Factors Influencing Fertility & Menstrual App Uptake by Diverse Women

A Stated Choice Experiment

E.C.I. Hogervorst

TU Delft

Factors Influencing Fertility & Menstrual App Uptake by Diverse Women

A Stated Choice Experiment

by

E.C.I. Hogervorst

to obtain the degree of Master of Science
at the Delft University of Technology,

Student number:	4676467	
Project duration:	September, 2023 - May, 2024	
Thesis committee:	Dr. E.J.E.(Eric) Molin,	TU Delft, First Supervisor & Committee Chair
	Dr. S. (Saba) Hinrichs-Krapels,	TU Delft, Second Supervisor
	Dr. C. (Caroline) Figueroa,	TU Delft, Advisor

Preface

The topic of this thesis sprouted from my interest in female health and eagerness to understand and employ Discrete Choice Modelling. Looking back at the past 9 months, the combination of this topic and method perhaps did not simplify the execution of this thesis. As the targeted sample group and employed method caused some obstacles along the way. However, I am very glad that I got to dive into an important and relevant subject. As of the start of this year the Dutch minister of Medical Health called raised the issue of female health, I was reminded again of the importance of this topic. Additionally, I met such motivated and knowledgeable women all concerned with this topic, which was such a nice motivation.

The past few months would not have been possible without the help of a few people pushing me and supporting me in various ways. I would like to thank my graduation committee for their support and guidance. First of all, I would like to express my gratitude to the chair of my committee, Eric Molin. Whenever I sent any question or part of my report your way, I would receive very helpful feedback very quickly. I believe this has immensely helped my thesis to move forward. More importantly, when I lost faith in the quality of my research, you reassured me. Secondly, I would like to thank Caroline Figueroa for always being available for feedback. Mostly, I would like to thank you for introducing this interesting topic to me and sharing your extensive knowledge on diversity in health app design. Lastly, I would like to thank Saba Hinrichs-Krapels, for making time in your busy schedule to complete my committee and providing constructive feedback. You often knew which questions to ask to help me improve my work. Additionally, I would like to thank Fanchao Liao, for selflessly making the time in her schedule to assist me with my model when I was stuck.

Additionally, I would like to thank my parents Louk and Mike for supporting me and trying to understand my thesis whenever I needed help. More so, I want to thank you for your unwavering support during my whole student career. Furthermore, I would like to thank all my friends for their interest in my thesis. Especially, I want to thank my friends Fleur and Puck, who started their thesis research in September as well. Our brainstorming sessions in the beginning have finally paid off. Also I want to thank my old roommate Lotte for always being available to discuss DCM-related issues. I look forward to our graduation lunch at Huszar. Lastly, I want to thank my roommates for putting up with me neglecting my cooking and cleaning tasks.

This report signifies the ending of my student years, which I have enjoyed so much. The TPM faculty has always felt like home and studying Engineering and Policy Analysis has kept me interested in and knowledgeable on a broad range of topics. I am very grateful for the past 7 years.

E.C.I. Hogervorst
Amsterdam, May 2024

Summary

Recently, the deficiency present in female-specific health has been emphasized.Knowledge and consequently treatments with regards to female-specific symptoms and chronic deceases have been lacking. Previous research has shown the opportunity of female health apps to decrease this discrepancy. Female health apps simultaneously contribute to women's health understanding and awareness, while also supplying data for necessary female-specific health research. However, this requires uptake and engagement among consumers. More importantly, it requires uptake among diverse women subgroups of the population. Otherwise, the apps likely contribute to increased health disparities present between different socio-demographic groups and biased female health research.

Therefore, research uncovering the influence of app factors on female health app uptake, while accounting for the diversity in the female population is necessary. The specific female health apps under scrutiny in this research are fertility & menstrual tracking apps (MTA), due to their relative popularity, diverse potential consumer group and importance in women's lives. Consequently, the following research question is answered in this research to contribute to closing the knowledge gap of MTA preference among diverse women:
'To what extent do various factors influence fertility & menstrual tracking app uptake by women from diverse backgrounds?'

The Discrete Choice Modelling method is use to predict uptake, estimate the extent to which factors influence app choices and determine the influence of personal characteristics on preferences. A Mixed Logit (ML) model and Latent Class Choice Model (LCCM) are estimated to determine trade-offs consumers make when choosing to download a MTA. Personal characteristics, such as socio-demographics, attitudes and MTA familiarity are included to determine their influence on MTA preferences. Furthermore, the LCCM identifies the potential consumer segments present among potential MTA consumers.

These models are estimated based on data retrieved from an online Discrete Choice Experiment (DCE), distributed through a survey. The survey also contained questions eliciting MTA use, attitudes towards MTAs and additional app characteristics excluded from the DCEs. Within the DCEs respondents were asked to choose between two app alternatives that varied in costs, accuracy, privacy level, communication style, ease of use, extent of personalisation, and extra function availability. These attributes were derived from a literature review that synthesized research focused on factors influencing MTA uptake and factors influencing general mHealth uptake by people with a low Socio-Economic Position (SEP). To predict actual MTA uptake, respondents were asked if they would download their chosen app or prefer to download no app at all.

The estimated LCCM identified two consumer segments present. These are the non-adopter and the adopter segment. The non-adopter segment choose to not download a MTA often, regardless of the MTA's features. Consequently, no MTA preferences were estimated for that segment. The adopter segment was mostly influenced by accuracy, followed by costs and then privacy. Communication and personalisation influence the choice to a moderate extent and ease of use and extra functions have inconsiderate influence. A respondent that is younger than 40, is familiar with MTAs, positive towards MTAs and values female health knowledge as important, is likely to belong to the adopter sector. When no segment distinction within the consumer group was made, the estimated ML predicted similar preferences to the LCCM, except costs is more important than accuracy and personalisation more important than communication. Additionally, this model revealed that age, ethnicity, education and attitude towards MTA significantly influence the choice to download a MTA. Furthermore, Dutch respondents are less influenced in their decision for a MTA by the level of personalisation than non-Dutch respondents are.Lastly, the app characteristic medical grounding was deemed important by respondents, to a higher extent by respondents inexperienced with MTA use.

In conclusion, app developers should focus on cost, accuracy and privacy, as they influence uptake MTA to a great extent. The most important factor being either cost or accuracy depending on the willingness to adopt in the consumer group. Additionally, medical grounding of the app should be focused on, especially when targeting inexperienced MTA consumers. Furthermore, age, education, ethnicity and MTA attitude should also be considered when developing and marketing a MTA. Lastly, when targeting non-Dutch respondents, more emphasize should be put on ensuring a high level of personalisation. Further research should focus on 1) a more diverse and larger sample group 2) measuring ease of use differently 3) influencing factors that are not app-specific such as policies, societal context or recommendations 4) influence of other background variables 5) the reasoning behind choices.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
ASC	Alternative Specific Constants
DCE	Discrete Choice Experiment
DCM	Discrete Choice Modelling
FABM	Fertility Awareness Based Method
FTA	Fertility Tracking Application
LCCM	Latent Class Choice Model
MARS	Mobile Application Rating Scale
MASUN	Method of App Selection Based on User Needs
ML	Mixed Logit
MNL	Multinomial Logit
MTA	Menstrual Tracking Application
PRAT	Patient Readiness Assessment for Telehealth Tool
RUT	Random Utility Theory
SEP	Social Economic Position
SP	Stated Preference
TAM	Technology Acceptance Model

1 Introduction

1.1. Problem Introduction

An increasing amount of research highlights a deficiency in female-specific knowledge within medical research. This deficiency arises from medical research mostly including male subjects in human studies and subsequently generalizing the findings to the population as a whole (Angum et al., 2020; Maas & Appelman, 2010; Peterlin et al., 2011). Due to the lack of female-specified healthcare research, accurately diagnosing female-typical symptoms and diseases is more challenging for healthcare professionals (Maas & Appelman, 2010). A telling example is endometriosis, a condition whose symptoms have been dismissed as period symptoms due to lack of awareness and knowledge (Becker et al., 2017).

This gender disparity extends to digital healthcare, with technical products or services designed to improve healthcare often disregarding female-specific health issues. For example, when Apple launched Apple Health, claiming to enable measuring all desired health metrics, they failed to include period-tracking for an entire year (Tiffany, 2018).

As a response the past few years have seen the rise of FemTech, which is the overarching term for digital services and products designed to improve female healthcare. As articulated by Corbin (2020) FemTech 'not only assists women in understanding their own bodies, but also enhances scientific knowledge and research about the female population as a whole'. This technological development continues to grow and consequently, its market worth is expected to rise to 60 billion dollar by 2027 (Eijkholt, 2021). Next to technological products, such as wearable breastpumps, a big share of FemTech is mobile health (mHealth), which are health applications on mobile devices. These apps often use personal input data and algorithms to provide personalized medical information focused on fertility, pregnancy and reproductive, post-natal, hormonal, menstrual and sexual healthcare. An example is Glow, a menstrual and fertility tracking app that uses AI and data to provide personalized information about ovulation, pregnancy and women's[1] health (Almeida et al., 2022). Despite this promising development, research has pointed out that FemTech app implementation comes with challenges. According to research, challenges range from privacy and regulatory concerns due to sensitive data use to medical reliability and accuracy of the information provided. (Corbin, 2020; Faubion, 2021; Freis et al., 2018; Gross et al., 2021; Gutterman, 2023; Jacobs & Evers, 2023).

Additionally, FemTech products and services have been critiqued on the basis of reinforcing social inequalities (Gutterman, 2023; Hendl & Jansky, 2022). Research studies regarding these apps often lack diverse respondent samples, resulting in under-representation of women from diverse backgrounds. (Pratap et al., 2020). As a consequence, research finds that adequate consideration of the unique health needs and digital capacities of these women is lacking (Corbin, 2020; Figueroa et al., 2023; Hughson et al., 2018). This disparity between diverse women's health needs and the digital health tools available is reflected in the low mHealth app uptake among women from minority groups Figueroa et al. (2021) and Hughson et al. (2018).

