support with handling depts, financial abuse, and job seeking. Example services: financial administration support, free pass for low-income individuals, and budget and debt counselling.
- **Transport and services**: aimed to support with mobility issues. Example services: transport pass (free or reduction), community transport, and disability parking permits.
- **Personal care**: aimed to support with daily routines, such as washing, dressing, and getting out and about. Example services: general practitioner, home care, pedicure, and ergotherapy.
- **Child-rearing**: aimed to support with looking after your child, child's development or wellbeing. Example services: domestic abuse support, day care, after school care, preschools, playgroups, street coaches, and youth worker.

Subsequently, lifestyle coaches from the 'Gezondheidsplein' allocated appropriate care and social services to the twelve service domains. The specification of the domains and their corresponding services had been previously conducted and was not part of this study.

It should be noted that the outcome variable of this pilot study was the service domain, not the specific service. In practice, the lifestyle coach advises on the best-suited service. In the scoring table, lifestyle coaches were instructed to first specify the service domain and then provide the details of the specific service itself. However, predicting the specific advice was beyond the scope of the pilot study.

3.7 Data preparation

The output of the PH_Q scoring table and CTcue was captured in separate Excel files. For authority reasons, the researcher did not have any influence on the compilation of data from the PH_Q and scoring table. Data other than the study variables section described data was excluded. The output of the queries used was directly exported from CTcue. Before data was exported from CTcue a variable selection was made. Subsequently, all files were combined in one dataset.

The study encountered a large number of missing data, especially in data extracted with CTcue. No complete cases occurred. Variables with >50% missing data were excluded from the analyses. Excluded variables were blood pressure systolic (54%), blood pressure diastolic (54%), total cholesterol, LDL and HDL cholesterol (Appendix 5).

3.8 Statistical analyses

A descriptive analysis was conducted on the patient's characteristics (socio-demographics, behavioural, and medical conditions). The results were reported in terms of the mean ± standard deviation (SD) for continuous predictors and as frequency n (%) for categorical variables. The Chi-square test (χ^2-test) was used to examine the relationship between categorical predictors and outcome variables. In addition, logistic regression was performed to examine the strength and direction of the relationship between predictors indicating a significant association. For continuous predictors, logistic regression was employed to examine the relationship with outcome variables.

3.8.1 χ^2-test

The χ^2-test is a test to examine whether predictors indicate a significant association with the outcome variables (Equation 1). The results of the X^2-test are reported as the X^2. In all tests, the null hypothesis stated that there is no relationship between the predictor and the service domain. Predictors with a p-value $p < 0.05$ were marked as significant.

$$\textit{Equation 1}: x^2 = \sum \frac{(O-E)^2}{E}$$

Where χ2 is the test statistics, O is the observed number of frequencies, and E is the expected number of frequencies.

3.8.2 Logistic regression

In logistic regression, the outcome variable represents the probability of an event occurring, such as receiving health advice in a certain service domain or not. This outcome is coded as a binary variable, with a value of 1 indicating that the patient received the health advice in a certain service domain and a value of 0 indicating that the patient did not receive the health advice in that service domain.

> Y_i = 1 ⇔ patient i received health advice in a certain service domain i.e., 'Lifestyle behaviours'
>
> Y_i = 0 ⇔ patient i did not receive health advice in a certain service domain.

Univariate logistic regression was used to identify significant predictors associated with the service domain (Equation 2). Multivariate logistic regression was then performed to analyse the relationship between these significant predictors (the result of univariate analyses) and the outcome variable (Equation 3). Subsequently, odds ratios (OR) were obtained by exponentiating the coefficient.

$$\textit{Equation 2}: P_{(Y)} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

Where P is the probability of getting health advice in a certain service domain, and i is the regression coefficient of the X and Xi predictor.

$$\textit{Equation 3}: P_{(Y)} = \frac{e^{\Sigma\beta_i+\beta_i x}}{1+e^{\Sigma\beta_i+\beta_1 x}}$$

Where P is the probability of getting health advice in a certain service domain, and i is the regression coefficient of X and Xi predictors.

One assumption in logistic regression is linearity between the logit of the outcome and each continuous predictor. This assumption was assessed using a visual interpretation of the scatterplots. Variables that did not meet the assumption of linearity were excluded. Additionally, logistic regression assumes there is no multicollinearity between predictors. Variance inflation factor (VIF) was used to indicate multicollinearity.

For each OR the 95% confidence interval is given within parentheses. The OR indicates the strength and direction of the association between the six dimensions of positive health and the service domains. OR's greater than 1 indicate a positive association between the predictor and the outcome. Conversely, ORs less than 1 indicate a negative association. OR's equal to 1 suggest no association. For both χ2 and logistic regression, p-values are reported as the significance threshold. The reported p-values were solely reported as a threshold to determine whether the variables, based on the univariate analyses, should be included in the multivariate analysis. The p-value was not taken into account when analysing the strength and direction of the association. Therefore, outcomes that are not significant according to the p-values are still reported.

3.9 Analysis Pilot study

The previously described method was used to gain a better understanding of the primary data. The first step was analysing the referral process. This was done by evaluating the process with referrers, lifestyle coaches and the program manager of the

‘Gezondheidsplein’. Next, the inclusion process as described in section 3.4 was analysed. Both processes were visualised in a swimming lane flowchart. After patient inclusion, data collection started as described in section 3.5. The data collection process was evaluated to determine how data collection can be improved to ensure data quality. The analyses of the referral, inclusion and data collection processes are presented in part I of the results section. Next, the results of the exploratory statistical analyses on the primary data are analysed (results II). Finally, the insights from the pilot study were critically evaluated to modify the initial explorative study design and a revised design is proposed (results III).

4 Results I - Process analyses Pilot study

4.1 Referral process

The referral begins at the ward or outpatient clinic, where the nurse (specialist) or physician identifies eligible patients (Figure 1). Nurses play a vital role in referral, as they have close and regular interactions with patients. Due to this close involvement, nurses are well-positioned to identify eligible patients. Therefore, physicians and especially nurses should possess the necessary competencies to recognise patients who need support in improving their lifestyle. However, from brief dialogues with ward nurses, it became clear that they were not directly aware of the 'Gezondheidsplein' and their role as referrers. Nurses tend to miss or omit signals when their role in referring is unclear. Moreover, uncertainty about the referral role might result in a lack of accountability leading to disengagement (78).

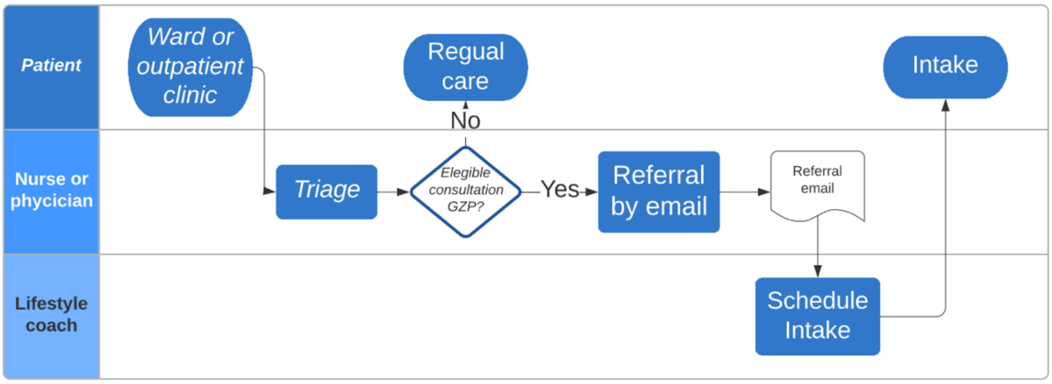


Figure 1: Swimming Lane chart referral process

4.2 Documentation process 'Gezondheidsplein'

In total 133 patients (≥ 18 years old) were referred to the 'Gezondheidsplein' (Appendix 6). Seven patients (5.3%) had a no-show for the first appointment. Patients were called by the lifestyle coaches and appointments were rescheduled. After a second attempt, 7 out of 13 patients ultimately attended the rescheduled appointment. Reasons for no-shows were not documented.

The deliberate adherence to broad inclusion criteria aimed to encompass a substantial number of patients within the study cohort. However, in retrospect, a considerable portion of the population had to be excluded from the study because they did not meet two essential inclusion criteria regarding the PH_Q and the scoring table, respectively 46.5% and 42.5% of the patients. Most of these patients visited the 'Gezondheidsplein' shortly after the outpatient clinic opened in March. According to the program manager, at that time, the processes around the consultation were not well established. Moreover, the limited presence of the program manager hampered their ability to direct this process. Brief interviews with lifestyle coaches revealed that they do not consider the PH_Q as the primary source of input for giving advice. Instead, one lifestyle coach emphasised that the information and observations obtained during the interview are the main input for giving advice. This could explain why consultations proceeded without the completed PH_Q.

Additionally, lifestyle coaches were requested to specify the given advice in a scoring table directly after the consultation with the patient. Particularly, during the start-up phase of the 'Gezondheidsplein', the lifestyle coaches did not consistently fill out the provided scoring tables. The scoring tables are accessible via a link that directs the lifestyles coaches to the form. Filling out the scoring table was not seen as part of the lifestyle coach's primary task. Their primary tasks involve identifying needs and providing advice. Filling out the scoring table should be incorporated into the process since it was observed that filling out the form was not fully ensured within the process during the study period. Moreover, it was found that filling out the form was not intuitive. On multiple occasions, the form was inaccurately edited by lifestyle coaches, necessitating the developers to redraft the form.

4.3 Recruitment process

In total, 33 patients were eligible to the study of which 24 (72.7%) provided verbal consent to extract data from the patient's EHR after being approached by the researchers (Appendix 6). As described in section 3.4 patients who did not object to being contracted were called by the researcher to ask for verbal consent. Most patients saw but did not read the e-mail with the attached PIF. Consequently, the intention and methodology of the study had to be explained during this call. Calling each patient is infeasible for a larger patient cohort as the average duration of phone calls to patients was 4 minutes. Moreover, the 7-day objection period set by the LHC creates a delay in inclusion and shortens therefore the inclusion period.

4.4 Estimated sample size

The 'Gezondheidsplein' distinguishes between two types of consultations: intake consultations lasting 60 minutes, and follow-up consultations lasting 30 minutes. A lifestyle coach estimates that by the end of 2023, a realistic number of consultations based on the available working hours (30 hours allocated for consultations per week) would be around 10-15 intake consultations and 20-40 follow-up consultations per week. Considering the capacity of the "Gezondheidsplein", there is potential to increase the number of referrals, thereby enabling the inclusion of a greater number of patients.

According to a projection by the programme manager, most departments will have started referring patients by the end of 2024, and relevant staff will have received training on preventive care. At present, predicting the exact number of patients referred by then remains a challenge. However, based on existing figures on referrals and inclusions, it becomes possible to estimate a sample size to be achieved by the end of 2023.

Based on the lifestyle coach's estimation, the 'Gezondheidsplein' has a weekly capacity of 10-15 intake consultations. Assuming enough patients are eligible for consultations, it is conceivable that 40 - 60 patients could receive Health advice each month. Thus, within the timeframe from July to the end of December (6 months), the potential range would be 240 - 360 patients. However, the findings reveal that approximately 30% of patients do not provide consent of using their data in the study. Consequently, by the end of this year, given the completion of all requisite documents (PH_Q and scoring table), a projected 168 - 252 patients could still be eligible for inclusion. Note that this number could be higher if more departments and outpatient clinics start referring patients.

4.5 Data quality

Data was collected from documentation of the 'Gezondheidsplein' (PH_Q, scoring table, excel file) and patients 'EHRs. The retrospective nature of the study affected the quality of the data. In terms of accessibility, data was only accessible by requesting the necessary data from the lifestyle coaches. Currently, the 'Gezondheidsplein' does not have its environment in the EHR; hence, data can only be accessed by employees directly involved with the outpatient clinic.

Access to data from the EHR could be obtained in three ways, namely: directly through HIX, CTcue, or by submitting a request to BIT. Due to time constraints, it had been decided to access the data via CTcue. However, the choice of this software affected the quality of the data. Patient history and assessments are stored in the EHR in structured and unstructured form. In general, clinicians prefer to use unstructured data whereas those reusing data for scientific research prefer structured data. Structured data was easily accessible as CTcue has features to report patient data in a structured way using forms and pre-defined query terms. In contrast, obtaining unstructured data was challenging.

Unstructured documentation contained more detailed information. Creating a specific query was, therefore, necessary for finding and extracting the desired data. Depending on how different individuals construct a query, distinct datasets can be retrieved.

The results of the query used to collect data showed that there was currently no consistent way of reporting. Details about whether a patient smokes cigarettes or drinks alcohol or not can be reported in forms with 'Yes' or 'No'. In practice, such information is reported in unstructured free-text fields. Similarly, numerical data such as blood pressure, BMI or cholesterol can be reported in forms with corresponding units (mmHg, kg/m^2, mmol/L). However, these measurements were often found in free text fields.

Timeliness is another crucial factor affecting data quality. In this context, the standard retrieval of the latest available data by CTcue is important. Up-to-date information is essential for accurate analysis. By collecting the latest data, CTcue improves the reliability and usability of the information collected, thus contributing to better data quality. However, measurements such as blood pressure and cholesterol were outdated (>1 year old) and therefore labelled as missing data.

In terms of completeness, the dataset had 2 (8.33%) complete cases (Figure 2). Figure 2 shows the pattern of missing data. The left vertical axe represents the number of cases, and the right vertical axe represents the number of missing data. The bottom horizontal axe represents the number of missing entries with respect to the variables on the top horizontal axe. For example, the figure illustrates that data on cholesterol, and consequently on LDL and HDL, were missing in 19 cases. Moreover, the figure shows that in 13 cases, data on blood pressure were missing.

