

Data in de zorg

Naam: Rogier Krom S2167271

Begeleiders:
Universiteit Twente: I. ten Klooster MSc
Dr. S.M. Kelders
Politecnico di Milano: E. Tauro

Datum: 5-7-2022

“Het genereren van persona’s als onderdeel in de ontwikkeling van een Early Warning System”

Voorwoord

Beste lezer,

voor u ligt de scriptie 'Het genereren van persona's als onderdeel in de ontwikkeling van een Early Warning System'.

Dit onderzoek is uitgevoerd in het kader van de bachelor opdracht van de studie Gezondheidswetenschappen aan de Universiteit Twente. Het afstudeeronderzoek is een onderdeel van het Power4FitFoot project waarin er een Early Warning System voor diabetespatiënten met een verhoogde kans op voetulcers en/of amputaties wordt ontwikkeld. Het onderzoek bestaat uit een kwantitatief onderzoek en is uitgevoerd in de periode februari 2022 tot juli 2022.

In dit voorwoord wil ik graag mijn begeleiders I. ten Klooster MSc en Dr. S.M. Kelders van de Universiteit Twente en E. Tauro van de Politecnico di Milano bedanken voor de prettige en leerzame samenwerking, feedback en begeleiding tijdens het schrijven van de scriptie. Daarnaast wil ik mijn medestudenten Diede Paarlberg en Jesper Wallinga bedanken met wie ik samen de onderzoeksgroep 'Data in de zorg' vormde. Ook wil ik Podotherapie Segerink in Deventer bedanken voor het bieden van een praktijk waar ik de vragenlijst kon afnemen voor de pilot.

Ik wens u veel leesplezier toe,

Rogier Krom

Enschede, juli 2022

Samenvatting

Achtergrond: Diabetes is een chronische ziekte met een hoge ziektelast, daarbij komt dat diabetespatiënten vaak last krijgen van een of meerdere complicaties. Een voorbeeld van een complicatie is diabetisch voetulcus, een wondje aan de voeten die grote gevolgen kan hebben zoals amputaties van de onderste ledematen. Het Power4FitFoot project focust zich op de ontwikkeling van een Early Warning System voor voetulcers en/of amputaties. Dit Early Warning System kan aan de hand van voorspellende factoren hoge risico patiënten een vroegtijdige waarschuwing geven. Echter moet bij de ontwikkeling van een nieuwe eHealth technologie rekening worden gehouden met de eindgebruiker van de technologie. In de ontwikkeling kunnen persona's, fictieve weergaven van de eindgebruikers, een goed beeld geven van de eisen waaraan het Early Warning System uiteindelijk moet voldoen.

Doel: Het doel van dit onderzoek is doormiddel van kwantitatieve data en clusteranalyses groepen met dezelfde karakteristieken te genereren binnen een diabetespopulatie. Deze groepen oftewel clusters geven de data weer die uiteindelijk de basis vormen voor de persona's. Aan de hand van de persona's kan een aanbeveling worden gegeven voor de ontwikkeling van het Early Warning System.

Methode: Het onderzoek is uitgevoerd aan de hand van literatuuronderzoek en een vragenlijst. De vragenlijst is opgebouwd uit verschillende domeinen die in de literatuur relevant werden gevonden binnen de generaties van persona's. De domeinen zijn getest doormiddel van in eerder onderzoek gevalideerde meetinstrument. De populatie werd verworven op verschillende sociale media en ook een in de wachtruimte van een podotherapie praktijk is er een kans geboden waar de patiënten mee konden werken aan het onderzoek. Kwantitatieve data van 24 personen is verzameld voor de uiteindelijke analyse. De Principal Component Analysis en Gower Analysis werden uitgevoerd om de clusters van elkaar te onderscheiden. Doormiddel van de Silhouet-scores van de twee methodes werd het aantal clusters en de beste methodiek voor de clustering bepaald. De uiteindelijke clusters werden daarna getest op het vertonen van significante verschillen tussen de clusters door de Mann-Whitney U-test.

Resultaten: Uit de Silhouet-analyse is gebleken dat de PCA-methode met een aantal van twee clusters binnen de onderzoekspopulatie de meeste significante verschillen vertoonden.

Conclusie: Door de gegevens van de gegenereerde clusters in beeld te brengen werden persona kaarten ontworpen. Deze kaarten werden gepersonaliseerd met namen en foto's van de mogelijke personen zodat er een duidelijk beeld wordt verkregen van de eindgebruikers. Daarnaast bevatten de persona kaarten ook de resultaten hoe deze persona scoort op de karakteristieke domeinen.

Aanbevelingen: De persona's bevatten elk hun eigen kenmerken in omgang met bijvoorbeeld technologie, voetverzorging en medicatie inname. Tijdens de ontwikkeling van een Early Warning System moet er dus rekening mee gehouden worden dat beide groepen ook andere eisen stellen aan de technologie. In vervolg onderzoek kan de opgestelde clusteranalyse worden toegepast op een grotere populatie om meer clusters in de diabetespopulatie te onderscheiden.

Inhoudsopgave

Voorwoord	1
Samenvatting.....	2
1. Inleiding.....	5
2. Theoretisch raamwerk	8
2.1 Introductie	8
2.2 Procedure bij de ontwikkeling van systemen	8
2.2.1 Multidisciplinaire aanpak.....	8
2.3 CeHReS roadmap	8
2.3.1 Fases in de roadmap.....	9
2.3.2 Persona's in de CeHRes roadmap.....	9
2.4 Segmentatie van de doelgroep.....	10
2.4.1 Factoren van persona's.....	10
2.4.2 Opgestelde domeinen	11
3. Methode	12
3.1 Onderzoeksdesign.....	12
3.2 Respondenten.....	12
3.3 Materialen	12
3.3.1 Leefstijl en socio-economisch	13
3.3.2 Zelf gerapporteerde gezondheid.....	13
3.3.3 Kwaliteit van leven	13
3.3.4 Gezondheidsvaardigheid	13
3.3.5 Gezondheidsbetrokkenheid.....	14
3.3.6 Therapietrouwheid.....	14
3.3.7 Chronische aandoening en voetverzorging	14
3.3.8 Pilot.....	15
3.4 Procedure	15
3.5 Analyse	15
4. Resultaten.....	17
4.1 Data Cleaning.....	17
4.2 Exploratieve analyse/Silhouet-analyse.....	17
4.3 Clusteranalyse PCA.....	18
4.3.1 Tabel PCA	18
4.3.2 Resultaten cluster 1	18
4.3.3 Resultaten cluster 2	18
4.4 Significantie.....	19

5. Discussie	20
5.1 Antwoord op de onderzoeksvraag	20
5.2 Koppeling met de gevonden literatuur	21
5.3 Reflectie op de resultaten	23
5.3.1 Lege responses	23
5.3.2 Waardering van vragen.....	23
5.3.3 Verschil in silhouet-scores	24
5.4 Conclusie.....	24
Referenties	26
Appendices	30
Appendix 1: Informatiebrief deelname.....	30
Appendix 2: Vragenlijst (Nederlandse versie)	31
Appendix 3: Uitnodigingsbrief social media	52
Appendix 4: Jupyter script 'Data cleaning'	54
Appendix 5: Verklaring categorieën en tot stand komen scores.....	60
Appendix 6: Jupyter script 'Gower-Method'	62
Appendix 7: Jupyter script 'PCA'	64
Appendix 8: Missende data	66
Appendix 9: Resultaten PCA(tabel 4)	67

1. Inleiding

Door de groeiende vergrijzing en stijgende zorgkosten is de zorg één van de grootste uitgavenposten [1]. Wereldwijd en ook in Nederland leven steeds meer mensen die chronisch ziek zijn. In 2015 hadden er alleen al in Nederland naar schatting 5,3 miljoen mensen een chronische ziekte [2]. Een chronische ziekte eist veel van een patiënt, zo speelt de ziekte een grote rol in het dagelijks leven. Ook is het groeiend aantal chronisch zieken een uitdaging voor de zorgverleners. Zij moeten namelijk bepalen welke behandeling het beste bij een patiënt past en hoe ze de benodigde ondersteuning aan patiënten kunnen aanbieden, echter door de combinatie van het tekort aan zorgverleners en de stijging van de chronisch zieken wordt het bieden van de juiste zorg een steeds grotere uitdaging [3].

Een ernstige chronische ziekte is diabetes mellitus [4]. Momenteel hebben in Nederland 1,2 miljoen mensen diabetes, jaarlijks komen daar 52.000 mensen bij en naar schatting leven er 1,1 miljoen Nederlanders met prediabetes, een voorfase van diabetes type 2 [5]. De meest voorkomende vormen van diabetes zijn diabetes type 1 en diabetes type 2. Diabetes type 2 is de meest voorkomende vorm van diabetes, echter kan een gezonde levensstijl helpen om diabetes type 2 te voorkomen [5]. Onder diabetespatiënten heeft 40 tot 56 procent van de mensen last van één of meerdere chronische complicaties van diabetes, zoals hart- en vaatziekten, oogafwijkingen en nieraandoeningen [5]. De bijkomende aandoeningen onder diabetes patiënten in verhouding tot de gehele populatie kan gemeten worden met de gestandaardiseerde morbiditeitsratio (SMR) [6]. Een SMR van boven de 1,0 betekent dat binnen de diabetes populatie een bepaalde aandoening vaker voorkomt dan de prevalentie in de Nederlandse bevolking. Van de diabetes type 2 patiënten ervaart ongeveer 15-25% (SMR 1,15-1,25) een voetzweer. Daarbij komt dat het risico op een amputatie tien tot twintig keer hoger bij diabetes type 2 patiënten dan bij mensen zonder diabetes type 2 [7].

Binnen de gezondheidszorg is de afgelopen jaren zelfmanagement binnen het zorgproces van een patiënt toegenomen [8]. Dit is een gevolg van de personeelstekorten in de zorgverlening [3]. Zelfmanagement kan hierbij een oplossing bieden. Zelfmanagement is gebaseerd op de mogelijkheden van een individu om zijn symptomen, behandeling, fysieke en psychosociale gevolgen en veranderingen in leefstijl die verbonden zijn aan leven met een chronische aandoening, te managen [9]. Bij diabetespatiënten is het zelfmanagement van de patiënt bijvoorbeeld gerelateerd aan het preventief verzorgen van de voeten om voetulcers te voorkomen [10]. Samen met de stijgende behoefte van zelfmanagement stijgt ook de vraag naar technologische oplossingen [11]. Technologie gebaseerde diensten, zoals mobiele toepassingen of monitoring op afstand, zorgen voor betere toegang tot zorg en doorlopende ondersteuning voor patiënten [12]. Om de zorg op lange termijn betaalbaar, kwalitatief goed en schaalbaar te houden is er een grote vraag naar slimme, vaak technologische, oplossingen. De ontwikkeling van een Early Warning System speelt in op de stijgende vraag naar technologische oplossingen die het zelfmanagement binnen de zorg ondersteunen. Binnen het Power4FitFoot project wordt er een Early Warning System [13] ontwikkeld die een vroegtijdige waarschuwing geeft wanneer een patiënt een verhoogde kans heeft op een voetzweer en/of amputaties. In Power4FitFoot wordt de verschuiving naar gepersonaliseerde gezondheidszorg weerspiegeld in de nadruk op zelfmanagement in plaats van ziektebeheer, waardoor patiënten in staat worden gesteld een proactieve rol te spelen in de beslissingen

over hun gezondheid. Diabetespatiënten zullen dus zelf steeds meer regie krijgen in de behandeling van de ziekte, de controle van hun voeten, de voetverzorging en de medicatie inname.

Door de digitalisering van patiëntendossiers, de toename van beschikbare databronnen en meerdere technieken kan er steeds meer data over diabetes worden verzameld. Aan de hand van deze data kunnen verscheidene prognostische modellen worden ontwikkeld waarmee het risico op een voetzweer of amputatie kan worden voorspeld. Prognostische modellen worden gedreven door kunstmatige intelligentie (AI) [14]. Dit is de technologie die computermodellen gebruiken om intelligent gedrag te. Dergelijke modellen kunnen worden toegepast voor prioritering van routinezorg [7], het behalen van specifieke gezondheidsdoelen, het toewijzen van middelen in de gezondheidszorg of het verminderen van de zorgkosten [15] of voor procedures zoals prognose, diagnose en behandelplanning [16]. Zo wordt er bijvoorbeeld in de behandeling van borstkanker gebruik gemaakt van prognostische modellen om zo het verloop van de ziekte en de daarbij beste gepersonaliseerde behandeling te kunnen bieden [17]. De prognostische modellen kunnen worden gebruikt bij de ontwikkeling van een Early Warning System [18]. Het Power4FitFoot programma concentreert zich op patiënten met diabetes die een verhoogde kans hebben op de ontwikkeling van een voetzweer en/of amputaties. Binnen het Power4FitFoot project wordt gebruik gemaakt van zogenaamde 'Bigdata' om een gepersonaliseerde benadering van hoge risicofactoren op het krijgen van een voetzweer en/of amputaties te koppelen aan een systeem die deze factoren kan identificeren en voorspellen.

Om de kwaliteit van de zorg te verbeteren wordt er steeds meer ingezet op persoonlijke patiëntgerichte zorgoplossingen [19]. Als onderdeel van de ontwikkeling van een Early Warning System kan de generatie van persona's een belangrijk aspect zijn. Persona's zijn een fictieve weergave van personen gebaseerd op werkelijke data [20]. Deze persona's worden gegenereerd om een representatief beeld van een primaire eindgebruikersgroep te krijgen. Persona's worden aanbevolen als onderdeel van een gebruikersgericht proces voor het ontwerpen en implementeren van IT voor de gezondheid van de consument, in het bijzonder als een eenvoudig instrument voor de communicatie met de vele belanghebbenden die betrokken zijn bij het ontwerp van IT voor de gezondheid [21]. Binnen de 'Center for eHealth Research (CeHRes) Roadmap' [22], een ontwikkelingsaanpak voor eHealth interventies, speelt de ontwikkeling van persona's al een belangrijke rol. Door middel van deze data gedreven persona's kan communicatie over specifieke behoeftes en gedrag van patiënten worden afgestemd. De toepassing van gebruikersprofielen en persona's heeft namelijk niet veel aandacht gekregen in het onderzoek naar informatica in de gezondheidszorg [23]. Door de behoeftes en eigenschappen van de eindgebruikers zo goed mogelijk mee te nemen bij de ontwikkeling van een nieuwe model kan de vraag naar een dergelijk model worden verhoogd.

Zelfmanagement heeft potentie, maar lang niet alle patiënten reageren op de implementatie van nieuwe interventies [24]. Zo zijn er verschillende barrières bij de eindgebruikers zoals tijd, kennis, wantrouwen of kosten waardoor de zelfmanagement oplossingen niet werken [24]. Als onderdeel van de ontwikkeling van een Early Warning System kunnen er persona's binnen de diabetes populatie worden onderscheiden, om zo bij de ontwikkeling van de technologie

rekening te kunnen houden met de uiteindelijke eindgebruikers en hun mogelijke barrières ten opzichte van het Early Warning System. Het doel van dit onderzoek is daarom het creëren van persona's binnen de diabetes populatie om deze uiteindelijk te kunnen gebruiken bij de ontwikkeling van een Early Warning System. De onderzoeksvraag in dit onderzoek luidt als volgt: 'Welke persona's zijn er te onderscheiden bij diabetespatiënten?'