Initially, multiple sources emphasized the potential of mHealth to reduce health disparities, by facilitating enhance self-awareness, providing knowledge and lowering healthcare access barriers

[1]In the survey only respondents that identified as women participated, which is why the term women is used throughout this paper. However, the author acknowledges that not all and not only women menstruate.

(Ramos et al., 2021; Sarkar et al., 2016) However, low uptake of menstrual apps by women from diverse backgrounds does not only counteract this opportunity, but potentially exacerbates health disparities already present. As primary prevention and monitoring through mHealth relies on consumer engagement (Nittas, Mütsch, & Puhan, 2020), the potential of mHealth apps to enhance female health knowledge and improve women's health, is dependent on uptake among all women. Due to limited access, low digital literacy, lack of inclusion in research and consequential unknown mHealth preferences, current female-focused mHealth apps fail to meet the needs of demographic groups impacted by disparities (Smith & Magnani, 2019).

Increasing uptake requires closing the gap between available mHealth apps and women's needs. Therefore, knowledge on women's mHealth preferences and desires are required, as understanding low app uptake causes is imperative. Furthermore, a thorough understanding of the diversity among possible consumer group preferences is necessary. This will contribute to reducing barriers for app uptake and tailoring app preferences to the diverse needs of women. By doing so, mHealth apps opportunities to contribute to improving the health of all women will increase.

1.2. Scope

Due to time and resource constraints, it is important to clearly define the scope of this research. First of all, it is crucial to specify the category of female health app under investigation. FemTech, or female health apps, encompasses a wide variety of apps, such as menopause, pregnancy, menstruation, ovulation, post-partum or breast self-check apps (Gambon et al., 2020). However, given significant differences between these apps and their user demographics, this research concentrates on one type of female health app. Fertility and menstrual tracking apps, i.e.: apps that focus on menstruation and ovulation by supplying information and providing tracking features, are chosen for their extensive potential user group, primarily women within fertile age ranges (ages 12 to +/- 45), compared to the user group of for example post-partum apps. Also, fertility and menstrual tracking apps, are currently among the most popular FemTech apps.

Throughout studies the terms Fertility Tracking Application (FTA) and Menstrual Tracking Application (MTA) are used interchangeably. However, the term FTA is sometimes also used to describe apps focused on Fertility Awareness Based Methods (FABM). The main aim of FABM apps is to provide a natural alternative to conception methods. These apps are often more extensive and require more user input and knowledge than regular FTAs and MTAs. As those apps do not fall under this research scope, the term MTA is used in this research to indicate fertility & menstrual tracking apps.

Additionally, a clear definition of women from minority groups considered in this research is required. Existing research identifies subsets such as: Women from racial, linguistic or ethnic minorities, as well as older, disabled, low-income, low educated, transgender or homosexual women (Corbin, 2020; Figueroa et al., 2021; Hankivsky et al., 2010; Jacobs & Evers, 2023). Within the Netherlands specifically, The Dutch ministry of Health, Welfare and Sport, addresses health disparities experienced by minority groups, based on socio-economic status (Volksgezondheid en Zorg, n.d.). Substantiated by Shavers (2007) they indicate socio-economic status with education level, income and profession. Taking a similar approach, the research paper by (Hengst et al., 2023) looks into facilitators and barriers regarding mHealth adoption amongst people with a low social economic position (SEP). They define this as 'vulnerable groups who experience health disparities' and name indicators such as low income, low education and ethnic minorities. Given the similarities between this research and the Dutch ministry's definition, this study will now refer to this subset of the target group as *women with a low SEP*.

Throughout this research the terms factors, characteristics, facilitators, barriers and features are used to describe various aspects influencing app uptake. The term factors is used to denote everything influencing app uptake, from personal characteristics such as age to societal context to app features. It also encapsulates facilitators and barriers. An app characteristic is used to describe anything 'surrounding' an app. It is therefore not a feature of the app harbors, but for example the reputation an app has. An app feature is a literal function the app does or does not possess, such as the ability to track your ovulation. Personal characteristics are used to describe socio-demographic characteristics, attitudes and perceptions and MTA user experience.

Lastly, the term 'uptake' is defined as 'the act of downloading and installing a smartphone app' aligning with the definition proposed by Szinay et al. (2020). Measuring actual engagement requires research conducted over an extended period, which is not feasible within the limited timeframe of this study.

1.3. Knowledge Gap Definition

1.3.1. Literature Overview

In the current body of research factors influencing health app uptake have been extensively examined across diverse consumer groups, aiming to bridge the gap between apps and consumer preferences (Abelson et al., 2017; Bidmon et al., 2014; Goodman et al., 2023; Krebs & Duncan, 2015; Mohamed et al., 2011; Nijland et al., 2011; Or & Karsh, 2009). These papers encompass different types of mHealth apps, a variety of research methods and a broad range of consumer groups, including the elderly, indigenous and specific patient groups.

To obtain an general overview of the amount of research on mHealth uptake, the systematic review by Szinay et al. (2020) scrutinized 41 papers under examination revealing 18 factors influencing uptake. Multiple of the reviewed papers looked into mHealth uptake with regards to diverse populations and specifically minorities. Not mentioned in the systematic review is the paper by Sarkar et al. (2016), who underscore the significance of diverse populations with regards to uptake barriers. However, among all of these papers, the female population is underrepresented and research specifically focused on MTAs is missing.

Regarding fertility and period-tracking apps specifically, there is multiple research on women's use of and attitude towards these apps (Al-Rshoud et al., 2021; Broad et al., 2022; Gonçalves et al., 2021; Karasneh et al., 2020; Moglia et al., 2016; Starling et al., 2018; Worsfold et al., 2021). However, most of these papers elicit women's attitudes and ideal app preferences but do not investigate their actual influence on uptake or distinguish these preferences between their subjects. Specifically focusing on consumer preferences and important app factors is the paper by Starling et al. (2018). They conduct a survey asking women to rank app characteristics from 'not at all important' to 'very important'. The respondents were categorized based on their previous knowledge of these apps and possible interactions between the results and the categories were examined. The precise influence of app characteristics on uptake was not within this research's scope. Additionally, interactions with personal characteristics were not included. Gambier-Ross et al. (2018) look into MTAs specifically and recommends that apps should be aimed at a wider audience. Consequently they state that 'research should explore the diverse relationships between different subgroups of women and FTAs [MTAs]'

Research that did focus on subgroup preferences with regards to female health uptake is the research by Mora et al. (2020). They look into pregnancy app preferences of Latin women, by engaging the target group in the app design. The research established a culturally sensitive pregnancy app design. It incorporated culturally specific advice like nutritional advice based on common Latin diets. Additionally, Hughson et al. (2018) performed a narrative review on the impact of pregnancy apps for culturally and linguistic diverse women and found that the uptake among these women is lower. They mention that 'existing research is inconclusive regarding whether, or to what extent, factors related to social disadvantage influence pregnancy app uptake, use, and utility'.

The extent to which certain app factors influence uptake and the impact of differing personal characteristics on uptake has been investigated for health apps in general by among others Nittas, Mütsch, and Puhan (2020), Szinay et al. (2021), and Xie and Or (2023). These papers all employed Discrete Choice Modelling to determine the exact influence of various factors, such as app characteristics and socio-demographics on app uptake.

However, research targeted at fertility and menstrual apps specifically is required due to the unique features of these apps. Firstly, these apps collect very sensitive data, such as sexual activity, sexual orientation and female health specific indicators (Faubion, 2021). Furthermore, these apps serve as a form of contraception, hence the dependability, accuracy and reliability of the apps are determent in app, according to (Starling et al., 2018). Lastly, it is also important to note that the impact of menstrual apps in general remain largely unstudied compared to health apps in general (Corbin, 2020; Taylor, 2021).

In conclusion, research specified to female health and menstrual apps explored attitudes, user experiences and obstacles regarding mHealth app uptake. Additionally, influences of various factors have been explored for mHealth in general. However, to the best of the author's knowledge, to what extent which exact factors influence the decision of women from different backgrounds to specifically download a fertility app remains unclear. Some factors have been identified, but their precise influence on actual uptake is not clear. Furthermore, preferences and factor influence differences between groups remain under-researched. This knowledge gap will be defined more precisely in the following section, leading up to the research question aiming to fill that knowledge gap.

1.3.2. Research Questions

As previous studies have shown, foundational research on factors influencing MTA uptake has been executed. However, the extent to which these factors influence uptake is unclear. Additionally, multiple user groups have been investigated, however no comparisons have been made. Therefore, it remains unclear which MTA features influence uptake for differing women. It is inconclusive if uptake is influenced by differing app preferences and to what extent.

Taking all of this into account, it is clear that there is a knowledge gap concerning the following questions: What menstrual app features (e.g.: amount of data collection) influence women's decisions to download and actively use a MTA? What other app factors (e.g.: attitude towards MTAs) influence women's uptake of a MTA? Additionally, do women from different backgrounds have different factors influencing their choice to download a MTA?

Answering these question is imperative to bridge the knowledge gap in understanding factors influencing preferences and thus adoption. Additionally, exploring whether women have distinct preferences will provide valuable insights to target female subsets specifically. Consequently, both these aspects will be considered when answering the following main research question:

'To what extent do various factors influence fertility & menstrual tracking app uptake by women from diverse backgrounds?'