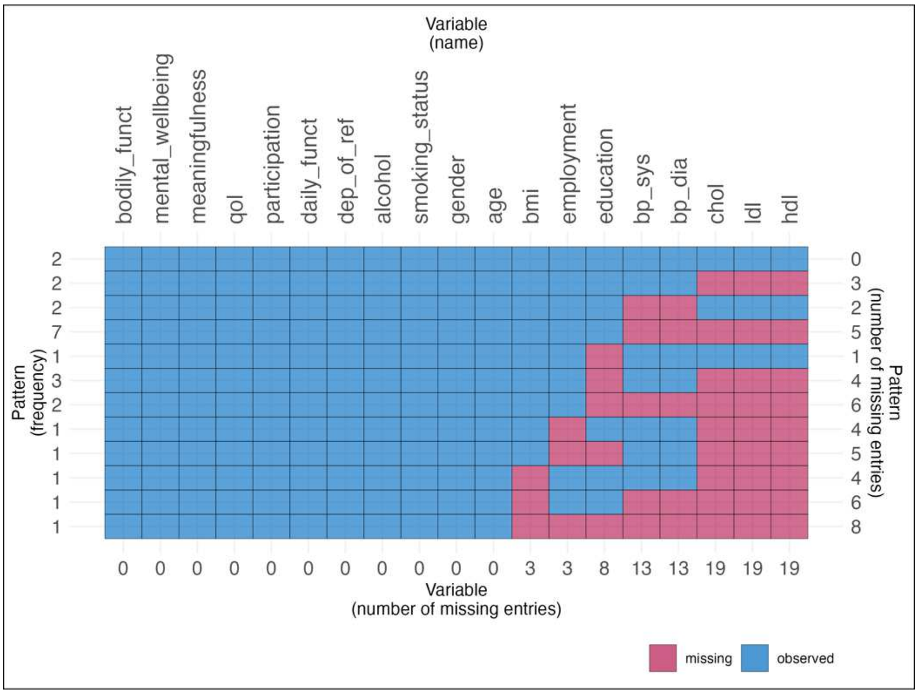


Figure 2: Missing data pattern plot

5 Results II - Data analyses Pilot study

5.1 Patient cohort characteristics

The patient cohort visiting the 'Gezondheidsplein' and participating in this pilot study (n=24) had an average age of 49 ± 17.7. Among the patients, 91.7 % were female. At the time of inclusion, most patients 47.6% were employed. Regarding behavioural characteristics, most patients who visited the "Gezondheidsplein" were smokers or former smokers (66.7%). Nine patients (37.5%) reported alcohol consumption. Patients were referred from 6 departments. Nine patients (37.5%) were referred from the gynaecology department, while from the oncology and lung departments, both 5 patients were referred (20.8%). All referred from gynaecology received health advice in the service domain 'lifestyle behaviour'. Table 2 provides a comprehensive description of the study population, encompassing socio-demographic, behavioural, and medical condition characteristics.

Table 2: Patient cohort characteristics (n=24). Missing data is not included in the table.

Socio-demographic		
Age (Mean ± SD)		49 ± 17.7
Gender = Female (%)		22 (91.7)
Education (%)	Low	4 (25.0)
	Medium	7 (43.8)
	High	5 (31.2)
Employment (%)	Employed	10 (47.6)
	Unemployed	4 (19.0)
	Volunteering	2 (9.5)
	Retired	5 (23.8)
Degree of urbanisation (%)	Not urbanised	1 (4.2)
	Hardly urbanised	2 (8.3)
	Moderately urbanised	11 (45.8)
	Strongly urbanised	9 (37.5)
	Extremely urbanised	1 (4.2)
Behavioural		
Smoking status (%)	Never	8 (33.3)
	Current	8 (33.3)
	Stopped	8 (33.3)
Alcohol = yes (%)		9 (37.5)
Medical condition		
Body Mass Index (Mean ± SD)		34.66 ± 8.16
Cholesterol (Mean ± SD)		4.78 ± 0.65
Low-density lipoprotein (Mean ± SD)		2.55 ± 1.09
High-density lipoprotein (Mean ± SD)		1.40 ± 0.17
Department of referral (%)	Cardiology	2 (8.3)
	DIVAN	2 (8.3)
	Gastrointestinal Oncology	1 (4.2)
	Gynaecology	9 (37.5)
	Pulmonology	5 (20.8)
	Oncology	5 (20.8)

The patient's cohort PH scores obtained from the PH_Q are summarised in Table 3. Among the six dimensions of positive health, the included patients scored the lowest in the dimension of bodily functioning and the highest in the dimension of meaningfulness.

Table 3: Population's positive health scores obtained from the PH_Q

Dimension	Score (Mean ± SD)
Bodily functioning	5.67 ± 1.31
Daily functioning	6.75 ± 1.59
Mental well-being	7.00 ± 1.62
Quality of life	7.08 ± 1.53
Participation	7.17 ± 1.17
Meaningfulness	8.04 ± 1.12

The results show advice was given in five of the twelve service domains. Of the 24 patients, 13 received advice in the service domain 'Nutrition', 9 in the service domain 'Lifestyle behaviours', and 8 in the service domain 'Physical activity', with the fewest advice in service domain with the fewest advice in service domain 4 (n=6). The advice within the service domain 'sleep/relaxation' were excluded as there were just 2 events. No advice was given in the other domains. Appendix 7 includes a table with the advice per patient characteristic with respect to the service domains.

5.2 Statistical analyses

5.2.1 Association between categorical predictors and the service domains

To examine the relationship between categorical characteristics and the service domains, X^2-tests were conducted. The results reveal a significant association between 'Smoking status' and the service domain 'Lifestyle behaviours' (χ^2, = 13.87, df = 2, p<0.001). Similarly, a significant association was found between 'Smoking status' and the service domain 'Nutrition' (χ^2 = 9.40, df = 2, p<0.001). No other variables showed a significant association with the service domains. Table 4 summarises the results obtained from the χ^2-tests.

Table 4: Results Chi-square tests for categorical variables

	Lifestyle behaviours		Mental well-being		Nutrition		Physical activity	
Characteristics	**χ2**	**P-value**	**χ2**	**P-value**	**χ2**	**P-value**	**χ2**	**P-value**
Sociodemographic								
Gender	0	1.000	0	1.000	0.75	0.387	1.70	0.192
Education	0.97	0.615	1.37	0.504	2.96	0.228	5.20	0.074
Employment	1.71	0.636	5.08	0.166	4.16	0.245	0.53	0.913
Degree of urbanisation	2.47	0.650	2.13	0.712	5.69	0.223	2.93	0.569
Behavioural								
Smoking status	13.87	**< 0.001***	4.00	0.135	9.40	**< 0.001***	2.63	0.269
Alcohol consumption	0	1.000	0.00	1.000	0.10	0.751	0.20	0.655
Medical condition								
Department of referral	3.09	0.686	3.60	0.825	3.56	0.829	7.46	0.382

*Significant (p < 0.05)

Since the χ^2-test lacks information on association strength and direction, significant variables were analysed using univariate logistic regression. No significant association was found between smoking status and the service domain 'Lifestyle behaviour'. The analyses revealed a negative association between 'Smoking status' and the services domain 'Nutrition' (OR = 0.02, 95% CI: 0.00 - 0.26, p<0.05). In this study, however, we are predominantly interested in positive associations, where the presence of certain characteristics increases the likelihood of health advice in a specific domain.

5.2.2 Association between continuous predictors and the service domains

The relationship between continuous variables and the service domains of the health advice was assessed by conducting univariate logistic regression analyses. The analysis revealed that patients with a higher score on the bodily function dimension of positive health were less likely to receive health advice within the service domain 'Nutrition' (OR = 0.44, 95% CI: 0.17 - 0.91, $p<0.05$) (Table 5). In addition, patients with a higher BMI were found to be less likely to receive health advice within the service domain 'Lifestyle behaviours' (OR = 0.65, 95% CI: 0.34 - 0.86, $p<0.05$).

The univariate logistic regression analyses with the PH-scores as predictor and 'Nutrition' as outcome revealed a negative association between all the PH-scores and 'Nutrition'. These findings suggest that for patients with lower scores on the dimensions 'Bodily functions', 'Meaningfulness', 'Quality of Life', 'Participation', and 'Daily Functioning', the likelihood of receiving health advice in the 'Nutrition' domain increases compared to patients with a higher score.

Counterintuitive results were found in the analyses with the PH scores as a predictor and 'Lifestyle behaviour' as the outcome. It is important to highlight that higher scores on the PH dimensions are considered positive. For every 1-point improvement on the Bodily Functions score, the likelihood of receiving health advice in the 'Lifestyle Behaviour' domain increases by 1.85. Similar results were observed for the dimensions of 'Mental well-being', 'Quality of Life', 'Participation', and 'Daily Functioning'.

Table 5: Results from the univariate analyses examining the associations between continuous predictors and the service domains.

	Lifestyle behaviours		Mental well-being		Nutrition		Physical activity	
predictor	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**
Positive health								
Bodily functions	1.85 (0.92 – 4.56)	0.118	1 (0.48 – 2.15)	1.000	0.44 (0.17 – 0.91)	**0.049***	0.59 (0.27 – 1.18)	0.157
Mental well-being	1.56 (0.86 – 3.57)	0.207	0.71 (0.36 – 1.27)	0.259	0.93 (0.53 – 1.58)	0.796	0.80 (0.43 – 1.38)	0.423
Meaningfulness	1.09 (0.51 – 2.48)	0.811	0.79 (0.31 – 1.87)	0.595	0.58 (0.21 – 1.26)	0.207	1.11 (0.50 – 2.57)	0.793
Quality of life	1.35 (0.75 – 2.94)	0.376	0.95 (0.51 – 1.88)	0.875	0.70 (0.32 – 1.25)	0.290	0.80 (0.41 – 1.43)	0.454
Participation	1.98 (0.91 – 5.28)	0.114	0.70 (0.29 – 1.61)	0.415	0.65 (0.29 – 1.34)	0.266	0.83 (0.37 – 1.75)	0.615
Daily functioning	1.71 (0.89 – 4.66)	0.193	0.81 (0.45 – 1.47)	0.460	0.88 (0.48 – 1.49)	0.649	0.93 (0.53 – 1.67)	0.782
Sociodemographic								
Age	1.06 (1.00 – 1.13)	0.053	0.95 (0.87 – 1.00)	0.100	0.96 (0.91 – 1.00)	0.133	1.00 (0.94 – 1.04)	0.709
Medical condition								
BMI	0.65 (0.34 – 0.86)	**0.042***	1.04 (0.92 – 1.20)	0.549	1.09 (0.97 – 1.25)	0.171	1.12 (0.99 – 1.31)	0.094

*Significant ($p < 0.05$)

5.2.3 Multivariate logistic regression

Multivariate logistic regression is used to obtain a more comprehensive understanding of this relation. The multivariate analysis revealed that smoking status remained significant for current smokers and service domain 'Nutrition' (OR = 0.02, 95% CI: 0.00 - 0.35, $p<0.05$). Hence, for current smokers, the odds of health advice in the service domain

‘Nutrition’ were 0.03 times the odds of patients who never smoked (Table 6). The ORs were adjusted for the effects of other predictors in the model.

Table 6: Results of the multivariate logistic regression examining the association between univariate significant predictors and the service domain 'Nutrition'.

		Nutrition	
		Adjusted OR (95%CI)	**p-value**
Bodily function		0.47 (0.14 – 1.16)	0.141
Smoking status	Current	0.03 (0.00 – 0.35)	**0.010***
	Stopped	0.21 (0.01 – 2.62)	0.268

6 Results III - Revised study design

6.1 Study design

As described in section 4.5 the pilot had to deal with poor data quality in terms of accessibility, completeness, and timeliness. Adopting a prospective research design can significantly enhance data quality. A major disadvantage of using retrospective clinical data is the high variability, as there are often no standardized methods used for data collection. With a prospective design, data collection is more purposeful. The key advantage of such a design is that the researcher maintains control over the data collection process from the beginning on. Furthermore, the researcher can exert quality control throughout the study. For optimal fitting of the model to the population, a heterogeneous sample with high variability in predictor variables is preferred. A heterogeneous cohort makes it difficult to select candidate predictors that match each case. The use of a prospective study design limits the presence of unmeasurable candidate predictors. Candidate predictors can be more accurately tailored to the patient cohort. In the initial design, the candidate predictors cholesterol, LDL, and HDL were excluded, as it appeared that they were not part of routine clinical care for the patients or not relevant to report. The pilot study faced a large number of missing data and adopting a prospective design may be an appropriate alternative to ensure improved data quality.

6.2 Data splitting

Before model development and variable selection, the existing dataset should be split into two, namely a training dataset and a testing set. The training set will be used for estimating coefficients, assessing collinearity, stepwise selection, and model evaluation. The test set should only be used to assess the performance of the final model. To avoid bias, it is crucial not to use any of the samples that are part of the test set during model development. A standardly used ratio for splitting data is 80:20 train-test split. That is, 80% of the data is for training and 20% for testing (79).

However, the results of the pilot study show that the proportions of outcomes were unevenly distributed. For example, in the domains of lifestyle behaviours and physical activity, the ratios were 63:37 and 66:33, respectively. Therefore, the stratified random sample method should be used to split the data. This is done by selecting samples at random within the classes of the outcome variable, i.e., patients with health advice in the service domain 'Lifestyle behaviours' and patients without advice in that domain. This ensures an equal distribution of the outcome within both sets.

6.3 Sample size

In general, sample size calculations are a key component of quantitative studies. However, early exploratory pilot studies with limited data might not require such calculations, as there may not be enough data available to perform such calculations (80). The drawback is that the obtained results are not reliable, valid, and likely not representative of the actual population. To obtain an estimate of the minimum sample size, rules of thumb can be employed. A commonly accepted rule of thumb for logistic regression is to have a minimum of 10 events per variable (EVP). This minimum is required for the accurate estimation of regression coefficients (81). However, if the number of events is insufficient for certain domains, these might be excluded from the model. In the pilot, this rule of thumb is only met in the service domain 'Nutrition' as there were 13 events.

In the initial design 20 predictor variables were selected to predict the outcome variable (1 of the 12 service domains). Assuming balanced data, the probability of getting advice in 1 of the 12 service domains is 8.33%. To achieve the required 10 EVP a minimum of 10 patients per outcome variable are needed. That is a minimum of 120 patients per variable is required to meet the EVP rule. In theory, therefore, the pilot study required a sample size of 2400 (20*120). This calculation does not yet consider data splitting (section 6.2). Applying the standard 80:20 train-test split, a total sample size of n=3000 (2400*100/80) patients is required. However, in practice, out of the total of 133 referred patients, 78 had consultations between January 12, 2023, and June 30, 2023. Furthermore, the estimation of the sample size (as seen in section 4.4) indicates that, in the most favourable scenario, an additional 252 patients could still be included by the end of 2023. Therefore, it can be concluded that achieving a sample size of n=3000 is not feasible in the short term. To narrow the significant gap between the currently feasible and the required sample size, a thorough evaluation of the set of candidate variables and the referral process is imperative.

