Computers in Human Behavior Reports 20 (2025) 100872

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Full length article

Integrating data-driven methods and expert knowledge to develop personas: Balancing automation and multi-disciplinary validation

Rosa Lilia Segundo Díaz [a,*], Sevda Ece Kizilkilic [b,c,d], Wim Ramakers [a], Dominique Hansen [e,f], Paul Dendale [b,c], Karin Coninx [a]

[a] Human-Computer Interaction and eHealth, Faculty of Sciences, Hasselt University, Agoralaan Building D, Diepenbeek, 3590, Belgium
[b] Faculty of Medicine and Life Sciences, Hasselt University, Agoralaan Building D, Diepenbeek, 3590, Belgium
[c] Heart Centre Hasselt, Jessa Hospital, Stadsomvaart 11, Hasselt, 3500, Belgium
[d] Faculty of Medicine and Health Sciences, Ghent University, Corneel Heymanslaan 10, Ghent, 9000, Belgium
[e] REVAL (Rehabilitation Research Centre), Hasselt University, Wetenschapspark 7, Diepenbeek, 3590, Belgium
[f] BIOMED (Biomedical Research Institute), Hasselt University, Agoralaan Building C, Diepenbeek, 3590, Belgium

ARTICLE INFO

Keywords:
Data-driven personas
Clustering
Validation
eHealth
CVD
UCD

ABSTRACT

Data-driven personas are increasingly used to inform design decisions. Various methods are published to produce personas based on data collected from projects of different types and scales, each with a specific focus. This study aims to create a set of personas using data collected from a prior randomised controlled trial (RCT), which will be instrumental in designing future eHealth applications to support individuals with cardiovascular disease (CVD). Our method followed five phases for designing personas: (Phase I) expert analysis and variable selection, (Phase II) clustering, (Phase III) expert validation, (Phase IV) persona optimisation, and (Phase V) final check. To ensure that personas accurately reflected the patients, we employed the k-prototype algorithm to cluster mixed data and we focused on validation with colleagues, including medical colleagues, physiotherapists, a psychologist and Human-Computer Interaction (HCI) experts. Seven different personas resulted from the clustering. A validation step involved a multidisciplinary team that assessed the personas' realism, giving an average rating of 8.0 out of 10. Based on their feedback, three of the personas were slightly updated. The final descriptions of all seven personas incorporated the clustered data and the proposed changes after the validation. We concluded that data-driven approaches and expert-based refinement to develop personas is an effective method for understanding the target population. This study highlighted the importance of validation, revealing that creating personas cannot be fully automated, as this may result in losing essential characteristics that only experts can identify. Future research includes demonstrating the practical use of personas.

1. Introduction

Cardiovascular diseases (CVD) are the leading cause of death not only in EU (Eurostat, 2024) but also worldwide (WHO, 2024). They also cause a decrease in quality of life and cost €210 billion a year in lost productivity and healthcare provision (Wilkins et al., 2017). A known way to significantly improve the prognosis of CVD is to modify behavioural risk factors, such as smoking, unhealthy diet, physical inactivity, stress, and lack of sleep (Wilkins et al., 2017). To this end, mHealth/eHealth technologies are viewed as a valuable opportunity to assist patients with CVD in managing their health and controlling their health condition. Although it is known that eHealth apps are more likely to be used if the needs, desires and context of end-users are considered during design and development, designers currently have limited access to patients in the user-centred design (UCD) process used to design these apps. This is due to several reasons, among which the workload in rehabilitation centres and the increasing complexity of regulations to achieve ethical committee approval, including privacy concerns (General Data Protection Regulation, GDPR) and safety checks (Medical Device Regulation, MDR). This limitation to meet a considerable group of representative patients restricts the designers' focus to only a few features that a limited number of patients can provide. To address this issue, designers use personas as part of the eHealth design and development process, but in addition to or even instead of conducting focus groups, interviews or surveys specifically designed for this purpose, they re-use data from previous studies. Such an approach lowers the costs and time to collect user needs, and allows creating

* Corresponding author.
E-mail address: rosalilia.segundodiaz@uhasselt.be (R.L. Segundo Díaz).

https://doi.org/10.1016/j.chbr.2025.100872
Received 25 June 2025; Received in revised form 29 September 2025; Accepted 11 November 2025
Available online 14 November 2025

personas representing populations we already know. In the research effort we describe in this article, we used a structured and iterative approach of five phases to design a set of personas using data from a former study. Our personas aim to enhance the realisation process of future eHealth applications for similar target groups, and can also be used by medical CVD researchers and rehabilitation experts to guide the design of interventions. We emphasised a rigorous validation process involving a multidisciplinary team of experts. This team brought together different perspectives and expertise to ensure that the personas were strongly representative.

2. Related work

According to the international UCD standard (ISO 9241-210, 2019), the design process starts by empirically defining users and their context. In this process, personas are a traditional UCD tool commonly used in product marketing, software development, and other fields where systems, services, or products are designed for human interaction. Personas are "archetypes" of intended users (Cooper, 1999), and in the context of eHealth applications their description consists of relevant information for the design and evaluation of prototypes (Holden et al., 2017). More specifically, personas are used to ensure that the designs take into account the main needs of all intended users instead of what caregivers/researchers have in mind and what a few represented patients can inform. Personas have faced criticism for lacking scientific rigour (Chapman & Milham, 2006). However, as noted by Floyd et al. (2008), personas are a design technique, and it is essential to assess the diverse methods and contexts in which personas are used and how those methods are applied appropriately. According to Grudin (2006), the effectiveness of personas lies in our natural ability to create detailed representations of people, whether they are real or fictional.

In healthcare, the use of (data-driven) personas is relatively new but evolving not only to become a significant tool in the design of robust eHealth applications but also in the design of patient-centred interventions (Engelmann et al., 2023). In patients with CVD, researchers have created personas to improve the usability and accessibility of technology (ten Klooster et al., 2022), to support medication adherence (Haldane et al., 2019), and to tailor medical interventions (Engelmann et al., 2023; Vosbergen et al., 2015).

Personas are often designed intuitively, that is, with basic information about the target group obtained through the research team. They are based on designers' or caregivers' assumptions, adding certain ages, certain computer skills, and other demographic attributes, but most likely, the personas do not represent the target group. In other cases, researchers conduct user need studies to base the creation of personas on that data, which can include multiple potential sources of data, such as qualitative, quantitative, or mixed data (Salminen et al., 2021). Previous studies have recognised the importance of using mixed data types from several sources to include relevant characteristics that accurately describe the target (Haldane et al., 2019; Holden et al., 2017; Jansen et al., 2021; LeRouge et al., 2013; ten Klooster et al., 2022; Vosbergen et al., 2015). Those sources ranged from interviews focused on gathering patients' needs and preferences to reusing already available data collected during the UCD process or project trajectory. Gathering patients' needs and preferences usually requires significant time and effort and can be expensive (Patkar & Seyff, 2023). Reusing already available data may reduce the cost and time needed to collect data (Salminen et al., 2021), given that it can be considered a secondary analysis of data collected in a previous study (Holden et al., 2017; ten Klooster et al., 2022). Data selection from various sources presents another significant challenge. LeRouge et al. (2013) proposed a conceptual user model that includes technical, demographic, and healthcare-specific factors. Other studies have attempted to apply this model (Breeman et al., 2021; Dol et al., 2016, 2023; ten Klooster et al., 2022). However, it has been observed that while these studies aimed to classify their variables according to the model, each one collected different variables. Additionally, some variables mentioned in the model, including demographic factors (e.g., marital status, children), technical aspects (e.g., technology usage) and healthcare considerations (e.g., strategies for coping with the disease), are often not captured for various reasons. For example, researchers do not want to burden participants with additional questions, so they limit the number of questionnaires to capture only the needed data. Another reason is that different studies use different ad-hoc questionnaires, which may not include the same variables. Other authors mention the creation of biopsychosocial personas (Haldane et al., 2019; Holden et al., 2017; Li et al., 2024), which include biological, psychological and social domains and subdomains (i.e., demographics, medical status, functional status, psychological status, technology, healthcare system, social context, and economic context) to maintain health or recover from a disease. According to Li et al. (2024), there is a consensus that eHealth personas should incorporate biopsychosocial domains. However, it remains undecided which subdomains should be included and how to identify them for different personas and health management objectives.