2. Theoretisch raamwerk

2.1 Introductie

Het opstellen van de eisen is een cruciaal onderdeel in de ontwikkeling van een nieuwe technologie [25]. De eisen zijn de basis in de ontwikkeling van de technologie en beschrijven wat een technologie beoogd te doen, welke gegevens er moeten worden opgeslagen of verwerkt, wat de technologie moet weergeven en welke ervaringen het moet bieden aan de gebruiker. Om tot een antwoord op de onderzoeksvraag te komen is het van belang om kennis te hebben over zowel de procedure bij de ontwikkeling van systemen, de definitie en het gebruik van persona's en de manier waarop persona's worden gegenereerd.

2.2 Procedure bij de ontwikkeling van systemen

In systeem analyses en ontwerp processen van systemen worden op grote schaal verschillende modellen gebruikt [23]. Studies hebben bewezen dat modellen ontwikkelaars en belanghebbenden in staat stellen om de gegevensverwerking en interactie tussen het systeem en de uiteindelijke gebruiker te visualiseren [26].

2.2.1 Multidisciplinaire aanpak

Power4FitFoot volgt een participatieve multidisciplinaire ontwikkelingsaanpak: onderzoekers (geneeskunde, computer- en gedragswetenschappen) werken samen met patiënten, zorgverleners en industriële bedrijven om gepersonaliseerde zelfmanagementproducten en -diensten te ontwikkelen om diabetische voetzweren of amputaties te voorkomen. Bij de ontwikkeling van een technologie wordt het opstellen van eisen vaak overgelaten aan ingenieurs die een technologie gedreven aanpak hanteren [25]. Echter kan het potentieel van een technologie binnen de gezondheidszorg alleen ten volle worden benut wanneer de technologie wordt ontwikkeld door een multidisciplinair team dat een mensgerichte aanpak toepast waarbij rekening wordt gehouden met de specifieke context waarin de technologie zal worden gebruikt [25]. Een multidisciplinair team bestaat uit een groep met verschillende behoeftes, expertises of belangen, waardoor verschillende invalshoeken meedenken in het ontwerpproces.

Het eindresultaat van de multidisciplinaire aanpak in meerdere stappen voor de ontwikkeling van vereisten voor eHealth interventies is een ontwerpdocument met functionele en niet-functionele vereisten, een lijst van waarden van de stakeholders, en eindgebruikersprofielen in de vorm van persona's [25].

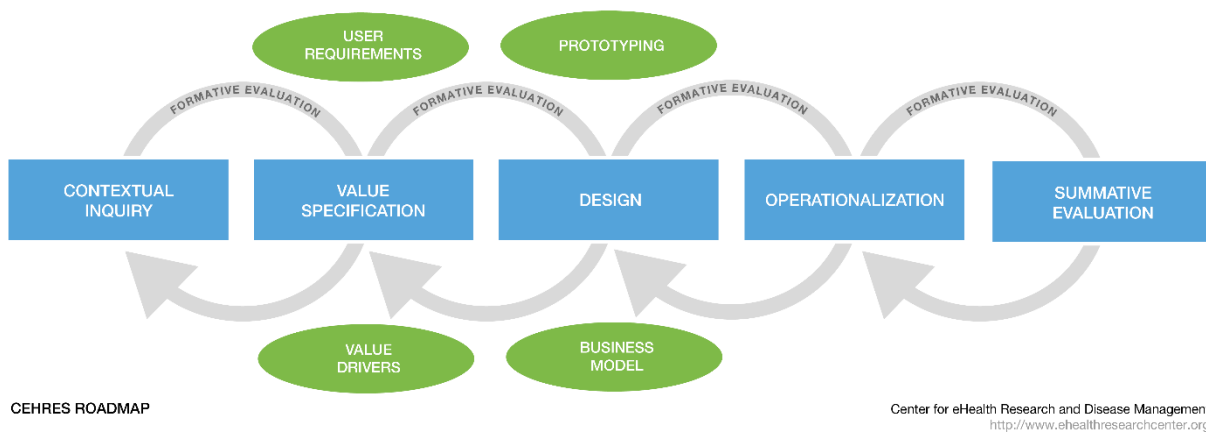
2.3 CeHReS roadmap

Om tot een afstemming tussen technologische, menselijke en contextuele factoren te komen is er een holistische aanpak nodig. Een holistische aanpak wil het geheel benadrukken afhankelijk van de onderlinge onderdelen. Een dergelijke benadering zou rekening houden met de financiële aspecten, het management en de technologie bij het ontwerpen, implementeren en evaluatie van gezondheid technologieën [27]. Een voorbeeld van een holistische aanpak is de CeHRes roadmap' [25]. De CeHRes roadmap is een ontwikkelingsaanpak voor eHealth interventies om waarde toevoegende en duurzame technologieën voor de gezondheidszorg te creëren.

2.3.1 Fases in de roadmap

De CeHRes roadmap [16] bestaat uit vijf verschillende fases (zie figuur 1):

1. *'Contextual inquiry'*: er wordt informatie verzameld over de gebruikscontext, de aangewezen eindgebruikers en de professional die de technologie moeten toepassen.
2. *'Value specification'*: de gegevens uit het contextonderzoek worden vertaald in de waarden van de belanghebbenden en de vereisten voor de technologie.
3. *'Design'*: prototypes van de technologie worden ontwikkeld op basis van de eisen en daarna getest.
4. *'Operationalisation'*: de definitieve versie van de technologie wordt geïntroduceerd samen met de aanvullende middelen (bv. gebruikersondersteuning).
5. *'Summative evaluation'*: ten slotte worden de acceptatie en het effect van de technologie geëvalueerd.



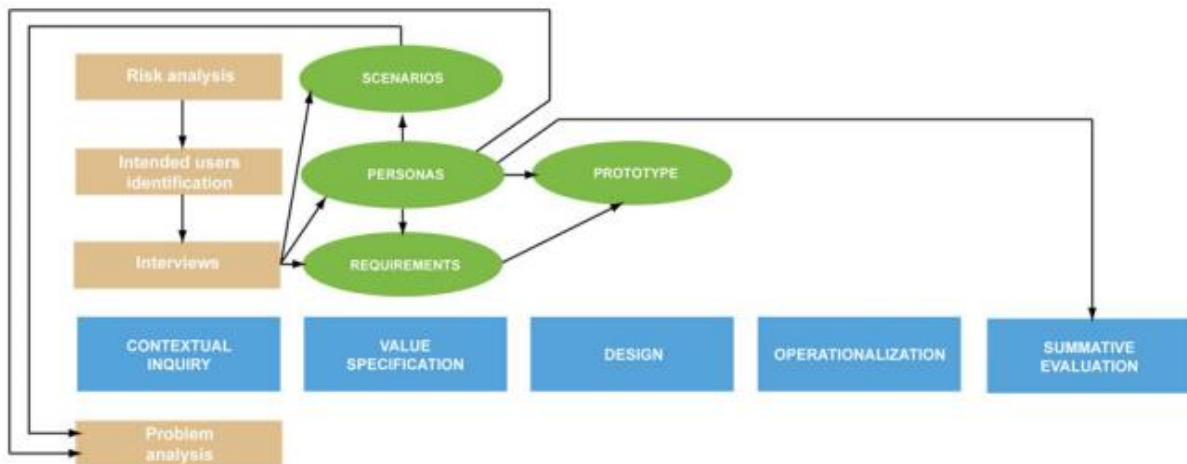
Figuur 1: CeHRes roadmap [14]

Binnen dit onderzoek wordt er voornamelijk gekeken naar de fases 'contextual inquiry' en 'value specification'.

2.3.2 Persona's in de CeHRes roadmap

Binnen de holistische aanpak van het ontwikkelen van nieuwe (eHealth) technologieën via de CeHRes roadmap, kan de ontwikkeling van persona's een belangrijke rol spelen. De persona's kunnen dienen als inspiratie voor functioneel ontwerp en interface- en interactieontwerp, maar ze kunnen ook worden gebruikt om na te denken over de bestaande organisatiestructuren achter de zorg [22]. Persona's dienen als levendige samenvatting van de eindgebruikers. Ze kunnen het ontwerp informeren en inspireren stakeholders om de juiste probleemanalyses te creëren. In figuur 2 is te zien waar in het ontwerpproces de persona's aan bod komen en welke functies de persona's hebben in het ontwerpproces. Zo zorgt de creatie van de persona's voor een duidelijk beeld van de eisen waaraan een technologie moet voldoen, kunnen er scenario's voor de technologie worden ontwikkeld en kan er een beeld van een prototype worden geschetst. Zo zullen wat oudere mensen vaker moeite hebben met het aanleren van nieuwe technologieën dan jongeren, hebben diabetespatiënten die al langer zijn gediagnosticeerd een hogere kans op de ontwikkeling van voetulcers en hebben hoger opgeleide mensen over het algemeen wat meer kennis van (medische) begrippen. Daarnaast

zijn persona's erg belangrijk tijdens de evaluatie van een technologie, deze eindgebruikers weten namelijk het beste of een technologie daadwerkelijk doet wat het beoogd te doen.



Figuur 2: Persona's in the CeHRes roadmap [20]

2.4 Segmentatie van de doelgroep

Bij doelgroep segmentatie gaat het erom binnen een populatie homogene subpopulaties (segmenten) te identificeren en deze te profileren [28]. In deze fase is het identificeren van populatie segmenten het hoofddoel. Om populatie segmenten te identificeren, moeten er verschillende determinanten in de populatie worden onderscheiden. Deze determinanten kunnen bijvoorbeeld worden opgedeeld in: kennis, houdingen en gedrag in een bepaalde context [25]. In het meest ideale scenario wordt het opsplitsen van een populatie gedaan op basis van grote reeksen kwantitatieve data [29]. Kwantitatief onderzoek kan worden ingezet om feiten te verzamelen die kunnen worden gegeneraliseerd naar een grotere populatie. Uiteindelijk vormen de segmenten in de populatie clusters, dit zijn groepen personen met overeenkomende kenmerken.

2.4.1 Factoren van persona's

Het profileren van de eindgebruikers van een eHealth technologie is een lastige taak [25]. Het gebruik van een eHealth technologie is namelijk erg afhankelijk van in hoeverre de technologie voldoet aan de behoeftes van de verschillende populaties. Dit is erg lastig in heterogene populatie waarin culturele achtergronden, vaardigheden en handicaps verschillen [30]. In deze heterogene populaties zijn er dus veel beweegredenen om een technologie wel of juist niet te gebruiken. Om te kunnen omgaan met de heterogeniteit van de eindgebruikerspopulatie van een eHealth technologie, is het een belangrijke stap in de ontwikkeling om publiekssegmenten te identificeren, profileren en ontwerpen [25].

De factoren die in de methode van dit onderzoek worden gepresenteerd, zijn afgeleid van profiel- en persona literatuur, literatuur over bruikbaarheid van technologie en literatuur over gezondheidszorg [31]. De karakteristieken van de persona's kunnen worden opgedeeld in gemeenschappelijke demografische kenmerken die voor een persona groep van belang zijn, deze omvatten: de voorkennis en ervaring van gebruikers, ervaring (vaardigheid in het zoeken op websites), fysieke kenmerken (algemene gezondheid, lichaamsbeweging), cognitieve

kenmerken (intelligentie), sociale en fysieke omgeving (thuisituatie, vrienden en/of familie) en baan, taak en vereisten [31]. Hoe deze onderwerpen terugkomen in de domeinen van de vragenlijst is verder verklaard in de methode van dit onderzoek.

2.4.2 Opgestelde domeinen

Uit de literatuur komen de volgende domeinen in de vorming van de persona's naar voren:

1. Leefstijl en socio-economisch: binnen dit domein worden persoonsgegevens verzameld om een profiel te kunnen schetsen van de respondent.
2. De zelf gerapporteerde gezondheid is een domein van vragen om zonder medische testen toch een aantal gezondheid gegevens over de respondenten te kunnen verzamelen. Er wordt gevraagd naar een aantal gezondheidskenmerken die de respondent zelf kan weten/meten.
3. Kwaliteit van leven [32]: wordt gedefinieerd als de kwaliteit van het bestaan en tevredenheid met de levensomstandigheden. Deze maatstaf is ook instaat om onderscheid te maken tussen bepaalde groepen en datgene wat normatief is. Binnen dit onderzoek wordt de kwaliteit van leven opgedeeld in drie domeinen namelijk kwaliteit van leven op fysiek gebied, relationeel gebied en persoonlijk gebied [33]. Dit wordt om de kwaliteit van leven nauwkeuriger te meten.
4. Gezondheidsvaardigheid [34] wordt ingezet om het volledige potentieel in gebruikersgerichte eHealth te bereiken, er moet worden gestreefd naar een match tussen de eHealth en de vaardigheden en kennis van de uiteindelijke gebruiker.
5. Gezondheidsbetrokkenheid [35] geeft inzicht in de patiëntbetrokkenheid ten opzichte van een ziekte. De uitkomsten van deze test kunnen worden gebruikt om op maat gemaakte interventies te ontwikkelen en veranderingen te beoordelen na zelfmanagement interventies.
6. Therapietrouwheid [36]: de therapietrouw is een belangrijk indicator bij patiënten met een chronische aandoening om de optimale resultaten van een behandeling te kunnen bereiken. therapietrouwe patiënten ervaren betere gezondheidsresultaten ten opzichte van de patiënten die niet trouw zijn aan de voorschriften van hun medicatie.
7. Chronische aandoening en voetverzorging [37]: de kennis en gedrag rondom de voetverzorging van diabetespatiënten is een belangrijk kenmerk binnen de diabetespopulatie. Vroegtijdige herkenning en behandeling van risicofactoren voor voetulcers en amputaties kan het ontstaan van ongunstige resultaten voorkomen of vertragen [10]. Een aantal voorbeelden van risicofactoren zijn: het verlies van gevoel, bewijs van verhoogde druk (eelt), misvorming, perifere vaatlijden, geschiedenis van ulcers of ernstige nagelpathologie [10].

3. Methode

3.1 Onderzoeksdesign

De onderstaande methode beschrijft een kwantitatief onderzoek om een antwoord te kunnen geven op de onderzoeksvraag: 'Welke persona's zijn er te onderscheiden bij diabetespatiënten?' De persona's zullen worden weergegeven in de vorm van een personakaart die de verschillende eigenschappen van het cluster weergeeft[38]. Dit onderzoek is goedgekeurd door de ethische commissie van de BMS faculteit van de Universiteit Twente, met het goedkeuringsnummer 220052.

3.2 Respondenten

De sampling strategy binnen dit onderzoek werd gedaan volgens de convenience sampling strategy [39]. Convenience sampling houdt in dat een deel van de bevolking werd geïnccludeerd aan de hand van vrijwillige deelname aan het onderzoek. De respondenten konden het makkelijkst worden bereikt en de data kon relatief eenvoudig worden verzameld. De doelgroep die beoogt werd te bereiken zijn diabetespatiënten. De inclusiecriteria om in aanmerking te komen voor de deelname aan het onderzoek waren: de desbetreffende persoon moet de diagnose diabetes hebben en moet 18 jaar of ouder zijn. Door middel van de convenience sampling werden 24 participanten geïnccludeerd voor onderzoek.

Het verwerven van de populatie vond plaats bij 'podotherapie Segerink' te Deventer. Er is gekozen voor een podotherapeutisch centrum omdat diabetisch patiënten tijdens hun ziekte regelmatig in aanraking komen met een podotherapeut voor de behandeling en verzorging van hun voeten. Verder werd er doormiddel van sociale media een oproep gedaan tot deelname aan het onderzoek. Dit gebeurde via de sociale media Facebook en LinkedIn en de website 'diabetestrefpunt.nl'. De respondenten konden reageren via een uitnodigingsbrief met daarin uitleg over het onderzoek en de inclusiecriteria(appendix 3: 'Uitnodigingsbrief social media').