6.4 Reduction candidate variables

Derived from the estimated feasible sample size, the maximum number of variables can be estimated. The estimated sample size is approximately 275 patients (24 included patients + 252 eligible patients by the end of 2023). The results of the pilot study showed that only 5 domains were represented in the selected patient cohort. Assuming a balanced dataset, the probability of getting advice in 1 of the 5 service domains is 20%. To achieve the required 10 EVP a minimum of 10 patients per outcome variable are needed. That is a minimum of 50 patients per variable is required to meet the EVP rule. In theory, therefore, 5 candidate predictors (275/50= 5.5) could be selected based on the estimated sample size. In the initial design 20 candidate predictors were selected, this means that 15 candidate predictors should be excluded. Experts' opinion and assessing for collinearity can be used to shrink the number of candidate variables.

6.4.1 Expert's opinion

Accurate results in model building require proper predictor selection. Inappropriate predictor selection can introduce biased relationships. Moreover, every redundant variable unnecessarily increases the required samples size. Aligning the candidate predictors accurately with the patient cohort increases the yield of a study with a similar study design. Consulting experts such as physicians or lifestyle coaches about what variables should be in the model enhances the candidate predictor selection. Observations from the pilot study indicate that dialogue with the patient and observations during the consultation serve as the primary inputs for the advice provided. The lifestyle coaches do not consciously consider specific characteristics when giving advice; instead, they rely on their knowledge and expertise. Currently, the advice of the lifestyle coaches is regarded as the golden standard. Hence, involving the lifestyle coaches in the selection of variables becomes crucial. Conducting focus group interviews with physicians, nurses and lifestyle coaches is an effective method for adequately pre-selecting and reducing the number of candidate predictors. Furthermore, a selection of clinically relevant outcome variables should be part of the interview. As a result, the specification of the domains needs to be carefully reassessed.

Reducing the outcome variable from 12 to 5 service domains has a significant effect on the required sample size. Instead of a minimum of 120 patients per variable, a minimum of 50 patients per variable is required to meet the EVP rule. In theory, therefore, the revised study requires a sample size of 1000 (20*50). This calculation does not yet consider data splitting (section 6.2). Applying the standard 80:20 train-test split, a total sample size of n=1250 (1000*100/80) patients is required.

6.4.2 Collinearity

Assessing for collinearity is a statistical approach to reduce the number of predictors. To examine the individual relationship between predictors and the outcome, predictors should be tested for collinearity. Collinearity occurs when a pair of predictor variables are correlated or associated with each other. The presence of collinearity is expressed in the VIF. A VIF > 10 indicates a strong correlation between predictors (82). Correlating predictors do not affect the predictive power or reliability of the full model but only influence the estimates of individual predictors. A group of correlating predictors combined may predict the outcome. However, the influence of individual predictors is not clear. For the complexity and interpretability of the model, selecting fewer variables is preferred. Therefore, it suffices to include only one of the correlating variables in the model because the other variables do not add any additional information. Combining multiple predictors variables is also a possibility to avoid collinearity and lower the required sample size.

Collinearity is foreseen between diastolic blood pressure and systolic blood pressure, two interrelated variables. These variables can be used to derive the mean blood pressure, a consolidated metric. Furthermore, an expectation of collinearity between cholesterol, specifically LDL and HDL, exists. It is possible that LDL and HDL might be redundant, and their inclusion as potential predictors could be reconsidered. Consolidating diastolic and systolic blood pressure and excluding LDL and HDL yields a requisite sample size of n=1060. Regrettably, this figure remains unattainable within the current context. Therefore, it is advisable to initiate efforts aimed at increasing the sample size through incremental inclusions.

6.5 Referral process

In the context of 5 clinically relevant service domains, meeting the EVP rule requires an increase of 50 patients for each additional predictor variable. To attain an increase in the number of inclusions, several actions should be considered. Increasing the number of inclusions begins at the ward and outpatient clinic levels. As previously mentioned, physicians and nurses play a key role in the referral process. Hence, these key stakeholders must have a comprehensive understanding of their role. The 'Gezondheidsplein' program manager emphasized the importance of nurses' awareness in the referral process. Fostering this awareness starts with ensuring the visibility of the "Gezondheidsplein" This entails integrating physicians and nurses into the vision and aspirations of the "Gezondheidsplein". The benefits for both the patient and the caregiver should be delineated. Furthermore, the program manager mentioned that nurses are not necessarily educated to work with a focus on preventing diseases but rather on treating diseases. Both viewpoints require different skill sets. In preventive healthcare, the emphasis is on addressing illness rather than promoting health. This change in perspective is also evident in how definitions of health have evolved over the years, as discussed earlier at the start of Chapter 2.

Conducting workshops is an appropriate strategy to initiate this paradigm shift. The workshops can contribute to promoting the Gezondheidsplein, inform stakeholders about its objectives, and impart necessary skills to medical professionals. Once the added value of the 'Gezondheidsplein' is clear, attention can be focused on the current situation. Presently, an average of 16 patients per month are referred, whereas the 'Gezondheidsplein' possesses significantly greater capacity. Together with the department concerned, an analysis must be undertaken to determine the reasons behind the current lower-than-anticipated referral rate. Once these factors are elucidated, efforts can be directed towards identifying the requisite changes needed to increase the referral rate. To determine whether the proposed changes are an improvement, the 'Gezondheidsplein' should provide departments and other stakeholders with comprehensive feedback on outcomes, encompassing both successes and failures. Based on the feedback, it should be assessed whether the changes lead to the expected result. If little change is visible, another evaluation will have to take place. After this, newly proposed changes can be put into practice.

6.6 Variable selection

After determining the set of candidate variables, the right variable selection method should be applied. In practical and statistical terms, fewer predictors are preferable for simpler, more interpretable, and practical models (83). However, essential variables must not be excluded to preserve accuracy and relevance. A commonly used variable selection method in medical applications is the stepwise backwards selection method (84). One advantage of this method is that multiple models with varying predictors can be examined easily with the use of statistical software. Another advantage is the reproducibility of results which helps model validation. Stepwise backward selection starts with a full model, which includes all selected candidate variables. The candidate predictors should be selected by experts' opinions and conducting a literature search.

Under the condition that the assumptions are not violated (section 3.8.2), the selection process starts with the construction of the full model. If the outcome is binary and the relationship between the patient's characteristics and the service domain is significant, multivariate logistic regression should be used as a statistical method to fit the model. Next, the Akaike information criterion (AIC) should be calculated for all predictors in the full model. The predictor with the lowest AIC should be excluded from the model. If the AIC of the model of the excluded variable is lower than the AIC of the previous model, it indicates that the exclusion of that predictor has resulted in a greater improvement in the model's fit compared to the increase in model complexity Again, the AIC should be calculated. Iteratively, the predictors with the lowest AIC should be excluded one by one until further removal of predictors no longer leads to a substantial improvement in the AIC. The model with the lowest AIC indicates the predictors for the final model.

6.7 Model evaluation

As described, the final model is developed based on the training set. The next step is to evaluate the performance of the model using the unseen data from the test set. The aim of this is to determine the extent to which the patterns and relationships found can be applied beyond the training data. As the selected predictors serve as input for the final algorithm for the platform, it is important to evaluate the extent to which the selected predictors can predict the service domain.

6.7.1 Confusion matrix

A confusion matrix could be used for the evaluation. This tabular representation of the model is used to compare predicted outcomes with the actual outcomes for binary outcomes (event and non-events) (Table 7). The table lists the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) Several performance metrics can be derived from the confusion matrix. Accuracy reflects the proportion of correct predictions by dividing the number of correct predictions (TP + TN) by the total number of predictions. The model aims to predict the events i.e., the right service domain. Therefore, it is also relevant to look at the proportion of the number of correct positive predictions compared to the total number of positive observations (positive predictive value), also known as the precision of the model. One last relevant metric that can be obtained from the confusion matrix is sensitivity. This metric defines the probability that the service domain is predicted correctly for all samples with a positive observation. Sensitivity is calculated by TP/(TP+FN). n the other hand, specificity defines the number of true negative classifications among the total number of negative observations. Specificity is calculated by TN/(TN + FP).

Table 7: Confusion matrix showing predictions versus the actual observations.

		Observation			
		True	*False*		
Prediction	*True*	True positive	False positive	→	Positive predictive value
	False	False-negative	True negative	→	Negative predictive value
		↓	↓		
		Sensitivity	Specificity		

6.7.2 Receiver operating characteristics curve

A cut-off for the predicted probability for binary outcomes is needed to classify patients as either positive or negative. In classification, a patient is classified as positive if the prediction is higher than the cut-off; otherwise, they are classified as negative. In predictive modelling, there is a trade-off between sensitivity and specificity. Increasing the cut-off results in decreasing sensitivity and increasing specificity. On the contrary, decreasing the cut-off results in increasing sensitivity and decreasing specificity. The receiver operating characteristic (ROC) curve is used to plot the trade-off results between sensitivity and specificity by various cut-offs ranging from 0 to 100%. However, the ROC curve itself is not informative.

6.7.3 Area under the curve

To evaluate the performance of the model, the area under the ROC curve, also known as the area under the curve (AUC) is calculated. Strong-performing models have higher ROC curves closer to the upper left corner and therefore a larger AUC. A perfect model, with an AUC of 1, predicts the outcome correctly for every patient. A model with an AUC of 0.5, represented by a 45-degree line, does not perform better than a model that randomly predicts the outcome.

7 Discussion

This study explored the relationship between personal characteristics and the service domain of a 'Positive Health' based advice of care and social services aimed at secondary prevention. The results of the pilot study revealed the first associations, including their direction and strength. Significant associations were found for 'Smoking status' and the service domains 'Lifestyle behaviours' and 'Nutrition' as an outcome of the χ^2-tests. The univariate logistic regression analyses for 'Smoking status' and the service domain 'Lifestyle behaviours' showed no relationship between these variables. Univariate logistic analyses revealed significant associations between 'Bodily functioning' and the service domain 'Nutrition' and 'BMI' and the service domain 'Lifestyle behaviours'. Patients with a higher score on the bodily function dimension of PH were less likely to receive health advice within the service domain 'Nutrition'. Additionally, patients with a higher BMI were less likely to receive health advice in the service domain 'Lifestyle behaviours'. The findings from the pilot study have been used, together with the observations from the process analyses, to revise the initial study design. An extended dataset with a minimum of 1060 samples is necessary to increase the statistical power, validity, and reliability of the results. The observations from the process analyses indicate that optimisation of the inclusion process may contribute to achieving a larger sample size. Data quality will improve by imputing missing data and improved variable selection. Imputing missing data offers several advantages, including preserving data completeness, producing unbiased estimates, enhancing statistical power, and improving model performance. In addition, using a more appropriate method for candidate predictor selection will result in a better alignment of predictors and the characteristics of the patient cohort. Finally, internal validation should be done to assess the performance and reliability of the model on unseen data.

7.1 Process analyses Pilot study

As the process analysis shows, the referral process typically starts on the ward (except for a small number of walk-ins). Physicians and nurses play a crucial role in patient referral due to their close and regular interactions with patients. Their involvement puts them in a good position to identify eligible patients. The findings of the analysis of the referral process align with Kemppainen et al's (85) integrative review, which extensively studied the role of nurses based on existing studies published between 1998 and 2011. Kemppainen et al identified four major competencies of health promotion. The concept of health promotion extends beyond merely recognising eligible patients and is mostly beyond the scope of the referral role. However, developing some of the competencies related to health promotion can contribute to scaling up the referrals to the 'Gezondheidsplein' and, consequently, increasing the sample size. For instance, a comprehensive understanding of health across different age groups, and epidemiology and disease processes of various conditions is essential. Furthermore, nurses must possess various communication skills to gather accurate information relevant to referrals and to motivate patients to participate in lifestyle improvement. Jallinoja et al. (86) conducted a questionnaire study to investigate the perception of healthcare professionals' roles in managing lifestyle-related diseases. Physicians and nurses are aware of their role in providing information and encouraging patients to make lifestyle changes. Another finding of this study is that 45% of the physicians and 42% of the nurses estimated that they lack the skills for lifestyle counselling. In addition, physicians (61%) and nurses (50%) reported workload is too high to identify the patient's needs. Geense et al (87) identified barriers to delivering health promotion activities. Lack of skills or time to discuss lifestyle and lack of evidence of the effectiveness of lifestyle improvement activities were mentioned by general practitioners and practice nurses as barriers. While these observations were mainly made in primary care, they may also contribute to the lower-than-expected referrals observed in the pilot.

Furthermore, the observations emphasize the crucial role of the 'Gezondheidsplein' as a healthcare facility.

The next step in the process was the documentation of information at the 'Gezondheidsplein'. Documentation starts when the referral email is received. Referrals made using free text fields, such as emails, may contain less detailed information compared to those made using an electronic referral system. The use of structured text fields can enhance the quality and completeness of referral information(88). Enhancing the quality and accuracy of referral information improves understanding of the patient's condition, medical history, and specific concerns, thereby enhancing data quality.

7.2 Data analyses Pilot study

Although the pilot study is small and explorative, a relevant significant association between 'Bodily functioning', one of the six dimensions of PH, and the service domain 'Nutrition' was found. The characteristic 'Bodily functioning' encompasses the assessment of patients regarding the extent to which the patient feels healthy, feels fit and can physically exercise. Patients with a higher score on the bodily function dimension of PH were less likely to receive health advice within the service domain 'Nutrition' (OR = 0.44, 95% CI: 0.17 - 0.91, $p<0.05$). In reviewing the literature, no data was found on the association between dimensions of PH and interventions aimed to improve PH scores. A possible explanation for this is that the PH_Q has not undergone validation as a measurement tool. Consequently, it would be inappropriate to use it for studying the effects of potential interventions. Regarding the outcome of the analysis, several studies (89–93) confirm that poor nutrition is associated with impaired bodily functions.

7.3 Univariate selection prior to model selection

Applying appropriate variable selection is an important prerequisite in building a model and ensuring the validity of the results. In the initial study design, the selection of candidate variables was done by using univariate analyses. Variables that showed a significant association ($p<0.05$) with the outcome variable were included in the final model. However, employing this method is discouraged, as it could decrease the performance of the model since non-significant predictors can still add value to a model (94). Furthermore, models built using this method showed low validity in new populations (95). Univariate analyses do not account for interactions between different predictors. However, individual predictors that show a weak association in the univariate analysis can potentially add value to the model when combined. Applying higher p-values ($p<0.25$) is a suggested option that can help avoid this problem partly. However, this can lead to an overfitted model, especially with a small sample size (96). In a simulation study, Hafermann et al (97), examined model performance where two different p-value thresholds were used during variable selection. In addition, the performance was examined using the backwards selection method. The univariate selection with a p-value of 0.2 showed the highest true positive rate, followed by univariate analyses with a p-value of 0.05. The worst true positive rate was found for the backwards selection. However, a simulated study may not accurately reflect real-world practice.