For the development of design personas, there are different approaches, from manual persona development (MPD) to data-driven persona development (DDPD) (Salminen et al., 2020, 2021). There is ongoing discussion about which approach is more appropriate. MPD tends to be subjective and is therefore criticised for its lack of objectivity and rigour, but also its high cost, lack of scaling, non-representative data, and expiration as users may change their behaviours (Salminen et al., 2020). On the other hand, DDPD offers statistical values that may provide a more objective analysis. However, without further analysis and relying only on the statistics, it may introduce biases, undesired generalisations into the personas, ignoring minority groups and inclusivity (Jansen et al., 2021; Salminen et al., 2020). To address the DDPD problems, Salminen et al. (2021) recommended that personas should be co-created by HCI experts and future users. Nevertheless, access to patients in the healthcare field is not always possible.

Another step in the process of creating personas for eHealth suggested in the literature is to validate personas once they have been created (Holden et al., 2017). If personas have been created using qualitative data, they can be validated using quantitative data and vice versa. Other techniques, such as interviews with real users (Li et al., 2024; Vosbergen et al., 2015) or validations within an interdisciplinary team (Olivares et al., 2020) have also been suggested. However, there has also been criticism of the lack of examples of specific data, how this validation proves accuracy and how the authors define success in validation (Chapman & Milham, 2006).

In this study, we co-created and validated personas with medical colleagues, physiotherapists, a psychologist, and HCI experts. This decision was made because the access to patients due to the workload in rehabilitation centres and the increasing complexity of regulations to obtain ethical committee approval makes it challenging to start focus groups or other techniques with actual patients. Additionally, our approach for a data-driven persona creation process aimed to incorporate data from a previous randomised controlled trial (RCT) that comprises qualitative and quantitative data. Our process involved five phases and different steps that combined data-driven methods (i.e., clustering, large language model), expert validation, and individual discussions with experts, resulting in a final narrative set of personas for future eHealth design efforts.

3. Methods

Different methods have been proposed for the creation of data-driven personas. Based on multiple examples and guidelines (Holden et al., 2017; ten Klooster et al., 2022; Williams et al., 2021), we created personas from a holistic perspective and identified relevant characteristics for our target population. The research approach used to develop the personas, illustrated in Fig. 1, identifies five phases namely,

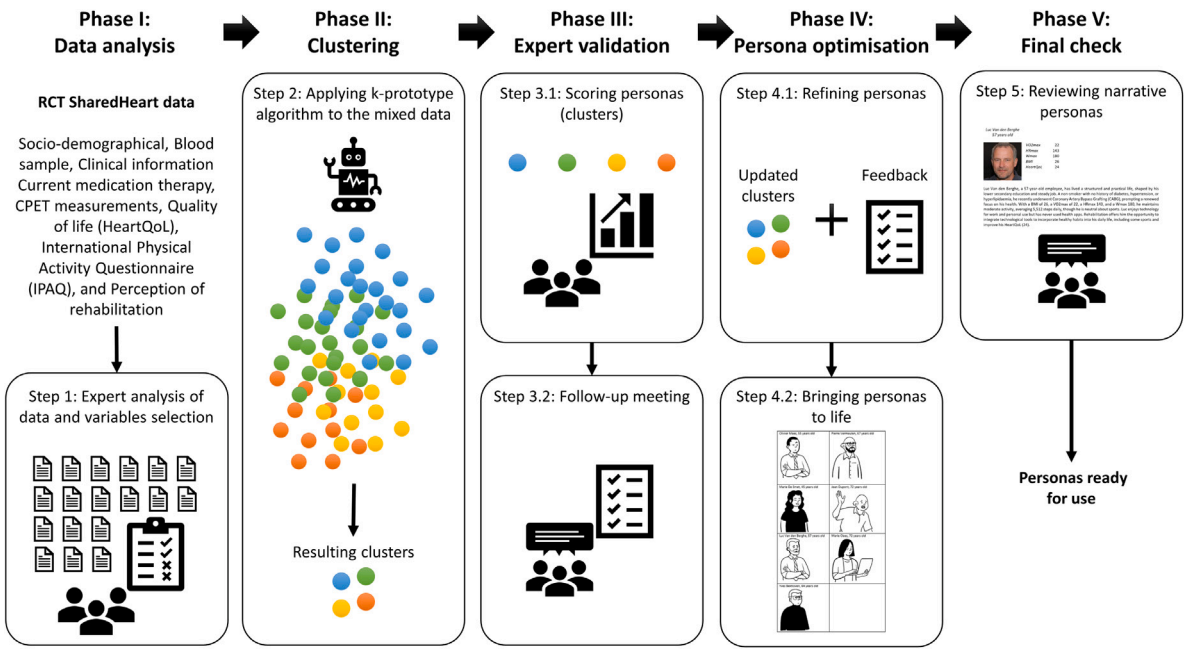

Fig. 1. Creation process of the Personas.

Table 1
Domains and variables selected to be included in the clustering; VO2max: maximum rate of oxygen consumption; HRmax: maximum heart rate; Wmax: maximal work capacity; BMI: Body Mass Index; HeartQoL: Quality of Life.

Background	**Medical Measures**
Age	VO2max
Gender	HRmax
Education	Wmax
Occupation	BMI
Health Risks	HeartQoL
Smoking	**Technology**
Diabetes	HealthApps Experience
Hyperlipidaemia	Computer Use
Hypertension	Tablet Use
Peripheral Vascular Disease	Smartphone Use
Indication for Cardiac Rehabilitation	Computer Purpose
	Tablet Purpose
Type of device/surgery	Smartphone Purpose
Physical Activity	Tracking Exercise
Enjoy Sports	Enjoy Technology
Average Steps	

data analysis (Phase I), clustering (Phase II), expert validation (Phase III), persona optimisation (Phase IV), and final check (Phase V). A series of steps were conducted throughout the various phases, indicated by a numerical sequence to show the order of tasks. We will discuss the different steps in the following sections.

3.1. Phase I: Data analysis

Data collected in a former RCT (Kizilkilic et al., 2025) with the SharedHeart application described in Bonneux et al. (2022) was used in the development of the personas. The SharedHeart study successfully explored a hybrid shared decision-making intervention for physical activity, in which the participating persons with CVD were supported by a mobile app for home-based physical activity in addition to supervised training sessions in the hospital's rehabilitation centre. Socio-demographical information, a blood sample, clinical information, current medication therapy, and cardiopulmonary exercise testing (CPET) measurements were collected for all patients at the start of the study. The standard questionnaires Quality of Life (HeartQoL, Frederix et al., 2017) and International Physical Activity Questionnaire (IPAQ, Craig et al., 2003) were also taken from the patients. In addition, an ad hoc questionnaire was used to probe their perception of rehabilitation in this shared decision-making context.

From those questionnaires, a medical colleague (cardiologist in training) and HCI experts selected 26 relevant variables, shown in Table 1, that include the different characteristics and preferences of CVD patients who participated in the RCT. The selected variables include demographic information, cardiovascular risk factors, level of physical activity, medical measures and self-reported technology literacy. Other variables that may change over time, such as weight, cholesterol, and triglycerides, were not deemed relevant to this analysis because their fluctuations do not directly influence the design or evaluation of technology tools. The selected variables aim to ensure that the personas reflect multiple aspects of the cardiac patients and their willingness to use technology in rehabilitation. As shown in Table 1, the included variables were grouped in different domains based on the examples of previous studies (ten Klooster et al., 2022; Williams et al., 2021). The "Medical measures" listed reflect the focus of the SharedHeart study on physical activity and the patients' general quality of life in relation to their CVD, whereas the "Health Risks" bring in a broader medical perspective.

3.1.1. Participants' data from the RCT

Seventy out of eighty patients from the RCT were included in the analysis because they had complete information on the selected variables. Of the 70 patients, 87% were male. The mean age of the participants was 62.9 years (with a median of 63, a standard deviation of 11.0, and an age range from 20 to 82 years). Most patients (60%) had completed lower or secondary education, 29% had earned a bachelor's degree, and 11% had completed a master's degree. Thirty-three patients (47%) were retired, 25 (36%) were employees, 8 (11%) were business owners, and 4 (6%) were unemployed. Table 2 summarises participants' data, including the selected variables and their descriptive statistics.

3.2. Phase II: Clustering

Two HCI researchers (R.W. and C.K.) categorised the answers to the open questions in the demographic questionnaire (i.e., occupation and

Table 2
Baseline demographics and variables used in the clustering.