Een aantal eigenschappen van de onderzoekspopulatie die geprobeerd werd te bereiken zijn: diabetes komt voor in de gehele populatie, diabetes komt net iets vaker voor bij mannen dan bij vrouwen [40]. De gemiddelde leeftijd waarop diabetes gediagnosticeerd wordt ligt bij diabetes type 1 rond de 35 jaar en bij diabetes type 2 rond de 61 jaar [40]. De uiteindelijke bereikte onderzoekspopulatie is te zien in tabel 1.

Karakteristieken respondenten	n = 24
Leeftijd in jaren, mean (min-max; SD)	54,8 (22-81; 18,25)
Geslacht, n(%)	
Man	14 (58%)
Vrouw	10 (42%)

Tabel 1: Karakteristieken respondenten

3.3 Materialen

Gegevens over de diabetes populatie werden verzameld aan de hand van een vragenlijst (appendix 2: 'Vragenlijst'). Deze vragenlijst is onderdeel van de methode in dit onderzoek en past goed bij de onderzoeksvraag omdat de persona's in dit onderzoek uiteindelijk worden onderscheiden doormiddel van kwantitatieve gegevens op verschillende domeinen en

factoren. De vragenlijst (appendix 2: 'Vragenlijst') is opgebouwd uit verschillende onderdelen van bestaande vragenlijsten, dit wordt verder verklaart in de uitleg van de domeinen. De onderdelen zijn vertaald naar het Nederlands om de voertaal van de onderzoekspopulatie aan te houden. Bij de vertaling is een 'double back translation' vanuit de Engelse taal toegepast[41]. Bij de 'double back translation' werden twee vertalingen van individuele personen samengevoegd tot één definitieve vragenlijst. Ter controle werd deze tekst daarna door de onderzoeker weer vertaald naar de originele taal.

De vragenlijst (appendix 2: 'Vragenlijst') bestaat uit de domeinen opgesteld in hoofdstuk 2.4.2 'Opgestelde domeinen'. Waar ook de definitie van de begrippen wordt gegeven zoals staat beschreven in de literatuur. Welke domeinen worden gemeten met welke vragen en hoe de domeinen terugkomen in de vragenlijst (appendix 2: 'Vragenlijst') wordt hieronder beschreven.

3.3.1 Leefstijl en socio-economisch

In deze sectie van de vragenlijst werden persoonsgegevens zoals geslacht, werkstatus, relatiestatus en opleiding gevraagd [42]. Daarnaast werd er gevraagd de behoefte voor hulp bij het gebruik van technische hulpmiddelen.

3.3.2 Zelf gerapporteerde gezondheid

De vragen omtrent de zelf gerapporteerde gezondheid gingen over gewicht, lengte en cholesterolgehalte. Maar ook gedragsmatige vragen die te maken hadden met alcohol- en/of rookgedrag. Zo werden de persoonsgegevens onder andere verzameld met de vragen: 'Wat is uw geslacht?', 'Wat is uw leeftijd?', 'Wat is uw lengte en gewicht?' en 'Wat is uw opleidingsniveau?'

3.3.3 Kwaliteit van leven

De kwaliteit van leven werd gemeten door middel van de QOLS-vragenlijst (QOLS) [43][33]. Hierin werd de kwaliteit van leven van een chronische patiënt getest aan de hand van 11 stellingen die moeten worden beantwoord op een Likert-schaal. Een voorbeeld van zo'n stelling is: 'Hoe voelt u zich momenteel over, uw gezondheid – lichamelijk fit en vitaal zijn'. Deze schaal verschilt van blij, tevreden, meestal voldaan, gemengd, meestal onvoldaan, ontevreden tot ongelukkig (appendix 2: 'Vragenlijst'). De kwaliteit van leven is opgedeeld in 3 secties, namelijk de kwaliteit van leven met betrekking tot fysieke, relationele en persoonlijke tevredenheid (zie appendix 5: 'Verklaring categorieën en tot stand komen scores'). De vragen binnen dit domein zijn getest doormiddel van de Cronbach's alfa test [33], een maat voor de betrouwbaarheid van vragenlijsten. Daaruit kwam dat de QOLS een interne betrouwbaarheid van $\alpha = .82$ tot $.92$, oftewel goed tot zeer goed.

3.3.4 Gezondheidsvaardigheid

Gezondheidsvaardigheid (eHLA) [34] werd opgedeeld in twee delen. Ten eerste een deel waar de respondenten moesten antwoorden op een aantal stellingen doormiddel van een Likert-schaal (appendix 2: 'Vragenlijst'). Een voorbeeld van een stelling is: 'Op een schaal van zeer gemakkelijk tot zeer moeilijk, hoe gemakkelijk vindt u het om informatie te vinden over de behandeling van ziekten waarover u meer wilt weten?'

Het tweede deel van de gezondheidsvaardigheid wordt gemeten door middel van medische begrippen. Aan de respondent wordt dan gevraagd in hoeverre diegene bekend is met het medische begrip. Dit wil dus niet zeggen dat diegene te maken moet hebben gehad met dit begrip maar puur of diegene de betekenis weet.

3.3.5 Gezondheidsbetrokkenheid

Gezondheidsbetrokkenheid werd gemeten door middel van de 'Patient Health Engagement scale (PHE)' [35]. Deze vragenlijst bestond uit vijf stellingen over de gezondheidsbetrokkenheid. De stellingen uit deze vragenlijst waren voor elke vraag hetzelfde, namelijk: 'Als ik aan mijn ziekte denk dan, ...'. Deze stellingen worden beantwoord op een 7-punts Likertschaal die varieert per stelling uit de vragenlijst. Een voorbeeld van een Likertschaal is 'voel ik mij verdwaaasd, voelt het problematisch, ben ik bewust of voel ik mij rustig'. Tussen de antwoordmogelijkheden waren ook lege opties toegevoegd voor wanneer een patiënt niet kan kiezen tussen twee stellingen.

Om de validiteit van de PHE vragenlijst te beoordelen, werden de PHE-factorscores geëvalueerd in relatie tot de PAM- en MMAS-4-scores [35]. Met als resultaat ($p < 0,001$) hogere niveaus van betrokkenheid gaven een actievere houding van de patiënt weer. Een t-test werd uitgevoerd om de PHE factor scores te meten voor patiënten met een gemiddelde of hoge therapietrouw aan medicatie ($p < 0.01$), dus betrokken patiënten hebben een gemiddelde of hogere therapietrouw

3.3.6 Therapietrouwheid

De therapietrouwheid (MARS-scale) [44] van een patiënt werd gemeten aan de hand van vijf stellingen omtrent het medicatiegebruik van een patiënt. Onder deze stellingen vielen bijvoorbeeld het vergeten van de inname, het wijzigen van de dosis of het stoppen van een kuur. Deze stellingen worden beantwoord op een 5-punts Likertschaal die varieert van: altijd, vaak, soms, bijna nooit tot nooit.

De MARS-5 vragenlijst werd beoordeeld op zijn betrouwbaarheid en validiteit. Met als resultaat een Cronbach's $\alpha=0,89$ voor diabetespatiënten en validiteit $p < 0,05$ binnen de diabetes groep [36].

3.3.7 Chronische aandoening en voetverzorging

Dit gedeelte van de vragenlijst begon met algemene vragen over diabetes, de complicaties, de duur van de ziekte en het medicijn gebruik.

Daarna werden er vragen gesteld over de voetverzorging [37] van de patiënt. Hoe beter een patiënt scoort op dit onderdeel des te beter de voetverzorging. In dit onderdeel van de vragenlijst wordt er gevraagd naar de verzorging van de voeten en het dragen van de juiste materialen aan de voeten. De antwoordmogelijkheden in deze sectie zijn 'Ja' of 'Nee', een voorbeeld van een vraag is: 'Lukt het u om uw voeten goed af te drogen na het wassen of douchen?' De interne consistentie van de Foot_care domein werd beoordeeld met Cronbachs alfa (0,731) en de test-hertestbetrouwbaarheid ($r = 0,714$, $p = 0,001$) voor alle onderdelen binnen dit domein waren de testen in voorgaand onderzoek aanvaardbaar [45].

3.3.8 Pilot

De vragenlijst is getest doormiddel van een pilot, om een mogelijkheid te creëren om de vragenlijst te verbeteren maar ook om duidelijkheid van de vragen te testen bij de respondenten. Na deze pilot zijn een aantal vragen op een andere manier geformuleerd of zijn er een aantal aspecten toegevoegd in de vraagstelling. Zo bleek het hebben van een partner die kon helpen met de voetverzorging een erg belangrijk aspect bij bijvoorbeeld het wassen van de voeten of het knippen van de nagels. Dit werd dan gedaan door de partner en niet door de patiënt zelf of een specialist.

3.4 Procedure

Voor de deelname aan het onderzoek moesten de deelnemers een informatieformulier doorlezen en de voorwaarden ondertekenen (appendix 1: 'Informatiebrief deelname'). De vragenlijsten zijn volledig anoniem en de gegevens zullen vergrendeld worden opgeslagen. Alle voorwaarden voor de deelname aan het onderzoek zijn te lezen in het informatieformulier.

De vragenlijsten (appendix 2: 'Vragenlijst') werden ingevuld door de respondenten in het programma Qualtrics. De eerste tien respondenten binnen dit onderzoek zijn verworven bij 'podothérapie Segerink' te Deventer. Na de afspraken van de diabetespatiënten bij dit podotherapeutisch centrum is er in de wachtruimte de kans geboden om mee te werken aan dit onderzoek. Hierbij was een onderzoeker aanwezig om uitleg te geven over het onderzoek of ondersteuning te bieden bij technische mankementen. Tegelijkertijd kon de onderzoeker achterhalen wat de patiënten vonden van de vragenlijst door vragen te stellen en observaties te doen.

3.5 Analyse

Binnen dit onderzoek werd er gewerkt in Jupyter met de programmeertaal Python. De data analyse werd uitgevoerd in vier stappen, de stappen binnen dit onderzoek zijn als volgt:

1. Data cleaning
2. Exploratieve analyse/Silhouet-analyse
3. Cluster analyse
4. Significantie van de clusters

Deze 4 stappen zijn hieronder beschreven.

Bij de (1) data cleaning werden de outliers aangepast of verwijderd. De data kan zowel cijfermatig of in letters worden aangeleverd, voor de analyse is het van belang dat de data binnen een kolom dezelfde vorm heeft. Daarom werd numerieke en categorische data onderscheiden waarna de data in de juiste vorm werd gezet. Daarnaast werden vragenlijsten waarbij er minder dan 90% van de items werden ingevuld verwijderd. Van de totaal 71 vragen moest dus totaal 90% van de vragen zijn beantwoord om geïnccludeerd te worden (zie appendix 8: 'Missende data'). Aan dit criterium voldeden 24 respondenten. Het script dat gebruikt is voor de verwerking in dit proces is te vinden in appendix 4 ('Jupyter script data cleaning'), hierin valt stap voor stap te lezen welke stappen er zijn gevolgd.

Bij de (2) silhouet analyse werden de silhouet-scores berekend na de vorming van de clusters [46]. Een silhouet-score is de methode van interpretatie en validatie van consistentie binnen

clusters van gegevens. Het berekent de afstand tot het dichtstbijzijnde cluster voor elke individu, van al deze afstanden werd een gemiddelde genomen. Waarbij de waarde 1 perfect is en -1 het slechtste scenario. De waarde 1 geeft aan dat een individu perfect binnen het cluster past, een waarde van 0 geeft aan dat er overlappende clusters bestaan en bij een waarde van -1 is de individu bij een verkeerd cluster ingedeeld. Omdat het doel is een grotere populatie samen te vatten in kleine clusters zal dus een laag aantal clusters worden gekozen. Voor de bepaling van de methode is er vergeleken welke methode de meeste significante verschillen geeft in de vragen en domeinen van de opgestelde clusters. De methode met de meeste significante verschillen in de vragen en domeinen tussen de clusters zal worden gekozen bij de analyse.

Vervolgens (3) werd deze clusteranalyse uitgevoerd en de participanten ingedeeld in clusters. Twee methodes (appendix 6: 'Jupyter script 'Gower-Method'' & appendix 7: 'Jupyter script 'PCA') voor het onderscheiden van de persona's in dit onderzoek zijn Gower[47][48] en Principal Component Analysis (PCA) [49], deze methodes zijn gekozen omdat er in de vragenlijsten wordt gewerkt met zowel numerieke als categorische variabelen.

Ten slotte (4) werd er door middel van de Mann-Whitney U test getest of clusters significant verschilden. De significantie is getest op de losstaande vragen die niet aan een domein werden gekoppeld en aan de domeinen (Appendix 5: 'Verklaring categorieën en tot stand komen scores'). Een p-waarde van kleiner dan 0,05 geeft een significant verschil aan tussen beide groepen. Vervolgens werden de persona's beschreven aan de hand van de variabelen die significant van elkaar verschillen.

4. Resultaten

De resultaten zullen behandeld worden aan de hand van de vier stappen: 1) Data cleaning 2) Exploratieve analyse/Silhouet-analyse 3) clusteranalyse 4) significantie van de clusters.

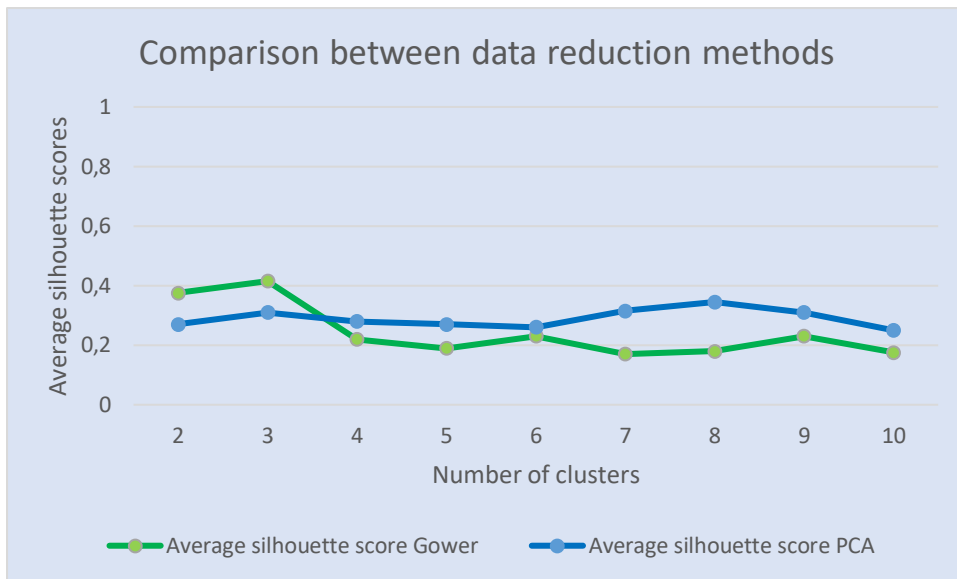
4.1 Data Cleaning

Tijdens het proces van de data cleaning wordt er bepaald welke respondenten bruikbaar zijn voor de resultaten. In totaal zijn er 43 personen verworven echter hadden negentien respondenten een voortgang van minder dan 90% in de vragenlijst waardoor ze zijn geëxcludeerd in dit onderzoek.

Voor de overige missende data werd gekeken of deze vragen verplichte vragen waren of dat vragen met bijvoorbeeld open tekstoptyes werden overgeslagen door de respondent. Mochten vragen door de respondenten niet beantwoord zijn, werd door de computer het mediaan van alle respondenten op die vraag berekend (zie appendix 4: 'Jupyter script 'Data cleaning'', IN[11]) en werd deze waarde ingevuld voor de respondent waar deze waarde nog miste. De missende data moest worden aangevuld om uiteindelijk de respondenten toe te kunnen wijzen aan een cluster.