Although univariate selection is relatively simple and applicable, it also has some limitations. It may lead to incorrect estimates, overestimation of the model's performance, and inappropriate variable selection due to confounding variables (98). Serval authors suggest using backwards selection (96,99,100). The main advantage of this method is that standard errors, p-values, and confidence intervals can be adjusted for confounding variables and interactions. However, backward selection also has some limitations. Van der Weele (94) argues that estimates and confidence intervals are no longer valid when applying the most common approach of using the same data for variable selection and fitting the model. Another drawback is that the sample size must be large enough to fit the full model with all candidate predictors included. Therefore, candidate predictors must be selected adequately. It is important to include only variables known to be associated with the outcome or likely to be so.

7.4 Traditional models vs machine learning

In our study, a traditional approach was used to examine the relationship between personal characteristics and service domains of health advice. Logistic regression was proposed as a statistical model to determine what variables should be selected as input for the algorithm. A classification algorithm, such as logistic regression, can also be used to predict whether patients received certain health advice in a specific domain or not. Although logistic regression is an efficient way to examine the relationship between independent variables and dichotomous outcomes, its performance is highly dependent on several factors. As described in the previous section, adequate variable selection is required to ensure the performance of the model. Additionally, statistic assumptions have to be met and tested first. Moreover, choosing the right model-building approach is essential, as is validating the results (102).

An alternative approach that is gaining popularity in the field is machine learning (ML). Multiple studies demonstrate ML's potential in disease prediction and claim to outperform traditional statistical modelling (103). Commonly used ML methods include Classification and Regression Trees (CART), Random Forest (RF), Support Vector Machines (SVM), and Artificial Neural Networks (ANN). ML offers several advantages over traditional biostatistical methods, including improved flexibility and scalability, making it suitable for various tasks, such as diagnoses and classification. Moreover, ML algorithms can be used for different types of data, such as demographic information, laboratory results, and free text notes from doctors, enabling accurate predictions for disease risk, diagnosis, prognosis, and treatment decisions (104). However, some studies have shown that the incremental predictive performance over traditional methods (such as regression) could be limited, while others show no advantage over ML (105–109).

Our final algorithm aims to predict the service that best suits the patient, a typical classification problem. Supervised ML is commonly used in classification problems. Supervised algorithms first perform analytical tasks to recognise patterns in the training data. In the training data, the outcome variables are known. The algorithm is then validated on a test set where the outcome variables are unknown. The algorithm predicts the outcomes for the test set or the target population. In healthcare, supervised learning techniques are increasingly adopted in clinical prediction. On the contrary, unsupervised machine learning is adopted less frequently. In contrast to supervised algorithms, unsupervised algorithms do not use labels or previous results. Their main goal is to find patterns and extract the hidden structure of data without using predefined labels. Unsupervised ML encompasses clustering techniques and dimensionality reduction and is employed for data analysis, stratification, reduction, and clustering of unclassified data. Since unsupervised ML is not used for prediction, it is not directly relevant to the development of our final algorithm.

However, unsupervised ML can be applied to identify patterns in the data in the EHR. Unsupervised ML cluster methods typically use algorithms like K-Means and Mean-shift to group unclassified data into independent clusters. In practice, these methods can be used to classify patients who are potentially eligible for a consult at the 'Gezondheidsplein' into clusters. In the future, when the digital platform and the algorithm are fully developed and implemented, unsupervised ML may also be used to cluster patients who have received health advice.

7.5 Artificial Intelligence

AI is an umbrella term in which various techniques, such as ML and Natural Language Processing (NLP), are combined to simulate human behaviours such as learning, judgment, and decision-making (110). Over the past decade, there has been increased interest in applying AI in medicine. AI is increasingly being applied in clinical practice. For example, AI can help optimize healthcare processes by supporting clinical documentation and decision-making, image analysis, medical device automation, and patient monitoring (111). The advancement in AI has the potential to support healthcare personnel in reducing processing times and enhancing the quality of patient care in clinical practice. It is conceivable to think if AI doctors will replace other healthcare workers. Multiple studies show that AI is capable of equalling or outperforming physicians' performance. Yet it is expected that AI will not replace physicians. However, it may be that physicians who do not use AI will be replaced by those who don't (112). A study conducted in 8 German university hospitals on physicians' expectations regarding the application of AI revealed that physicians foresee the future of medicine as a blend of human expertise and AI (113). This is in line with the observation of our study in which lifestyle coaches indicate that the human aspect remains necessary in giving advice. According to them, it is difficult to capture the observations made during the consultation in a model. Expectations are that AI will not replace physicians on a large scale, but rather enhance their efforts to care for patients. Possibly, the field of focus will shift, and physicians will concentrate more on actions that demand human skills such as empathy, persuasion, and perspective-taking ability (114). Shakeri argues that adopting AI in clinical practices will provide more flexibility and allow healthcare professionals to focus on patient interaction (115).

From the patient's perspective, AI will perhaps impact the patient-doctor relationship. AI has the potential to stimulate shared decision-making (SDM), a result of increased patient autonomy (116). On the other hand, the implementation of AI could potentially lead to a new form of paternalism, wherein the algorithm determines what is best for the patient (117). In terms of SDM, patients' values and preferences should be taken into account, not just clinical data. Every patient is unique and therefore the most effective solution for one patient might not be the most suitable for another patient. In the context of SDM, it is crucial to consider the patient's values, preferences, and goals during the development of the model (118). Currently, there is limited research examining the expectations of patients regarding the application of AI. Richardson et al (119) conducted a focus-group study on patient perspectives about the use of AI in healthcare. Participants were positive towards the application of AI tools in healthcare but emphasized that the tool must undergo robust testing and ensure accuracy. In addition, participants wanted the option to choose AI involvement in their care and preferred not to solely rely on AI decision-making without understanding the rationale. Furthermore, they expect supervision and regulatory protection against potential harm.

7.6 Strengths

Our study had several strengths. Firstly, this explanatory study is the pioneering effort to commence data collection to develop a platform that provides personalized advice to patients regarding lifestyle interventions for secondary prevention. The patient inclusion and data collection methods have been analysed and evaluated. The valuable insights gained, are significant for future studies as they enable researchers to apply more efficient inclusion and data collection methods. Optimising the inclusion process not only leads to a larger sample size but also allows for better utilisation of the capacity of the 'Gezondheidsplein'. Consequently, more patients may benefit from the expertise of lifestyle coaches and receive personalised advice to improve their health.

Secondly, the initial exploration of the data in the pilot study has yielded insights into data quality and the applicability of various statistical analysis techniques. The suggestions based on the primary dataset can be taken into account to ensure enhanced data quality in the future. Furthermore, the initial statistical analyses indicate that logistic regression can be employed to examine the relationships between personal characteristics and health advice, as coefficients serve as good indicators of variable importance. The advantages of logistic regression lie in its applicability and relatively simple implementation. Moreover, the classification algorithm is easy to train and can effectively classify unknown outcomes.

7.7 Limitations

As with most studies, the design of the current study and the results of the pilot study are subject to limitations. The design of the initial study may have inherent limitations, such as the choice of data collection methods, sample size, and statistical methods employed. These factors may affect the generalizability of the results to a broader population and the reliability and validity of the findings.

Due to the limited sample size and skewness of the data, the results of the pilot study are not reliable. Consequently, it is difficult to determine the relationship between personal characteristics and health advice solely based on the pilot study. As a result, determining the appropriate input variables for the final algorithm used in the platform becomes challenging. Additionally, it is difficult to conclude about the appropriateness of the initial study design based on these limited results.

In addition, the revised study builds upon the initial study design, with proposed optimisations in processes, modification of the variable selection method and model evaluation. However, no new statistical methods were introduced in the revised design. Despite the advantages of using logistic regression, it is essential to acknowledge its disadvantages. One significant drawback is that it requires meeting several statistical assumptions. For instance, there must be linearity between the predictors and outcome variables. As the pilot study demonstrated, this assumption may not always hold in real-world scenarios. Variables that do not satisfy this assumption cannot be included in the analysis, even though they might potentially enhance the model's performance and predictive capability.

Furthermore, the data quality was not optimal. The dataset is characterised by many missing data. Due to the substantial amount of missing data, imputing missing values was not reliable. It appeared that the selected variables did not match the patient cohort. This could have been prevented by involving lifestyle coaches in the selection of candidate predictors. In the current design, candidate predictors were identified solely by consulting existing literature. However, there was limited availability of studies examining the relationship between patient characteristics and the recommendation of a secondary prevention intervention. Due to the lack of prior clarity regarding the specific reason for referral, the decision was made to select candidate predictors for the four most prevalent non-communicable diseases. The assumption was made that the patient’s condition was associated with the provided health advice. Therefore, risk factors of four major NCDs are selected as candidate predictors in predicting health advice. In the practical context, these conditions were not generally considered grounds for patient referral in most of the cases. Notably, smoking cessation, dietary counselling, and weight control emerged as the most frequently recommended lifestyle changes.

Moreover, data quality was compromised using CTcue for data extraction from the EHR. In CTcue, creating a specific and efficient query is crucial for effective data extraction. If the query lacks specificity, it may result in incomplete data retrieval or potentially overlooking relevant data within the EHR. Additionally, the extraction software does not indicate whether the data is non-existent or inaccessible due to reasons such as authorisation or privacy issues. Therefore, it is difficult to determine the reasons for missing data.

In light of these limitations, it is important to interpret the study's results cautiously and consider the potential implications of using the results as input for the final algorithm of the digital platform. Future research should address these limitations. The revised study design provides a foundation for optimizing patient inclusion, improving data quality, and selecting appropriate candidate predictors. Further work is required to better understand the relationship between personal characteristics and service domains. It may be of added value to consider other algorithms to determine the variables to be included in the algorithm. Therefore, future studies should focus on comparing other techniques for selecting the input variables for the final algorithms. In addition, future investigations should focus on which algorithm performs best in predicting the advice for care or social services in terms of secondary prevention.

References

1. Ministerie van Volksgezondheid, Welzijn en Sport. Integraal Zorg Akkoord - Samen werken aan gezonde zorg. 2022 p. 117.

2. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: Analyzing the current state-of-research. J Bus Res. 2021 Feb;123:557–67.

3. European Commission. Public Health. [cited 2023 Aug 17]. Overview Non-communicable diseases. Available from: https://health.ec.europa.eu/non-communicable-diseases/overview_nl#:~:text=Niet%2Doverdraagbare%20ziekten%20zoals%20hart,oorzaken%20van%20vermijdbare%20vroegtijdige%20sterfte.

4. Schultz WM, Kelli HM, Lisko JC, Varghese T, Shen J, Sandesara P, et al. Socioeconomic Status and Cardiovascular Outcomes: Challenges and Interventions. Circulation. 2018 May 15;137(20):2166–78.

5. Larsson K, Lisspers K, Ställberg B, Johansson G, Gutzwiller FS, Mezzi K, et al. Treatment Patterns, Socioeconomic Status and Clinical Burden in Mild COPD: A Swedish Real-World, Retrospective Cohort Study, the ARCTIC Study. Int J Chron Obstruct Pulmon Dis. 2022 Jun;Volume 17:1409–21.

6. Pizzato M, Martinsen JI, Heikkinen S, Vignat J, Lynge E, Sparén P, et al. Socioeconomic status and risk of lung cancer by histological subtype in the Nordic countries. Cancer Med. 2022 Apr;11(8):1850–9.

7. Stojanović M, Cvetanović G, Anđelković Apostolović M, Stojanović D, Rančić N. Impact of socio-demographic characteristics and long-term complications on quality of life in patients with diabetes mellitus. Cent Eur J Public Health. 2018 Jun 30;26(2):104–10.

8. Murtagh S, McCombe G, Broughan J, Carroll Á, Casey M, Harrold Á, et al. Integrating Primary and Secondary Care to Enhance Chronic Disease Management: A Scoping Review. Int J Integr Care. 2021 Feb 9;21(1):4.

9. Ekwueme DU, Halpern MT, Chesson HW, Ashok M, Drope J, Hong YR, et al. Health Economics Research in Primary Prevention of Cancer: Assessment, Current Challenges, and Future Directions. JNCI Monogr. 2022 Jul 5;2022(59):28–41.

10. Van De Glind IM, Heinen MM, Geense WW, Mesters I, Wensing M, Van Achterberg T. Exploring the Range of Lifestyle Interventions Used in Dutch Health Care Practice: A Qualitative Description. Health Promot Pract. 2016 Mar;17(2):235–43.

11. Aggarwal M, Ornish D, Josephson R, Brown TM, Ostfeld RJ, Gordon N, et al. Closing Gaps in Lifestyle Adherence for Secondary Prevention of Coronary Heart Disease. Am J Cardiol. 2021 Apr;145:1–11.

12. Rees K, Takeda A, Martin N, Ellis L, Wijesekara D, Vepa A, et al. Mediterranean-style diet for the primary and secondary prevention of cardiovascular disease. Cochrane Heart Group, editor. Cochrane Database Syst Rev [Internet]. 2019 Mar 13 [cited 2023 Jul 16];2019(3). Available from: http://doi.wiley.com/10.1002/14651858.CD009825.pub3

13. Loftus TJ, Filiberto AC, Li Y, Balch J, Cook AC, Tighe PJ, et al. Decision analysis and reinforcement learning in surgical decision-making. Surgery. 2020 Aug;168(2):253–66.

14. Wang C, Redgrave J, Shafizadeh M, Majid A, Kilner K, Ali AN. Aerobic exercise interventions reduce blood pressure in patients after stroke or transient ischaemic attack: a systematic review and meta-analysis. Br J Sports Med. 2019 Dec;53(24):1515–25.

15. Verheggen E. Het sociale domein decentraal en integraal, maar wat betekent dat in de praktijk? Tijdschr Voor Gezondheidswetenschappen. 2018 Dec;96(8):334–6.

16. Vos H. Huisarts en complexe zorgvraag: Ervaring met integraal samenwerken tussen de eerste lijn, de gemeente en welzijn. Tijdschr Voor Gezondheidswetenschappen. 2019 Feb;97(1–2):20–2.