The following sub-questions will contribute to finding the answer to the research question:

- *What barriers and facilitators for uptake of health apps in general and MTAs in particular have been identified?*

  The answer to this question provides an overview of barriers identified in previous research for MTA uptake and health app uptake by diverse women.
- *How do women from diverse backgrounds currently utilize, perceive and approach MTAs?*

  The answer to this question will supplement the previous question by supplying an insight into females' attitudes towards menstrual tracking apps and the current (lack of) use of these apps.
- *To what extent do certain MTA attributes influence app uptake?*

  Based on the identified attitudes, perceptions and barriers, related app attributes will be selected. The relative influence of these app attributes on how much an app is preferred will be measured through Discrete Choice Experiments. It will quantify to what extent an attribute influences possible uptake compared to the other attributes.
- *To what extent do fertility app attribute preferences and uptake vary based on personal characteristics?*

  Lastly, this question will look into the heterogeneity of the established attribute preferences of the previous question. It will show if women's personal characteristics have significant interaction with their preferences regarding menstrual tracking app attributes. Additionally, potential consumer segments present in the possible MTA consumer group will be identified, by means of a Latent Class Choice Model, to establish their specific preferences. Hence, this question will provide the answer if women have diverse MTA needs.

1.4. Societal Relevance

The knowledge gained in this research, provides insight into the factors influencing uptake of menstrual apps by women from diverse backgrounds. Concretely, the final recommendations will entail app factors to focus on, based on the quantified level of their influence on app uptake and potential variations based on women's personal characteristics. Thereby the results will facilitate a knowledge-based foundation for the development of mHealth tools that align with the needs of diverse women. Aligned with these recommendations, the apps have a higher chance of being adopted by a wider and more diverse female user group.

In practice, these recommendations can assist with designing, deploying, marketing and distributing menstrual apps. These apps can be more accessible for all the different subsets within the female population. Or the recommendations can assist with designing culturally or socio-economically sensitive apps specifically tailored towards certain population subsets. A notable example is the work of Mora et al. (2020), who designed a pregnancy app for the Latino community specifically.

Currently, available apps are often provided by organisations with commercial aims or organisations with both profitable and medical objectives. Such as Ovia Health, a digital platform for family well-being (Ovia Health, 2021). For commercial actors these recommendations support targeted designs, which provides the opportunity for increased uptake or targeting a desired user group specifically. Supplying concrete factors to focus on, might encourage these app developers to be aware of inclusive app development. Similarly, the implications of this research are valuable for medically focused institutions. For instance, The Dutch National Institute for Public Health and Environment, which conducts independent health research and advises the government. Currently participating in the Horizon Europe program, which is concerned with gender equity in research and innovation, one of their important goals is female representation in health knowledge (van Rede, 2021). As discussed in sections 1.1 of this chapter, deployment of MTAs can assist with enhancing female health knowledge on research level as well as improving women's knowledge of their personal health. Moreover, the implications of this research extent to governmental organisations. Such as the Dutch Ministry of Health, Welfare and Sports who aims to decrease gender-based and socio-economically based health disparities prevalent in the Dutch healthcare system (Rijksinstituut voor Volksgezondheid en Milieu, 2018).

The efforts of this research contribute to addressing health disparities in two key ways. Firstly, female health apps have the potential to lower barriers to healthcare access, often experienced by women facing challenges such as childcare responsibilities or financial constraints, more prevalent among women from minority groups (Figueroa et al., 2023). These apps can easily distribute trustworthy medical information, allow for the opportunity to talk to healthcare professionals directly and provide women with the ability to track and self-manage their health (data)(Ramos et al., 2021; Sarkar et al., 2016; Taylor, 2021). Thereby establishing more knowledge, ownership and awareness among women. Secondly, a higher uptake of menstrual apps by diverse women leads to more representative input data (Corbin, 2020). As data gathered by these apps is used for female-specific healthcare research (Erickson et al., 2022). Ultimately contributing to closing gender healthcare disparities and the knowledge gap.

1.5. Method Justification

In this research the influence of certain app attributes and their interaction with personal characteristics on fertility app uptake will be measured by means of a focus group and a survey. The survey will contain questions with regards to Discrete Choice Modelling (DCM). As DCM has been used in previous research regarding health app uptake on multiple occasions, it is proven to be an apt method for this topic (Nittas, Mütsch, & Puhan, 2020; Szinay et al., 2021; Xie & Or, 2023). A short reasoning for the application on this research specifically is explained in this section.

Firstly, a thorough understanding of consumers' motivations for uptake requires establishing to what extent which factors determine fertility app uptake. Bretschneider (2015) and Haile et al. (2018) mention that complexity and unnecessary additions of MTAs can lead to cessation. Therefore, it is necessary to establish which factors and features should be prioritised. Consequently it is necessary to not just elicit preferred features or functions, but determine the trade-offs made between app features by consumers. Additionally, answering the research question requires precisely determining uptake. So uptake should be measured or predicted and additionally factors relations with uptake have to be clear. Furthermore the uptake needs to be measured for multiple consumer groups and allow detection of differences within preferences. Hence, the research method should enable a comparative result on the extent to which all these different factors influence female health app uptake.

The DCM theorem assumes that consumers choose the app that provides the highest satisfaction (or utility which is further explained in 3). Thereby, the consumer makes trade-offs between different app factors, such as offered features, reviews or how the app is presented in the app store (Szinay et al., 2021). Some factors hold more significance for consumers than others, exerting a greater influence on their decision (Xie & Or, 2023). For example, if app A provides more languages but offers lower privacy levels than app B, a trade-off between available languages and privacy levels is made. Through Discrete Choice Experiments (DCE), these trade-offs are recreated in hypothetical choice situations (this is explained thoroughly in chapter 3). By systematically evaluating these factors across various alternatives, statistical analyses can ascertain the level of importance assigned by consumers to specific factors (Xie & Or, 2023).

This allows for systematic comparison of various factors and creates a quantified image of how they are weighed against each other, while incorporating personal characteristics and being able to predict uptake in various scenarios.

1.6. Report Outline

This research report is structured as follows: Firstly, a review of the literature provides an overview of relevant attitudes, perceptions and factors in chapter 2. Secondly, the methodology chapter 3 elaborates on the theory behind DCM, the experimental set-up and model specification choices. Next up, the survey design is discussed in chapter 4 including the experimental design and all the questions included in the survey. After cleaning the data and discussing the respondents' statistics in chapter 5, the model estimations are displayed in the results chapter 6. Which discusses the weights of the factors and their heterogeneity. The model is applied to various scenarios in chapter 7. Lastly, implications, recommendations, limitations and future research opportunities are discussed in the conclusion and discussion chapter 8.

2

Literature Review & Conceptual model

This chapter supplies an overview of the current knowledge on fertility app uptake by means of a literature review. After shortly discussing the literature review approach that was taken, women's perceptions with regards to MTA are elaborated on, leading to the section discussing identified barriers and facilitators. Additionally, barriers and facilitators for mHealth uptake among diverse populations that have been identified in research are explored. The last section contains the conceptual model, which shows the proposed attributes and levels, feedback retrieved from the focus group and the improvements made.

2.1. Literature Review

2.1.1. Literature Review Approach

This literature review was conducted in December 2023. The aim of the literature review was to obtain an overview of the knowledge available in research regarding MTA uptake and to retrieve identified attributes influencing MTA uptake. The search for literature was done by using Google Scholar, PubMed and Scopus with variations of the following search query:

> (fertility OR menstru* OR "period tracking") AND (apps OR mhealth OR "mobile applications") AND ("user experiences" OR attitudes OR perceptions OR preferences OR understanding OR approach) AND NOT soil[2]

In total 250 sources were found. After excluding duplicates across the literature databases, the following exclusion criteria were applied:

- Published before 2013. This is based on the earliest launch of a popular MTA, which is Clue in 2013 (Clue, 2024).
- Focus on the technical and scientific aspects of MTAs. Such as, the algorithms or specific privacy issue mitigations.
- Not focused on fertility or menstrual tracking apps.

This resulted in 34 articles. Those articles were scanned by reading the abstracts and conclusions to exclude articles that did not incorporate the user/consumer experience. Snowballing was applied to avoid missing the most important resources. This resulted in 15 suitable articles that were reviewed.

Additionally, due to the lack of research focusing on people with a low SEP mentioned previously, the selected papers were supplemented by papers investigating mHealth apps in general. First, the paper by Szinay et al. (2020) mentioned in the literature overview was used to obtain the most important papers and establish the following search string.

> ("factors influencing" OR influences OR barriers OR facilitators) AND (uptake OR use OR adoption) AND (mhealth OR "mobile health applications" OR "health app") AND (

[2]The term fertility often resulted in agricultural related papers and was therefore added as a constraint

consumer OR user) AND ("low socioeconomic" OR "low income" OR "low education" OR "ethnic minority" OR minority OR "developing" OR "underprivileged") AND NOT (covid)

The following exclusion criteria were used.

- Targeting too specific consumer groups or consumer groups that did not align with or were not similar to the one under scrutiny in this research
- Focusing on a certain mHealth application that harbored significantly different features and objectives than MTAs
- Focusing on the developer, stakeholder or a healthcare worker point of view instead of the consumer point of view.

This resulted in 7 suited papers. For an overview of all eventually included papers from the two searches above, see appendix B.

2.1.2. Menstrual Tracking Applications

Before the perceptions and attitudes with regards to MTAs are discussed, a well rounded understanding of the apps under scrutiny is established. Currently, an abundance of fertility apps is available in app stores, with Moglia et al. (2016) identifying 1116 apps. With some of the functionalities differing between the apps, most of these apps harbor at least the following functions: Tracking of the cycle/menstruation and the prediction of ovulation, the fertile window and the upcoming menstruation (Moglia et al., 2016; Patel et al., 2023; Zwingerman et al., 2019). MTAs often contain a calendar on which the user can indicate days on which the user menstruated. Upcoming possible ovulation dates, menstruation and fertile days are often also marked on those calendars. Furthermore, other data like period-related symptoms or weight can be tracked and the apps often supply information on menstruation and ovulation (Moglia et al., 2016; Zwingerman et al., 2019). See image 1 for an example of a MTA and the different features it offers.