4.2 Exploratieve analyse/Silhouet-analyse

In figuur 3 zijn de resultaten af te lezen van de silhouet-analyse. Er wordt gekeken naar de hoogste score en het daarbij horende aantal clusters. De Gower-methode geeft een silhouet-score van 0,375 bij twee clusters en een score van 0,415 bij drie clusters, de PCA-methode geeft voor het zelfde aantal clusters respectievelijk de scores 0,27 en 0,31. Bij het opstellen van de silhouet-scores kwam naar voren dat voor een populatie van 24 respondenten drie clusters te veel is, bij een analyse van drie clusters waren er altijd twee groepen die geen significante verschillen toonden van elkaar en dus erg overeen kwamen. Bij zowel de Gower- en PCA-methode ontstonden tussen de twee gegenereerde clusters wel significante verschillen. Bij de PCA-methode werden op meer domeinen significantie verschillen waargenomen dan bij de Gower-methode. Dit resulteert in kiezen voor de PCA-methode met twee clusters in het vervolg voor de persona creatie.



Figuur 3: Silhouet scores Gower en PCA

4.3 Clusteranalyse PCA

4.3.1 Tabel PCA

In tabel 4 (Appendix 9: 'Resultaten PCA') staan de gemiddelde antwoorden per vraag per cluster weergegeven, deze antwoorden zijn gegenereerd na het doen van de PCA (appendix 7: 'Jupyter script 'PCA)'). De vragenlijst is te lezen in appendix 2 ('Vragenlijst'), per vraag is ook de antwoordmogelijkheid te zien met de bijbehorende code. Om een voorbeeld te geven: als een respondent bij vraag Q1.3 het derde antwoord geeft (antwoordscore 3) wil zeggen dat de respondent op de vraag 'Wat is uw burgerlijke staat?' heeft geantwoord dat diegene getrouwd is. De opbouw van de scores en de betekenis van de scores is uitgelegd in appendix 5 ('Verklaring categorieën en tot stand komen scores').

4.3.2 Resultaten cluster 1

Binnen dit hoofdstuk worden de resultaten uit tabel 4 (Appendix 9: 'Resultaten PCA') per cluster behandeld. Cluster 1 wordt gekenmerkt door de gemiddelde leeftijd van 64 jaar, ze zijn vaak getrouwd (83%) en hebben over het algemeen een lager beroepsopleiding (4.0 (3.0; 4.0) gevolgd. De personen in cluster 1 hebben een BMI van 27,8. Deze groep personen heeft vaak steun nodig bij het gebruik van digitale hulpmiddelen, wanneer ze dit nodig hebben dan is er wel de mogelijkheid tot hulp.

Ze vinden gemakkelijk informatie over hun ziekte echter hebben ze significant minder kennis van de medische begrippen. Dagelijks zijn er gemiddeld 5,5 verschillende soorten medicatie nodig. Met een MARS-score van 24,5 (5=slecht, 25=perfect) hebben ze een nagenoeg perfecte therapietrouw. Dat deze groep goed gedrag vertoont rondom de gedragsmatigheid van de ziekte is ook terug te zien in het gedrag rondom de voetverzorging. Met een score van 9,0 uit 13 houden ze zich goed aan de regels die zijn opgelegd voor diabetes patiënten met een verhoogde kans op voetulcers en/of amputaties.

4.3.3 Resultaten cluster 2

De resultaten van cluster 2 staan beschreven in kolom drie van tabel 4 (Appendix 9: 'Resultaten PCA'). Dit cluster met een gemiddelde leeftijd van 47,5 jaar is een stuk jonger dan

het gemiddelde van de populatie. Binnen cluster 2 is diabetes type 1 vaker aanwezig dan diabetes type 2. Dit hoogopgeleide cluster is werkzaam en grotendeels nog vrijgezel of in een relatie. Dat dit cluster hoog opgeleid is vertaalt zich weer in een goede kennis van digitale hulpmiddelen, hoge kennis van medische termen en het vinden/begrijpen van de verkregen informatie omtrent de ziekte.

Een BMI van 24,3, geen verhoogde cholesterolgehalten samen met de beschrijving van een goede algemene gezondheid en tevredenheid over de kwaliteit van de fysieke gesteldheid maakt dat cluster 2 geen duidelijke gezondheidsproblemen heeft. Dit wordt ondersteund door het lage medicijngebruik van de groep.

Deze groep heeft een lage MARS-score(21/25) en lage Foot_care-score (7/13). De personen hebben te maken met weinig medicatievoorschriften waardoor ze zich vaak niet hoeven te houden aan strenge recepten. Voetcomplicaties van diabetes zijn vaak nog niet aanwezig, ze letten dan wel op hun voetverzorging maar niet in zo'n mate als de oudere groep waarbij hoogstwaarschijnlijk al wel voetproblemen zijn ontwikkeld.

4.4 Significantie

In de laatste kolom van tabel 4 (Appendix 9: 'Resultaten PCA') staat de 'p-value' oftewel de P-waarde, die de significantie aangeeft tussen de beide clusters, de significante p-waardes staan dikgedrukt in appendix 9 ('Resultaten PCA(tabel 4)').

Na de PCA zijn er twee clusters gegenereerd deze cluster zijn significant verschillend van elkaar op 13 van de 37 domeinen. Binnen het Leefstijl en socio-economisch domein verschillen de leeftijd ($p < 0,05$), burgerlijke staat ($p < 0,05$), opleidingsniveau ($p < 0,001$) en huidige werkstatus ($p < 0,05$) van elkaar. Daarnaast rapporteren de groepen significante verschillen in het cholesterolgehalte ($p < 0,05$), BMI ($p < 0,05$) en de staat van de algemene gezondheid ($p = 0,0336$). Ook verschillen de types diabetes ($p = 0,05$) van elkaar, cluster 1 heeft voornamelijk type 2 terwijl patiënten met type 1 diabetes een groot aandeel hebben in cluster 2. Het verschil in medicatie inname ($p < 0,001$) tussen de twee groepen is erg groot. Op de fysieke vragen ($p < 0,001$) binnen het domein kwaliteit leven is er een significant verschil te zien ten opzichte van de twee clusters, dit geldt ook voor de vragen over de kennis van medische begrippen ($p < 0,05$) binnen het domein gezondheidsvaardigheid. De beide gedragsmatige domeinen therapietrouw ($p < 0,05$) en voetverzorging ($p < 0,05$) is ook een duidelijk significant verschil te zien tussen beide clusters.

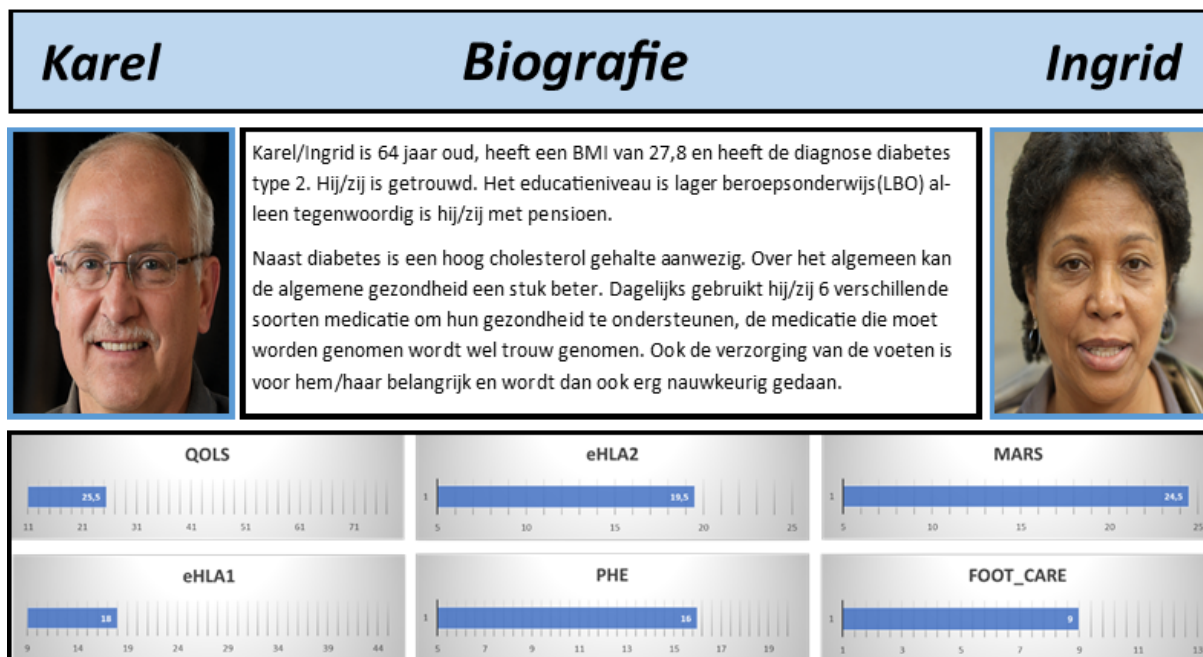
5. Discussie

In dit hoofdstuk zal de in de inleiding de gedefinieerde onderzoeksvraag worden beantwoord. Daarna zullen de gevonden resultaten worden gereflecteerd op de wetenschappelijke literatuur. Andere resultaten van dit onderzoek worden besproken, waarna wordt ingegaan op de mogelijke beperkingen van het onderzoek. Tot slot wordt een conclusie getrokken uit alle bevindingen en wordt advies gegeven over het gebruik van de gecreëerde persona's binnen de ontwikkeling van een Early Warning System.

5.1 Antwoord op de onderzoeksvraag

De vragenlijst bestond uit verschillende domeinen die gezamenlijk elk bepaalde kenmerken testen die karakteristiek kunnen zijn voor bepaalde clusters in een populatie. Het antwoord van de gedefinieerde onderzoeksvraag: 'Welke persona's zijn er te onderscheiden bij diabetespatiënten?', is weergegeven aan de hand van persona kaarten. De twee clusters zijn gevisualiseerd in figuur 4 en figuur 5 en hebben respectievelijk de namen Karel/Ingrid en Pim/Lotte gekregen. Het is dus mogelijk om aan de hand van kwantitatieve data clusters te genereren binnen een diabetespopulatie.

Cluster 1 wordt gepresenteerd door de persona kaart Karel/Ingrid (Figuur 4) en Cluster 2 door de persona kaart Pim/Lotte (Figuur 5). Karel/Ingrid is 64 jaar oud, heeft een BMI van 27,8 en heeft de diagnose diabetes type 2. De eigen gezondheid van Karel/Ingrid beschrijven ze als niet erg goed. Om de gezondheid op peil te houden is er relatief veel medicatie nodig. Pim/Lotte is 48 jaar oud, heeft een BMI van 24,3 en heeft de diagnose diabetes type 1. Ondanks de diagnose van deze chronische staat Pim/Lotte nog volop in het leven en begrijpt hij/zij veel van technologieën, medische begrippen en de omgang met een chronische ziekte. Naast deze verschillende eigenschappen verschillen de groepen erg in de therapietrouwheid, voetverzorging, en hun fysieke kwaliteit van leven. Zo heeft Karel/Ingrid een duidelijk hogere therapietrouw dan Pim/Lotte. Ook het gedrag rondom de voetverzorging is beduidend hoger voor Karel/Ingrid. Daarbij geeft de groep die gepresenteerd wordt door Karel/Ingrid aan dat zij over de kwaliteiten van hun fysieke gesteldheid meer ontevreden zijn dan Pim/Lotte. Beide groepen kunnen een potentieel eindgebruiker zijn van een Early Warning System, echter moet in de ontwikkeling wel rekening worden gehouden met de verschillen in kennis, gedrag en houding tussen beide groepen.



Figuur 4: Persona kaart 'Karel/Ingrid'



Figuur 5: Persona kaart 'Pim/Lotte'

5.2 Koppeling met de gevonden literatuur

De twee gecreëerde persona's komen overeen met eerder gevonden persona's binnen de diabetespopulatie volgens de literatuur [50], omdat uit de literatuur naar voren komt dat er onderscheid tussen de clusters kan worden gemaakt in de domeinen omtrent gedrag, kennis, gedrag ten opzichte van technologie, ziektebeeld en persoonsgegevens. In dit onderzoek is terug te zien dat de clusters op deze domeinen significant van elkaar verschillen. Bij de inzet van de creatie van persona's is het van belang om een duidelijk beeld te krijgen van wat het doel is van de technologie die ontwikkeld gaat worden en op basis van welke karakteristieken de persona's gebaseerd moeten worden [51]. Binnen dit onderzoek wordt er een

technologische applicatie ontwikkeld rondom het geven van een Early Warning System voor voetulcers en/of amputaties. Factoren gerelateerd aan de persona's binnen de ontwikkeling van een Early Warning System zijn daarom gezondheidsvaardigheid, therapietrouwheid en voetverzorging.

De clusters verschillen op een aantal belangrijke aspecten die ook in de ontwikkeling van een Early Warning System een rol kunnen spelen. Wanneer de persona's dienen als weergave van de eindgebruikers binnen de CeHRes roadmap [22], dus in het ontwikkel- en implementatieproces van nieuwe technologieën, kunnen er een aantal conclusies worden getrokken. Wanneer het Early Warning System een technologische applicatie wordt die de patiënten zelf moeten beheren moet er rekening worden gehouden met de technologische vaardigheden van de eindgebruiker. Persona 2 laat zien dat hij/zij zelfstandig goed om kan gaan met technologische hulpmiddelen. Dit is echter niet terug te zien bij persona 1 zij kunnen dit niet zelfstandig en hebben daarom hulp nodig bij het gebruik van technologische oplossingen. Bij de ontwikkeling van het Early Warning System moet er dus rekening gehouden worden met een groep die zich weet te redden met technologieën maar ook meteen groep die dit niet zelfstandig kan. In geval van een Early Warning kan persona 2 zelfstandig beslissen wat de vervolgstappen moeten zijn, is er een specialist nodig, kan ik dit zelf oplossen of hoe groot is het risico. Persona 1 heeft vaker hulp nodig en zal dus bij een early warning zelf niet goed weten wat hij/zij moet doen. Dit kan verwarring op paniek veroorzaken waardoor er moet worden nagedacht over wie de Early Warning krijgt. Bij persona 1 is het dus logischer om de waarschuwing via een arts of specialist te laten lopen of om voor de implementatie duidelijke voorlichting te geven over het gebruik van de toepassing [52].

Het Early Warning System is een preventieve interventie voor de ontwikkeling van voetulcers en/of amputaties. Voor persona 2 ligt dit risico bij de ontwikkeling van nieuwe voetulcers. Persona 2 geeft aan vaak nog geen voetzweren te hebben ontwikkeld. Daarentegen focust het Early Warning System zich niet alleen op het voorspellen van voetulcers maar ook op mogelijke amputaties. Binnen persona 1 hebben de personen vaker te maken gehad met voetulcers, dit is een voorspellende factor voor een hogere kans op amputaties [53] en het terugkeren van voetulcers. Verder zijn er tussen de clusters significante verschillen te zien op de zelfzorg domeinen: MARS en Foot_care. Persona 1 neemt zijn/haar medicatie veel trouwer en verzorgt zijn/haar voeten beter dan persona 2. Voor eventuele interventies wil dit zeggen dat het verbeteren van het gedrag binnen van persona 1 beter kan ten opzichte van de preventieve richtlijnen die in de literatuur zijn gesteld voor diabetespatiënten [10]. Tijdens de ontwikkeling van het Early Warning System moet er dus rekening gehouden worden met twee verschillende risicogroepen met elk hun eigen voorspellende factoren voor de ontwikkeling van voetulcers en/of amputaties.