17. Valentijn PP, Schepman SM, Opheij W, Bruijnzeels MA. Understanding integrated care: a comprehensive conceptual framework based on the integrative functions of primary care. Int J Integr Care [Internet]. 2013 Mar 22 [cited 2023 Jul 16];13(1). Available from: http://www.ijic.org/article/10.5334/ijic.886/

18. Pescheny JV, Pappas Y, Randhawa G. Facilitators and barriers of implementing and delivering social prescribing services: a systematic review. BMC Health Serv Res. 2018 Dec;18(1):86.

19. Van Vliet M, Doornenbal BM, Boerema S, Van Den Akker-van Marle EM. Development and psychometric evaluation of a Positive Health measurement scale: a factor analysis study based on a Dutch population. BMJ Open. 2021 Feb;11(2):e040816.

20. Huber M, Knottnerus JA, Green L, Horst HVD, Jadad AR, Kromhout D, et al. How should we define health? BMJ. 2011 Jul 26;343(jul26 2):d4163–d4163.

21. Huber M, Van Vliet M, Giezenberg M, Winkens B, Heerkens Y, Dagnelie PC, et al. Towards a 'patient-centred' operationalisation of the new dynamic concept of health: a mixed methods study. BMJ Open. 2016 Jan;6(1):e010091.

22. Chen, Chiang, Storey. Business Intelligence and Analytics: From Big Data to Big Impact. MIS Q. 2012;36(4):1165.

23. Galetsi P, Katsaliaki K, Kumar S. Values, challenges and future directions of big data analytics in healthcare: A systematic review. Soc Sci Med. 2019 Nov;241:112533.

24. Hassan N, Slight R, Weiand D, Vellinga A, Morgan G, Aboushareb F, et al. Preventing sepsis; how can artificial intelligence inform the clinical decision-making process? A systematic review. Int J Med Inf. 2021 Jun;150:104457.

25. Safaei M, Sundararajan EA, Driss M, Boulila W, Shapi'i A. A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. Comput Biol Med. 2021 Sep;136:104754.

26. Saleh Ibrahim Y, Khalid Al-Azzawi W, Hamad Mohamad AA, Nouri Hassan A, Meraf Z. Perception of the Impact of Artificial Intelligence in the Decision-Making Processes of Public Healthcare Professionals. Pandian S, editor. J Environ Public Health. 2022 Jul 13;2022:1–8.

27. Whicher D, Rapp T. The Value of Artificial Intelligence for Healthcare Decision Making—Lessons Learned. Value Health. 2022 Mar;25(3):328–30.

28. Demiris G, Iribarren SJ, Sward K, Lee S, Yang R. Patient generated health data use in clinical practice: A systematic review. Nurs Outlook. 2019 Jul;67(4):311–30.

29. World Health Organisation. Constitution. 2023. WHO Constitution. Available from: https://www.who.int/about/governance/constitution

30. Saracci R. The World Health Organisation needs to reconsider its definition of health. BMJ. 1997 May 10;314(7091):1409–10.

31. Bircher J. Towards a Dynamic Definition of Health and Disease. Med Health Care Philos. 2005 Nov;8(3):335–41.

32. Health Council of the Netherlands. European primary care. The Hague; 2002 p. 120. Report No.: 2004/20E.

33. World Health Organisation. Social determinants of health. [cited 1994 Jul 16]. Social determinants of health. Available from: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

34. Fitzpatrick T, Rosella LC, Calzavara A, Petch J, Pinto AD, Manson H, et al. Looking Beyond Income and Education. Am J Prev Med. 2015 Aug;49(2):161–71.

35. Rikard SM, Kim B, Michel JD, Peirce SM, Barnes LE, Williams MD. Identifying individual social risk factors using unstructured data in electronic health records and their relationship with adverse clinical outcomes. SSM - Popul Health. 2022 Sep;19:101210.

36. Walker RJ, Smalls BL, Campbell JA, Strom Williams JL, Egede LE. Impact of social determinants of health on outcomes for type 2 diabetes: a systematic review. Endocrine. 2014 Sep;47(1):29–48.

37. Cockerham WC, Hamby BW, Oates GR. The Social Determinants of Chronic Disease. Am J Prev Med. 2017 Jan;52(1):S5–12.

38. Gottlieb L, Sandel M, Adler NE. Collecting and Applying Data on Social Determinants of Health in Health Care Settings. JAMA Intern Med. 2013 Jun 10;173(11):1017.

39. Vest JR, Grannis SJ, Haut DP, Halverson PK, Menachemi N. Using structured and unstructured data to identify patients' need for services that address the social determinants of health. Int J Med Inf. 2017 Nov;107:101–6.

40. Baxter S, Johnson M, Chambers D, Sutton A, Goyder E, Booth A. The effects of integrated care: a systematic review of UK and international evidence. BMC Health Serv Res. 2018 Dec;18(1):350.

41. Valentijn PP, Boesveld IC, Van Der Klauw DM, Ruwaard D, Struijs JN, Molema JJW, et al. Towards a taxonomy for integrated care: a mixed-methods study. Int J Integr Care [Internet]. 2015 Mar 4 [cited 2023 Jul 16];15(1). Available from: http://www.ijic.org/article/10.5334/ijic.1513/

42. Dambha-Miller H, Simpson G, Hobson L, Roderick P, Little P, Everitt H, et al. Integrated primary care and social services for older adults with multimorbidity in England: a scoping review. BMC Geriatr. 2021 Dec;21(1):674.

43. Goodwin N. Understanding Integrated Care. Int J Integr Care. 2016 Oct 28;16(4):6.

44. Auschra C. Barriers to the Integration of Care in Inter-Organisational Settings: A Literature Review. Int J Integr Care. 2018 Jan 16;18(1):5.

45. Maruthappu M, Hasan A, Zeltner T. Enablers and Barriers in Implementing Integrated Care. Health Syst Reform. 2015 May 19;1(4):250–6.

46. Kozlowska O, Lumb A, Tan GD, Rea R. Barriers and facilitators to integrating primary and specialist healthcare in the United Kingdom: a narrative literature review. Future Healthc J. 5(1):64–80.

47. Booth AM, Wilson P. Evidence to inform the commissioning of social prescribing. 2015.

48. National Health Service. Social Prescribing. [cited 2023 Jul 16]. Social prescribing. Available from: https://www.england.nhs.uk/personalisedcare/social-prescribing/

49. Morse DF, Sandhu S, Mulligan K, Tierney S, Polley M, Chiva Giurca B, et al. Global developments in social prescribing. BMJ Glob Health. 2022 May;7(5):e008524.

50. Rempel ES, Wilson EN, Durrant H, Barnett J. Preparing the prescription: a review of the aim and measurement of social referral programmes. BMJ Open. 2017 Oct;7(10):e017734.

51. Vidovic D, Reinhardt GY, Hammerton C. Can Social Prescribing Foster Individual and Community Well-Being? A Systematic Review of the Evidence. Int J Environ Res Public Health. 2021 May 15;18(10):5276.

52. Barr JA, Tsai LP. Health coaching provided by registered nurses described: a systematic review and narrative synthesis. BMC Nurs. 2021 Dec;20(1):74.

53. Olsen JM, Nesbitt BJ. Health Coaching to Improve Healthy Lifestyle Behaviors: An Integrative Review. Am J Health Promot. 2010 Sep;25(1):e1–12.

54. Long H, Howells K, Peters S, Blakemore A. Does health coaching improve health-related quality of life and reduce hospital admissions in people with chronic obstructive pulmonary disease? A systematic review and meta-analysis. Br J Health Psychol. 2019 Apr 29;bjhp.12366.

55. Racey M, Jovkovic M, Alliston P, Ali MU, Sherifali D. Diabetes health coach in individuals with type 2 diabetes: A systematic review and meta analysis of quadruple aim outcomes. Front Endocrinol. 2022 Dec 16;13:1069401.

56. Chi CL, Nick Street W, Robinson JG, Crawford MA. Individualized patient-centered lifestyle recommendations: An expert system for communicating patient specific cardiovascular risk information and prioritizing lifestyle options. J Biomed Inform. 2012 Dec;45(6):1164–74.

57. Nam Y, Kim Y. Individualized Exercise and Diet Recommendations: An Expert System for Monitoring Physical Activity and Lifestyle Interventions in Obesity. J Electr Eng Technol. 2015 Nov 1;10(6):2434–41.

58. Patel S, Craigen G, Pinto Da Costa M, Inkster B. Correction: Opportunities and Challenges for Digital Social Prescribing in Mental Health: Questionnaire Study. J Med Internet Res. 2021 Mar 29;23(3):e29041.

59. JOY. About. [cited 2023 Jul 16]. Joy App. Available from: https://www.thejoyapp.com/about.

60. Access. Social Prescribing software. [cited 2023 Jul 16]. Social prescribing software. Available from: https://www.theaccessgroup.com/en-gb/health-social-care/software/social-prescribing/

61. Budreviciute A, Damiati S, Sabir DK, Onder K, Schuller-Goetzburg P, Plakys G, et al. Management and Prevention Strategies for Non-communicable Diseases (NCDs) and Their Risk Factors. Front Public Health. 2020 Nov 26;8:574111.

62. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak. 2019 Dec;19(1):211.

63. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study. Circulation. 2008 Feb 12;117(6):743–53.

64. Zafari H, Langlois S, Zulkernine F, Kosowan L, Singer A. AI in predicting COPD in the Canadian population. Biosystems. 2022 Jan;211:104585.

65. Ma X, Wu Y, Zhang L, Yuan W, Yan L, Fan S, et al. Comparison and development of machine learning tools for the prediction of chronic obstructive pulmonary disease in the Chinese population. J Transl Med. 2020 Dec;18(1):146.

66. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of Chronic Obstructive Pulmonary Disease (COPD) in Asthma Patients Using Electronic Medical Records. J Am Med Inform Assoc. 2009 May 1;16(3):371–9.

67. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. J Am Coll Cardiol. 2014 Jul;63(25):2935–59.

68. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ. 2007 Jul 21;335(7611):136.

69. Spathis D, Vlamos P. Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. Health Informatics J. 2019 Sep;25(3):811–27.

70. Meijers WC, De Boer RA. Common risk factors for heart failure and cancer. Cardiovasc Res. 2019 Apr 15;115(5):844–53.

71. Jamshidi E, Asgary A, Setareh S, Casutt A, Gonzalez M, Bianchi MP, et al. Medical and Personal Characteristics Can Predict the Risk of Lung Metastasis. Clin Oncol. 2023 Jun;35(6):e362–75.

72. Conroy R. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J. 2003 Jun;24(11):987–1003.

73. Mikkelsen B, Williams J, Rakovac I, Wickramasinghe K, Hennis A, Shin HR, et al. Life course approach to prevention and control of non-communicable diseases. BMJ. 2019 Jan 28;l257.

74. Tran KB, Lang JJ, Compton K, Xu R, Acheson AR, Henrikson HJ, et al. The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2022 Aug;400(10352):563–91.

75. Rijnstate. Rijnstate jaarverslag 2022 [Internet]. 2023 [cited 2023 Aug 17]. Available from: https://www.rijnstate.nl/over-rijnstate/nieuws/2023/rijnstate-presenteert-jaarverslag-2022/

76. Communnity of care. Community of care Midden-Gelderland [Internet]. [cited 2023 Aug 17]. Available from: https://communityofcare.nl

77. dRural. Proposal dRural. 2021.

78. Iriarte-Roteta A, Lopez-Dicastillo O, Mujika A, Ruiz-Zaldibar C, Hernantes N, Bermejo-Martins E, et al. Nurses' role in health promotion and prevention: A critical interpretive synthesis. J Clin Nurs. 2020 Nov;29(21–22):3937–49.

79. Joseph VR. Optimal ratio for data splitting. Stat Anal Data Min ASA Data Sci J. 2022 Aug;15(4):531–8.

80. Jones SR. An introduction to power and sample size estimation. Emerg Med J. 2003 Sep 1;20(5):453–8.

81. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996 Dec;49(12):1373–9.

82. Curto JD, Pinto JC. The corrected VIF (CVIF). J Appl Stat. 2011 Jul;38(7):1499–507.

83. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating [Internet]. Cham: Springer International Publishing; 2019 [cited 2023 Aug 4]. (Statistics for Biology and Health). Available from: http://link.springer.com/10.1007/978-3-030-16399-0

84. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Community Health. 2020 Feb;8(1):e000262.

85. Kemppainen V, Tossavainen K, Turunen H. Nurses' roles in health promotion practice: an integrative review. Health Promot Int. 2013 Dec;28(4):490–501.

86. Jallinoja P, Absetz P, Kuronen R, Nissinen A, Talja M, Uutela A, et al. The dilemma of patient responsibility for lifestyle change: Perceptions among primary care physicians and nurses. Scand J Prim Health Care. 2007 Jan;25(4):244–9.

87. Geense WW, Van De Glind IM, Visscher TL, Van Achterberg T. Barriers, facilitators and attitudes influencing health promotion activities in general practice: an explorative pilot study. BMC Fam Pract. 2013 Dec;14(1):20.

88. Woodward M, De Pennington N, Grandidge C, McCulloch P, Morgan L. Development and evaluation of an electronic hospital referral system: a human factors approach. Ergonomics. 2020 Jun 2;63(6):710–23.

89. Fanelli SM, Kelly OJ, Krok-Schoen JL, Taylor CA. Low Protein Intakes and Poor Diet Quality Associate with Functional Limitations in US Adults with Diabetes: A 2005–2016 NHANES Analysis. Nutrients. 2021 Jul 27;13(8):2582.

90. Godos J, Grosso G, Castellano S, Galvano F, Caraci F, Ferri R. Association between diet and sleep quality: A systematic review. Sleep Med Rev. 2021 Jun;57:101430.

91. Gomez-Delgado F, Katsiki N, Lopez-Miranda J, Perez-Martinez P. Dietary habits, lipoprotein metabolism and cardiovascular disease: From individual foods to dietary patterns. Crit Rev Food Sci Nutr. 2021 May 31;61(10):1651–69.

92. Jayedi A, Soltani S, Abdolshahi A, Shab-Bidar S. Healthy and unhealthy dietary patterns and the risk of chronic disease: an umbrella review of meta-analyses of prospective cohort studies. Br J Nutr. 2020 Dec 14;124(11):1133–44.

93. Van Iersel LEJ, Beijers RJHCG, Gosker HR, Schols AMWJ. Nutrition as a modifiable factor in the onset and progression of pulmonary function impairment in COPD: a systematic review. Nutr Rev. 2022 May 9;80(6):1434–44.

94. Richter AN, Khoshgoftaar TM. A review of statistical and machine learning methods for modeling cancer risk using structured clinical data. Artif Intell Med. 2018 Aug;90:1–14.

95. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. J Thorac Dis. 2019 Mar;11(S4):S574–84.

96. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012 May 1;98(9):683–90.

97. Hafermann L, Becher H, Herrmann C, Klein N, Heinze G, Rauch G. Statistical model building: Background “knowledge” based on inappropriate preselection causes misspecification. BMC Med Res Methodol. 2021 Dec;21(1):196.

98. Talbot D, Massamba VK. A descriptive review of variable selection methods in four epidemiologic journals: there is still room for improvement. Eur J Epidemiol. 2019 Aug;34(8):725–30.

99. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009 Mar 31;338(mar31 1):b604–b604.

100. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol. 1996 Aug;49(8):907–16.

101. VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol. 2019 Mar 15;34(3):211–9.

102. Stoltzfus JC. Logistic Regression: A Brief Primer: LOGISTIC REGRESSION: A BRIEF PRIMER. Acad Emerg Med. 2011 Oct;18(10):1099–104.

103. Domínguez-Rodríguez S, Serna-Pascual M, Oletto A, Barnabas S, Zuidewind P, Dobbels E, et al. Machine learning outperformed logistic regression classification even with limit sample size: A model to predict pediatric HIV mortality and clinical progression to AIDS. Yaseen ZM, editor. PLOS ONE. 2022 Oct 14;17(10):e0276116.

104. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 2019 May;20(5):e262–73.

105. Nusinovici S, Tham YC, Chak Yan MY, Wei Ting DS, Li J, Sabanayagam C, et al. Logistic regression was as good as machine learning for predicting major chronic diseases. J Clin Epidemiol. 2020 Jun;122:56–69.

106. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019 Jun;110:12–22.

107. Song X, Liu X, Liu F, Wang C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. Int J Med Inf. 2021 Jul;151:104484.

108. Lynam AL, Dennis JM, Owen KR, Oram RA, Jones AG, Shields BM, et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. Diagn Progn Res. 2020 Dec;4(1):6.

109. Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a Machine Learning Model Applied to Harmonized Electronic Health Record Data for the Prediction of Incident Atrial Fibrillation. JAMA Netw Open. 2020 Jan 17;3(1):e1919396.

110. Zhang C, Lu Y. Study on artificial intelligence: The state of the art and future prospects. J Ind Inf Integr. 2021 Sep;23:100224.

111. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare [Internet]. Elsevier; 2020 [cited 2023 Aug 4]. p. 25–60. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128184387000022

112. Martinho A, Kroesen M, Chorus C. A healthy debate: Exploring the views of medical doctors on the ethics of artificial intelligence. Artif Intell Med. 2021 Nov;121:102190.

113. Maassen O, Fritsch S, Palm J, Deffge S, Kunze J, Marx G, et al. Future Medical Artificial Intelligence Application Requirements and Expectations of Physicians in German University Hospitals: Web-Based Survey. J Med Internet Res. 2021 Mar 5;23(3):e26646.

114. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. 2019;6(2):94–8.

115. Aminololama-Shakeri S, López JE. The Doctor-Patient Relationship With Artificial Intelligence. Am J Roentgenol. 2019 Feb;212(2):308–10.

116. Sauerbrei A, Kerasidou A, Lucivero F, Hallowell N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. BMC Med Inform Decis Mak. 2023 Apr 20;23(1):73.

117. Lorenzini G, Arbelaez Ossa L, Shaw DM, Elger BS. Artificial intelligence and the doctor–patient relationship expanding the paradigm of shared decision making. Bioethics. 2023 Jun;37(5):424–9.

118. McDougall J. Computer knows best? The need for value-flexibility in medical AI. J Med Ethics. 2018 Nov 22;45(3):156–60.

119. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. Npj Digit Med. 2021 Sep 21;4(1):140.

Appendix 1: Positive health questionnaire (PH_Q)

Rijnstate

Mijn positieve gezondheid

De vragenlijst geeft je inzicht in jouw ervaren gezondheid op de volgende zes dimensies:
- Hoe het lichamelijk met je gaat
- Hoe het mentaal met je gaat
- Of je jouw leven zinvol vindt
- Of je een prettig leven hebt
- Of je contact hebt met andere mensen
- Hoe je dagelijks leven eruit ziet

Uw antwoorden zijn zichtbaar voor de medewerkers van het gezondheidsplein in Rijnstate.

Sectie 1 ...

Laat ons eerst weten wie je bent

1

Wat is je naam (voor en achternaam) *

Voer uw antwoord in

2

Wat is je geboortedatum? *

Geef de datum op (d-M-yyyy)

3

Je vult deze vragenlijst in omdat je bent doorverwezen naar het gezondheidsplein. Heb je specifieke vragen of onderwerpen die je daar graag zou willen bespreken?

Voer uw antwoord in

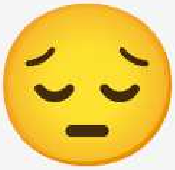
Sectie 2 ...

Mijn positieve gezondheid - Lichaam

De eerste vragen hebben betrekking op het lichaam

4

Voel jij je gezond? *



◯ Nee

◯ Een beetje

◯ Ja

5

Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

6

Voel jij je fit? *

- ◯ Nee

- ◯ Een beetje

- ◯ Ja

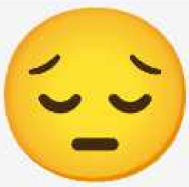
7

Wil je over dit antwoord praten in het gesprek? *

- ◯ Ja
- ◯ Nee

8

Kan je goed bewegen? *



- ◯ Nee

- ◯ Een beetje

- ◯ Ja

9

Wil je over dit antwoord praten in het gesprek? *

- ◯ Ja
- ◯ Nee

10

Welk cijfer geef je aan je lichaam *

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

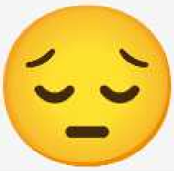
Sectie 3

Mijn positieve gezondheid - gevoel en gedachten

De komende vragen hebben betrekking op je gevoel en gedachten

11

Kan je dingen goed onthouden? *



◯ Nee

◯ Een beetje

◯ Ja

12

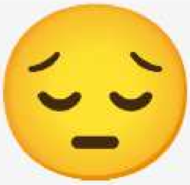
Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

13

Kan je goed nadenken? *



◯ Nee

◯ Een beetje

◯ Ja

14

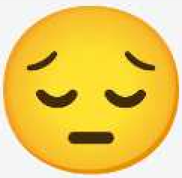
Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

15

Weet je wat je moet doen als het niet goed gaat? *



◯ Nee

◯ Een beetje

◯ Ja

16

Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

17

Welk cijfer geef je je gevoel en gedachten? *

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

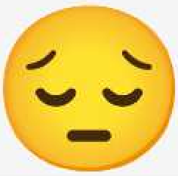
Sectie 4

Mijn positieve gezondheid - zinvol leven

De volgende vragen gaan over zinvol leven

18

Zijn er dingen die je graag wil doen in het leven? *



◯ Nee

◯ Een beetje

◯ Ja

19

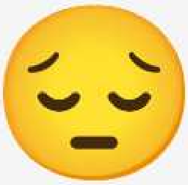
Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

20

Maak jij je zorgen over je toekomst? *



◯ Nee

◯ Een beetje

◯ Ja

21

Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

22

Vind je het leven zinvol? Welk cijfer geef je? *

◯ 1

◯ 2

◯ 3

◯ 4

◯ 5

◯ 6

◯ 7

◯ 8

◯ 9

◯ 10

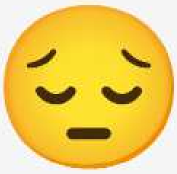
Sectie 5 ···

Mijn positieve gezondheid - kwaliteit van leven

Onderstaande vragen gaan over de kwaliteit van leven

23

Ben je gelukkig? *



◯ Nee

◯ Een beetje

◯ Ja

24

Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

25

Voel jij je goed? *

- Nee
- Een beetje
- Ja

26

Wil je over dit antwoord praten in het gesprek? *

- Ja
- Nee

27

Kan jij je leven aan? *

- Nee
- Een beetje
- Ja

28

Wil je over dit antwoord praten in het gesprek? *

- Ja
- Nee

29

Welk cijfer geef je jouw kwaliteit van leven? *

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

Sectie 6 ...

Mijn positieve gezondheid - meedoen

De onderstaande vragen gaan over meedoen

30

Heb je contact met andere mensen? *

◯ Nee

◯ Een beetje

◯ Ja

31

Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

32

Heb je mensen die je kunnen helpen? *

◯ Nee

◯ Een beetje

◯ Ja

33

Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

34

Heb je het gevoel dat je er erbij hoort? *

◯ Nee

◯ Een beetje

◯ Ja

35

Wil je over dit antwoord praten in het gesprek? *

◯ Ja

◯ Nee

36

Welk cijfer geef je aan meedoen? *

- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10

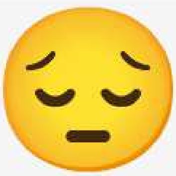
Sectie 7

Mijn positieve gezondheid - dagelijks leven

Onderstaande vragen hebben betrekking op het dagelijks leven

37

Weet je wat je kan en niet kan? *



◯ Nee

◯ Een beetje

◯ Ja

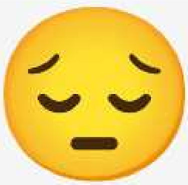
38

Wil je over dit antwoord praten in het gesprek? *

- ◯ Ja
- ◯ Nee

39

Weet je hoe je gezond kunt leven? *



◯ Nee

◯ Een beetje

◯ Ja

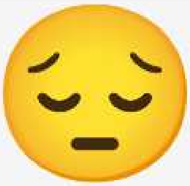
40

Wil je over dit antwoord praten in het gesprek? *

- ◯ Ja
- ◯ Nee

41

Kan jij jouw dag goed indelen? *



- ◯ Nee
- ◯ Een beetje
- ◯ Ja

42

Wil je over dit antwoord praten in het gesprek? *

- ◯ Ja
- ◯ Nee

43

Welk cijfer geef je aan het dagelijks leven? *

- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10

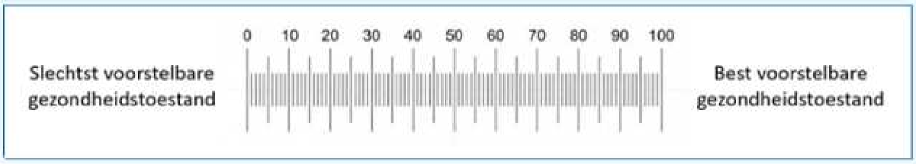
Gezondheidstoestand

Om mensen te helpen bij het aangeven hoe goed of hoe slecht een gezondheidstoestand is, hebben we een meetschaal (te vergelijken met een thermometer) gemaakt. Op de meetschaal hiernaast betekent "100" de beste gezondheidstoestand die u zich kunt voorstellen, en "0" de slechtste gezondheidstoestand die u zich kunt voorstellen.

We willen u vragen aan te geven hoe goed of hoe slecht volgens u uw eigen gezondheidstoestand vandaag is. Vul een getal in tussen 0 en 100 dat volgens u aangeeft hoe goed of hoe slecht uw gezondheidstoestand vandaag is.

44

Uw gezondheidstoestand vandaag: *



Aantal moet tussen 0 en 100 liggen

Sectie 9

Wensen

De volgende vragen gaan over jouw wensen om aan de slag gaan met je leefstijl. Selecteer het antwoord dat je het best vindt passen op dit moment. Er zijn meerdere antwoorden mogelijk.

45

Ik wil graag aan de slag met:

- [ ] Bewegen
- [ ] Voeding
- [ ] Slaap
- [ ] Ontspanning
- [ ] Hulp bij stoppen met roken
- [ ] Hulp bij alcoholgebruik
- [ ] Hulp bij geldzaken
- [ ] Hulp in huis
- [ ] Hulp bij sociale contacten
- [ ] Ik weet niet waarmee ik aan de slag wil
- [ ] Andere

46

Als ik een advies krijgen moeten ze rekening houden met:

- [ ] Geen dingen om rekening mee te houden
- [ ] Lichamelijke beperkingen
- [ ] Ziekte(n) die ik heb
- [ ] Hoeveel geld het mij gaat kosten
- [ ] Hoeveel tijd het mij gaat kosten
- [ ] Reisafstand
- [ ] Andere

47

Geeft u toestemming om uw gegevens anoniem te gebruiken voor de ontwikkeling van een nieuw platform? *

- ○ Ja
- ○ Nee

Sectie 10 ...

Mijn positieve gezondheid

Bedankt voor het invullen van de vragen!
Als u op verzenden klikt worden de resultaten opgeslagen en kunt u deze op het gezondheidsplein bespreken.

Appendix 2: Scoring table

Rijnstate

Positieve Gezondheid vragenlijst (Gezondheidsplein) (gedeelte voor leefstijlcoach)

Dit onderdeel wordt ingevuld door de leefstijlcoach.
De eerste vragen hebben betrekking op de gedeelde informatie voorafgaand aan het gesprek op het gezondheidsplein.
Dit zijn bijv. vragen over wat er al bekend was over deze patiënt met betrekking tot bijv. roken.
Het tweede deel heeft betrekking op het advies dat gegeven wordt. Dit gedeelte van de vragenlijst vul je dus in na of tijdens het gesprek. Middels categorieën en open velden kan aangegeven worden welke dienst of richting geadviseerd is op basis van het gesprek en de ingevulde vragenlijst gericht op positieve gezondheid.
Heel erg bedankt voor jullie tijd!
Hartelijke groet, Eva Smit en de dRural projectgroep: Jolien, Anne-Jet, Laura, Lonneke en Melanie

Sectie 1 ...

ACHTERGROND

Dit gedeelte omvat vragen om de informatie voorafgaand aan het gesprek in kaart te brengen

1. Naam medewerker gezondheidsplein

Voer uw antwoord in

2. Naam patiënt (voor en achternaam) *

Voer uw antwoord in

3. Geboortedatum patiënt *

Geef de datum op (d-M-yyyy)

4. Reden van de verwijzing of het verzoek aan het gezondheidsplein *

Voer uw antwoord in

5. Naam en afdeling van de verwijzing *

Voer uw antwoord in

6. Opleidingsniveau patiënt *

- ◯ Hoog (HBO/Universiteit)
- ◯ Midden (MBO/HAVO/VWO)
- ◯ Laag (basisonderwijs/VMBO/MAVO)

7. Werk patiënt *

Voer uw antwoord in

8. Heeft de patiënt toestemming verleend tot het delen van zijn gegevens voor terugkoppeling en/of doorverwijzing? *

- ◯ Ja
- ◯ Nee
- ◯ Onbekend

Sectie 2

ADVIES

Welk advies is gegeven naar aanleiding van het gesprek? Hieronder enkele domeinen met daarbij specifieke activiteiten. Probeer zo specifiek mogelijk aan te geven in welke richting het advies was.