Figure 1: Example MTA: Period Calendar

For an overview of functionalities of current MTAs, established by Ko et al. (2023) and Zwingerman et al. (2019) see appendix A. As the overview of Zwingerman et al. (2019) shows, the functionalities of MTAs differ but most apps offer tracking and prediction of the menstruation and fertility. The abundance of available features and functions is apparent in these overviews, which stresses the variety of available MTAs and the overwhelming amount of choices users have to make both between and within apps (Zwingerman et al., 2019).

Research has found that the quality of these apps varies significantly. As Zwingerman et al. (2019) found that 22% of the 140 apps they analysed contained serious inaccuracies. Moglia et al. (2016) even state that most free menstrual cycle tracking apps are inaccurate. They ascribed this to the lack of citing medical literature or involvement of healthcare professionals. The consequential inaccuracies manifest in the prediction functions and the health information supplied (Johnson et al., 2018). With some women counting on the predicted fertile window to either avoid or initiate intercourse and the possibility of pregnancy, reliable predictions are highly important (Dudouet, 2022; Zwingerman et al., 2019). The large amount of inaccurate MTAs is probably due to the lack of regulation and guidelines for the scientific standards these apps should comply with (Nair et al., 2023). The lack of regulation also influences the privacy policies of these apps. The apps gather and use data on sensitive and intimate topics, such as intercourse dates, which is often shared with companies, marketeers or researchers (Levy & Romo-Avilés, 2019). Some apps inform users how their data is used and possibly shared with additional actors, but many apps do not supply this information(Zwingerman et al., 2019). According to Essén et al. (2022) some countries are developing frameworks for health app guidelines, however, they also state that with regards to privacy more regulation is needed. Although this aspect of MTAs is currently a prominent topic in research (Alfawzan et al., 2022; Almeida et al., 2022), it entails various facets and angles, which are not the primary focus of this research and therefore not further explored here.

2.1.3. Perceptions and Use of Menstrual Apps

Perceptions and Attitudes with Regards to MTAs

Confirming the willingness to use these apps, Earle et al. (2021) discuss that 80 % of 1000 female survey respondents indicated their wish to use a MTA in the future. Similarly, Starling et al. (2018) point out an increasing interest and demand for MTA, with 76,9% respondents reporting their intention for future use. With Haile et al. (2018) looking into the use of a MTA in seven different countries (Ghana, Kenya, Nigeria en Rwanda, India, Jordan and Egypt), they state that the apps are attractive to a diverse population of users. They found that the apps can address unmet needs even in low resource settings. This was confirmed by Nair et al. (2023) who also found a positive outlook towards MTAs in India and a high awareness of MTAs. Gambier-Ross et al. (2018) found that participants who used other health apps, were more likely to use MTAs. Additionally, the ones who used contraception were less likely to use MTAs, which can be appointed to the regularity or lack of menstruation as a consequence of using contraception that discontinues bleeding. With regards to usefulness of and trust in MTAs, Patel et al. (2023) found that 70% of their sample reported partly trusting the information provided by MTAs. The sample found the apps useful as 83 % believed that the apps had improved their knowledge of their menstrual cycle. This is confirmed by Zhaunova et al. (2023) who reported 89 % and 85% of respondents stating their knowledge of respectively the menstrual cycle and pregnancy had improved. These improvements were stronger among frequent, premium and long-term users and differed based on education level. Similarly Nair et al. (2023) found that most users agreed or strongly agreed with improved understanding of their own body and knowledge with regards to their menstruation (66% and 56%).

Reasons for Use

The main reasons women use fertility tracking apps are: Trying to conceive, trying not to conceive (contraception) and to track and understand their cycle. Women state that they want to understand their sexual health, their body, and the recurrence of certain symptoms in certain phases of their cycle (Bretschneider, 2015; Epstein et al., 2017; Gambier-Ross et al., 2018; Levy & Romo-Avilés, 2019). Some women use the apps for medical reasons due to a sexual health condition such as endometriosis and/or to accurately inform their physician on their cycle and health (Bretschneider, 2015; Epstein et al., 2017; Gambier-Ross et al., 2018; Levy & Romo-Avilés, 2019). Other women use the apps to plan their sexual activity and prepare for their next period, which requires a high level of accuracy in the apps prediction of the fertile window and starting day of the menstruation (Blair et al., 2021; Bretschneider, 2015; Broad et al., 2022; Epstein et al., 2017; Gambier-Ross et al., 2018; Levy & Romo-Avilés, 2019). Additionally, some users use MTAs for the information provision on female health to improve their knowledge (Epstein et al., 2017; Gambier-Ross et al., 2018; Nair et al., 2023) In conclusion, women generally use MTAs either for awareness and planning surrounding their menstrual cycle or to (not) get pregnant.

Changing Reasons for Use

An important finding is that the reasons and motivations of women for MTA usage change according to their age and stage of life (Dudouet, 2022; Earle et al., 2021; Epstein et al., 2017; Levy & Romo-Avilés, 2019). The youngest users of these apps are in their teenage years, which is when most women first menstruate. For these users, the fertility-related functions are often less important and tracking or understanding their cycle is the main goal (Eschler et al., 2019). However, after a certain age the interest in their fertility and ovulation increases due to the desire for or apprehension of getting pregnant. Consequently, a larger part of women within this age category use the apps to determine their fertile window and plan their sexual activities accordingly (Dudouet, 2022). In an even later stage in women's lives the motivations change again due to no desire/inability to get pregnant and the apps are mainly used for cycle tracking (Haile et al., 2018). Around the age of 40-50 women enter their peri-menopause and the need for tracking might increase due to inconsistent and heavy menstruation. As women enter menopause the need for these apps declines again due to absence of menstruation and ovulation (Epstein et al., 2017). Gambier-Ross et al. (2018) found that in general an increase in age correlates with a decrease in MTA use. Additionally, Zhaunova et al. (2023) found that the reasons for use also differ based on education and income level, with participants with a lower education using the app to learn about their bodies and avoid pregnancy as opposed to higher educated participants using the app predominantly to get pregnant. Respondents from lower income countries were more likely to use the app to improve their sexual health as opposed to higher income country respondents using the apps for menstrual cycle and symptom tracking. A similar result was found by Haile et al. (2018) who found the same two reasons for use differing significantly per country and age.

2.1.4. Barriers and Facilitators for Menstrual App Uptake

MTAs contain a numerous amount of features and functionalities, which can simultaneously act as a barrier or facilitator based on its presence or absence. For instance, lack of personalising symptom tracking can be a barrier, whereas the ability to select which symptoms to track can act as a facilitator (Eschler et al., 2019; Gambier-Ross et al., 2018).

All the factors mentioned in the reviewed papers were synthesized and combined into themes. A list of all identified factors and corresponding themes can be found in appendix B. These are factors found to be important with regards to fertility tracking apps in particular. The most important factors are:

- Ease of Use / Complexity
- Information Communication
- Accuracy
- Evidence-based
- Privacy & Security
- Costs
- Design
- Discretion
- (Tracking) Features & Functions
- Personalisation
- Reminders & Notifications
- Reputation

Ease of use
Usability is one of the first app requirements respondents name in research on MTAs and mHealth uptake (Blair et al., 2021; Karasneh et al., 2020; Patel et al., 2023). Complex or confusing app interfaces or designs act as barriers for and lead to cessation of app use (Bretschneider, 2015; Epstein et al., 2017). Additionally, apps with an abundance of features and/or unnecessary functions are mentioned as a barrier (Bretschneider, 2015; Patel et al., 2023). For most MTAs daily input during the menstruation is required to prevent the predictions from being inaccurate. This can lead to what Bretschneider (2015) call 'tracker fatigue', describing the burden of tracking too many data points too often. Therefore, intuitive and easy data input is mentioned as a facilitator (Gambier-Ross et al., 2018; Moglia et al., 2016). These sources consider easy data input, changing settings and intuitive navigation as important usability aspects.

Evidence based & Accuracy
With 213 out of 326 (65,3%) participants naming accuracy of cycle prediction as the most valued quality of period tracking apps, this factor is deemed very important in the research by Patel et al. (2023). Incorrect predictions undermine the added value of using a MTA for tracking as opposed to traditional methods. Which is also found by Nair et al. (2023) to be a reason women refrain from using a MTA for their cycle tracking. Also referring to the apps' content in addition to their accuracy, Starling et al. (2018) mention the preferences of women for a medically accurate application. This entails the information supplied by the apps and the (lack of) scientific literature used to substantiate that information (Starling

et al., 2018). Furthermore, the (lack of) involvement of health care professionals is mentioned (Karasneh et al., 2020; Nair et al., 2023). Lastly, MTAs tested in published research are mentioned as important by many participants of the study by Starling et al. (2018). Karasneh et al. (2020) state that even though multiple apps label themselves as evidence-based, most of them are not tested in trials.

Privacy

Privacy encompasses the sharing of personal sensitive data with third parties and the transparency of data usage. The aforementioned wish to understand which data is used for what purpose, also applies to the privacy barriers potential users experience (Nair et al., 2023). The sensitive nature of the data is mentioned as a reason behind privacy concerns, as users require insight into why their sensitive data is necessary for the app's calculations (Levy & Romo-Avilés, 2019). In Patel et al. (2023) multiple participants stated being unsure of how their data was used.Similarly, Eschler et al. (2019) propose transparency as a way to improve evidence-based information provision and content. According to Ko et al. (2023) often apps do not provide a locking function and data protection of women's health apps is often poor. Additionally, they state that mHealth apps with sufficient data protection are often favored by consumers. Which is backed up by Starling et al. (2018). Interestingly, Epstein et al. (2017) found that women preferred a MTA over traditional tracking methods due to improved privacy, but only if the app's design was discrete (which will be further explained underneath *discretion*). Patel et al. (2023) mention that only a few participants expressed concerns regarding the privacy issue. They ascribe this to the lack of transparency and complex privacy policies, leading to lack of awareness of data privacy among users.