Zowel de literatuur als dit onderzoek laten zien dat segmentatie van de eindgebruikers in de vorm van persona's waarde toevoegt in het ontwikkelproces van gezondheidstechnologieën [54]. De persona's zijn namelijk een hulpmiddel in het ontwikkelproces van technologieën, daarom moet er goed worden gekeken naar de eigenschappen van elk cluster om de

technologie zo goed mogelijk aan te kunnen laten sluiten op de eisen van de persona's om de technologie uiteindelijk te laten slagen.

5.3 Reflectie op de resultaten

5.3.1 Lege responses

Met het lage aantal respondenten, waarvan slechts 24 respondenten werden geselecteerd voor de analyse, werden er in dit onderzoek slechts twee persona's gecreëerd. Waardoor het hoge aantal respondenten die lege antwoorden gaven bij de vragenlijst is veroorzaakt is lastig te verklaren. De respondenten hebben de informatiebrief wel gelezen, dit blijkt uit dat ze akkoord gaven op deelname aan het onderzoek en de manier van de databehandeling. Echter, bij het doorklikken naar de echte vragenlijst werd er onmiddellijk gestopt waardoor de voortgang bij een grote groep bleef steken op slechts 2%. Dit kan verklaart worden door de moeilijkheidsgraad van de vraagstelling of gevoelens van kwetsbaarheid bij de vragen. Duidelijk is in elk geval dat er een bepaalde barrière is geweest voor specifieke groep mensen om door te gaan met het invullen van de vragenlijst. Voor dit onderzoek kan dit betekenen dat er een beeld is verkregen vanuit een groep mensen die meer bereid zijn voor deelname aan wetenschappelijk onderzoek. Deze groep kan een vertekend beeld geven ten opzichte van de gehele populatie omdat ze wellicht verschillen in eigenschappen, gedrag of ziektebeeld. Dit kan ook gevolgen hebben voor de interne validiteit van het onderzoek omdat mogelijk niet de eigenschappen van de gehele diabetespopulatie in kaart worden gebracht.

De resultaten tussen de clusters toonden significante verschillen op meerdere domeinen, echter kan een hoger aantal respondenten de significantie tussen de clusters verhogen. Voor een hoger aantal respondenten daalt hoogstwaarschijnlijk de p-waarde van vragen tussen de clusters. De validiteit van het onderzoek kan op dit aspect binnen het onderzoek nog worden verhoogd omdat momenteel slecht een klein deel van de populatie heeft deelgenomen aan het onderzoek. Het doel van clustering is de gehele diabetespopulatie samen vatten in clusters, echter bij een kleine onderzoekspopulatie is het onduidelijk of deze representatief is voor de gehele populatie.

Naast een hogere significantie kan het stijgen van het aantal respondenten ook een hoger aantal clusters tot gevolg hebben. Bij een grotere onderzoekspopulatie is het dus mogelijk om significante verschillen te ontdekken in meerdere clusters. Echter bij een te hoog aantal clusters wordt het doel van clustering voorbij geschoten. Het doel is namelijk een grote populatie samen te vatten in een klein aantal persona's. Een grotere populatie bevat namelijk veel meer karakteristieken dan weer te geven zijn in slechts twee clusters.

5.3.2 Waardering van vragen

In de bestaande vragenlijst krijgen alle vragen en domeinen dezelfde waarde in de generatie van clusters. Echter kan in vervolg onderzoek gekeken worden of er vragen/domeinen zijn die belangrijker zijn in de onderscheiding van de clusters. In de literatuur ('2.4.1 Factoren van persona's') kwam al naar voren dat de verzorging van de voeten en het dragen van de juiste materialen aan de voeten erg belangrijke factoren zijn in de ontwikkeling van voetulcers. In vervolgonderzoek kan aan dit domein een hogere waardering worden gehangen waardoor de

persona's uiteindelijk nog meer geselecteerd kunnen worden op bijvoorbeeld een hogere kans op voetulcers.

Uit de literatuur is ook gebleken dat de ontwikkeling van persona's een belangrijk onderdeel kunnen zijn van het ontwerpproces van nieuwe technologieën. Wanneer er uit onderzoek komt dat de technologische kennis of technologische vaardigheden belangrijk zijn in het ontwikkel proces dan kunnen vragen gerelateerd aan deze onderwerpen een hogere waardering krijgen waardoor je onderscheid krijgt in de clusters op dit relevante onderwerp.

5.3.3 Verschil in silhouet-scores

Zoals in de resultaten al vermeld staat de silhouet-score voor de afstand tussen een individu en het gemiddelde van een cluster. Met 1 als perfecte score en -1 als slechtste score. Er wordt dus gestreefd naar een hoge silhouet-score echter kan een hoge score niet altijd gelijk worden aangenomen als het beste aantal clusters.

Tijdens een onderzoek moet er bij de silhouet-score rekening gehouden worden met twee belangrijke voorwaarden voor het kiezen van een methode en het daarbij horende aantal clusters. Het eerste punt is dat een cluster wel uit een relevant aantal individuen moet bestaan, een cluster van één of twee individuen kan geen cluster worden genoemd. Het kan voorkomen dat een hoog aantal clusters ook een hoge silhouet-score krijgt, zie figuur 3 'Silhouet scores methodes'. In dit figuur krijg de PCA de hoogste silhouet-score bij een aantal van acht clusters, echter acht clusters bij een populatie van 24 respondenten schiet het doel van clustering voorbij. Het volgende punt waar op gelet moet worden bij de silhouet-analyse is dat een hogere silhouet-score bij een methode niet gelijk wil zeggen dat deze methode de beste methode is. Net als in dit onderzoek gaf de PCA met een lagere silhouet-score een hogere statistische relevantie tussen de clusters dan de Gower-methode.

5.4 Conclusie

Aanbevolen wordt om voor beide persona's een verschillende visie in te nemen als uitgangspunt bij de ontwikkeling van een Early Warning System. Persona 1 beschikt over meer risicofactoren doordat zij al vaker te maken hebben gehad met het krijgen van voetulcers. Persona 2 heeft daarentegen een hogere behoefte aan een preventief model, zij zijn vitaal en willen graag voorkomen dat zij in de toekomst te maken krijgen met voetulcers en/of amputaties.

Voor het vervolgonderzoek kunnen er meerdere clusters gecreëerd worden binnen een doelpopulatie. Bij een grotere populatie is het namelijk aannemelijk dat er meer meerdere significante verschillende karakteristieken op de vragen binnen de clusters ontstaan. Deze verschillende clusters hebben elk hun eigen eigenschappen die gebruikt kunnen worden in ontwikkelprocessen waarin de eindgebruiker een onderdeel speelt. Daarom wordt ook aanbevolen om de vragenlijst (appendix 2: 'Vragenlijst') verder te verspreiden om zo meer respondenten te verwerven.

Dit onderzoek kan voort worden gezet doormiddel van de Qualtrics vragenlijst en het het opgesteld Jupyter script (appendix 4: 'Jupyter script 'Data cleaning'', appendix 6: 'Jupyter script 'Gower-Method'' en appendix 7: 'Jupyter script 'PCA)'). In het script is het mogelijk om de verkregen data uit de vragenlijst (appendix 2: 'Vragenlijst') in te laden. Het script geeft

automatisch het aantal clusters met de daarbij horende silhouet-score en de gegevens per cluster.

Referenties

- [1] Reguliere zorgkosten stijgen weer na coronadip | Nieuwsbericht | Zorginstituut Nederland n.d. <https://www.zorginstituutnederland.nl/actueel/nieuws/2021/10/04/reguliere-zorgkosten-stijgen-weer-na-coronadip> (accessed March 14, 2022).
- [2] Heijmans M, Lemmens L, Otten W, Havers J. Zelfmanagement door mensen met chronische ziekten: kennissynthese van onderzoek en implementatie in Nederland. 2015.
- [3] Stijgende tekorten aan zorgpersoneel - ZonMw n.d. <https://www.zonmw.nl/nl/actueel/nieuws/detail/item/stijgende-tekorten-aan-zorgpersoneel/> (accessed June 21, 2022).
- [4] Sowah RA, Bampoe-Addo AA, Armoo SK, Saalia FK, Gatsi F, Sarkodie-Mensah B. Design and Development of Diabetes Management System Using Machine Learning 2020. doi:10.1155/2020/8870141.
- [5] M.C. Poortvliet, C.T.M. Schrijvers, C.A. Baan. Diabetes in Nederland. RIVM Rapp 260322001/2007 n.d. <https://rivm.openrepository.com/bitstream/handle/10029/16486/260322001.pdf?sequence=1&isAllowed=y> (accessed February 17, 2022).
- [6] Nielen M. Diabetes mellitus in Nederland. Prevalentie en incidentie: heden, verleden en toekomst. | Nivel 2020. <https://www.nivel.nl/nl/publicatie/diabetes-mellitus-nederland-prevalentie-en-incidentie-heden-verleden-en-toekomst> (accessed March 1, 2022).
- [7] Beulens JWJ, Yauw JS, Elders PJM, Feenstra T, Herings R, Slieker RC, Moons KGM, Nijpels G, van der Heijden AA. Prognostic models for predicting the risk of foot ulcer or amputation in people with type 2 diabetes: a systematic review and external validation study. *Diabetologia* 2021;64:1550–62. doi:10.1007/S00125-021-05448-W.
- [8] Heijmans M, Lemmens L, Otten W, Havers J. Zelfmanagement door mensen met chronische ziekten: kennissynthese van onderzoek en implementatie in Nederland. 2015.
- [9] Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. *Patient Educ Couns* 2002;48:177–87. doi:10.1016/S0738-3991(02)00032-0.
- [10] Association AD. Preventive Foot Care in Diabetes. *Diabetes Care* 2004;27:s63–4. doi:10.2337/DIACARE.27.2007.S63.
- [11] Greenwood DA, Gee PM, Fatkin KJ, Peoples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. *J Diabetes Sci Technol* 2017;11:1015–27. doi:10.1177/1932296817713506.
- [12] Powers MA, Bardsley JK, Cypress M, Funnell MM, Harms D, Hess-Fischl A, Hooks B, Isaacs D, Mandel ED, Maryniuk MD, Norton A, Rinker J, Siminerio LM, Uelman S. Diabetes Self-management Education and Support in Adults With Type 2 Diabetes: A Consensus Report of the American Diabetes Association, the Association of Diabetes Care & Education Specialists, the Academy of Nutrition and Dietetics, the American Academy of Family Physicians, the American Academy of PAs, the American Association of Nurse Practitioners, and the American Pharmacists Association. *Diabetes Care* 2020;43:1636–49. doi:10.2337/DCI20-0023.
- [13] Muralitharan S, Nelson W, Di S, McGillion M, Devereaux PJ, Barr NG, Petch J. Machine Learning–Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review. *J Med Internet Res* 2021;23(2):E25187 <https://www.jmir.org/2021/2/E25187> 2021;23:e25187. doi:10.2196/25187.

- [14] Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Personalized Medicine Review Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis n.d. doi:10.3390/jpm10020021.
- [15] Vogenberg FR. Predictive and Prognostic Models: Implications for Healthcare Decision-Making in a Modern Recession. *Am Heal Drug Benefits* 2009;2:218.
- [16] Beerlage-De Jong N, Van Gemert-Pijnen L, Wentzel J, Hendrix R, Siemons L. Technology to support integrated Antimicrobial Stewardship Programs: a user centered and stakeholder driven development approach n.d. doi:10.4081/idr.2017.6829.
- [17] Min N, Wei Y, Zheng Y, Li X. Advancement of prognostic models in breast cancer: a narrative review. *Gland Surg* 2021;10:2815. doi:10.21037/GS-21-441.
- [18] Gillies CE, Taylor DF, Cummings BC, Ansari S, Islim F, Kronick SL, Medlin RP, Ward KR. Demonstrating the consequences of learning missingness patterns in early warning systems for preventative health care: A novel simulation and solution. *J Biomed Inform* 2020;110:103528. doi:10.1016/J.JBI.2020.103528.
- [19] Karimi M, Tsiachristas A, Looman W, Stokes J, Galen M van, Rutten-van Mólken M. Bundled payments for chronic diseases increased health care expenditure in the Netherlands, especially for multimorbid patients. *Health Policy (New York)* 2021;125:751–9. doi:10.1016/J.HEALTHPOL.2021.04.004.
- [20] Pruitt J, Grudin J. *Personas: Practice and Theory* 2003.
- [21] Holden RJ, Kulanthaivel A, Purkayastha S, Goggins KM, Kirpalani S. Know thy eHealth user: Development of biopsychosocial personas from a study of older adults with heart failure. *Int J Med Inform* 2017;108:158–67. doi:10.1016/J.IJMEDINF.2017.10.006.
- [22] Van Velsen L, Van Gemert-Pijnen L, Nijland N, Beaujean D, Van Steenberghe J. *Personas: The Linking Pin in Holistic Design for eHealth* n.d.
- [23] LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. *Int J Med Inform* 2013;82:e251–68. doi:10.1016/J.IJMEDINF.2011.03.006.
- [24] Zhang Y, Li X, Luo S, Liu C, Xie Y, Guo J, Liu F, Zhou Z. Use, Perspectives, and Attitudes Regarding Diabetes Management Mobile Apps Among Diabetes Patients and Diabetologists in China: National Web-Based Survey. *JMIR MHealth UHealth* 2019;7. doi:10.2196/12658.
- [25] Van Velsen L, Wentzel J, Van Gemert-Pijnen JEW. Designing eHealth that Matters via a Multidisciplinary Requirements Development Approach. *JMIR Res Protoc* 2013;2(1)E21 <https://www.researchprotocols.org/2013/1/e21> 2013;2:e2547. doi:10.2196/RESPROT.2547.
- [26] Demiris G, Rantz MJ, Aud MA, Marek KD, Tyrer HW, Skubic M, Hussam AA. Older adults' attitudes towards and perceptions of "smart home" technologies: A pilot study. *Med Inform Internet Med* 2004;29:87–94. doi:10.1080/14639230410001684387.
- [27] Van Gemert-Pijnen JEW, Nijland N, Van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, Seydel ER. A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies. *J Med Internet Res* 2011;13(4)E111 <https://www.jmir.org/2011/4/e111> 2011;13:e1672. doi:10.2196/JMIR.1672.
- [28] Boslaugh SE, Kreuter MW, Nicholson RA, Naleid K. Comparing demographic, health status and psychosocial strategies of audience segmentation to promote physical activity n.d. doi:10.1093/her/cyg138.