9. In welke categorie heb je advies gegeven? (kies hier de belangrijkste, later is er een optie om nog op andere categorieën reactie te geven) *

Selecteer uw antwoord

10. Je hebt in de categorie bewegen een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Wandelen, zelf
- [ ] Wandelclub
- [ ] Fietsen
- [ ] Beweegoefeningen thuis
- [ ] Sport
- [ ] Personal trainer
- [ ] Fysiotherapie
- [ ] Gespecialiseerde fysiotherapie
- [ ] Gecombineerde Leefstijl Interventie
- [ ] Andere

11. Ruimte voor toelichting op het advies (welke informatie, type sport, type fysiotherapie etc.)

Voer uw antwoord in

12. Je hebt in de categorie mentaal welbevinden een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Dagopvang
- [ ] Huisarts
- [ ] Rouwtherapeut
- [ ] Psycholoog
- [ ] POH
- [ ] Psychiater
- [ ] Maatschappelijk werker
- [ ] Humanistisch raadsvrouw
- [ ] Bewegen
- [ ] Andere

13. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

14. Je hebt in de categorie voeding een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Samen eten
- [ ] Maaltijdvoorzieningen
- [ ] Diëtist
- [ ] Gespecialiseerd diëtist
- [ ] Gecombineerde Leefstijl Interventie
- [ ] Algemene informatie over gezond eten
- [ ] Voedingssupplementen
- [ ] Andere

15. Ruimte voor toelichting op het advies (welke informatie, type diëtist etc.)

Voer uw antwoord in

16. Je hebt in de categorie wonen / huishouden een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- ☐ Digitaal programma of informatie
- ☐ Begeleid wonen
- ☐ Wijkcoach
- ☐ Woning corporatie
- ☐ Thuiszorg
- ☐ Dag invulling
- ☐ Mantelzorg
- ☐ Andere

17. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

18. Je hebt in de categorie sociale omgeving / contacten een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Clubjes / verenigingen
- [ ] Vrijwilligerscentrale
- [ ] Samen eten
- [ ] Buurthuis
- [ ] Dag invulling
- [ ] Mantelzorg
- [ ] Andere

19. Ruimte voor toelichting op het advies (welke informatie, welke clubjes etc.)

Voer uw antwoord in

20. Je hebt in de categorie slapen / ontspanning een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Algemene informatie slaaphygiëne
- [ ] Ontspanningstechnieken
- [ ] Meditatie
- [ ] Lezen
- [ ] Muziektherapie
- [ ] Acupunctuur
- [ ] Huisarts
- [ ] Andere

21. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

22. Je hebt in de categorie persoonlijke verzorging een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Huisarts
- [ ] Thuiszorg
- [ ] Pedicure
- [ ] Ergotherapie
- [ ] Andere

23. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

24. Je hebt in de categorie gewoontes een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Verslavingszorg
- [ ] Coach
- [ ] Telefonisch consult
- [ ] Andere

25. Ruimte voor toelichting op het advies (welke informatie, welke gewoonte/verslaving etc.)

Voer uw antwoord in

26. Je hebt in de categorie financieel / werk een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Wijkcoach
- [ ] Hulp bij administratie
- [ ] Gelrepas
- [ ] Digitale ondersteuning
- [ ] Werk zoeken
- [ ] Budgetcoach
- [ ] Andere

27. Ruimte voor toelichting op het advies (welke informatie, coaching etc.)

Voer uw antwoord in

28. Je hebt in de categorie vervoer een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Vervoerpas
- [ ] Regiotaxi
- [ ] Gehandicaptenparkeerkaart
- [ ] Leerlingen vervoer
- [ ] Automaatje
- [ ] Andere

29. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

30. Je hebt in de categorie fysiek welbevinden een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Huisarts
- [ ] Apotheek
- [ ] Fysiotherapie
- [ ] Acupuntuur
- [ ] Andere

31. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

32. Je hebt in de categorie opvoeden een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Huiselijk geweld
- [ ] Consultatiebureau
- [ ] Kindercoaches
- [ ] Peuterspeelzalen
- [ ] Jongerenwerk
- [ ] Straatcoaches
- [ ] Kindertelefoon
- [ ] Onderwijs
- [ ] Zelfvertrouwen
- [ ] Andere

33. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

34. Op welke manier zou aan dit domein gewerkt moeten worden? *

Selecteer maximaal 2 opties.

- [ ] Zelf aan de slag
- [ ] Aan de slag met hulp van vrienden / familie
- [ ] Aan de slag met professionele hulp

35. In welke categorie heb je daarnaast nog advies gegeven? *

Indien meerdere categorieën: vul hier de categorie in die hierna het meest belangrijk is

Selecteer uw antwoord

36. Je hebt in de categorie bewegen een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Wandelen, zelf
- [ ] Wandelclub
- [ ] Fietsen
- [ ] Beweegoefeningen thuis
- [ ] Sport
- [ ] Personal trainer
- [ ] Fysiotherapie
- [ ] Gespecialiseerde fysiotherapie
- [ ] Gecombineerde Leefstijl Interventie
- [ ] Andere

37. Ruimte voor toelichting op het advies (welke informatie, type sport, type fysiotherapie etc.)

Voer uw antwoord in

38. Je hebt in de categorie mentaal welbevinden een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- ☐ Digitaal programma of informatie
- ☐ Dagopvang
- ☐ Huisarts
- ☐ Rouwtherapeut
- ☐ Psycholoog
- ☐ POH
- ☐ Psychiater
- ☐ Maatschappelijk werker
- ☐ Humanistisch raadsvrouw
- ☐ Bewegen
- ☐ Andere

39. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

40. Je hebt in de categorie voeding een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Samen eten
- [ ] Maaltijdvoorzieningen
- [ ] Diëtist
- [ ] Gespecialiseerd diëtist
- [ ] Gecombineerde Leefstijl Interventie
- [ ] Algemene informatie over gezond eten
- [ ] Voedingssupplementen
- [ ] Andere

41. Ruimte voor toelichting op het advies (welke informatie, type diëtist etc.)

Voer uw antwoord in

42. Je hebt in de categorie wonen / huishouden een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Begeleid wonen
- [ ] Wijkcoach
- [ ] Woning corporatie
- [ ] Thuiszorg
- [ ] Dag invulling
- [ ] Mantelzorg
- [ ] Andere

43. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

44. Je hebt in de categorie sociale omgeving / contacten een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Clubjes / verenigingen
- [ ] Vrijwilligerscentrale
- [ ] Samen eten
- [ ] Buurthuis
- [ ] Dag invulling
- [ ] Mantelzorg
- [ ] Andere

45. Ruimte voor toelichting op het advies (welke informatie, welke clubjes etc.)

Voer uw antwoord in

46. Je hebt in de categorie slapen / ontspanning een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- ☐ Digitaal programma of informatie
- ☐ Algemene informatie slaaphygiëne
- ☐ Ontspanningstechnieken
- ☐ Meditatie
- ☐ Lezen
- ☐ Muziektherapie
- ☐ Acupunctuur
- ☐ Huisarts
- ☐ Andere

47. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

48. Je hebt in de categorie persoonlijke verzorging een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Huisarts
- [ ] Thuiszorg
- [ ] Pedicure
- [ ] Ergotherapie
- [ ] Andere

49. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

50. Je hebt in de categorie gewoontes een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Verslavingszorg
- [ ] Coach
- [ ] Telefonisch consult
- [ ] Andere

51. Ruimte voor toelichting op het advies (welke informatie, welke gewoonte/verslaving etc.)

Voer uw antwoord in

52. Je hebt in de categorie financieel / werk een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Wijkcoach
- [ ] Hulp bij administratie
- [ ] Gelrepas
- [ ] Digitale ondersteuning
- [ ] Werk zoeken
- [ ] Budgetcoach
- [ ] Andere

53. Ruimte voor toelichting op het advies (welke informatie, coaching etc.)

Voer uw antwoord in

54. Je hebt in de categorie vervoer een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Vervoerpas
- [ ] Regiotaxi
- [ ] Gehandicaptenparkeerkaart
- [ ] Leerlingen vervoer
- [ ] Automaatje
- [ ] Andere

55. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

56. Je hebt in de categorie fysiek welbevinden een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Huisarts
- [ ] Apotheek
- [ ] Fysiotherapie
- [ ] Acupuntuur
- [ ] Andere

57. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

58. Je hebt in de categorie opvoeden een advies gegeven. Wat heb je geadviseerd? (meerdere opties mogelijk) *

- [ ] Digitaal programma of informatie
- [ ] Huiselijk geweld
- [ ] Consultatiebureau
- [ ] Kindercoaches
- [ ] Peuterspeelzalen
- [ ] Jongerenwerk
- [ ] Straatcoaches
- [ ] Kindertelefoon
- [ ] Onderwijs
- [ ] Zelfvertrouwen
- [ ] Andere

59. Ruimte voor toelichting op het advies (welke informatie etc.)

Voer uw antwoord in

60. Op welke manier zou aan dit domein gewerkt moeten worden? *

Selecteer maximaal 2 opties.

- [ ] Zelf aan de slag
- [ ] Aan de slag met hulp van vrienden / familie
- [ ] Aan de slag met professionele hulp

61. In welke categorie heb je daarnaast nog advies gegeven? *

Meerderde antwoorden mogelijk, vul hier alle categorieën in waarin je een advies hebt gegeven die je hierboven nog niet genoemd hebt.

- [ ] Bewegen
- [ ] Mentaal welbevinden
- [ ] Voeding
- [ ] Wonen / Huishouden
- [ ] Sociale omgeving / contacten
- [ ] Slapen / ontspanning
- [ ] Persoonlijke verzorging
- [ ] Gewoontes (roken, alcohol, gokken, gamen, drugs etc.)
- [ ] Financieel / werk
- [ ] Vervoer
- [ ] Fysiek welbevinden (bijv. bij pijn)
- [ ] Opvoeden
- [ ] Geen

62. Je hebt mogelijk advies gegeven in meerdere categorieën. Vul hieronder de belangrijkste (maximaal 5) domeinen in waar je over hebt gesproken, begin met de belangrijkste voor deze patiënt en eindig met de minst belangrijke.

de categorieën: bewegen, mentaal welbevinden, voeding, wonen / huishouden, sociale omgeving / contacten, slaap / ontspanning, persoonlijke verzorging, gewoontes, financieel / werk, vervoer, fysiek welbevinden, opvoeden

Voer uw antwoord in

63. Wellicht behoeft het gegeven advies iets meer toelichting dan de gekozen categorieën of opties. Vul hier gerust meer informatie in over het gesprek en het advies

Voer uw antwoord in

Appendix 3: Patient information form

Informatie- en toestemmingsformulier voor het gebruik van uw zorggegevens voor wetenschappelijk onderzoek

Naam studie: Persoonlijk gezondheidsadvies voor secundaire preventie op basis van patiënt gegevens.

Geachte heer/mevrouw,

U heeft een gesprek gehad met een leefstijlcoach bij het Gezondheidsplein.

In ons ziekenhuis zijn we niet alleen betrokken bij de behandeling van patiënten, maar proberen we ook de zorg te verbeteren door het uitvoeren van wetenschappelijk onderzoek.

Wij vragen u of wij uw medische gegevens mogen gebruiken voor wetenschappelijk onderzoek. De gegevens zullen worden verzameld door onderzoekers van Rijnstate. Als u toestemming geeft voor het gebruik van uw medische gegevens kunt u dit aangegeven bij uw leefstijlcoach van het Gezondheidsplein. Voordat u dit beslist krijgt u hieronder uitleg over het onderzoek.

U heeft een advies gekregen van de leefstijlcoach. We onderzoeken of we dit advies ook automatisch kunnen geven via een website. Om een advies te kunnen geven dat bij de patiënt past hebben we persoonlijke informatie nodig. In dit onderzoek verzamelen we deze informatie.

We bewaren deze persoonsgegevens:

- geslacht
- woonplaats
- leeftijd
- burgerlijke staat: of u getrouwd, niet getrouwd, geregistreerd partner bent.
- opleiding
- werk en inkomen
- afkomst
- dingen om rekening mee te houden in het advies (lichamelijk, financieel, of (reis) tijd)

We bewaren deze gegevens over uw gezondheid:

- of u rookt
- uw alcohol gebruik
- uw slaappatroon
- uw eetpatroon
- uw beweegpatroon
- de afdeling in het ziekenhuis waardoor u bent doorgestuurd
- of u een ziekte of aandoening heeft
- tijd sinds het vaststellen van uw ziekte of aandoening
- uw medicijn gebruik
- uw bloeddruk
- uw bloedsuikerwaarden
- uw lengte en gewicht

U kiest zelf of u meedoet. Zo niet, verandert uw zorg niet.

Om uw privacy te beschermen geven wij uw gegevens een code. Aan deze code is te zien om welke patiënt het gaat. Dat is nodig om de juiste patiënt aan de juiste gegevens te koppelen tijdens het verzamelen van de gegevens. Op dat moment kunnen we nog weten waar de gegevens vandaan komen. Daarom is uw toestemming nodig. Na het verzamelen van de gegevens wordt de code verwijderd en blijft de data anoniem. Uw naam wordt dan niet meer gebruikt. De onderzoeksgegevens zijn bij publicatie in een wetenschappelijk tijdschrift niet naar u te herleiden.

We bewaren uw gegevens 10 jaar in het ziekenhuis. Hierna worden ze vernietigd.

Meer informatie over uw rechten bij verwerking van gegevens

Voor algemene informatie over uw rechten bij verwerking van uw persoonsgegevens kunt u kijken op de website van de Autoriteit Persoonsgegevens. Bij vragen of klachten over het gebruik of de verwerking van uw gegevens, of over uw rechten, kunt u contact opnemen met:

Functionaris -

Telefoon: -

Mail: -

Website: https://www.rijnstate.nl/praktische-informatie/mijn-rechten-en-privacy

Vragen

Vragen over het onderzoek kunt u stellen aan Enno Kivits, een van de onderzoekers.

E-mail: -

Met vriendelijke groet,

Enno Kivits, onderzoeker

Appendix 4: Query CTcue

CTcue is a software that is used to extract patient data from the Electronic Health Record (EHR) systems. The features Forms, Vital Signs, and Measurements were used to extract structured data. The 'Reports' feature was used for unstructured data.