Costs

The costs associated with MTAs can act as a barrier, with Patel et al. (2023) noting costs to be a reason for cessation of MTA use for 10% of their 375 participants. They further elaborated on this by mentioning that in-app purchases were not affordable but posed to the users very often. Participants stated being interested in the features associated with the costs, but not being able to pay for them and thus getting distracted by the subscription advertisements. Starling et al. (2018) reports that 50 % of their respondents rated costs as very important and 32% as somewhat important. It must be noted that most MTAs currently available are initially free of charge.

Design

The design factor encompasses the discrete designs, the user interface and the so-called 'pinkification' of MTA designs. The first will be explained in the discretion factor beneath. The second regards the use of visuals in designs and easy navigation design, which also influence the *ease of use* factor. Patel et al. (2023) and Karasneh et al. (2020) both mention the use of visuals for improved communication and navigation. The third aspect of 'pinkification' touches on the fact that some users stopped using apps that had gendered designs with a lot of pink and flowers (Epstein et al., 2017; Patel et al., 2023).

Discretion

This factor covers the concern MTA users have with regards to the outlook of the apps. Gambier-Ross et al. (2018) mention that women preferred using an MTA that was inconspicuous and looked similar to other health apps. This is seconded by Epstein et al. (2017). Where women mention that the tracking through MTA offers a more discrete tracking method as also mentioned in *privacy*. However, this is only the case if the app design is discrete, and does not have an obvious period-tracking associated name or design.

Functions

This factor entails the absence or presence of a function influencing the willingness of women to download or use a MTA. As there is a lot of variety among apps in available functions, which is mentioned in 2.1.2 and A, this will be touched upon quickly. An important feature is the presence of *reminders and notifications*. With Starling et al. (2018) finding that users often look for apps containing birth control reminders. Nair et al. (2023) reported users indicated the fertile phase reminder as essential. Epstein et al. (2017) discuss reminders for the start of the period as a desired feature. Lastly, some participants of the study by Bretschneider (2015) desired notifications to enter enough data. See B for more examples of functionalities that act as facilitators or barriers.

Personalisation
Due to previously mentioned changing motivations for use, women prefer an app that can be customised according to their needs. For example, women that have dealt with infertility do not want to be reminded of their fertile window or ovulation every month (Gambier-Ross et al., 2018; Levy & Romo-Avilés, 2019). Some apps have the ability to indicate the reason for use, whereby ovulation and fertile days are disregarded if fertility awareness is not the purpose of the user. Additionally, women would like to determine themselves which additional features or symptoms they prefer to track, to avoid too many unnecessary tracking features. Both customization of the symptom tracking list and ability to dis-/ enable elements are named as facilitators (Eschler et al., 2019). A similar reason for lower uptake, is the concern of women regarding the normative app settings and assumptions (Bretschneider, 2015; Epstein et al., 2017). Women name the fact that the apps assume a certain regularity in the menstrual cycle of the user. An example of this is the lack of a 'pause-function' or the ability to report being pregnant. The inability of apps to account for deviant bodies or needs, leads to app use cessation or prevents women with irregular, long or heavy cycles from MTA uptake (Epstein et al., 2017). The wish for personalisation also applies to reminders and notifications Gambier-Ross et al. (2018). When the app gives too many notifications or notifications that are not important for the user, it increases the chance of discontinuation. Therefore, the ability to adjust (the frequency of) reminders is a facilitator (Lupton, 2016; Starling et al., 2018). Withing the literature concerning MTAs personalisation is not linked to cultural influences, which this is the case in the literature reviewed later on in section 2.1.5.

Reputation
Lastly, this factor includes recommendations and reviews. Starling et al. (2018) asked users to rank different app characteristics, among which a section specified to the reputation of a MTA. A recommendation from a healthcare provider was the most important, followed by an app that had published research conducted on it. That was followed by a recommendation from a family member or friend, then the rating in the app store.Gambier-Ross et al. (2018) also found that healthcare providers' recommendations was a reason women starting using MTAs, just like app recommendations from mothers and friends.

Even though, almost all of the papers reviewed in the previous section collect socio-demographic data from their participants, the connection with the facilitators and barriers is often superficially executed. With regards to mHealth research in general, the effect of personal characteristics is investigated more extensively, by focusing on lower income countries, diverse users and user from low socio-economic positions. The following section discusses the important findings.

2.1.5. Barriers and Facilitators for mHealth Uptake Among Diverse Users

The paper by (Hengst et al., 2023), executes a systematic review of literature identifying barriers and facilitators for mHealth uptake by people with a low socio-economic position. This paper was published in September 2023 and therefore it provides a very comprehensive and accurate overview of identified barriers and facilitators. By scanning 13 articles, the researchers discovered 30 barriers and facilitators. These were categorised according to the classification of (Stowell et al., 2018). This classification consists of the intrapersonal, interpersonal, community, ecological and app-specific levels. The intrapersonal level contains factors such as *(Dis)trust*. The interpersonal level contains factors such as *Digital training*. The community and ecological level contain factors such as *Stakeholder collaboration* and *Policies* respectively. Lastly, the app-specific level contains app functions such as *Usability*. Even though, all of these factors were deemed significant in determining mHealth uptake, not all of the levels are relevant in this research. Due to the scope, factors should be directly influential by the means of most app developers. Ecological factors such as policies are complex and require a different research angle. Therefore, only the following app-specific level factors from the paper are considered in this research:

- Tailoring
- Usability
- Visuals
- Evidence-based
- Language
- Personalisation

- Patient-centered design
- Privacy and secure data sharing
- Frequency of reminders

See table 23 in appendix B for a short explanation of all the factors. Having established two lists of identified factors influencing MTA and mHealth uptake, the next section 2.2.1 provides the substantiated selection of the eventually chosen attributes.

2.2. Conceptual Model

This section displays the conceptual model development process. First it is explained how synthesizing the findings from the previous section led to the attribute and level development. Subsequently, each attribute is discussed individually. Afterwards, the improvements of the attributes are discussed. Finally, the improvements are applied and the eventual selection of attributes is shown.

2.2.1. Attribute & Level Development

In the previous section identified factors for mHealth uptake by women with a low SEP and for MTA uptake were presented. The selection of attributes need to be influential by app developers and relevant to answer the main research question and thus apply to MTAs and be relevant for diverse women. Consequently, the redundant factors in 2.1.4 and table 23 were included. These were: Evidence-based, Personalisation, Privacy & security, (Frequency of) reminders and Usability.

Additionally, the costs factor is included. Hengst et al. (2023) identify cost as a barrier, but on an ecological level and not on the app-specific level. In MTA research, costs is named as an important factor in many papers. Gambier-Ross et al. (2018) specifically states that the app developer should keep costs low and accessible. Thereby, defining costs as a factor that can be influenced by app developers. Next to that, including costs as an attribute will refrain respondents from making assumptions about app costs when for example an app is very accurate or highly evidence-based. Lastly, including costs as a factor allows for welfare measures like Willingness-to-Pay when analysing the model (Lancsar & Louviere, 2008).
The extra factors identified for people with a low SEP are: tailoring, visuals, language and patient-centered design. The attributes patient-centered design and tailoring are fairly similar. Patient-centered design is described in the research as the extent to which the target group is engaged in the app development process. Tailoring is defined as the extent to which the information and app are adapted to the characteristics, needs and preferences of the whole target group. Both factors apply to design considerations for the target group. Therefore, these factors will be reduced to one factor, namely *tailoring*, measuring the amount to which the design is tailored towards the target group. Furthermore, the factors language and visuals are named in one breath to tackle low literacy by supplying easily understandable language without medical terms supported by the use of visuals (Hengst et al., 2023). Therefore, these will be included in a combined attribute denoting the way information is provided. *Information provision* entails the way the app supplies the user with information. Including this attribute, results in the following first selection of 8 attributes:

- Costs
- Evidence-based
- Personalisation
- Privacy & Security
- Ease of Use
- (Frequency of) Reminders and Notifications
- Information Provision
- Tailoring

In the following sections, these attributes, their levels and how they are defined within the DCEs are clearly described.

Costs

Initially this attribute and the levels were based on the current real life costs of the most popular MTAs. As mentioned before these apps are initially free but require a monthly payment for the premium features.Therefore, this attribute was defined as *'The monthly price paid for full access to the entire app'*. The values of the levels were based on the costs of real life apps, with preservation of equidistance. Among the most popular apps are Flo, Clue, Eve, Glow and My Period Tracker. See table 1 for an overview of their costs.

Table 1: Premium Costs of Popular MTAs

App	Monthly Payment
Clue	€7,99
Flo	€4,99
Eve	€10,99
My Calendar	€4,49
Period Calendar	€3,49

Additionally, Zwingerman et al. (2019) examined multiple apps and found that subscription based apps ranged from €1,39 to €13,99 per month. As most apps are free, the level 0,00 will be included. The other two levels were determined based on the app costs range and the stated costs of the four popular apps above, resulting in the following levels of this attribute:

- €0,00
- €5,00
- €10,00

(Frequency of) Reminders & Notifications

This attribute is included in other research papers conducting DCE with regards to mHealth preferences. Nittas, Mütsch, and Puhan (2020) use the following definition for their attribute en its levels:

How would you prefer the times and frequency of your reminders to be set?

- I set the time and frequency of my reminders myself
- The app sets the times and frequency of reminders automatically, based on my data

This definition underlines the customisation and personalisation aspect of the frequency of reminders and notifications. With regards to notifications and reminders, women often express notifications and reminders as a facilitator, especially when they can be adapted to personal preferences (Gambier-Ross et al., 2018; Karasneh et al., 2020; Lupton, 2016). Likewise, too many or unnecessary notifications can act as a barrier (Gambier-Ross et al., 2018) Due to the emphasis on the customization of notifications and reminders in most papers and the importance of avoiding overlapping conceptualisation between attributes, this attribute is incorporated into the customization/personalisation attribute as one of the possible features that can be personalised.