Domains and variables	Patients (n = 70)
Demographics	
Age (y), mean (± SD)	62.9 (± 11.0)
Age min, max	20, 82
Sex (male), n (%)	61 (87.1)
Education	
Lower or secondary, n (%)	42 (60.0)
Bachelor, n (%)	20 (28.6)
Master, n (%)	8 (11.4)
Occupation	
Retired, n (%)	33 (47.1)
Employee, n (%)	25 (35.7)
Business owner, n (%)	8 (11.4)
Unemployed, n (%)	4 (5.7)
Health risks	
Smoking	
Ex-smoker, n (%)	37 (52.9)
No-smoker, n (%)	27 (38.6)
Smoker, n (%)	6 (8.6)
Diabetes (yes), n (%)	11 (15.7)
Hiperlipidemia (yes), n (%)	50 (71.4)
Hypertension (yes), n (%)	42 (60.0)
Indication for cardiac rehabilitation	
Percutaneous coronary intervention (PCI), n (%)	32 (45.7)
Ablation, n (%)	16 (22.9)
PM, ICD, CRT-D, CRT-P, n (%)	8 (11.4)
Coronary artery bypass grafting (CABG), n (%)	6 (8.6)
Heart Failure, n (%)	4 (5.7)
Heart valve repair or replacement, n (%)	4 (5.7)
Physical activity	
Enjoy sports	
I like it, n (%)	26 (37.1)
Neutral, n (%)	26 (37.1)
I like it very much, n (%)	9 (12.9)
I do not like it, n (%)	7 (10.0)
I do not like it at all, n (%)	2 (2.9)
Average steps, mean (± SD)	7911 (± 4269)
Medical measures	
VO2max, mean (± SD)	19.6 (± 5.6)
HRmax, mean (± SD)	126.6 (± 25.8)
Wmax, mean (± SD)	148.6 (± 55.7)
BMI, mean (± SD)	27.2 (± 4.5)
HeartQoL	27.8 (± 8.5)
Technology	
Health Apps Experience (yes), n (%)	27 (38.6)
Computer Use	
Daily, n (%)	46 (65.7)
A few times a week, n (%)	12 (17.1)
A few times a month, n (%)	8 (11.4)
Never, n (%)	4 (5.7)
Tablet Use	
Daily, n (%)	18 (25.7)
A few times a week, n (%)	10 (14.3)
A few times a month, n (%)	8 (11.4)
Never, n (%)	30 (42.9)
I do not know it, n (%)	4 (5.7)
Smartphone Use	
Daily, n (%)	64 (91.4)
A few times a week, n (%)	4 (5.7)
Never, n (%)	2 (2.9)
Computer Purpose	
Personal use, n (%)	42 (60.0)
Personal use, Work, n (%)	20 (28.6)
Work, n (%)	5 (7.1)
I do not use it, n (%)	3 (4.3)
Tablet Purpose	
Personal use, n (%)	29 (41.4)
Personal use, Work, n (%)	2 (2.9)
Work, n (%)	3 (4.3)
I do not use it, n (%)	36 (51.4)
Smartphone Purpose	
Personal use, n (%)	45 (64.3)
Personal use, Work, n (%)	21 (30.0)
Work, n (%)	2 (2.9)
I do not use it, n (%)	2 (2.9)
Tracking Exercise (yes), n (%)	25 (35.7)
Enjoy Technology	
I like it, n (%)	25 (35.7)
Neutral, n (%)	18 (25.7)
I like it very much, n (%)	16 (22.9)
I do not like it, n (%)	8 (11.4)
I do not like it at all, n (%)	3 (4.3)

(*continued on next page*)

Table 2 (*continued*).

education). We conducted Shapiro–Wilk tests to confirm the normal distribution of the included variables. VO2max and HeartQoL were not normally distributed ($p < 0.05$) and, therefore, transformed into log values before carrying out the cluster analyses.

Cluster analysis, or segmentation, is a statistical technique that groups records in large datasets based on available data, such as demographic, behavioural variables, among others. Different algorithms have been applied for designing personas, including hierarchical clustering, k-means clustering, latent semantic analysis (LSA), among others (Salminen et al., 2020). In our approach, the k-prototype algorithm was used due to the presence of both numerical and ordinal variables in the data, which other algorithms like k-means cannot handle (Huang, 1997). The k-prototype method is actually an extension of k-means clustering (Szepannek et al., 2024). The analyses were conducted using RStudio version 2024.09.0 Build 375 and the clustMixType version 0.4-2 package.

An exploration of the various approaches available to determine the optimal number of clusters was conducted. Initially, the elbow method was explored as a means of calculation. This method calculates the "withinss", which is a measure that shows how much samples in one cluster differ from one another. The smaller the withinss, the more similar the objects in the same cluster; the larger the withinss, the less similarity there is within the cluster. The method involves plotting the intra-cluster sum for different values and selecting the point where the slope decreases to form an elbow-like structure. We also explored the validation indices provided by the clustMixType package. Statistical advisors indicate no strict way to find the optimal number of clusters, so they advise to analyse the cluster's observations to ensure a balance of the data and conclude on the optimal number.

3.3. Phase III: Expert validation

After the clustering process, which was mainly performed by one of the co-authoring computer science/HCI researchers (S.D.R.L.), a multidisciplinary group of five rehabilitation experts (in alphabetical order: B.K., D.P., H.D., K.E. and V.F.) was invited to take part in the validation stage. They work in the rehabilitation centre where the above mentioned SharedHeart study took place, but only one of them was involved in the shared decision-making encounters with the participants. Validators assumed two different roles: one role was to evaluate the clusters, and the other was to review and reach a consensus on the final personas' characteristics.

3.3.1. Scoring personas (clusters)

Four rehabilitation experts (B.K., H.D., K.E. and V.F.) completed a questionnaire (included in Supplementary Material A) to validate the seven personas resulting from the clustering step: one cardiologist in training, two physiotherapists (one of which is paramedic head of the rehabilitation centre and the other one is a professor also involved in eHealth applications), and a psychologist guiding individual consults with patients in the hospital's rehabilitation centre. In the questionnaire we included an explanation of the purpose of the study, the applied methodology and the clustering results achieved by the algorithm. We

provided an example of a persona description to show how the personas will look like after the validation, and after elaboration once the final set of variables to describe and cluster the personas is fixed. The four rehabilitation experts rated the realism of each persona on a scale from 1 to 10, where 1 indicates "not realistic at all" and 10 indicates "very realistic." Additionally, they composed a rationale for their ratings and provided feedback on the selected clustering variables with proposals to possibly remove selected variables or add additional variables.

The suggestions of the four rehabilitation experts were integrated in the personas resulting from clustering, unless there was no consensus, and the average realism score for each persona was calculated.

3.3.2. Follow-up meeting

Two HCI experts (C.K. and S.D.R.L.) organised live, individual meetings with the validators. First, they commented on their own answers in the questionnaire and provided clarifications in case of questions from the HCI experts. Then they were informed about the previously collected validation results provided by their colleagues, including suggestions for changes in variables or their representative values, so they could comment on this information. This approach with short, individual conversations ensured equal participation of different rehabilitation disciplines in the decision process, while smoothly evolving towards a consensus.

The fifth rehabilitation expert (D.P.) who was involved in the validation is a senior cardiologist, having more than 30 years of experience and functioning as head of the cardiology department. He took a different role in the validation process of our personas, as he did not complete the questionnaire with the realism rating in the initial step due to practical reasons. However, this turned out to be a coincidental strength of the applied validation approach. He received the integrated validation results of his colleagues, commented on the selected variables and the representative values per persona and made a decision in case a consensus was not yet reached.

3.4. Phase IV: Persona optimisation

This phase involved a two-step process to refine the personas and turn them into compelling narrative descriptions.

3.4.1. Refining personas

HCI experts (C.K. and S.D.R.L.) conducted a thorough analysis of the seven clusters, taking into account the valuable feedback provided through the questionnaires and subsequent follow-up meetings. Our review specifically targeted the variables within each cluster that the validation experts highlighted as requiring adjustment. As a result of this analysis, these variables were considered within the clusters, recalculated and updated to ensure that they better reflected the insights gained during the validation process. This systematic approach enabled us to improve the accuracy and relevance of our personas.

3.4.2. Bringing personas to life

To bring the personas to life, different approaches are possible. The most traditional one would be to manually write narratives based on the results of the previous steps in our approach. However, in this research we explored the use of LLM. We used ChatGPT (GPT-4o Individual Free Plan) to generate initial narratives for the personas, which we refined and enhanced to ensure they describe our personas and resemble our target audience. With the standard configuration, ChatGPT was provided with detailed data from our updated clusters, and we requested for each cluster to create a persona description that reflected the characteristics provided. ChatGPT typically presents results in bullet point format. Therefore, we asked for a narrative description of the persona.

After that, an HCI researcher (S.D.R.L.), reviewed the provided descriptions. Insights and motivations gathered during the discussions with the validation experts were carefully integrated to ensure that each persona reflected the different user needs and behaviours. This process involved refining the language and details to create richer representations of the personas and provide a more comprehensive understanding of the target audience.

Table 3
Accumulated distances of all observations belonging to a cluster to their respective k-prototype.