Lifestyle behaviour

Alcohol consumption

- Forms - Description: Alcohol
    - Form response – Textual answer: No values
        - 23 results
- Reports – Content: Alcoholgebruik, Alcohol, Alcohol, Alcoholintox
    - –
        - 24 results

Synonyms used:

Alcoholgebruik: *Alcohol Consumption (204), Alcohol Drinking (3), Alcohol use (28), Ethanol Intake (1), Alcohol Intake (1753), Gebruik van Alcoholische Producten (1), Alcoholgebruik (10000), Gebruik Alcohol (1032), Alcohol Gebruik (10000), Alcoholconsumptie (3613), Alcohol Product Use (0), Alcoholic Drink Intake (0)*

Alcohol: *Alcohols (62), Alcohol (1000)*

Alcoholintox: *Drunk (12), Alcohol Intoxication (6), Alcoholintox (1499), Alcoholvergiftiging (22), Intoxicatie Alcohol (1358), Dronkenschap (372) Alcoholintoxicatie (7615), Alcohol Intoxicatie 3860), Dronken (10000), Gedronken (10000), Drinkt (10000), Drink (10000), Drinken (10000), , Dronk (10000), Alcoholic Intoxication (0), Drunkenness (0)*

Smoking status

- Forms - Description: Roken
    - Form response – Textual answer: No value
        - 22 results
- Reports – Content: Roken
    - -
        - 23 results

Synonyms used:

Roken: Smoking (520), Rookster (10000), Roker (10000), Roken (10000)

Medical conditions

Diastolic blood pressure

- Vital signs – Description: Diastolische bloedruk
    - -
        - 19 results
- Reports – Content: Bloeddruk, Diastolische Bloeddruk Normaal, Diastolische Bloeddruk Verhoogd, Diastolische Bloeddruk, Bloeddruk Diastolisch Abnormaal
    - -
        - 19 results

Synonyms used:

Bloeddruk: *Blood pressure (644), Bloeddruk (10000), Bloeddrukmeting (10000), Blood Pressure Determination (0), Blood Pressure taking (0), Rnrx Blood Pressure (0), Rnox Take Blood Pressure (0).*

Diastolische Bloeddruk: *Diastolische Bloeddruk (6), Bloeddruk Diastolisch Normaal (1), Normal Diastolic Blood Pressure (0), Normal Diastolic Arterial Pressure(0), Normal Diastolic Arterial (0), Diastolic Bp Normal (0)*

Diastolische Bloeddruk Verhoogd: *Diastolische Druk Verhoogd (3), Diastolische Bloeddruk Verhoogd (28), Bloeddruk Diastolisch Hoog (5), Increased Diastolic Blood Pressure (0), Increased Diastolic Arterial (0), Increased Diastolic Arterial Pressure (0), High Diastolic Arterial Pressure (0), High Diastolic Arterial (0), Blood Pressure Diastolic High (0), Diastolic Pressure Increased (0), Diastolic Bp Increased (0), Bloeddruk Diastolisch Verhoogd (0)*

Diastolische Bloeddruk: *Bloeddruk Diastolisch (96), Diastolische Bloeddruk (3965), Diastolic Blood Pressure Measurement (0), Blood Pressure Diastolic (0)*

Bloeddruk Diastolisch Abnormaal: *Abnormal Diastolic Arterial (0), Abnormal Diastolic Arterial Pressure (0), Abnormal Diastolic Blood Pressure (0), Bloeddruk Diastolisch Abnormaal (0)*

Systolic blood pressure

- Vital signs – Description: Systolische bloedruk
  - **-**
    - 19 results
- Reports – Content: Bloeddruk, Systolische Bloeddruk, Bloeddruk Systolisch Normaal, Bloeddruk Systolisch Abnormaal, tensie
  - **-**
    - 21 results

Synonyms used:

Bloeddruk: *Blood pressure (644), Bloeddruk (10000), Bloeddrukmeting (10000), Blood Pressure Determination (0), Blood Pressure taking (0), Rnrx Blood Pressure (0), Rnox Take Blood Pressure (0).*

Systolische Bloeddruk: *Blood Pressure Systolic (3), Systolische Bloeddruk (10000), Bloeddruk Systolisch (788), Systolisch Bloeddruk (95), Systolic Blood Pressure Measurement (0)*

Bloeddruk Systolisch Normaal: *Systolische Druk Normaal (6), Bloeddruk Systolisch Normaal (1), Normal Systolic Arterial Pressure (0), Normal Systolic Blood Pressure (0), Systolic Pressure Normal (0)*

Bloeddruk Systolisch Abnormaal: Abnormal Systolic Blood Pressure (0), Abnormal Systolic Arterial Pressure (0), Bloeddruk Systolisch Abnormaal (0)

Tensie: tensie (10000)

BMI

- Vital signs – Description: BMI, Unit: kg/m^2
  - **-**
    - 18 results
- Reports – Content: BMI
  - **-**

        ▪ 22 results

Synonyms used:

BMI: Body Mass Index (10000), Quetelet Index 1427), BMI (1000), Quetelet-index (1427), Body-mass Index (1000)

Total cholesterol

- Measurements – Description: Cholesterol, Unit: mmol/L
    - **-**
        - 16 results
- Reports – Content: Cholesterol
    - **-**
        - 13 results

Synonyms used:

Cholesterol: *Cholesterol (10000), Cholesterolanalyse (3), Cholesterol Measurement Test (0), Cholesterol Analyses (0), Cholesterol Measurement (0), Measurement of Cholesterol (0), Lab-based Chem Measurements Cholesterol (0)*

LDL cholesterol

- Measurements – Description: LDL_CHOLEST., LDL-cholestrol, LDL-chol, LDL chol, LDL, LDL-C, LDL Cholestrol.
    - **-**
        - 14 results
- Reports – Content: LDL, LDL, Ldl Increased, Ldl Low
    - **-**
        - 10 results

Synonyms used:

LDL: *Low Density Lipoprotein Cholesterol (14), Low Density Lipoprotein (19), Ldl Cholesterol (10000), LDLC (9359), Ldl-chol (10000), Familial Hypercholesterolemia (23), LDL (10000), Type II Hyperlipidemie (278), Familiale Hypercholesterolemie (7), Low Density Lipoprotein Cholesterol Level (0), Ldl Cholesterol Measurement (0), Low Density Lipoprotene (0), Low Density Lipoprotene Cholesterol (0).*

Ldl Increased: *Ldl Verhoogd (253), Verhoogd Ldl (3281), Ldl Cholesterol Verhoogd (45), Ldl Gestegen (151), Gestegen Ldl (91), Low Density Lipoprotein Increased (0), Ldl Increased (0), Raised Ldl (0), Low Density Lipoprotene Cholesterol Hoog (0), Low Density Lipoprotene Verhoogd (0)*

Ldl Low: *Ldl Laag (134), Laag Ldl (745), Verlaagd Ldl (24), Low Density Lipoprotein Decreased (0), Low Density Lipoprotein Cholesterol Low (0), Ldl Low (0), Low Density Lipoprotene Cholesterol Laag (0), Low Density Lipoprotene Verlaagd (0)*

HDL cholesterol

- Measurements – Description: HDL, HDL-cholesterol, HDL-C-hol, HDL chol, HDL-Chol, HDL-C, HDL Cholestrol
    - **-**
        - 14 results
- Reports – Content: HDL, Hdl Decreased, Hdl raised, Hdl Cholesterol normal, Hdl Cholestrol Inscreased
    - **-**

- 10 results

Synonyms used:

HDL: *High Density Lipoprotein (5), Hdl Cholesterol (10000), High Density Lipoprotein Cholesterol (4), HDL (10000), HDLC (750), High Density Lipoprotein Measurement (0), High Density Lipoprotein Cholesterol Level (0), Hdl Cholesterol Measurement (0), Hdl Measurement (0), Measurement of High Density Lipoproteins (0), High Density Lipoprotene Cholesterol (0), High Density Lipoprotene (0)*

Hdl Decreased: *Low Hdl (3), Hdl Cholesterol Verlaagd (722), Hdl Laag (97), Laag Hdl (2398), Verlaagd Hdl (405), Hdl Verlaagd (39), High Density Lipoprotein Decreased (0), High Density Lipoprotein Cholesterol Low (0), Hdl Decreased (0), High Density Lipoprotene Cholesterol Laag (0), High Density Lipoprotene Verlaging (0), High Density Lipoprotene Verlaagd (0)*

Hdl raised: *Verhoogd Hdl (131), Hdl Verhoogd (49), Hdl Gestegen (11), Gestegen Hdl (1), Increased Hdl, Hdl Raised (0), High Density Lipoprotein Cholesterol High (0), High Density Lipoprotein Increased (0), High Density Lipoprotene (0), Cholesterol Hoog (0), High Density Lipoprotene Verhoogd (0)*

Hdl Cholesterol normal: *Hdl Cholesterol Normaal (1330), Hdl Cholesterol Normal (0), High Density Lipoprotein Normal (0), High Density Lipoprotene Normaal (0)*

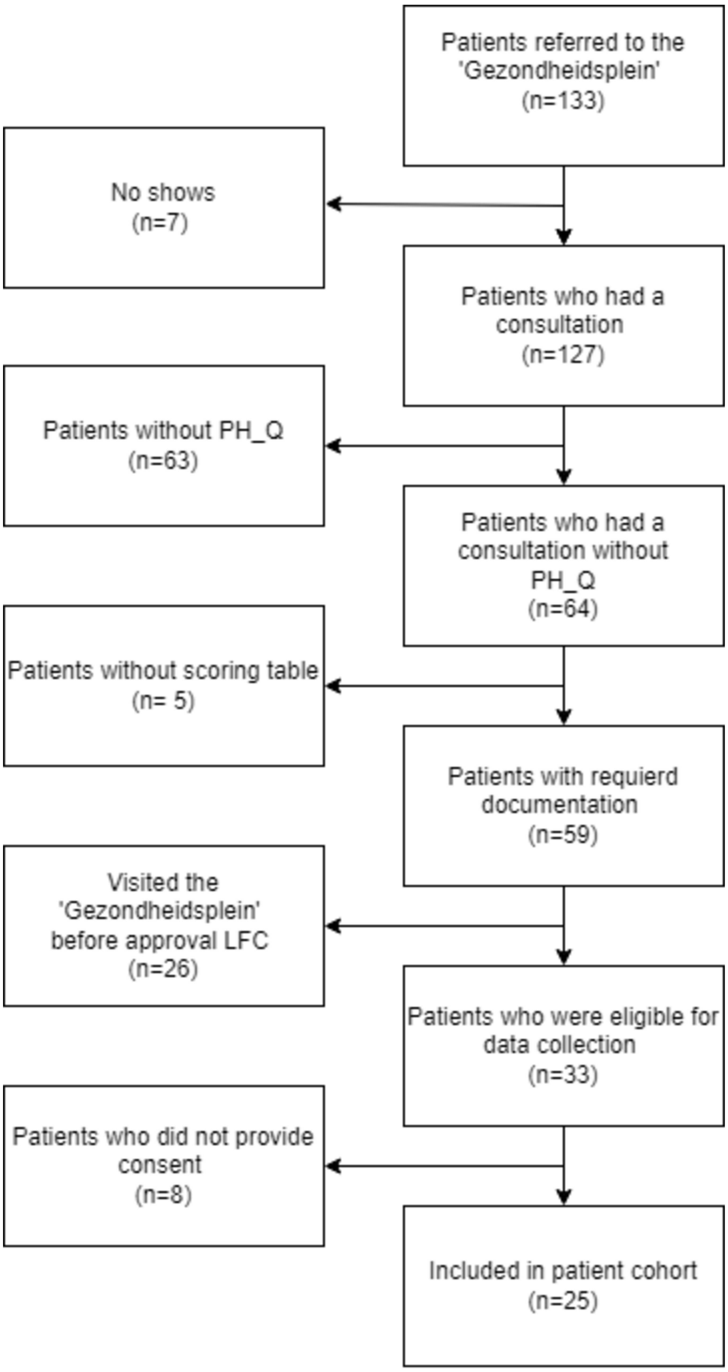
Hdl Cholestrol Inscreased: *Hdl-cholesterol Verhoogd (11), Verhoogd Hdl-cholesterol (45), Hdl Cholesterol Increased (0)*

Appendix 5: Missing data per predictor

Number and percentage of missing data per predictor (n=24).

Predictor	Number of missing values	Percentage of missing values
Sociodemographic		
Age	0	0%
Gender	0	0%
Education	8	33%
Employment	3	13%
Degree of urbanisation	0	0%
Behavioural		
Smoking status	0	0%
Alcohol	0	0%
Medical condition		
Department of referral	0	0%
Body Mass Index (BMI)	3	13%
Blood pressure systolic	13	54%
Blood pressure diastolic	13	54%
Total cholesterol	19	79%
LDL cholesterol	19	79%
HDL cholesterol	19	79%
Positive Health		
Bodily functioning	0	0%
Mental wellbeing	0	0%
Meaningfulness	0	0%
Quality of life	0	0%
Participation	0	0%
Daily functioning	0	0%

Appendix 6: Flowchart of patient inclusion



Appendix 7: Advice per patient characteristic

Characteristics	**Lifestyle behaviours n = 9 (37.5%)**	**Mental well-being n = 6 (25.0%)**	**Nutrition n = 13 (54.2%)**	**Physical activity n = 8 (33.3%)**
Sociodemographic				
Gender				
Female	8	6	13	6
Male	1	0	0	2
Education				
Low	2	0	2	2
Medium	2	2	6	1
High	1	1	2	4
Employment				
Employed	3	5	6	3
Unemployed	1	1	2	1
Retired	3	0	1	2
Voluntary	1	0	2	1
Degree of urbanisation				
Not urbanised	1	0	1	0
Hardly urbanised	1	0	2	1
Moderately urbanised	4	4	4	3
Strongly urbanised	3	2	6	3
Extremely urbanised	0	0	0	1
Behavioural				
Smoking status				
Never	0	4	7	4
Current	7	1	1	1
Stopped	2	1	5	3
Alcohol consumption				
No	6	4	9	4
Yes	3	2	4	4
Medical condition				
Department of referral				
Cardiology	1	0	1	1
DIVAN	0	0	1	1
Gastrointestinal oncology	1	0	0	0
Gynaecology	3	3	5	4
Lung	2	1	3	1
Oncology	2	2	3	1