Personalisation

Customizing the app and personalising its content is mentioned as a preferred feature in multiple research papers (Eschler et al., 2019; Karasneh et al., 2020; Ko et al., 2023). Based on those papers, this attribute is defined as *'The extent tot which the user can set up the app to her/his own liking'*. Aspects that are often mentioned with regards to customization are settings, reminders and notifications, symptoms and (tracking) features. Firstly, the reminders and notifications were included to include the previous attribute. The second aspect taken into consideration was the ability to enable or disable functions, which encapsulates all of the mentioned aspects above. Consequently, levels were defined as follows:

- The app offers no personalisation options

- The app offers personalisation of the reminders and notifications
- The app offers personalisation of the reminders and notifications and the ability to disable/enable tracking elements

Privacy & Security

There are two research papers that execute a DCE with regards to mHealth and include a privacy attribute. The first is the paper by Nittas, Mütsch, and Puhan (2020), who base their definition and levels on the level of control an user has on when and with whom their data is shared. The options include providing information and additional consent from the user. The second paper is by Folkvord et al. (2022), who define their levels more broadly with data protection or no information on data protection. t is unclear how the researchers defined their data protection attribute.

Being transparent and informing consumers about the use of sensitive data, is mentioned by multiple women as an important factor (Bretschneider, 2015; Eschler et al., 2019). A combination of these aspects - data protection information and user consent - is therefore used to define this attribute and its levels. The attribute is defined as: *'The extent to which the app is transparent with regards to its data use'*. The levels are:

- The app does not provide any information on how it handles data
- The app provides information
- The app provides information and requires consent

Ease of Use

The level selection for this attribute is inspired by the levels used by Xie et al. (2023) and Xie and Or (2023), who provide the option not easy, moderately easy and very easy to use. However, the term 'use' can encompass a broad range of app aspects. To further define 'use' the levels were defined based on the terms used by Gambier-Ross et al. (2018) and Moglia et al. (2016) in their papers when describing 'ease of use'. As a result the used definition is: *'The extent to which it is easy to navigate, put in data and change the settings in the app'.*

- Not easy to navigate through app, input data, change settings
- Moderately easy to navigate through app, input data, change settings
- Very easy to navigate through app, input data, change settings

Evidence-based

The attribute evidence-based encompasses multiple aspects of the scientific level of MTAs. As described in section 2.1.4, it can encompass (lack of) the involvement of health professionals, citing scientific literature or inaccuracies of information and/or prediction algorithms. When apps are tested in trials, this often comes down to the predictions and the efficacy of the apps for family planning (Karasneh et al., 2020). Focusing on that part of evidence-based apps, allows the levels to be defined as percentages, which increases simplicity in the experiments.

Johnson et al. (2018) have investigated the accuracy of fertility apps. Since apps often predict the menstruation, ovulation and fertile days, these three accuracy measures were considered. As most women rely on the fertile days and menstruation prediction more than on the precise day that they ovulate, the ovulation date was left out. For the accuracy attribute to be relevant for trade-offs a certain benchmark is necessary. For example, an accuracy of 30% offers users little to no benefits in using MTAs opposed to traditional tracking methods. Therefore, since the menstruation prediction rates have the highest values, those are used for this attribute. Consequently, this attribute is defined as: *'The extent to which the app correctly predicts the menstruation cycle'*. It includes the following levels:

- The app predicts the menstruation correct 65% of the time
- The app predicts the menstruation correct 80% of the time
- The app predicts the menstruation correct 90% of the time

Information Communication

This attribute is defined as *'The way in which provided information is communicated and how understandable the information is'*. This is based on findings by (Hengst et al., 2023), who emphasize on the use of medical terms in most mHealth, which requires a certain level of health literacy. The same research also underscores cooping with this by using visuals to improve understanding. This is supported by Patel et al. (2023), who found that good visuals improve navigation. Consequently, the following levels were defined:

- The medical information provided by the app contains medical terms and requires a certain level of health literacy
- The medical information provided by the app contains no medical terms and is easily understandable
- The medical information provided by the app contains no medical terms, is easily understandable and supported by images and video's

Tailoring

Lastly, the attribute *tailoring* has been defined before in section 2.2 as *'the extent to which the design is tailored towards the target group'*. The levels for this attribute will denote either the engagement of the target group or the absence of engagement. Further distinguishing different levels of tailoring is not useful, since tailoring is different per target group. For example, based on religions different adjustments would be made than if an app was tailored to women with a handicap.

- The app has been designed without engagement from the target group
- The app has been designed with engagement from the target group

2.2.2. Attribute & Level Improvement

As mentioned before, the attributes and levels were validated by means of a focus group. Furthermore, multiple meetings with a DCE expert and DEI expert provided feedback as well. The remarks will be addressed and improved in this section.

Focus Group

The focus group had a two-fold objective. Firstly, to gain a better understanding of the perception and attitude towards MTAs. Secondly, to discuss and verify the first proposition of attributes and levels. With the research objective in mind, the aim was to establish a diverse group of respondents. See table 24 for an overview of the socio-demographics of the 6 women participating in the focus group. To facilitate a diverse participant group, the focus group was organised through an online platform. Firstly, the participants were introduced to the topic and objective of the focus group. Afterwards, an open discussion guided by questions was initiated. This resulted in current use, motivations for use and necessary features. All participants showed interest in using MTAs, albeit with reservations about their added value to traditional tracking methods. Furthermore, the motivations for use and key considerations for use were in line with the findings from the literature review and thus confirmed previous results.

Secondly, the constructed attributes and levels were discussed to determine any missing attributes and confusing definitions or levels. While discussing the key features, no additional features were mentioned, indicating that there were no missing important attributes. Other improvements proposed in the focus group are discussed in sections 2.2.2 and 2.2.2. A more detailed outline of the focus group and the remarks made can be read in C.

Missing and Redundant Attributes

Firstly, it was suggested that the attribute *tailoring* would not be evident to the respondents when selecting an app. Meaning that the matter in which a target group is incorporated into the design development can not be directly noticed by an app user when using the app, when for example the *ease of use* is something that an app user notices while using the app. Additionally, the focus group remarked that this attribute was vaguely defined and confusing. Therefore, this attribute was left out of the selection. However, as noted by multiple papers involvement of the target group in the app development is of importance (Figueroa et al., 2023; Mora et al., 2020). Consequently, it was added to the survey in a different form, as can be read in section 4.4.2.

Additionally, it was pointed out that there was no attribute depicting the additional features that a premium membership would provide. Including such an option would provide the insight into the worth of additional features to different respondents. Also, the hypothetical choice-sets would resonate more with the real life MTAs. Furthermore, adding the attribute *extra functions* provided an opportunity to include the extra informative educational material often mentioned by women in research (Gambier-Ross et al., 2018; Karasneh et al., 2020; Zhaunova et al., 2023). In available apps such as Clue, Flo and My Calendar premiums users have access to informative articles on female health. Another additional function that multiple apps supply to premium users is a chat function. This function offers the user the opportunity to ask cycle related questions. The added attribute *extra functions* had following definition and corresponding levels:

The app contains the following extra function:

- None
- Information about female health
- Chat function for medical questions and information

Definition Improvements

Firstly, it was evident through both validation methods that the definition of the attributes and levels needed to be more concise. Consequently, almost all attributes and levels were improved with more concise definitions. *Evidence-based* and *information communication* were renamed *accuracy* and *communication*. Additional improvements that only entailed more define wording can be seen in 2. Two attributes were changed significantly, which were *costs* and *ease of use*.

Costs

All participants of the focus group noted that the chosen range for the cost attribute was not realistic. A payment of €10,00 per month was considered very expensive, even €5,00 per month was considered a high amount. The participants did mention being open to paying up to €5,00 once to gain access to the entire app. Based on these remarks, the attribute was reviewed. In reality the monthly subscription allows the users multiple different features. It is not possible to simulate that within the DCE. Within the focus group more realistic price levels were discussed, while considering the findings by Zwingerman et al. (2019), which found that the median price was €2,80. With preserving equidistance and the first level being €0,00, the highest level was established at €5,60. As two respondents stated that uneven numbers in the prices felt random, they were changed to €2,50 and €5,00.

Ease of use

This attribute was improved to ensure a more tangible definition. As easy to use is subjective, hard to measure and does not supply an actionable recommendation towards an app developer. One of the barriers of tracking apps that is often mentioned, also with regards to MTAs is the burden of data input (Bretschneider, 2015; Patel et al., 2023; Zwingerman et al., 2019). With regards to MTAs this concerns the registering of days that the user menstruates. Some apps provide easy data input by allowing just one simple click, others request the user to register multiple aspects thereby elongating the data input process (Bretschneider, 2015; Patel et al., 2023). Additionally, the focus group pointed out that participants would like a reminder to input their data. This would ensure not forgetting the registration and remove the burden of remembering. This mitigation is also mentioned in the literature to reduce the data input burden (Karasneh et al., 2020). This attribute was changed accordingly, which can be seen in table 2.

2.2.3. Final Attribute & Level Selection

After the changes named above, the final attribute selection was finished. The incorporated attributes and their corresponding levels can be seen in table 2.