Number of clusters	Sum of withinss
1	627.90
2	536.35
3	495.79
4	472.17
5	445.61
6	435.08
7	417.94
8	411.54
9	397.36
10	388.77

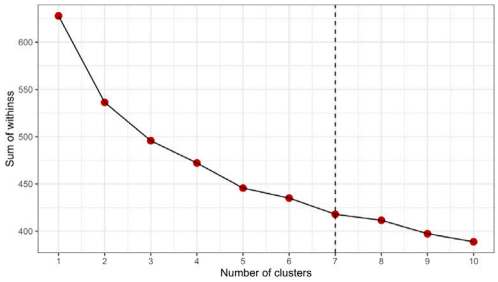

Fig. 2. Line plot to identify the elbow point.

3.5. Phase V: Final check

The last phase aims to review the narrative of personas and give closure to the design of personas. During this step, validators were provided with the final description of personas, which are brought to life through carefully selected images that give presence to the persona.

In this stage, it is not planned to make further modifications but to explain that the personas created here are the base, and they can evolve or adapt with new characteristics (e.g., motivations, personality traits, physical limitations) that align with their intended purpose.

4. Results

This study analysed data collected from 70 patients who were participants in a prior RCT. The relevant data is outlined in Table 2 located in Section 3.1.1. This dataset served as the basis for the subsequent stages of our methodology.

4.1. Clustering

4.1.1. Obtaining the optimal number of clusters

An exploration of the various approaches available to determine the optimal number of clusters was conducted. Initially, the elbow method was explored as a means of calculation. Table 3 shows the vector obtained with the sum of the distances from 1 to 10 clusters, and Fig. 2 the plot of these values. An elbow was identified at 7 clusters, which corresponded to the point at which the line plot exhibited the most pronounced curvature. However, it should be noted that this method is subjective and there is no definitive conclusion regarding the optimal number of clusters. Consequently, alternative methods were investigated to determine the optimal number.

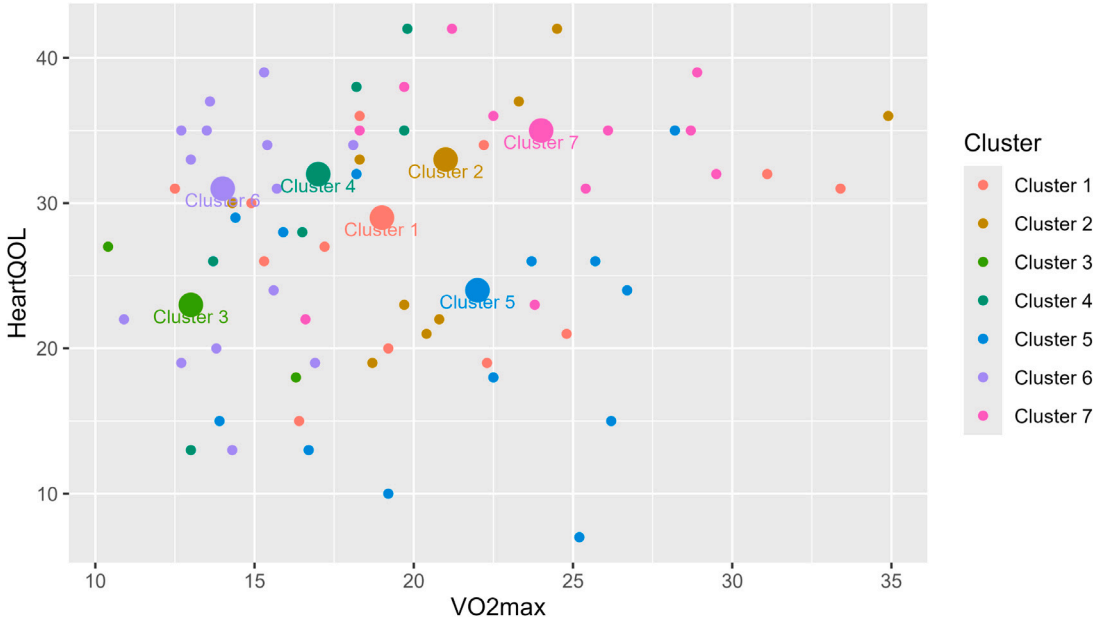

Fig. 3. Relationship between HeartQoL and VO2max in the clusters. The clusters are presented separately in Supplementary Material C.

Table 4
Number of clusters obtained with different validation indices.

Validation indices	Number of clusters
withinss	7
cindex	9
dunn	2
gamma	10
gplus	10
mcclain	10
ptbiserial	2
silhouette	10
tau	2

The validation indices provided by the clustMixType package were calculated. As can be seen in Table 4, a consensus has not been reached among the validation indices as the optimal number of clusters ranges from 2 to 10. After exploring the actual clusters obtained with a different amount of clusters, it was concluded that having seven clusters was the most appropriate choice. This conclusion was based on the observation that with fewer than seven clusters, the representation of the population was limited. It was also noted that with more than seven clusters, some tended to contain only one observation.

4.1.2. Personas clustering using k-prototype algorithm

The analysis of the 26 variables of 70 patients was conducted using the k-prototype algorithm and the selected optimal number of clusters. The results, as illustrated in Table 5, indicate that each of the seven clusters contains between 2 and 15 observations. Specifically, Persona 3 contains only two observations, while Persona 6 contains 15.

As described in the Methods section, we relied on the rehabilitation experts to validate the personas and evaluate how well they represent the target patient population. Nevertheless, as HCI experts we deemed it necessary to interpret the generated representative values for the selected variables per persona. This is not a reality check with respect to the target patients, but rather checking how the representative values per variable are situated in the range of values for that variable in the overall dataset. It intuitively indicates a level of coverage of the dataset, per variable, but without thorough consideration of the realism of the combination of variables' value per persona.

The generated personas represent patients between 40 and over 70 years old. 13% of the patients in the study were female, and the clustering reflects this demographic, with only Persona 3 representing this group. The seven personas include a mix of employees and retirees with educational degrees that tend to be lower or secondary school, except for a few with bachelor degrees. The under-representation of masters in the original data causes them to be absent in the algorithmically achieved personas.

Health risks such as hyperlipidaemia, and hypertension are prevalent. Most personas are ex-smokers, and diabetes is uncommon. In terms of the indication for rehabilitation due to an intervention or device, most personas had percutaneous coronary intervention (PCI), one had an ablation and the other had a pacemaker.

Physical activity levels differ significantly, with some enjoying sports and walking frequently, while others have low step counts. VO2max, HRmax, and Wmax values indicate varying fitness levels, with Persona 3 showing the lowest. The HeartQoL shows a perception of quality of life ranging from 23 to 35. Specifically, Persona 3 with a HeartQoL of 23 has a low step count and VO2max, while Persona 7 has the highest average step count and VO2max of 24, indicating how they perceive their quality of life. Fig. 3 shows the relationship between physical condition VO2max and the perception of quality of life with the HeartQoL score in each cluster. Similar graphs could be constructed for other selected variables the personas are based on, to visually inspect the positioning of the real patients compared to the cluster centre, while interpreting the generated data-driven personas.

Technology usage varies from persona to persona. While computers and smartphones are common, tablet usage differs significantly among personas. Additionally, a few of them utilise health apps or track their exercise routines. In summary, technology use and health conditions can vary based on factors such as age, occupation, and lifestyle.

4.2. Expert validation

4.2.1. Scoring personas (clusters)

The purpose of the validation by the rehabilitation experts is twofold: (1) to evaluate the realism of the combination of values per persona, and (2) to evaluate whether the coverage of a typical rehabilitation population is achieved by the combination of the seven personas.

Table 5
Clusters.