- [29] Slater MD, Flora JA. Health lifestyles: audience segmentation analysis for public health interventions. *Health Educ Q* 1991;18:221–33. doi:10.1177/109019819101800207.
- [30] van Velsen L, van der Geest T, ter Hedde M, Derks W. Requirements engineering for e-Government services: A citizen-centric approach and case study. *Gov Inf Q* 2009;26:477–86. doi:10.1016/J.GIQ.2009.02.007.
- [31] Lerouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. *Int J Med Inform* 2013;82:e251–68. doi:10.1016/j.ijmedinf.2011.03.006.
- [32] Felce D, Perry J. Quality of life: Its definition and measurement. *Res Dev Disabil* 1995;16:51–74. doi:10.1016/0891-4222(94)00028-8.
- [33] Burckhardt CS, Anderson KL. The Quality of Life Scale (QOLS): Reliability, validity, and utilization. *Health Qual Life Outcomes* 2003;1. doi:10.1186/1477-7525-1-60.
- [34] Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing Competencies Needed to Engage With Digital Health Services: Development of the eHealth Literacy Assessment Toolkit. *J Med Internet Res* 2018;20. doi:10.2196/JMIR.8347.
- [35] Graffigna G, Barello S, Bonanomi A, Lozza E. Measuring patient engagement: Development and psychometric properties of the patient health engagement (PHE) scale. *Front Psychol* 2015;6:274. doi:10.3389/FPSYG.2015.00274/ABSTRACT.
- [36] Hai A, Chan Y, Horne R, Hankins M, Chisari C. The Medication Adherence Report Scale: A measurement tool for eliciting patients' reports of nonadherence adherence, MARS-5, Medication Adherence Report Scale, self-report, validation 2019. doi:10.1111/bcp.14193.
- [37] Navarro-Flores E, Morales-Asencio JM, Cervera-Marín JA, Labajos-Manzanares MT, Gijon-Nogueron G. Development, validation and psychometric analysis of the diabetic foot self-care questionnaire of the University of Malaga, Spain (DFSQ-UMA). *J Tissue Viability* 2015;24:24–34. doi:10.1016/J.JTV.2014.11.001.
- [38] Tauro E, Gorini A, Caglio C, Gabanelli P, Caiani EG. COVID-19 and mental disorders in healthcare Personnel: A novel framework to develop Personas from an online survey. *J Biomed Inform* 2022;126. doi:10.1016/J.JBI.2022.103993.
- [39] Emerson RW. Convenience Sampling, Random Sampling, and Snowball Sampling: How Does Sampling Affect the Validity of Research?: <https://Doi-OrgEzproxy2UtwenteNL/101177/0145482X1510900215> 2015;109:164–8. doi:10.1177/0145482X1510900215.
- [40] Diabetes in cijfers n.d. <https://www.diabetesfonds.nl/over-diabetes/diabetes-in-het-algemeen/diabetes-in-cijfers> (accessed June 21, 2022).
- [41] Ozolins U, Hale S, Cheng X, Hyatt A, Schofield P. Translation and back-translation methodology in health research – a critique. <https://DoiOrg/101080/1473716720201734453> 2020;20:69–77. doi:10.1080/14737167.2020.1734453.
- [42] Aan De Stegge WB, Abu-Hanna A, Bus SA. Development of a multivariable prediction model for plantar foot ulcer recurrence in high-risk people with diabetes. *BMJ Open Diab Res Care* 2020;8:1207. doi:10.1136/bmjdr-2020-001207.
- [43] Burckhardt CS, Woods SL, Schultz AA, Ziebarth DM. Quality of life of adults with chronic illness: a psychometric study. *Res Nurs Health* 1989;12:347–54. doi:10.1002/NUR.4770120604.

- [44] Fialko L, Garety PA, Kuipers E, Dunn G, Bebbington PE, Fowler D, Freeman D. A large-scale validation study of the Medication Adherence Rating Scale (MARS). *Schizophr Res* 2008;100:53–9. doi:10.1016/J.SCHRES.2007.10.029.
- [45] Antohe I, Popa AD, Antohe I, Popa AD. Ten Rules of Foot Care—The Development of a Short Questionnaire for Patients with Diabetes. *J Diabetes Mellit* 2021;11:83–96. doi:10.4236/JDM.2021.113007.
- [46] sklearn.metrics.silhouette_score — scikit-learn 1.1.1 documentation n.d. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html (accessed June 13, 2022).
- [47] Klooster I ten, Wentzel J, Sieverink F, Linssen G, Wesselink R, van Gemert-Pijnen L. Personas for Better Targeted eHealth Technologies: User-Centered Design Approach. *JMIR Hum Factors* 2022;9(1)E24172 <https://HumanfactorsJmirOrg/2022/1/E24172> 2022;9:e24172. doi:10.2196/24172.
- [48] What is K-Medoids Clustering and When should you use it instead of K-Means | by Ali Soleymani | Medium n.d. <https://medium.com/@ali.soleymani.co/beyond-scikit-learn-is-it-time-to-retire-k-means-and-use-this-method-instead-b8eb9ca9079a> (accessed June 15, 2022).
- [49] Bro R, Smilde AK. Principal component analysis. *Anal Methods* 2014;6:2812–31. doi:10.1039/C3AY41907J.
- [50] LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. *Int J Med Inform* 2013;82. doi:10.1016/J.IJMEDINF.2011.03.006.
- [51] Van Velsen L, Wentzel J, Ewc J, Gemert-Pijnen V. Designing eHealth that Matters via a Multidisciplinary Requirements Development Approach n.d. doi:10.2196/resprot.2547.
- [52] Vancea M, Solé-Casals J. Population Aging in the European Information Societies: Towards a Comprehensive Research Agenda in eHealth Innovations for Elderly. *Aging Dis* 2016;7:526. doi:10.14336/AD.2015.1214.
- [53] Beulens JWJ, Yauw JS, Elders PJM, Feenstra T, Herings R, Slieker RC, Moons KGM, Nijpels G, van der Heijden AA. Prognostic models for predicting the risk of foot ulcer or amputation in people with type 2 diabetes: a systematic review and external validation study. *Diabetologia* 2021;64:1550–62. doi:10.1007/S00125-021-05448-W.
- [54] Gulliksen J, Göransson B, Boivie I, Blomkvist S, Persson J, Cajander Å. Behaviour and Information Technology Key principles for user-centred systems design 2010. doi:10.1080/01449290310001624329.

Appendices

Appendix 1: Informatiebrief deelname

Q0.1 Geachte meneer/mevrouw,

U wordt vriendelijk uitgenodigd om deel te nemen aan een vragenlijst als onderdeel van het Power4FitFoot onderzoek. Voordat u beslist of u deelneemt, is het belangrijk om te begrijpen waarom het onderzoek wordt uitgevoerd en wat het onderzoek inhoudt. Neem de tijd om de volgende informatie zorgvuldig door te lezen voordat u beslist of u deelneemt aan het onderzoek en bespreek dit desgewenst met anderen.

OVER HET ONDERZOEK

Dit onderzoek wordt gedaan door Rogier Krom onder begeleiding van Iris ten Klooster van de Faculteit Behavioural, Management and Social Sciences van de Universiteit Twente. Het doel van dit onderzoek is het creëren van persona's binnen een diabetespopulatie voor het ontwikkelen van een Early Warning System van een diabetische voet. Persona's zijn een fictieve weergave van bepaalde groepen binnen de grotere groep van diabetes patiënten. Het onderzoek is goedgekeurd door de ethische commissie van de Universiteit Twente (BMS). Alle gegevens worden geanonimiseerd en kunnen daarna niet aan u als patiënt gekoppeld worden. Alle gegevens die worden verzameld zullen volledig anoniem in een beveiligde omgeving worden opgeslagen. De gegevens zullen enkel en alleen voor de doeleinden van het Power4FitFoot onderzoek worden gebruikt.

Het invullen van de onderstaande vragenlijst zal ongeveer 10 tot 15 minuten in beslag nemen. Als u vragen heeft over het onderzoek, neem dan contact op met de onderzoeker: Rogier Krom (r.f.h.krom@student.utwente.nl) of bekijk de website (<https://www.power4fitfoot.nl/>). Mocht u niet akkoord gaan dan wordt u gelijk doorverwezen naar het einde van deze vragenlijst.

Alvast bedankt!

	Ja, akkoord (1)	Nee, niet akkoord (2)
Ik heb de bovenstaande informatie gelezen en ga akkoord met de deelname aan het onderzoek (1)	<input type="radio"/>	<input type="radio"/>
Ik ga akkoord met de manier waarop de verzamelde data wordt gebruikt en wordt opgeslagen (2)	<input type="radio"/>	<input type="radio"/>

Power4FitFoot: Personas binnen een diabetes populatie

Start of Block: Informatie onderzoek Power4FitFoot

Q1.1 Wat is uw geslacht?

- Man (1)
 - Vrouw (2)
 - Anders (3)
-

Q1.2 Hoe oud bent u? (schrijf uw leeftijd op als een getal)

Q1.3 Wat is uw burgerlijke staat?

- Vrijgezel (1)
 - In een relatie (2)
 - Getrouwd (3)
 - Gescheiden of uit elkaar (4)
 - Weduwe (5)
 - Anders (6)
-

Q1.4 Wat is uw hoogst behaalde opleidingsniveau?

- Geen formeel onderwijs (1)
 - Basisonderwijs (2)
 - Lager beroepsonderwijs (LBO) (3)
 - Middelbaar beroepsonderwijs (MBO) (4)
 - Mavo (5)
 - Havo (6)
 - Vwo / Gymnasium (7)
 - Hoger beroepsonderwijs (Hbo) (8)
 - Wetenschappelijk onderwijs (Wo) (9)
 - Anders, namelijk... (10) _____
-

Q1.5 Wat is uw huidige werksituatie?

- Werkloos (1)
 - Gepensioneerd (2)
 - Vrijwilliger (3)
 - Werkzaam (4)
-

Q1.6 Wat is uw postcode? (1234 _ AB)

Q1.7 Heeft u steun nodig bij het gebruik van digitale hulpmiddelen?

- Nooit (1)
 - Zelden (2)
 - Soms (3)
 - Vaak (4)
 - Altijd (5)
-

Q1.8 Krijgt u de steun die u nodig heeft bij het gebruik van digitale hulpmiddelen?

- Nooit (1)
- Zelden (2)
- Soms (3)
- Vaak (4)
- Altijd (5)

End of Block: 1. Leefstijl en socio-economisch

Start of Block: 2. Zelfgerapporteerde gezondheid

Q2.1 Noteer uw geschatte lengte (in cm).

Q2.2 Gelieve uw gewicht bij benadering op te geven (in kg).

Q2.3 Heeft u een hoog cholesterolgehalte?

- Ja (1)
- Nee (2)
- Weet ik niet (3)

Q2.4 Wat is uw gedrag met betrekking tot roken?

- Nooit gerookt (1)
- Ex-roker (2)
- Roker, 20 of minder sigaretten per dag (3)
- Roker, 20 of meer sigaretten per dag (4)

Display This Question:

If Q2.4 = 1

Q2.5 Bent u een niet-roker die de meeste dagen thuis of op het werk aan rook wordt blootgesteld?

- Ja (1)
- Nee (2)

Q2.6 Wat is uw alcoholconsumptie?

- 0 consumpties per week (1)
- 1-7 consumpties per week (2)
- 8-14 consumpties per week (3)
- 15 of meer consumpties per week (4)

Display This Question:

If Q1.1 = 1

Q2.7.1 Drinkt u als man 5 of meer drankjes in één keer op een tweewekelijkse of meer frequente basis?

- Ja (1)
 - Nee (2)
-

Display This Question:

If Q1.1 = 2

Q2.7.2 Drinkt u als vrouw 3 of meer drankjes in één keer op een tweewekelijkse of meer frequente basis?

- Ja (1)
 - Nee (2)
-

Q2.8 Hoe vaak doet u aan lichaamsbeweging?

- Minder dan een keer per week (1)
 - Gemiddeld één keer per week (2)
 - Gemiddeld 2 tot 3 keer per week (3)
 - Gemiddeld 4 tot 5 keer per week (4)
 - Gemiddeld 5 keer of vaker per week (5)
-

Q2.9 Hoe zou u uw gezondheid over het algemeen beschrijven?

- Slecht (1)
- Redelijk (2)
- Goed (3)
- Heel goed (4)
- Uitstekend (5)

End of Block: 2. Zelfgerapporteerde gezondheid

Start of Block: 3. Kwaliteit van leven

Q3.1

Lees elk item en selecteer het antwoord die het beste beschrijft hoe u zich op dit moment voelt over de gegeven items. Als een bepaalde situatie of relatie niet voor u van toepassing is, vul dan in of u er tevreden of ontevreden over bent dat u deze relatie niet hebt of dat deze situatie niet bij u van toepassing is.

Hoe voelt u zich momenteel over,

	Blij (1)	Tevreden (2)	Meestal voldaan (3)	Gemengd (4)	Meestal onvoldaan (5)	Ontevreden (6)	Ongelukkig (7)
materiële gemakken thuis, voedsel, gemakken, financiële zekerheid (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
uw gezondheid - lichamelijk fit en vitaal zijn (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
uw relaties met ouders, broers en zussen & andere - communiceren, bezoeken, helpen (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
kinderen krijgen en opvoeden (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
uw nauwe relaties (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
uw goede vrienden (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
leren - naar school gaan, het verbeteren van begrip, extra kennis krijgen (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
jezelf begrijpen - je mogelijkheden kennen en beperkingen - weten waar het leven om draait (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
uw werk - baan of thuis situatie (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
jezelf creatief uitdrukken (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
uw onafhankelijkheid (bv. in staat zijn om dingen alleen te doen) (11)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: 3. Kwaliteit van leven

Start of Block: 4. Health literacy

Q4.1 Op een schaal van zeer gemakkelijk tot zeer moeilijk, hoe gemakkelijk vindt u het om:

	Zeer gemakkelijk (1)	Gemakkelijk (2)	Normaal (3)	Moeilijk (4)	Zeer moeilijk (5)
informatie te vinden over de behandeling van ziekten waarover u meer wilt weten? (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
uit te zoeken waar u professionele hulp kunt krijgen als u ziek bent? (bijvoorbeeld door een arts, bij de apotheek, door een psycholoog) (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
te begrijpen wat de arts tegen u zegt? (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
de instructies van uw arts of apotheek op te volgen? (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
informatie te vinden over wat u kunt doen aan geestelijke gezondheidsproblemen zoals stress en depressie? (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
op basis van informatie in de media beslissen hoe u zich wilt beschermen tegen ziekte? (bijv. kranten, brochures, internet of andere media) (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
informatie te vinden over gewoonten die goed zijn voor je mentale welzijn? (bijv. meditatie, lichaamsbeweging, wandelingen, enz.) (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
begrijpelijk gezondheidsadvies van familie of vrienden te krijgen? (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
te beoordelen hoe uw dagelijkse gewoonten uw gezondheid beïnvloeden? (bijv. drink- en eetgewoonten, lichaamsbeweging, enz.) (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4.2 Op een schaal van helemaal niet bekend tot zeer bekend, hoe bekend bent u met de volgende medische termen:

	1. Helemaal niet bekend (1)	2. Enigszins bekend (2)	3. Matig bekend (3)	4. Bekend (4)	5. Zeer goed bekend (5)
Revalidatie (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De milt (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ziekenhuisopnames (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Een ademhalingsmasker (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Autisme (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: 4. Health literacy

Start of Block: 5. Gezondheidsbetrokkenheid

Q5.1 Kies de zin die u het beste weergeeft. Als u niet kunt kiezen tussen twee zinnen, kunt u de lege optie ertussen kiezen.

	voel ik een black-out (1)	(2)	raak ik in paniek (3)	(4)	ben ik bewust (5)	(6)	ben ik positief (7)
Als ik aan mijn ziekte denk dan, (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q5.2 Kies de zin die u het beste weergeeft. Als u niet kunt kiezen tussen twee zinnen, kunt u de lege optie ertussen kiezen.

	voel ik mij verdwaasd (1)	(2)	voelt het problematisch (3)	(4)	ben ik mij bewust (5)	(6)	voel ik mij rustig (7)
Als ik aan mijn ziekte denk dan, (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q5.3 Kies de zin die u het beste weergeeft. Als u niet kunt kiezen tussen twee zinnen, kunt u de lege optie ertussen kiezen.

	voel ik me overweldigd door emoties (1)	(2)	voel ik mij angstig elke keer als er een nieuw symptoom opduikt (3)	(4)	ben ik gewend geraakt aan mijn ziekte/conditie (5)	(6)	ervaar ik ondanks mijn ziekte samenhang en continuïteit in mijn leven (7)
Als ik aan mijn ziekte denk dan, (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q5.4 Kies de zin die u het beste weergeeft. Als u niet kunt kiezen tussen twee zinnen, kunt u de lege optie ertussen kiezen.

	ben ik erg ontmoedigd door mijn ziekte (1)	(2)	voel ik me angstig als ik mijn ziekte probeer te beheersen (3)	(4)	voel ik mijn aangepast aan mijn ziekte (5)	(6)	ben ik over het algemeen positief over de toekomst en gezondheidsconditie (7)
Als ik aan mijn ziekte denk dan, (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q5.5 Kies de zin die u het beste weergeeft. Als u niet kunt kiezen tussen twee zinnen, kunt u de lege optie ertussen kiezen.

	voel ik mij totaal onderdrukt door mijn ziekte (1)	(2)	ben ik overstuur wanneer ik te maken krijg met nieuwe symptomen (3)	(4)	heb ik mijn ziekte geaccepteerd (5)	(6)	kan ik ondanks mijn ziekte zin gaan geven aan het leven (7)
Als ik aan mijn ziekte denk dan, (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: 5. Gezondheidsbetrokkenheid

Start of Block: 6. Therapietrouwheid

Q6.1 Veel mensen vinden een manier om hun medicijnen te gebruiken die bij hen past. Dit kan afwijken van de instructies op het etiket of van wat hun arts had gezegd. Hier zijn enkele manieren waarop

mensen hebben gezegd dat ze hun medicijnen gebruiken. Kruis voor elke verklaring het vakje aan dat het beste op u van toepassing is.