Table 2: Selected attributes and levels

Factor	Description	Levels
Costs	The price paid for full access to the entire app is:	• €0,00 • €2,50 • €5,00
Privacy & Security	This attribute entails the extent to which the app communicates with users on the use of their data. The app provides the following transparency on data use:	• None • Provides information • Provides information + asks for consent
Communication	This attribute entails the way the app communicates with the users. The app communicates with its users by means of:	• Medical terms • Simple language • Simple language + icons and images
Accuracy	This attribute entails the extent to which the predictions of the menstruation are correct. The percentage of times the app predicts the menstruation correct:	• 65% • 80% • 95%
Ease of Use	This attribute entails the effort needed to input cycle tracking data. Tracking the cycle is done by:	• Entering multiple data points • Entering a single data point • Entering a single data point + reminder notification
Personalisation	This attribute entails the extent to which the user can set up the app to his/her own liking. The app allows me to customize:	• Nothing • Notifications • Notifications + dis-/enabling functions
Extra function	This attribute depicts additional functions the app can contain. The app contains the following additional function:	• None • Information on female health • Chatfunction for medical questions and information

3
Methodology

In this chapter, the methodology underlying the research is explained. First Discrete Choice Modelling, Discrete Choice Experiments and the method approach are introduced. Secondly, the choice task design and the requirements to be considered are discussed. Following that, the experimental design and the decisions to be made regarding the design are broken down. The following sections discuss the survey construction and distribution. The last two sections explain how the model was estimated and interpreted based on the DCE data.

3.1. DCM

Discrete Choice Modelling (DCM) is a method that aims to determine factors at play in decision-making (Bernasco & Block, 2013). DCM studies consumer preferences based on the trade-offs they make when choosing between alternatives. When presented with the choice between two or more alternatives with different attributes, it is assumed the consumer makes trade-offs between these attributes. Studying these trade-offs provides an insight into the influence these attributes have on the consumer preferences. Within DCM there are two ways to examine consumer choice behaviour. The first is through Revealed Preferences (RP) and the second is through Stated Preferences (SP). RP determines consumer preferences based on real-life consumer choices. The data is acquired based on the choices the consumer has made regarding existing products or services. SP data is based on choices made in a hypothetical, though real-life simulating, situation. It elicits consumer preferences based on the choice they state they would make. Naturally, RP is more significant because it is based on actual choices made, while SP relies on consumers stating which choice they think they would make. However, RP data on current fertility app purchases is not available to the researcher. Furthermore, available data most likely does not distinguish app purchases based on women's backgrounds. Due to the lack of data on women's choices regarding fertility apps, RP was not an option for this research. The chosen method SP acquires necessary data through Discrete Choice Experiments (DCE), where the respondent is provided with multiple hypothetical choice situations (i.e.: choice-sets).

3.1.1. DCE

DCEs are conducted through surveys and distributed among the targeted respondents. The experiments facilitate a situation that is similar to the real life choice situation and consist of multiple choice-sets presented to the respondent and optionally an opt-out option. These choice-sets contain two or more apps (alternatives) with similar factors (attributes) (e.g.: privacy level of the app). These attributes will have certain levels (e.g.: high, medium or low level of privacy) that vary between the alternatives and will change with each new choice-set. With every new choice-set the respondent is asked to choose one of the alternatives. Due to the varying levels of the attributes, the respondent is forced to make constant trade-offs between the attributes of the alternatives. Multiple choices then allow for calculation of 'the relative strength of preferences for improvements in certain attributes' (Szinay et al., 2021). In practice, this indicates to what extent certain attributes are more influential than others in determining app preferences.

3.1.2. DCM Approach

With regards to the approach to be taken, papers researching preferences in health care contain fairly similar steps(Ryan, 2000; Vicente, 2022). The steps to be followed are: (1) Qualitative work for the identification of key attributes (2) survey and scenario development, (3) piloting and survey adjustments, (4) survey administration and data collection, (5) choice data analysis. Similarly, Trapero-Bertran et al. (2019), looking into the attributes that should be included in DCE related to health technologies, list the following steps to conduct DCE: 1) choice task design 2) experimental design 3) conduct 4) analysis. In this research the approach existed of a combination of the two approaches named above:

1. Qualitative research
    (a) Literature review
    (b) Focus group
2. Choice task & experiment design
3. Survey construction
4. Survey piloting
5. Survey distribution
6. Model estimation
7. Model analysis

3.2. Qualitative Research & Choice Task Design

3.2.1. Attribute Development

Maintaining good practice used by Nittas, Mütsch, Braun, et al. (2020) to identify possible attributes, an extensive literature review was executed in 2. Articles published on barriers and facilitators, experiences, motivations, etc. regarding MTAs were reviewed to obtain a list of identified factors important in MTA uptake. Since the available literature does not cover factors specific to app uptake by minorities or communities with a low SEP, the paper by Hengst et al. (2023) was additionally reviewed. The paper executes a systematic review on identified barriers and facilitators for mHealth uptake by people with a low SEP. Subsequently, the results from that research are compared and synthesized to the factor list based on the MTA uptake literature. Finally, this results in a list of factors identified as relevant in both cases.

When selecting the relevant attributes among all the identified factors, there are a few guidelines for good attribute selection which can be followed. According to Molin (2023b) attributes need to be 1) the most important ones for the respondents and 2) relevant for policy and design. Furthermore, the attributes should 3) not have conceptual overlap with each other and be 4) uni-dimensional (Trapero-Bertran et al., 2019). The last requirement entails that an attribute measures and describes one aspect of a characteristic (Mandeville et al., 2014). Lastly, it is common practice to include no more than 10 attributes to manage the cognitive load for participants, with most DCEs having between 5-7 attributes (Szinay et al., 2021).

3.2.2. Level Development

After the attribute selection, their corresponding levels are constructed based on the same literature, supplemented by literature conducting DCEs with regards to mHealth app uptake. This results in a first proposition of the conceptual model, which depicts the selected attributes with corresponding levels. Regarding level selection Molin (2023b) names the following requirements: 1) the range should be wide and include all values of existing alternatives and future alternatives and 2) preserving equidistance. The first requirement ensures that there are considerable differences between the levels. If that is not ensured, respondents might disregard the levels of those attributes (Szinay et al., 2021). The second requirement assures orthogonality between attributes (Molin, 2023b), which is useful when interpreting the estimated effects (Kløjgaard et al., 2012). While adhering to these requirements, the combinations of all the attribute values should make sense. Likewise, Ryan (2000) emphasizes that the levels should be plausible and actionable to encourage participants to take it seriously. Lastly, Szinay et al. (2021) add that an attribute level balance is preferred, to avoid attributes with more levels from obtaining a higher relative importance. This means that preferably all the attributes have the same amount of levels.

3.3. Focus Group

The first selection of attributes, was validated through meetings with the supervisors and a focus group. This ensures both expert feedback and target group feedback are acquired. The focus group was executed to determine if any essential app attributes were not included into the conceptual model and to ensure that the attributes and levels included were understandable and well defined. During the focus group, firstly an open discussion with regards to MTA use and the ideal MTA was initiated. After that the list of attributes and levels was discussed. The participants were asked to consider if they deemed the selected attributes relevant, if the most important attributes were included and lastly, if the attributes and levels were clearly defined. The comments and remarks were used to improve the attributes and levels. The elaborate focus group description and explanation can be found in C.

3.4. Experimental Design

When constructing the experimental design a number of choices are made according to the characteristics of the experiment. Szinay et al. (2021) name the following 6 considerations: (1) the analytical model specification, (2) including only main effects or interaction effects as well, (3) if the design is labeled or unlabeled, (4) the number of choice tasks and blocking options to be used, (5) which type of experimental design to use, and (6) achieving attribute-level balance. Additionally, they also name the option to include a ’opt-out’ or ’choose neither’ option with the choice-sets.

Analytical Model Specification

The specification of the analytical model used to fit the data and obtain the parameter values, is dependent on the research objective and available data. Different choice models make different assumptions to estimate the results and can acknowledge different data characteristics. Further explanation of the chosen analytical model can be found in 3.7.

Main Effects and Interaction Effects

The choice model can incorporate main effects and interaction effects while estimating preferences. Interaction effects of the factors display if the preference for a certain factor modifies based on the level of another. An often used example is that the comfort level of a train station might influence the preference for waiting time (Molin, 2023b). As this is not the focus of this research interaction effects between attributes will not be included. However, the interaction of attributes with personal characteristics is relevant to the research aim and will therefore be included. This can show if for example the education level of a respondent interacts with the level of privacy a MTA provides.

Labeled or Unlabeled

The hypothetical choice-sets can contain labeled or unlabeled attribute. In this research the respondents will see different presentations of the same hypothetical app. Which means that the alternatives are unlabeled, i.e. app A and app B. Meaning the alternatives will not contain specific attributes belonging to that specific alternative only. They contain generic attributes that are identical among the two alternatives. Only the levels of the attributes will vary between the two alternatives. Due to the use of unlabeled alternatives and generic attributes, the choice-sets can be constructed sequentially. This means that the alternatives are constructed first and placed into choice-sets afterwards.

Choice-set Design

The selected attributes and levels have to be combined into alternatives for the choice-sets. Depending on the amount of attributes and levels used, the number of choice-sets may need to be reduced with experimental designs (Bernasco & Block, 2013; Train, 2009). Including all combinations of level attributes, i.e.: a full-factorial design, leads to too many choice-sets. For example, if an DCE has 5 attributes with each 3 levels, a full factorial design requires $3^5 = 243$ choice-sets (Lancsar & Louviere, 2008). Therefore, fractional factorial designs are typically used. There are three types of fractional factorial designs, random designs, orthogonal designs and efficient designs. As the name suggests, random designs are randomly selected from full factorial designs. Efficient designs minimize standard errors, but require prior assumptions with regards to the parameters for the attributes. Because such information is not available,this research used an orthogonal design to construct the experimental design. Orthogonal designs preserve orthogonality between the attributes. This means that the correlations between the attributes are zero (Lancsar & Louviere, 2008). This results in low standard errors and more reliable parameters.

To obtain an orthogonal design, the software Ngene is used, which constructs the alternatives and choice-sets. The result was checked for dominant alternatives. Dominant alternatives entail alternatives of which all the attribute levels are better than the attribute levels of the other alternative. In that case there is no trade-off to be made between attributes because the dominant alternative is an easy and obvious choice. If a dominant alternative is discovered, this choice-set can be removed dependent on the impact on orthogonality and the extend to which the alternative is dominant.