Clustering							
Cluster	Persona 1	Persona 2	Persona 3	Persona 4	Persona 5	Persona 6	Persona 7
Cluster Size	12	11	2	6	13	15	11
Background							
Age	50–59 years	60–69 years	40–49 years	>= 70 years	50–59 years	>= 70 years	60–69 years
Gender	Man	Man	Woman	Man	Man	***Man ->Woman***	Man
Education	bachelor	lower or secondary	lower or secondary	Bachelor	lower or secondary	lower or secondary	lower or secondary
Occupation	Employee	Retired	Employee	Retired	Employee	Retired	Employee
Health risks							
Smoker	Ex-smoker	Ex-smoker	Ex-smoker	Ex-smoker	No-smoker	Ex-smoker	Ex-smoker
Diabetes	No	No	No	No	No	Yes	No
Hyperlipidaemia	Yes	Yes	Yes	No	No	Yes	Yes
Hypertension	Yes	Yes	Yes	No	No	Yes	Yes
Peripheral Vascular Disease	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indication for Rehabilitation							
Type of device/surgery	PCI	Ablation	PM, ICD, CRT-D, CRT-P	PCI	***PCI ->CABG***	PCI	PCI
Physical activity							
Enjoy Sports	I like it	I like it	I do not like it	I like it	Neutral	Neutral	I like it
Average Steps	8466	8544	3368	5527	5512	5830	11 920
Medical measures							
VO2max	19	21	13	17	22	14	24
HRmax	121	129	115	110	143	111	134
Wmax	156	165	63	112	180	98	182
BMI	***28 ->33***	25	26	25	26	27	27
HeartQoL	29	33	23	32	24	31	35
Technology							
Health Apps Experience	Yes	No	No	Yes	No	No	No
Computer Use	Daily	Daily	Never	Daily	Daily	Daily	A few times a month
Tablet Use	Never	Daily	Daily	Daily	Never	Never	A few times a week
Smartphone Use	Daily	Daily	Daily	Daily	Daily	Daily	Daily
Computer Purpose	Personal use, Work	Personal use	I do not use it	Personal use	Personal use, Work	Personal use	Personal use
Tablet Purpose	I do not use it	Personal use	I do not use it	Personal use	I do not use it	I do not use it	I do not use it
Smartphone Purpose	Personal use, Work	Personal use	Personal use	Personal use	Personal use, Work	Personal use	Personal use
Tracking Exercise	No	No	No	Yes	No	No	Yes
Enjoy Technology	I like it	I like it very much	I do not like it at all	I like it very much	I like it	I like it	Neutral

Validators scored the realism of the seven personas (clusters). Table 6 shows that the overall average realism score was 8, with average realism score for each Persona ranging from 7.5 (Persona 1 and 3) to 9 (Persona 2). Remarks from validators included that Persona 1 has a low VO2max and HRmax considering the active step pattern. For Persona 6, they raised concerns similar to those of the previous Persona, with quite a normal number of steps but relatively poor results in its physical testing. A major comment over all seven Personas was that their BMI is similar and relatively low, so patients with obesity were lacking in the personas.

Another underrepresented group was women, whereas future rehabilitation technology should be inclusive. Therefore, we asked the rehabilitation experts to suggest which persona characteristics are typical of a female patient. Experts identified personas 4 and 6 as having characteristics relevant to describing a female patient. Experts explained that Persona 4 has a relatively passive profile, is older, and their medical measures can be used for a woman. Persona 6 has several risk factors, and their medical measures may also be used to describe a female patient.

Regarding patients rehabilitating after cardiac surgery, one expert remarked that there currently is no representation of these patients in the personas, though they count for 20% of the patients in their rehabilitation centre. For this case, it was decided to analyse the clusters to update one of the personas from PCI to CABG.

Validators were asked to mention which characteristics in the personas were less relevant and what additional characteristics they would add. They considered the variables shown in Table 1 relevant to identify their patients. Less relevant characteristics – from their perspective – include technology-related variables, vascular disease, age ranges and HRmax. Technology-related variables might not be relevant from the medical point of view, but the aim of these personas is creating technological tools (e.g., applications, games) that patients can use to support health behaviour change and improve their quality of life. Thus inherently technology has to be included. Vascular disease was observed to be present for all personas, which diminishes the discriminative value of this health characteristic over all personas. The age ranges used are not typical for patients suffering from these pathologies, but they indicated that it is unlikely to be a significant problem for future design efforts. In this case, it was decided to use ranges for clustering because the clusters did not cover the full range of patient ages when using the actual ages, and by using ranges we were able to cover almost all ranges except for a few patients younger than 40 (a minority in the data). They also mentioned that HRmax is not always critically important so it could be omitted in future clustering when applying similar methods for data-driven personas to data sets of other studies.

Regarding additional characteristics, some validators requested to include the complete medical background such as history of cardiac

Table 6
Validators V1-V4 assigned a realism score to the personas.

Clusters	V1	V2	V3	V4	Average realism score (1–10)
Persona 1	7	7	8	8	7.5
Persona 2	9	9	9	9	9.0
Persona 3	8	8	8	6	7.5
Persona 4	8	8	8	9	8.3
Persona 5	9	8	7	8	8.0
Persona 6	8	7	8	8	7.8
Persona 7	8	9	7	8	8.0
Total					**8.0**

events and other medical problems. Physiotherapists suggested to include orthopaedic or physical problems that could be a factor in the rehabilitation program. From a psychological point of view, the validators suggested to include motivations and personal goals for following a rehabilitation program, the meaning of their heart condition and how they cope with it.

While orthopaedic and psychological data were not collected during the study, they were considered in the final persona design, as well as the remarks from the validation process and follow-up meetings to create the persona descriptions.

4.2.2. Follow-up meeting

The follow-up meetings took place about a week after all the validators submitted their persona scores. We aimed to hold these meetings as soon as possible to allow them to revise their responses and engage in further discussion. The meetings were individual to ensure equal participation of different rehabilitation disciplines in the decision process.

The first meeting was with the psychologist. The expert was very interested in our approach and even started working with the persona example we provided within the questionnaire to create a theoretical classification of the personas. The expert explained the sketch representing the classification to clarify how the persona was interpreted. In that explanation, HeartQoL was included as one element that shows personas' intrinsic and extrinsic motivations and their willingness to use eHealth tools. It was also mentioned that our personas should reflect how patients give meaning to their heart disease and the way they cope with it motivationally.

In the second meeting, we asked the physiotherapist about his need for a complete patient record including the history of medical events. The expert clarified that it is necessary to know if the patient already has a heart problem or co-morbidities and what exercises could be prescribed in case of orthopaedic problems. We also clarified that we only needed certain variables to describe the personas but that they would still have access to the complete file when dealing with actual patients using the mHealth/eHealth applications realised based on the personas. The meeting with the other physiotherapist/researcher was integrated into another research meeting and did not result in any additional remarks other than his written remarks from the scoring.

The last meeting was with the cardiologist in training. The discussion with the expert was about the variables of vascular disease and age groups. The expert noted that peripheral vascular disease was considered during the RCT, but this factor does not significantly affect rehabilitation as opposed to other health characteristics. Regarding age ranges, it was mentioned that there could be some combinations between age and rehabilitation (e.g., ablation is mainly done in older patients). However, it was confirmed that we could work with age ranges to cover most patients. From their responses in the scoring questionnaire, we proposed some changes to the personas and asked them whether they agreed or suggested other personas to be changed. Their responses were considered in Step 4.1, Refining personas, and described in Section 4.3.1.

4.3. Persona optimisation

4.3.1. Refining personas

In this step, we consider the remarks from the validation questionnaires and discussions in the follow-up meetings to update the personas clusters. We aimed to preserve the integrity of the data-driven persona process by incorporating expert feedback in a balanced manner to minimise changes and potential bias. To achieve this, we used data from the cluster itself to determine where changes should be made. This approach enabled us to preserve the integrity of the data within the cluster while updating our personas to reflect our target patients better.

Since BMI was a major concern in the discussion with the experts and overweight is a significant risk factor for cardiac patients, we analysed the clustering data. We looked for the persona with the highest BMI and therefore recalculated Persona 1's BMI by averaging the BMI values above 28 within that group. Validators agreed that a BMI of 33 is above overweight.

Due to the underrepresentation of women in our personas, another change was proposed. Validators suggested to change Persona 4 or 6. Based on the cluster data and feedback from validators, the gender of Persona 6 was changed to female. Specifically, in this cluster, there were 3 out of 9 women, and the experts noted that this persona has several risk factors yet maintains a lifestyle with a relatively normal level of physical capacity, so those characteristics can be used to describe a female patient.

Another important remark was about the variable indication for rehabilitation. Validators mentioned that they missed indications such as CABG within the personas. Based on the analysis of clusters, Persona 5 had the most patients with CABG. Therefore, it was updated with the variable indication for rehabilitation set to CABG. Table 5 shows in bold italic the changes made to the different clusters.

4.3.2. Bringing personas to life

We used the updated data in Table 5 and the feedback from the validation in Section 4.2 to bring the personas to life. Personas descriptions were created based on the data contained in each cluster. We used ChatGPT for the initial description, and from there, we fine-tuned the language, added other characteristics such as motivations, and looked for an image that represented our personas. Personas 5 and 6 are presented below to illustrate the type of narrative we used, and the reader can find the complete list with the integrated image in the Supplementary Material B. Fig. 4 shows images of our personas obtained from ThisPersonNotExist.org as AI-generated faces.