	Altijd (1)	Vaak (2)	Soms (3)	Bijna nooit (4)	Nooit (5)
Ik vergeet mijn medicijnen in te nemen (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik wijzig de dosis van mijn medicijnen (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik stop met de inname van mijn medicijnen voor een bepaalde tijd (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik beslis om een dosis van mijn medicijnen over te slaan (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik neem minder medicijnen dan staan voorgeschreven (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: 6. Therapietrouwheid

Start of Block: 7. Chronische aandoening

Q7.1 Welke chronische aandoening heeft u? (meer dan één kan van toepassing zijn).

- Cardiovasculaire aandoening (1)
 - Oncologische aandoening (2)
 - Aandoening aan de luchtwegen (3)
 - Diabetes (4)
 - Neurologische aandoening (5)
 - Aandoening aan het immunitaire systeem (6)
 - Anders, namelijk.. (7) _____
-

Q7.2 Hoe lang geleden werd bij u een chronische aandoening vastgesteld?

- Minder dan 1 jaar geleden (1)
 - Tussen 1 en 5 jaar geleden (2)
 - Tussen 6 en 10 jaar geleden (3)
 - Meer dan 10 jaar geleden (4)
-

Q7.3 Hoeveel verschillende geneesmiddelen gebruikt u per dag?

Q7.4 Welk type diabetes heeft u?

- Type 1 (1)
 - Type 2 (2)
 - Geen idee (3)
-

Q7.5 Heeft u voorheen of momenteel last van een of meerdere van de volgende complicaties?

- Retinopathie / oogafwijking (1)
 - Diabetische voetulcus / voetzweer (2)
 - Nefropathie / nierschade (3)
 - Perifeer vaatlijden / vernauwing in de doorbloeding van de benen (4)
 - Neuropathie / zenuwaandoening (5)
-

Q7.6 Inspecteert u uw voeten elke dag?

- Ja (1)
 - Nee (2)
-

Q7.7 Wast u elke dag uw voeten met water op kamertemperatuur (handwarm water, het water dient ten alle tijden kouder te zijn dan 37 graden)?

Ja (1)

Nee (2)

Q7.8 Lukt het u om uw voeten goed af te drogen na het wassen of douchen?

Ja (1)

Nee (2)

Q7.9 Controleert u de ruimtes tussen uw tenen?

Ja (1)

Nee (2)

Q7.10 Gebruikt u hydraterende crème voor droge huid van uw voeten?

Ja (1)

Nee (2)

Q7.11 Knipt u uw teennagels recht af?

Ja (1)

Nee (2)

Q7.12 Controleert u de binnenkant van uw schoeisel?

Ja (1)

Nee (2)

Q7.13 Loopt u soms op blote voeten?

Ja (1)

Nee (2)

Q7.14 Loopt u wel eens op sokken zonder het dragen van schoenen?

Ja (1)

Nee (2)

Q7.15 Stelt u uw voeten wel eens bloot aan warme objecten, waardoor uw voeten kunnen verbranden?

Ja (1)

Nee (2)

Q7.16 Vraagt u om een controle bij een specialist als er een blaar, snijwond of zweer ontstaat?

Ja (1)

Nee (2)

Q7.17 Lukt het u om de juiste schoenen aan te schaffen?

Ja (1)

Nee (2)

Q7.18 Heeft u een partner of dicht bijstaande die u kan helpen met de verzorging van uw voeten?

Ja (1)

Nee (2)

End of Block: 7. Chronische aandoening

Start of Block: 8. Meewerken aan interview

Q8.1 Voor het vervolg van dit onderzoek zijn wij, de onderzoekers van Universiteit Twente, op zoek naar diabetes patiënten die mee willen werken aan een interview. Tijdens dit interview wordt er onderzocht wat de behoeften zijn met betrekking tot een Early Warning System voor een diabetische voet onder diabetes patiënten.

Bent u bereid om mee te werken aan het vervolgonderzoek en beschikbaar voor een interview?

Ja (1)

Nee (2)

Display This Question:

If Q8.1 = 1

Q8.2 Wat is uw email-adres waar de onderzoeker u op kan bereiken voor het plannen van een interview?

End of Block: 8. Meewerken aan interview

Appendix 3: Uitnodigingsbrief social media

Beste meneer/mevrouw,

Bent u of kent u iemand met diabetes? Dan kunt u mij heel erg helpen! Voor mijn bachelorscriptie doe ik binnen het Power4FitFoot-project onderzoek naar kenmerken van diabetespatiënten. Binnen dit project wordt er een vroegtijdig waarschuwingssysteem voor complicaties (voetzweren en amputaties) ontwikkeld. Door kenmerken van diabetespatiënten te onderzoeken zorgen wij ervoor dat deze goed aansluit bij de doelgroep.

Wilt u input leveren aan het Power4FitFoot onderzoek?

Wij vragen u om een vragenlijst in te vullen door te klikken op de volgende link: https://utwentebbs.eu.qualtrics.com/jfe/form/SV_bP1EcRcN9wm5QCq Het invullen van de vragenlijst duurt ongeveer 10-15 minuten.

Wie kunnen de vragenlijst invullen?

U kunt de vragenlijst invullen als u:

- ouder dan 18 jaar bent en,
- diabetes type 1 of type 2 hebt.

Voor meer informatie over dit onderzoek, kijk gerust op <https://www.power4fitfoot.nl/> de website van Power4FitFoot. Delen van dit bericht zou ik zeker waarderen!

Appendix 4: Jupyter script 'Data cleaning'

```
IN [1]
import pandas as pd
import warnings
import math
from sklearn.preprocessing import OneHotEncoder
import numpy as np
from scipy import stats

IN [2]
df = pd.read_csv('Data004.csv')

df = df.drop([0, 1], axis=0)
df = df.drop(['StartDate', 'EndDate', 'Status', 'IPAddress', 'Progress',
             'Duration (in seconds)', 'Finished', 'RecordedDate', 'ResponseId',
             'RecipientLastName', 'RecipientFirstName', 'RecipientEmail',
             'ExternalReference', 'LocationLatitude', 'LocationLongitude',
             'DistributionChannel', 'UserLanguage', 'Q0.1_1', 'Q0.1_2',
             'Q1.4_10_TEXT', 'Q7.1_7_TEXT', 'Q8.1', 'Q8.2'], axis=1)

df = df.reset_index(drop=True)

IN [3]
def convert_to_number(x):
    try:
        x = float(x)
    except:
        x = int(''.join(c for c in x if c.isdigit()))

    if math.isnan(x):
        x = 0

    return x

IN [4]
def one_hot_encoding_fixed(df, n=0, prefix=None):

    col_names = [f"{prefix}{'.'}{level+1}" for level in range(n)]
    x = pd.DataFrame(0, index=np.arange(len(df)), columns = col_names)

    for row in range(df.shape[0]):
        temp = df.iloc[row]
        try:
            math.isnan(temp)
            continue
        except:
            temp = temp.split(',')
            for i in temp:
                i = int(i)
                x.iloc[row, i-1] = 1

    return x

IN [5]
warnmsg = "Issue in row {} at question {}, please perform manual checking."
```

```

for row in range(df.shape[0]):

    #Check Q2.1 and Q2.2 (height and weight). They have to be in the correct
    range and in numeric form.

    #Q2.1
    temp = df.loc[row, 'Q2.1']
    #Check if number
    temp = convert_to_number(temp)
    if temp < 3:
        temp = temp*100
    #If height in weird range, raise warning
    if (temp < 50 or temp > 250):
        warnings.warn(warnmsg.format(row, 'Q2.1'))

    df.loc[row, 'Q2.1'] = int(temp)

    #Q2.2
    temp = df.loc[row, 'Q2.2']
    #Check if number
    temp = convert_to_number(temp)
    #If weight range wrid, raise warning
    if (temp < 40 or temp > 200):
        warnings.warn(warnmsg.format(row, 'Q2.2'))

    df.loc[row, 'Q2.2'] = int(temp)

    #Q2.5
    temp = float(df.loc[row, 'Q2.5'])
    if math.isnan(temp):
        temp = -99
    df.loc[row, 'Q2.5'] = int(temp)

    #Q7.3
    temp = df.loc[row, 'Q7.3']
    #Check if number
    temp = convert_to_number(temp)
    #If weird range raise warning
    if (temp <= 0 or temp > 20):
        warnings.warn(warnmsg.format(row, 'Q7.3'))
    df.loc[row, 'Q7.3'] = int(temp)

IN [6]
#Q7.1
q7_1 = one_hot_encoding_fixed(df.loc[:, 'Q7.1'], n=7, prefix="Q7.1")
df = pd.concat([df, q7_1], axis=1)
df = df.drop('Q7.1', axis=1)

#Q7.5
q7_5 = one_hot_encoding_fixed(df.loc[:, 'Q7.5'], n=5, prefix="Q7.5")
df = pd.concat([df, q7_5], axis=1)
df = df.drop('Q7.5', axis=1)

```



```

### TEMPORARY ###
df = df.drop('Q1.6', axis=1)

IN [7]
#Removed rows 6/20/24 due to high amount of missing val
df = df.drop([6,20,24,27,28,30,31,34,35,36], axis=0)
df = df.reset_index(drop=True)

IN [8]
#Q2.7.1 and Q2.7.2
q271 = df.loc[:, 'Q2.7.1']
q272 = df.loc[:, 'Q2.7.2']
q27 = []
for i in range(len(q271)):
    if math.isnan(float(q271[i])):
        q27.append(float(q272[i]))
    else:
        q27.append(float(q271[i]))

q27 = pd.DataFrame(q27, columns=["Q2.7"])
df = pd.concat([df, q27], axis=1)
df = df.drop(['Q2.7.1', 'Q2.7.2'], axis=1)

#BMI (Q2.1 + Q2.2)
h = df.loc[:, 'Q2.1']
w = df.loc[:, 'Q2.2']
BMI = []
for i in range(len(h)):
    temp = w[i] / (h[i]/100 * h[i]/100)
    BMI.append(temp)

BMI = pd.DataFrame(BMI, columns=["BMI"])
df = pd.concat([df, BMI], axis=1)

IN [9]
### MISSING DATA HANDLING ###
for row in range(df.shape[0]):
    print(df.iloc[row,:].isnull().sum())

IN [10]
type_dict = {'Q1.1' : 'object',
             'Q1.2' : 'float',
             'Q1.3' : 'object',
             'Q1.4' : 'float',
             'Q1.5' : 'object',
             'Q1.7' : 'float',
             'Q1.8' : 'float',
             'Q2.3' : 'object',
             'Q2.4' : 'float',
             'Q2.5' : 'object',
             'Q2.6' : 'float',
             'Q2.8' : 'float',
             'Q2.9' : 'float',
             'Q3.1_1' : 'float',
             'Q3.1_2' : 'float',
             'Q3.1_3' : 'float',
             'Q3.1_4' : 'float',

```

'Q3.1_5' : 'float',
'Q3.1_6' : 'float',
'Q3.1_7' : 'float',
'Q3.1_8' : 'float',
'Q3.1_9' : 'float',
'Q3.1_10' : 'float',
'Q3.1_11' : 'float',
'Q4.1_1' : 'float',
'Q4.1_2' : 'float',
'Q4.1_3' : 'float',
'Q4.1_4' : 'float',
'Q4.1_5' : 'float',
'Q4.1_6' : 'float',
'Q4.1_7' : 'float',
'Q4.1_8' : 'float',
'Q4.1_9' : 'float',
'Q4.2_1' : 'float',
'Q4.2_2' : 'float',
'Q4.2_3' : 'float',
'Q4.2_4' : 'float',
'Q4.2_5' : 'float',
'Q5.1_1' : 'float',
'Q5.2_1' : 'float',
'Q5.3_1' : 'float',
'Q5.4_1' : 'float',
'Q5.5_1' : 'float',
'Q6.1_1' : 'float',
'Q6.1_2' : 'float',
'Q6.1_3' : 'float',
'Q6.1_4' : 'float',
'Q6.1_5' : 'float',
'Q7.2' : 'float',
'Q7.3' : 'float',
'Q7.4' : 'object',
'Q7.6' : 'object',
'Q7.7' : 'object',
'Q7.8' : 'object',
'Q7.9' : 'object',
'Q7.10' : 'object',
'Q7.11' : 'object',
'Q7.12' : 'object',
'Q7.13' : 'object',
'Q7.14' : 'object',
'Q7.15' : 'object',
'Q7.16' : 'object',
'Q7.17' : 'object',
'Q7.18' : 'object',
'Q7.1.1' : 'object',
'Q7.1.2' : 'object',
'Q7.1.3' : 'object',
'Q7.1.4' : 'object',
'Q7.1.5' : 'object',
'Q7.1.6' : 'object',
'Q7.1.7' : 'object',
'Q7.5.1' : 'object',
'Q7.5.2' : 'object',
'Q7.5.3' : 'object',

```

        'Q7.5.4' : 'object',
        'Q7.5.5' : 'object',
        'Q2.7' : 'object',
        'BMI' : 'float'
    }

df = df.astype(type_dict)

IN [11]
for column in df.columns:
    feature = df.loc[:,column]
    if feature.isnull().sum() > 0:

        # Numerical variables
        if type_dict[column] == 'float':
            median_val = np.nanmedian(feature.astype(float), axis=0)
            for i in range(len(feature)):
                if pd.isnull(df.loc[i,column]):
                    df.loc[i,column] = median_val

        # Categorical variables
        else:
            mode_val = stats.mode(feature)
            for i in range(len(feature)):
                if pd.isnull(df.loc[i,column]):
                    df.loc[i,column] = int(mode_val.mode[0])

#Q3.1_4
#Q3.1_7
#Q5.1_1
#Q5.3_1
#Q5.4_1
#Q5.5_1
#Q7.16