Persona 1 is Olivier Maes, 55 years old. He is a former smoker dealing with high cholesterol and hypertension. He recently underwent a PCI and participates in a cardiac rehabilitation program. He averages 8,466 steps daily, has a VO2max of 19 ml/kg/min, a HRmax of 121 bpm, a Wmax of 156 W, a BMI of 33 kg/m^2, and a HeartQoL score of 29. Olivier daily uses a computer and smartphone and regularly employs health apps for monitoring. He is open to new health-tracking devices. Olivier is committed to an active lifestyle and improving his health through sports and technology. Additional tracking devices that reinforce his motivation could benefit his rehabilitation.

Elise Claes represents Persona 6. She is a 72-year-old retired woman, manages diabetes, hyperlipidaemia, and hypertension after undergoing a PCI. She walks daily but does not track her activity. Though she enjoys technology, using her computer and smartphone daily, she has no experience with health apps and has never used a tablet. Her BMI (27 kg/m^2), HRmax (111 bpm), and VO2max (14 ml/kg/min) indicate moderate physical limitations, though she remains independent and mobile as shown in her HeartQoL score (31). While neutral about sports, she is open to new technology that improves her well-being. Digital health solutions that reinforce her motivation and are designed for ease of use could benefit her.

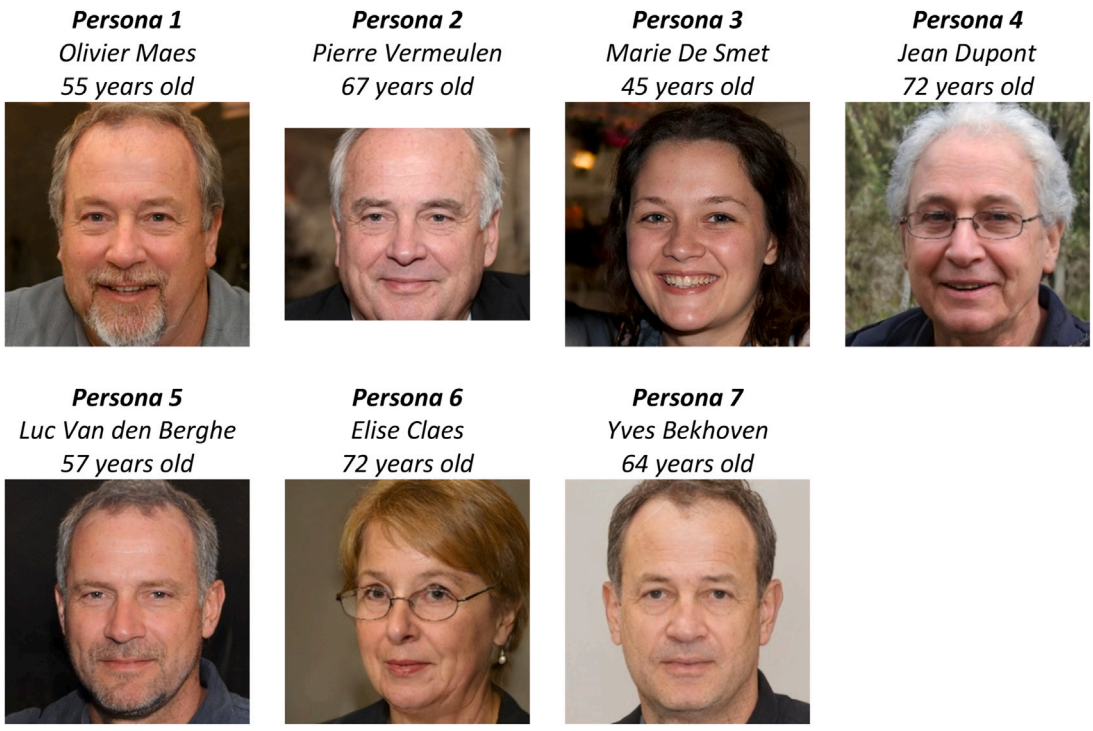

Fig. 4. SharedHeart personas. Images are AI-generated faces courtesy of ThisPersonNotExist.org.

4.4. Final check

The validators were given the final description of the personas, which were brought to life through carefully selected images, as shown in the previous section. As mentioned earlier, there is no intention to make any further changes. However, it was explained that the personas can be extended to include additional characteristics related to the context in which the persona will be used.

5. Discussion

The aim of this study was to design a set of personas to facilitate innovation in eHealth by making available the main needs and characteristics of the intended users. Our approach involved data from a prior RCT and 5 phases with a series of steps for designing personas: (Phase I) expert analysis of data and variables selection, (Phase II) clustering, (Phase III) expert validation, (Phase IV) persona optimisation, and (Phase V) final check. In the following paragraphs, we discuss the main findings and implications and describe limitations and future work.

5.1. Strengths of the approach

Our approach demonstrates advantages, including cost and time efficiency and a commitment to consistency, quality, and trustworthiness. Regarding cost and time efficiency, as mentioned before, conducting interviews, focus groups, surveys, or other methods that involve actual patients, to understand their needs and characteristics is time-consuming and expensive. Additionally, these studies are subject to increasing ethical and other regulations that make it difficult for researchers to conduct them. In contrast, using data collected during the UCD process, previous studies or throughout the project may reduce the cost and time needed to collect data to create personas for future designs. Previous research has demonstrated the feasibility of using data collected in previous studies to build their personas (Holden et al., 2017; ten Klooster et al., 2022; Williams et al., 2021). In our experience, the time to prepare the protocol to get information about patients, the ethical request and response, the time for recruitment and performing a user needs study could take between 6 and 12 months. Therefore, using data from a previous RCT (or intervention study) speeds up the process when the RCT's target group is the same or similar to the new target group, because the data is already there, and the following steps as outlined in this article are straightforward once a researcher or practitioner familiarised with them. However, the reuse of available data must be done with careful consideration of privacy and GDPR. It is recommended that an opt-in option be incorporated into the informed consent forms, e.g., "In addition to the study I am participating in, I agree that researchers may use the data collected for additional studies on the same disease as mine". This was the case in the informed consent of our previous SharedHeart RCT. Following this approach ensures transparency and enhances participant understanding regarding the use of their data. Additionally, the use of automated tools such as algorithms and LLM technologies also speeds up the initial design, which the involved experts will later improve.

Another advantage is the consistency that data provides. While creating personas manually often lacks clear guidelines, leading to inconsistencies, the use of data and clustering techniques ensures consistency in the resulting groups and personas. That consistency means that if two experts design personas based on these data clusters, they will likely produce similar outcomes. On the other hand, the results will likely differ significantly if the same experts create personas from scratch based solely on their assumptions and perspectives.

Persona validation is one of our key contributions to the field, as it provided more quality to the designed personas and built trust in our experts. While previous studies have highlighted the need to conduct such validations, we have identified only a limited number of examples of this practice and even fewer demonstrating the methodologies used. LeRouge et al. (2013) reported a review of the personas by the research team but did not describe the validation methods used. Olivares et al. (2020) conducted a validation study that included two phases: internal validation by researchers and external validation by clinicians. After each phase, the personas were improved. Vosbergen et al. (2015) conducted a validation study, asking patients whether they identified with one of the five personas. Likewise, Williams et al. (2021) tested their designed archetypes by using them in a serious

game to assess their acceptability, accessibility, and alignment with people's experiences. Participants provided feedback, indicating that at least one archetype was related to them or someone they knew. Li et al. (2024) surveyed 95 breast cancer patients to validate personas in weight management. They found that 51.58% of the patients involved identified with one of the five personas, while 48.42% found two different personas relevant to them. In contrast, Holden et al. (2017) conducted non-parametric tests to compare clusters and, in that way, validate their clusters.

Our validation approach was somewhat similar to that of Olivares et al. (2020), who worked with expert researchers and clinicians. In their validation, they conducted a 60-minute focus group meeting. They evaluated the personas individually for over 20 min and then had an open discussion. In our approach, we emailed the questionnaire (included in the Supplementary Material A) to give them enough time to analyse the personas and give each one a realism score. Then, we conducted individual discussions to ensure equal participation from various rehabilitation disciplines in the decision process. Their consistent and high realism scores (mean = 8.0) indicated that they found the personas representative of their actual patients. The two-step approach of the validation, with the questionnaire followed by the meetings, ensured the multidisciplinary perspective and facilitated equal-opportunity participation of the different caregiver roles. The quantitative realism score urged the validators to a concrete decision on the value of each persona, whereas the open questions and the meetings were inviting to bring in discipline-specific comments on their scores or the personas in general. The cardiologists and physiotherapists, for instance, provided valuable comments on the values per selected variable in the personas, and on the combination of values in relation to the indication for rehabilitation. A physiotherapist was very much aware of the impact of the personas on future application designs. Via the question for the full medical history, he mainly focused on safety for the patients that would be using an application prescribing exercise. The psychologist was familiar with the concept of ''archetypes'' to represent different personalities from a psychological point of view, and intuitively reflected on coverage of the patient population with the personas. He started sketching the diversity in patients with respect to the way they give meaning to their heart disease from the point of view of qualitative research and how they motivationally cope with it (Callebaut et al., 1995). Patients experience going through a heart disease as a disruption of physical and psychological integrity, with loss of autonomy and control. Some cope in an open, constructive way with their disease (health and lifestyle as a challenge, a new chance), others experience their heart disease as a frightening threat (controlling coping). The qualitative approach shows that patients can make use of technology and devices from completely different motivational drives (enthusiastic enjoying technique versus clinging to technique out of fear).

The validation process showed that the creation of personas cannot be totally automated because of the risk to lose important characteristics that only experts know from their expertise and interaction with patients. Furthermore, as our experts were actively engaged in the process, the quality of the created personas improved. We integrate qualitative variables as discussed with the psychologist, regarding coping with the disease and motivation to rehabilitate. For example, Luc (Persona 5) appears to cope better with the disease and is more intrinsically motivated than Marie (Persona 3). Luc has a structured and practical approach to life, which likely helps him adapt to rehabilitation more effectively. In contrast, Marie sticks rigidly to what she considers safe and familiar out of fear, indicating that she struggles more with coping with the disease. Luc has recently undergone CABG, prompting a renewed focus on his health. His willingness to integrate new habits and use technology suggests openness to change. Meanwhile, Marie is more resistant to technology and less engaged in physical activities. Her sedentary lifestyle indicates a lower level of motivation. These additions confirm that using real patient data generates greater engagement in persona design and, hopefully, fosters greater confidence in their use. Finally, we recommend that the researchers reflect on whether they need informed consent from the people they invite as validators for the validation process. Generally, consent is not required if the validators are professionals who perform this task as part of their job on the project. However, consent may be necessary if background information or personal details are collected from validators or in preparation for certain scientific publications.

In summary, our approach is consistent with the findings of other studies (ten Klooster et al., 2022; Williams et al., 2021) that recommend combining quantitative and qualitative data to create more detailed and contextualised persona descriptions. Additionally, it also highlights that we should focus on patients' typical problems, such as emotions and motivations, which are typically not collected in all studies as standardised questionnaires. Addressing these factors, we could create more fine-tuned personas that enhance the robustness of future designs (Dol et al., 2023).

5.2. Limitations and future work

The presented study acknowledges limitations, such as potential biases in the selection of variables and algorithms, and a relatively small sample. We created personas from a holistic perspective based on several guidelines (Holden et al., 2017; LeRouge et al., 2013), previous works (ten Klooster et al., 2022; Williams et al., 2021), and on our expertise. We selected the relevant data from the available datasets. With the help of medical colleagues, we analysed the datasets, selected the variables that could accurately describe our target patients, and we classified the data through a comprehensive analysis and discussions with our colleagues. When classifying the data into the different domains based on the different examples, it was clear that the health-related data is typically available, while the personal preferences of patients are less represented and less standardised. Consistent with existing literature, this research found that to have more fine-tuned personas, it is necessary to collect person-related data in the studies (ten Klooster et al., 2022). Additionally, this data collection should be more standardised (Holden et al., 2017). While there are ongoing efforts to standardise digital and health literacy questionnaires (Scherrenberg et al., 2023), the burden on participants when completing these questionnaires complicates their application. However, future research will attempt to include questionnaires and qualitative techniques with psychological factors that might help identify motivations and coping strategies in patients with heart disease.

With the final set of variables selected, we performed various analyses to find the optimal number of clusters and obtained a final set of seven clusters for designing personas. Our approach involved mixed data using a k-prototype algorithm that can both process numerical and categorical (ordinal) data. In this analysis, all variables were assigned equal weight in the clustering process. An alternative approach could involve identifying which factors are more critical for our personas and applying different weightings to prioritise those specific characteristics in the clustering process.

This study had a relatively small sample size, which presents some challenges for cluster analysis studies. Despite choosing a stopping rule of 7 clusters to have 7 clusters with an average of 10 observations, one major drawback was the emergence of one small cluster consisting of only two observations. There was also little variability in our sample due to the inclusion in the RCT with regard to certain variables, such as gender (i.e., 13% female). Interestingly, the cluster with only two observations represented the only female group in the clustering. Future work could include repeating the methodology in more extensive studies or combining data from similar studies to increase the sample for the clustering. Additionally, there is a need for increased transparency in the persona creation process. Making the results available and sharing the approaches, methods, and data used is essential. These actions

might also increase the amount of data for future persona creation and, at the same time, improve the processes.

Another line of research is to demonstrate the practical value of personas by putting them into practice. To this end, the personas reported in this paper are currently being used to design an eHealth application focusing on different kinds of physical activity. The target population and the focus of the application – supporting patients to strive for and maintain an active lifestyle – are similar to the former RCT study in which the information was collected. As the researcher of this new project was not involved in the former RCT, the personas represent their future users, which leads to efficiency gains in the initial analysis and design stages. The researcher added new characteristics to complete the personas for the project context, for example, their actual level of physical activity. We have selected three personas to guide the design, and the remaining personas will be used in the (initial) evaluation phases to perform an expert assessment of early UCD artefacts such as task-/dialogue models and low-fidelity prototypes. It should be noted that personas should not replace real, representative patients in the usability engineering process. In our project, they are involved at the latest when an interactive low- or high fidelity prototyping is ready for usability testing.

6. Conclusion

This research introduces a detailed methodology for designing personas that help inform the design of eHealth applications and interventions for CVD patients. In this study, we created seven multidimensional personas by clustering both quantitative and qualitative data. For that, we used data previously collected in an RCT, and we ensured the privacy of participants while enhancing the accessibility of information about patients. After clustering, a validation phase involved a multidisciplinary team assessing the realism of these personas and providing suggestions for improvements. Our approach demonstrates the need for more standardised data on patients' personal characteristics, and the validation process revealed that creating personas cannot be fully automated, as this may result in the loss of essential characteristics that only experts can identify through their interactions with patients. The next step in this research is to demonstrate the practical use of data-driven personas.

CRediT authorship contribution statement

Rosa Lilia Segundo Díaz: Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Sevda Ece Kizilkilic:** Writing – review & editing. **Wim Ramakers:** Writing – review & editing. **Dominique Hansen:** Writing – review & editing. **Paul Dendale:** Writing – review & editing. **Karin Coninx:** Writing – review & editing, Conceptualization.

Declaration of the use of AI

During the preparation of this work the first author used Grammarly in order to improve readability and language of the work. After using this tool, the author reviewed and edited the content as needed and take full responsibility for the content of the publication. Additionally, within the methodology of this paper, ChatGPT and ThisPersonNotExist.org tools were used to generate narratives from data and AI-generated faces, respectively.

Funding

This research and the SharedHeart study were supported by H2020 CoroPrevention (grant 848056). The design and development of the SharedHeart applications were supported by UHasselt Special Research Fund (grant BOF18DOC26).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank all validators, including Kim Bonné and Frank Vandereyt, in addition to the co-authors, for their valuable contribution to this work, in particular for insightful discussions and valuable feedback during the validation process.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chbr.2025.100872.

Data availability

Data will be made available on request.

References

Bonneux, C., Hansen, D., Dendale, P., & Coninx, K. (2022). The SharedHeart approach: Technology-supported shared decision making to increase physical activity in cardiac patients. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering: vol. 431 LNICST*, (pp. 469–488). Springer Science and Business Media Deutschland GmbH, http://dx.doi.org/10.1007/978-3-030-99194-4_29.

Breeman, L. D., Keesman, M., Atsma, D. E., Chavannes, N. H., Janssen, V., van Gemert-Pijnen, L., Kemps, H., Kraaij, W., Rauwers, F., Reijnders, T., Scholte op Reimer, W., Wentzel, J., Kraaijenhagen, R. A., & Evers, A. W. (2021). A multi-stakeholder approach to ehealth development: Promoting sustained healthy living among cardiovascular patients. *International Journal of Medical Informatics*, *147*, Article 104364. http://dx.doi.org/10.1016/J.IJMEDINF.2020.104364.

Callebaut, J., Janssens, M., Lorre, D., & Hendrickx, H. (1995). *The naked consumer: The secret of motivational research in global marketing*. Censydiam Inst.

Chapman, C. N., & Milham, R. P. (2006). The personas' new clothes: Methodological and practical arguments against a popular method. In *Proceedings of the Human Factors and Ergonomics Society* (pp. 634–636). SAGE PublicationsSage CA: Los Angeles, CA, http://dx.doi.org/10.1177/154193120605000503.