```

```

IN [12]
# CALCULATION OF INDEXES
QOLS_phys = pd.DataFrame(df.loc[:,['Q3.1_1', 'Q3.1_2',
'Q3.1_11',]].sum(axis=1), columns=["QOLS_phys"])
QOLS_relations = pd.DataFrame(df.loc[:,['Q3.1_3', 'Q3.1_4', 'Q3.1_5',
'Q3.1_6']].sum(axis=1), columns=["QOLS_rel"])
QOLS_personal = pd.DataFrame(df.loc[:,['Q3.1_7', 'Q3.1_8', 'Q3.1_9',
'Q3.1_10']].sum(axis=1), columns=["QOLS_pers"])
eHLA_2 = pd.DataFrame(df.loc[:,['Q4.1_1', 'Q4.1_2', 'Q4.1_3', 'Q4.1_4',
'Q4.1_5', 'Q4.1_6', 'Q4.1_7', 'Q4.1_8', 'Q4.1_9']].sum(axis=1),
columns=["eHLA_2"])
eHLA_3 = pd.DataFrame(df.loc[:,['Q4.2_1', 'Q4.2_2', 'Q4.2_3', 'Q4.2_4',
'Q4.2_5',]].sum(axis=1), columns=["eHLA_3"])
MARS = pd.DataFrame(df.loc[:,['Q6.1_1', 'Q6.1_2', 'Q6.1_3', 'Q6.1_4',
'Q6.1_5',]].sum(axis=1), columns=["MARS"])
FOOT_CARE = pd.DataFrame(0, index=np.arange(df.shape[0]),
columns=["FOOT_CARE"])
QOLS = pd.DataFrame(pd.concat([QOLS_phys, QOLS_relations, QOLS_personal],
axis=1).sum(axis=1), columns=["QOLS"])

for i in range(df.shape[0]):

```

```

temp = df.loc[i, ['Q7.6', 'Q7.7', 'Q7.8', 'Q7.9', 'Q7.10', 'Q7.11',
'Q7.12', 'Q7.13', 'Q7.14', 'Q7.15', 'Q7.16', 'Q7.17', 'Q7.18']]
valsum = 0
for j in range(len(temp)):
    if j==7 or j==8 or j==9:
        if temp[j]=='2':
            valsum = valsum + 1
    else:
        if temp[j]=='1':
            valsum = valsum + 1
    FOOT_CARE.iloc[i] = valsum

PHE = pd.DataFrame(0, index=np.arange(df.shape[0]), columns=["PHE"])
for i in range(df.shape[0]):
    temp = df.loc[i, ['Q5.1_1', 'Q5.2_1', 'Q5.3_1', 'Q5.4_1', 'Q5.5_1']]
    valsum = 0
    for j in range(len(temp)):
        t = math.ceil(temp[j]/2)
        valsum = valsum + t
    PHE.iloc[i] = valsum

# ADDING INDEXES
df = pd.concat([df, QOLS_phys, QOLS_relations, QOLS_personal, QOLS, eHLA_2,
eHLA_3, MARS, PHE, FOOT_CARE], axis=1)
df = df.drop(['Q3.1_1', 'Q3.1_2', 'Q3.1_11', 'Q3.1_3', 'Q3.1_4', 'Q3.1_5',
'Q3.1_6', 'Q3.1_7', 'Q3.1_8', 'Q3.1_9', 'Q3.1_10',
'Q4.1_1', 'Q4.1_2', 'Q4.1_3', 'Q4.1_4', 'Q4.1_5', 'Q4.1_6',
'Q4.1_7', 'Q4.1_8', 'Q4.1_9',
'Q4.2_1', 'Q4.2_2', 'Q4.2_3', 'Q4.2_4', 'Q4.2_5',
'Q6.1_1', 'Q6.1_2', 'Q6.1_3', 'Q6.1_4', 'Q6.1_5',
'Q5.1_1', 'Q5.2_1', 'Q5.3_1', 'Q5.4_1', 'Q5.5_1',
'Q7.6', 'Q7.7', 'Q7.8', 'Q7.9', 'Q7.10', 'Q7.11', 'Q7.12',
'Q7.13', 'Q7.14', 'Q7.15', 'Q7.16', 'Q7.17', 'Q7.18'], axis=1)

df

IN [13]
# # df.to_csv('Data_clear_300522.csv', index=False)
df.to_pickle('Data_clear_010622')
df.to_csv('Data_clear_010622.csv', index=False)

```

Appendix 5: Verklaring categorieën en tot stand komen scores

Appendix 4, regel 'IN[12]'

Categorie:	Vragen:	Score:
QOLS		
QOLS_phys	Q3.1_1, Q3.1_2, Q3.1_11 Deze vragen omvatten stellingen gerelateerd aan het fysieke aspect van de persoon.	Antwoord mogelijkheden (punten): 1. Blij (1) 2. Tevreden (2) 3. Meestal voldaan (3) 4. Gemengd (4)
QOLS_rel	Q3.1_3, Q3.1_4, Q3.1_5, Q3.1_6 Deze vragen omvatten stellingen gerelateerd aan het relationele aspect van de persoon.	5. Meestal onvoldaan (5) 6. Ontevreden (6) 7. Ongelukkig (7)
QOLS_pers	Q3.1_7, Q3.1_8, Q3.1_9, Q3.1_10 Deze vragen omvatten stellingen gerelateerd aan persoonlijke aspecten van de persoon.	Des te lager de score van de QOLS des te meer tevreden de persoon is over verschillende aspecten in het kwaliteit van leven.
HLA		
eHLA1	Q4.1_1', 'Q4.1_2', 'Q4.1_3', 'Q4.1_4', 'Q4.1_5', 'Q4.1_6', 'Q4.1_7', 'Q4.1_8', 'Q4.1_9'	Antwoord mogelijkheden (punten): 1. Zeer gemakkelijk (1) 2. Gemakkelijk (2) 3. Normaal (3) 4. Moeilijk (4) 5. Zeer moeilijk (5) Score 9 geeft aan dat patiënt makkelijk informatie kan vinden over zijn/haar ziekte, een score van 45 geeft aan dat diegene erg moeilijk tot informatie over de ziekte komt.
eHLA2	Q4.2_1', 'Q4.2_2', 'Q4.2_3', 'Q4.2_4', 'Q4.2_5'	Antwoord mogelijkheden (punten): 1. Helemaal niet bekend (1) 2. Enigszins bekend (2) 3. Matig bekend (3) 4. Bekend (4) 5. Zeer goed bekend (5) Score 5 geeft slechte kennis van medische begrippen en 25 geeft goede medische kennis.

PHE	Q5.1_1, Q5.2_1, Q5.3_1, Q5.4_1, Q5.5_1	<p>Likert-scale 1-7, gecodeerd (score/2= antwoord naar boven afronden) geeft minimum is 1 en maximum is 4. Dit doordat er 4 antwoord mogelijkheden zijn met lege opties ertussen.</p> <p>5 punten = patiënt ervaart veel problemen door de ziekte 20 punten = patiënt ondervindt weinig hinder door de ziekte</p>
MARS	Q6.1_1', 'Q6.1_2', 'Q6.1_3', 'Q6.1_4', 'Q6.1_5	<p>Antwoord mogelijkheden (punten):</p> <ol style="list-style-type: none"> 1. Altijd (1) 2. Vaak (2) 3. Soms (3) 4. Bijna nooit (4) 5. Nooit (5) <p>5 punten = slechte therapietrouwheid 25 punten = perfecte therapietrouwheid</p>
FOOT_CARE	<p>Q7.6', 'Q7.7', 'Q7.8', 'Q7.9', 'Q7.10', 'Q7.11', 'Q7.12', 'Q7.13*', 'Q7.14*', 'Q7.15*', 'Q7.16', 'Q7.17', 'Q7.18</p> <p>*deze vragen zijn andersom gesteld waardoor ze omgezet zijn naar ja → 0 en nee→1</p>	<p>Antwoord mogelijkheden (punten):</p> <ol style="list-style-type: none"> 1. Ja (1) 2. Nee (2) <p>Gecodeerd naar Ja → 1 en nee → 0. Geeft 13 punten een perfecte voetverzorging, 0 punten totaal geen voetverzorging.</p>

Tabel 2: Verklaring categorieën en tot stand komen scores

Appendix 6: Jupyter script 'Gower-Method'

```
IN [1]
import pandas as pd
import gower
import numpy as np
from sklearn.preprocessing import MinMaxScaler, minmax_scale
from scipy.cluster.hierarchy import linkage, dendrogram, fcluster
from scipy.spatial.distance import squareform, pdist
from sklearn.feature_selection import SelectPercentile, f_classif
from sklearn.metrics import silhouette_score
from scipy.stats import mannwhitneyu, kruskal

IN [2]
df = pd.read_pickle('Data_clear_010622')
df.dtypes

IN [3]
scaler = MinMaxScaler()

df_scaled = scaler.fit_transform(df)
df_scaled = pd.DataFrame(df_scaled, columns = df.columns)
df_scaled

IN [4]
cat = []
types = df.dtypes
for i in range(len(types)):
    if types[i] != 'object':
        cat.append(False)
    else:
        cat.append(True)

IN [5]
gw = gower.gower_matrix(df_scaled, cat_features=cat)
gw = pd.DataFrame(gw)
gw

IN [6]
def gower_calculator(df, cat_var):

    r,c = df.shape
    output = []
    for row in range(r):
        temp_out = []
        for i in range(r):
            sum_compare = 0
            for column in range(c):
                if cat_var[column]:
                    #Calculate DICE coefficient
                    coeff = np.where(df_scaled.iloc[row,column] ==
df_scaled.iloc[i,column],0,1)
                else:
                    #Calculate range-normalized manhattan distance
                    coeff =
pdist(df.iloc[[row,i],column].to_numpy().reshape(-1,1), 'cityblock')
                    coeff = coeff[0]
```

```
        sum_compare = sum_compare + coeff

        distance = (sum_compare/c)
        temp_out.append(distance)
    output.append(temp_out)

    return output
```

IN [7]

```
output = gower_calculator(df_scaled, cat)
output = pd.DataFrame(output)
output
```

IN [8]

```
Zd = linkage(squareform(gw), method='ward')
D = dendrogram(Zd)
```

IN [9]

```
y = pd.DataFrame(fcluster(Zd, 0.5, criterion='distance'),
columns=["Labels"])
y
```

IN [10]

```
silhouette_score(gw, y, metric="precomputed")
```

IN [11]

```
df = pd.concat([df, y], axis=1)
```


Appendix 7: Jupyter script 'PCA'

```
IN [1]
import pandas as pd
import gower
import numpy as np
from sklearn.preprocessing import MinMaxScaler, minmax_scale, RobustScaler
from scipy.cluster.hierarchy import linkage, dendrogram, fcluster
from scipy.spatial.distance import squareform, pdist
from sklearn.feature_selection import SelectPercentile, f_classif
from sklearn.metrics import silhouette_score
from scipy.stats import mannwhitneyu, kruskal
import pandas as pd
import gower
import numpy as np
from sklearn.preprocessing import MinMaxScaler, minmax_scale
from scipy.cluster.hierarchy import linkage, dendrogram, fcluster
from scipy.spatial.distance import squareform, pdist
from sklearn.feature_selection import SelectPercentile, f_classif
from sklearn.metrics import silhouette_score
from scipy.stats import mannwhitneyu, kruskal
from sklearn.decomposition import PCA
from sklearn_extra.cluster import KMedoids

IN [2]
df = pd.read_pickle('Data_clear_010622')
df.dtypes

IN [3]
df.fillna(0, inplace = True)

IN [4]
scaler = RobustScaler()

df_scaled = scaler.fit_transform(df)
df_scaled = pd.DataFrame(df_scaled, columns=df.columns)
df_scaled

IN [5]
pca = PCA(n_components=0.75, svd_solver='full')

df_pca = pd.DataFrame(pca.fit_transform(df_scaled))

IN [6]
df_pca

IN [7]
kmedoids = KMedoids(n_clusters=2, metric='sqeuclidean', method='pam',
init='k-medoids++', max_iter=1000, random_state=42)

kmedoids.fit(df_pca)

labels = kmedoids.predict(df_pca)

IN [8]
silhouette_score(df_pca, labels, metric="sqeuclidean")
```

```
IN [9]  
df['Labels'] = labels
```

Appendix 8: Missende data

#Respondent	missing values
1	3
2	1
3	2
4	8
5	2
6	3
7	23
8	2
9	2
10	2
11	3
12	3
13	3
14	3
15	1
16	3
17	3
18	2
19	2
20	3
21	55
22	2
23	3
24	3
25	71
26	2
27	2
28	71
29	71
30	54
31	71
32	71
33	55
34	30
35	71
36	71
37	71
38	71
39	71
40	71
41	71
42	71
43	71

Tabel 3: missende data

Appendix 9: Resultaten PCA(tabel 4)

Question	Cluster_1 (n = 12)	Cluster_2 (n = 12)	P-value
Q1.1	(1,) (75%)	(2,) (58%)	0.2138
Q1.2	64.0 (57.4; 74.0)	47.5 (24.4; 58.0)	0.0072
Q1.3	(3,) (83%)	(1,) (50%)	0.0441
Q1.4	4.0 (3.0; 4.0)	8.0 (8.0; 9.0)	< 0.001
Q1.5	(2,) (50%)	(4,) (83%)	0.0435
Q1.7	3.0 (1.0; 3.0)	1.0 (1.0; 2.1)	0.1503
Q1.8	4.5 (1.0; 5.0)	3.5 (1.0; 5.0)	0.8282
Q2.3	(1,) (58%)	(2,) (100%)	0.0019
Q2.4	(2,) (75%)	(1,) (58%)	0.0896
Q2.5	(-99,) (83%)	(2,) (58%)	0.0894
Q2.6	1.5 (1.0; 2.0)	2.0 (1.0; 2.6)	0.3584
Q2.8	2.5 (1.5; 4.0)	3.0 (2.0; 4.0)	0.3903
Q2.9	2.0 (2.0; 3.0)	3.0 (2.5; 3.0)	0.0336
Q7.2	4.0 (3.0; 4.0)	4.0 (2.5; 4.0)	0.5254
Q7.3	5.5 (3.0; 8.2)	1.5 (1.0; 2.6)	< 0.001
Q7.4	(2,) (92%)	(1,) (58%)	0.0272
Q7.1.1	(0,) (75%)	(0,) (92%)	0.5901
Q7.1.2	(0,) (92%)	(0,) (100%)	1.0000
Q7.1.3	(0,) (83%)	(0,) (100%)	0.4783
Q7.1.5	(0,) (92%)	(0,) (100%)	1.0000
Q7.1.6	(0,) (100%)	(0,) (92%)	1.0000
Q7.1.7	(0,) (75%)	(0,) (83%)	1.0000
Q7.5.1	(0,) (92%)	(0,) (75%)	0.5901
Q7.5.2	(0,) (67%)	(0,) (83%)	0.6404
Q7.5.3	(0,) (92%)	(0,) (83%)	1.0000
Q7.5.5	(0,) (75%)	(0,) (83%)	1.0000
Q2.7	(2,) (92%)	(2,) (67%)	0.3168
BMI	27.8 (26.0; 29.3)	24.3 (23.2; 27.3)	0.0193
QOLS_phys	8.5 (6.5; 10.0)	5.0 (5.0; 6.0)	< 0.001
QOLS_rel	7.5 (4.9; 10.0)	7.5 (5.0; 9.6)	0.7483
QOLS_pers	9.0 (7.0; 12.7)	8.0 (6.5; 10.0)	0.3999
QOLS	25.5 (20.9; 30.1)	20.5 (17.5; 24.7)	0.0933
eHLA_2	18.0 (15.9; 23.6)	16.0 (12.4; 19.6)	0.2127
eHLA_3	19.5 (14.0; 22.7)	23.5 (20.9; 25.0)	0.0119
MARS	24.5 (24.0; 25.0)	21.0 (17.9; 24.6)	0.0168
PHE	16.0 (13.4; 17.0)	16.5 (15.0; 18.0)	0.2797
FOOT_CARE	9.0 (8.0; 10.6)	7.0 (6.0; 8.0)	0.0310

Tabel 4: resultaten PCA