Cooper, A. (1999). *The Inmates are Running the Asylum*. Wiesbaden: Vieweg+Teubner Verlag, http://dx.doi.org/10.1007/978-3-322-99786-9_1, 17–17.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise*, *35*(8), 1381–1395. http://dx.doi.org/10.1249/01.MSS.0000078924.61453.FB.

Dol, A., Kulyk, O., Velthuijsen, H., van Gemert-Pijnen, J., & van Strien, T. (2016). Denk je zèlf! Developing a personalised virtual coach for emotional eaters using personas. *International Journal on Advances in Life Sciences*, *8*(3 & 4), 42–47.

Dol, A., van Strien, T., Velthuijsen, H., van Gemert-Pijnen, L., & Bode, C. (2023). Preferences for coaching strategies in a personalized virtual coach for emotional eaters: an explorative study. *Frontiers in Psychology*, *14*, Article 1260229. http://dx.doi.org/10.3389/fpsyg.2023.1260229.

Engelmann, P., Eilerskov, N., Thilsing, T., Bernardini, F., Rasmussen, S., Löwe, B., Herrmann-Lingen, C., Gostoli, S., Andréasson, F., Rafanelli, C., Pedersen, S. S., Jaarsma, T., & Kohlmann, S. (2023). Needs of multimorbid heart failure patients and their carers: a qualitative interview study and the creation of personas as a basis for a blended collaborative care intervention. *Frontiers in Cardiovascular Medicine*, *10*, Article 1186390. http://dx.doi.org/10.3389/fcvm.2023.1186390.

Eurostat (2024). *Occupational diseases statistics: Statistics Explained: Technical report december 2023*, (pp. 1–10). European Heart Network, URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cardiovascular_diseases_statistics.

Floyd, I. R., Jones, M. C., & Twidale, M. B. (2008). Resolving incommensurable debates: A preliminary identification of persona kinds, attributes, and characteristics. *Artifact*, *2*(1), 12–26. http://dx.doi.org/10.1080/17493460802276836.

Frederix, I., Solmi, F., Piepoli, M. F., & Dendale, P. (2017). Cardiac telerehabilitation: A novel cost-efficient care delivery strategy that can induce long-term health benefits. *European Journal of Preventive Cardiology*, *24*(16), 1708–1717. http://dx.doi.org/10.1177/2047487317732274.

Grudin, J. (2006). Why personas work: the psychological evidence. *The Persona Lifecycle*, 642–663. http://dx.doi.org/10.1016/B978-012566251-2/50013-7.

Haldane, V., Koh, J. J. K., Srivastava, A., Teo, K. W. Q., Tan, Y. G., Cheng, R. X., Yap, Y. C., Ong, P. S., van Dam, R. M., Foo, J. M., Müller-Riemenschneider, F., Koh, G. C. H., Foong, P. S., Perel, P., & Legido-Quigley, H. (2019). User preferences and persona design for an mhealth intervention to support adherence to cardiovascular disease medication in singapore: A multi-method study. *JMIR MHealth and UHealth*, *7*(5), http://dx.doi.org/10.2196/10465.

Holden, R. J., Kulanthaivel, A., Purkayastha, S., Goggins, K. M., & Kirpalani, S. (2017). Know thy ehealth user: Development of biopsychosocial personas from a study of older adults with heart failure. *International Journal of Medical Informatics*, *108*, 158–167. http://dx.doi.org/10.1016/J.IJMEDINF.2017.10.006.

Huang, Z. (1997). Clustering large data sets with mixed numeric and categorical values. In *Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 21–34).

ISO 9241-210 (2019). ISO 9241-210: Ergonomics of human–system interaction - human-centred design for interactive systems. *International organization for standardization*: *vol. 2*, (p. 32). International Organization for Standardization, URL https://www.iso.org/obp/ui#iso:std:iso:9241:-210:ed-2:v1:en.

Jansen, B. J., Jung, S. G., Salminen, J., Guan, K. W., & Nielsen, L. (2021). Strengths and weaknesses of persona creation methods: Guidelines and opportunities for digital innovations. In *Proceedings of the Annual Hawaii International Conference on System Sciences*: *vol. 2020-Janua*, (pp. 4971–4980). IEEE Computer Society, http://dx.doi.org/10.24251/HICSS.2021.604.

Kizilkilic, S. E., Ramakers, W., Falter, M., Scherrenberg, M., Bonneux, C., Pieters, Z., Milani, M., Hansen, D., De Pauw, M., Coninx, K., & Dendale, P. (2025). A digitally-supported shared decision making approach for patients during cardiac rehabilitation: a randomized controlled trial. *European Journal of Preventive Cardiology*, http://dx.doi.org/10.1093/eurjpc/zwaf537.

LeRouge, C., Ma, J., Sneha, S., & Tolle, K. (2013). User profiles and personas in the design and development of consumer health technologies. *International Journal of Medical Informatics*, *82*(11), e251–e268. http://dx.doi.org/10.1016/J.IJMEDINF.2011.03.006.

Li, X., Zhang, N., Yang, J., Geng, Z., Zhou, J., & Zhang, J. (2024). Weight management personas of breast cancer patients undergoing chemotherapy in China: a multi-method study. *BMC Medical Informatics and Decision Making*, *24*(1), 1–10. http://dx.doi.org/10.1186/S12911-024-02515-1.

Olivares, M., Pigot, H., Bottari, C., Lavoie, M., Zayani, T., Bier, N., Le Dorze, G., Pinard, S., Le Pevedic, B., Swaine, B., Therriault, P. Y., Thépaut, A., & Giroux, S. (2020). Use of a persona to support the interdisciplinary design of an assistive technology for meal preparation in traumatic brain injury. *Interacting with Computers*, *32*(5–6), 435–456. http://dx.doi.org/10.1093/IWCOMP/IWAB002.

Patkar, N., & Seyff, N. (2023). Data-driven persona creation, validation, and evolution. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*: *vol. 13975 LNCS*, (pp. 262–271). Springer, Cham, http://dx.doi.org/10.1007/978-3-031-29786-1_18.

Salminen, J., Guan, K., Jung, S. G., Chowdhury, S. A., & Jansen, B. J. (2020). A literature review of quantitative persona creation. In *Conference on human factors in computing systems - proceedings* (pp. 1–14). Association for Computing Machinery, http://dx.doi.org/10.1145/3313831.3376502.

Salminen, J., Guan, K., Jung, S. G., & Jansen, B. J. (2021). A survey of 15 years of data-driven persona development. *International Journal of Human-Computer Interaction*, *37*(18), 1685–1708. http://dx.doi.org/10.1080/10447318.2021.1908670.

Scherrenberg, M., Falter, M., Kaihara, T., Xu, L., van Leunen, M., Kemps, H., Kindermans, H., & Dendale, P. (2023). Development and internal validation of the digital health readiness questionnaire: Prospective single-center survey study. *Journal of Medical Internet Research*, *25*(1), Article e41615. http://dx.doi.org/10.2196/41615.

Szepannek, G., Aschenbruck, R., & Wilhelm, A. (2024). Clustering large mixed-type data with ordinal variables. *Advances in Data Analysis and Classification*, 1–19. http://dx.doi.org/10.1007/S11634-024-00595-5.

ten Klooster, I., Wentzel, J., Sieverink, F., Linssen, G., Wesselink, R., & van Gemert-Pijnen, L. (2022). Personas for better targeted ehealth technologies: User-centered design approach. *JMIR Human Factors*, *9*(1), http://dx.doi.org/10.2196/24172.

Vosbergen, S., Mulder-Wiggers, J. M., Lacroix, J. P., Kemps, H. M., Kraaijenhagen, R. A., Jaspers, M. W., & Peek, N. (2015). Using personas to tailor educational messages to the preferences of coronary heart disease patients. *Journal of Biomedical Informatics*, *53*, 100–112. http://dx.doi.org/10.1016/J.JBI.2014.09.004.

WHO (2024). *World health statistics 2024: monitoring health for the SDGs, Sustainable Development Goals* (pp. 1–96). Geneva: World Health Organization.

Wilkins, E., Wilson, L., Wickramasinghe, K., Bhatnagar, P., Leal, J., Luengo-Fernandez, R., Burns, R., Rayner, M., & Townsend, N. (2017). *European cardiovascular disease statistics 2017*. European Heart Network.

Williams, A. J., Menneer, T., Sidana, M., Walker, T., Maguire, K., Mueller, M., Paterson, C., Leyshon, M., Leyshon, C., Seymour, E., Howard, Z., Bland, E., Morrissey, K., & Taylor, T. J. (2021). Fostering engagement with health and housing innovation: Development of participant personas in a social housing cohort. *JMIR Public Health and Surveillance*, *7*(2), Article e25037. http://dx.doi.org/10.2196/25037.