Choice Tasks and Blocking

The number of choice-sets is dependent on the amount of attributes and levels. As mentioned before this may lead to too many choice-sets. Keeping in mind that participants have to fill in additional questions regarding personal characteristics, the aim is to keep the amount of choice tasks lower than 10 following the advice of Molin (2023a). This can be done by dividing the choice tasks into blocks, which are separate surveys containing a share of the choice-sets. These blocks are assigned to participants at random. These blocks were generated in the Ngene software.

Attribute-level Balance

Another requirement leading to more reliable and significant parameters, is attribute level balance. This entails that each attribute level appears an equal number of times, so that the participants sees all the levels an equal amount of times. When constructing the experimental design in Ngene, attribute level balance can be assured. Additionally, the attribute level combinations are checked to prevent impossible combinations. As this might result in confusion among respondents and cause unreliable responses (Lancsar & Louviere, 2008).

'Opt-out' Option

Additionally, allowing respondents a 'status quo' or 'opt-out' option next to the choice between the two alternatives is often applied in DCEs. According to Lancsar and Louviere (2008) providing a 'choose neither' option enhances the realism and congruence with consumer theory of the experiment. In real life the respondents might not download either one of the apps, due to both options not meeting their app requirements. Therefore, after choosing one of the two alternatives, the respondents were asked if they would download their preferred app. Asking that additional question allows for estimation of the uptake of these apps. By asking the respondent to make a choice between the alternatives first and then asking them if they would buy the app, an analysis of the trade-offs is still possible even if the consumer does not want to buy the app.

3.5. Survey Design

Finally, the choice modelling experiments need to be implemented in an appealing survey design. To ensure participation the survey must be conclusive, short, accessible and unambiguous (Nittas, Mütsch, Braun, et al., 2020). The survey included an opening statement, attitude and perception questions, the choice experiments and socio-demographic statistic questions. The language used has to be understandable for females from all population subsets. More detailed information on the survey can be found in 4.

Afterwards, the survey is piloted by administering the draft in one-on-one sessions with a small test group of women. The remarks concerning readability, understandability and the duration of the survey were recorded. Consequently, the design of the survey was adjusted accordingly, as can be seen in 4.

3.6. Conducting the Discrete Choice Experiment

After the survey has been improved, it was distributed among the target group. For this research the targeted respondents group consists of women, girls or people who ovulate and/or menstruate. Additionally, women who are currently pregnant and thus do not ovulate or menstruate, can be targeted as well. It is not necessary that the respondents have used a MTA before, since the aim of this research is to establish why the respondent would choose to download an app in the first place. Thirdly, with regards to the focus on women with diverse backgrounds, actively recruiting amongst diverse networks is necessary. For a precise description of the distribution see 4.6. The survey was accessible for approximately two weeks.

3.7. Choice Model Estimation

After closing the survey, the first step is to exploratively examine the data. Thereby, obtaining an overview of the gathered data. Then the respondents' individual personal characteristics were examined. This was done by means of SPSS. The data was descriptively analysed to obtain the composition of the sample group.

Thereafter, the model was estimated using the survey data. This was done with the Apollo package in R developed by Hess and Palma (2019), which can perform all commonly used analysis of discrete choice experiments (Bernasco & Block, 2013; Nittas, Mütsch, Braun, et al., 2020). In short, a model is estimated based on the choice data from the DCEs. Based on the choices made, the model derives values for multiple parameters, e.g. the attribute parameters, constant parameters or interaction effect parameters. The latter depict the influence of personal characteristics on respondents' choices. The interpretation of the parameters is dependent on the type of analytical model that is used. This is described in 3.8 and 6. The participant preferences are estimated with a modeling technique based on the random utility theory (RUT), described in 3.7.1, that assumes that a preference consists of a systematic and random component (Bernasco & Block, 2013; Train, 2009).

3.7.1. Random Utility Theory

The RUT states that utility (U) for individual i for choice j consists of a systematic (V) and a random component (ε). Equation 1 denotes this as follows:

$$U_{ij} = V_{ij} + \varepsilon_{ij}, j = 1, ..., J \tag{1}$$

The systematic component is also called observable utility (V), which consists of the utility derived from the observed factors. This is formulated like equation 2:

$$V_{ij} = X'_{ij} * \beta \tag{2}$$

X'_{ij} = is a vector of the attributes
β = is a vector of the coefficients to be estimated. These are the 'weights' of the attributes. These parameters depict the influence an attribute has on the total utility derived from an alternative.

The random component ε, also called the 'error term', captures everything that is unobservant for the researcher. Combining the two equations, we obtain the following utility function:

$$U_{ij} = X'_{ij} * \beta + \varepsilon_{ij}, j = 1, ..., J \tag{3}$$

The measured utility is a latent variable, meaning that indicators depicting utility are observed and not utility in itself. The observed indicator in this case are choices made by respondents. It is assumed that a respondent chooses an alternative if, and only if, its utility is higher than the utility of any other option in the set of alternatives. In other words, the respondent chooses the alternative that yields maximum utility (Lancsar & Louviere, 2008).

However, it is not possible to say with certainty that the alternative with the highest observed utility will be chosen. Due to the prevalence of unobserved factors, denoted with the error term, choices can only be predicted in a probability. The probability that a alternative i is chosen by a respondent is given in the following equation:

$$P(i) = P(V_i + \varepsilon_i > V_j + \varepsilon_j, \forall j \neq i) \tag{4}$$

It denotes the chance that the utility for alternative i is higher than the utility of any other alternatives j in the set of J alternatives (Lancsar & Louviere, 2008). This is the most general formulation of the discrete choice model.

3.7.2. Multinomial Logit Model

The choice probability function of discrete choice models is dependent of the distribution of the error term. When ε is distributed independently and identically (i.i.d) as extreme value type 1 across alternatives, choice situations and individuals, the following choice probability formula can be derived:

$$P(i) = P(V_i + \varepsilon_i > V_j + \varepsilon_j, \forall j \neq i) = \frac{\exp V_i}{\sum_{j=1...J} \exp V_i} \quad (5)$$

This probability choice function is the Multinomial Logit (MNL) Model (Chorus, 2022a). The MNL model is one of the more basic models of discrete choice modelling. It is often used as a starting point when predicting choice probabilities. Since it is relatively easily estimated (Lancsar & Louviere, 2008).

However, as Ben-Akiva et al. (2019) emphasize, this model has one strong limiting property, which is the assumption of Independence of Irrelevant Alternatives (IIA). This declares that the relative choice probabilities between the included relevant alternatives should not change when an additional alternative is added to the choice-set. As Chorus (2022a, 2022b) explains it: The IIA property is related to the underlying assumption of an independently distributed error term. However, if two or more alternatives have something in common influencing their utility for individuals, which is not measured in the observed utility, this common aspect is captured in the error term. Therefore, the error terms of the alternatives are thus not independent but correlated. Additionally, this also applies to the unobserved factors across individuals or sequences of choices over time (Chorus, 2022b; Train, 2009). With regards to this research, heterogeneity across individuals' preferences is assumed and also one of the main research objectives. Therefore, the IIA property is not preserved, and consequently the power of the model's predictions decreases and is less realistic. One of the models that copes with this limiting property is the Mixed Logit (ML) Model.

3.7.3. Mixed Logit Model

The ML deals with three shortcomings of the MNL model (Train, 2009) . Firstly, it can capture nesting effects. Nesting effects capture the correlation between unobserved factors. For example, imagine that a car, train and bus are used as alternatives and their measured attributes are time and costs. For respondents, the train and bus might have correlated unobserved factors due to both alternatives being public transport. These unobserved factors impact the choice of a respondent for a car or public transport alternative. Secondly, it allows for correlation in unobserved factors over time (i.e. it can deal with panel data). Thereby, accounting for the fact that a single decision-maker makes multiple choices over time and these might be correlated. Thirdly, it allows for random taste variation, i.e. preference heterogeneity across individuals (Lancsar & Louviere, 2008). Which makes it more apt for this research, considering the objective is to examine possible heterogeneity in app preferences among women.

Capturing the nesting effects is achieved by adding an additional error component depicting variation across individuals and their choices of common unobserved factors. The error term υ is assumed to be distributed normally ($N(0, \sigma_\upsilon)$). The estimated parameter σ_υ denotes the variation of the added error term. If $\sigma_\upsilon \approx 0$ then a MNL model would have sufficed due to little to no variation in unobserved utility. The MNL model predicts choice probabilities with the following equation:

$$P_{ij} = \int \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}} f(\upsilon_i)\, d\upsilon \quad (6)$$

Accounting for the taste heterogeneity is achieved by specifying a probability density function for the estimated βs. However, as the Latent Class Choice Model (LCCM) is also capable of capturing taste heterogeneity but simultaneously explaining potential heterogeneity, this aspect of the ML model will not be further explored. The LCCM and the way it captures taste heterogeneity is explained in the section down below.

3.7.4. Latent Class Choice Model

Another choice model that considers taste heterogeneity is the Latent Class Choice Model (LCCM). A LCCM assumes that the population consists of certain segments, which can be profiled based on a combination of certain characteristics. Each segment is similar within its traits and preferences, while different from the other segments. The LCCM model consists of two parts: a class membership model and a class-specific model. The class-specific model estimates the class-specific preference weights for the attributes. It estimates the choice behaviour per class and therefore can determine the attribute preferences. This is done in the same way as with the models described above, but separately per class. The class-membership part of the model estimates the the probability that a decision-maker belongs to a certain class based on the characteristics of the decision-maker. This looks as follows:

$$P_{j,iq} = \frac{\exp \beta_q X_{j,i}}{\sum_{j=1}^{I_n}} \tag{7}$$

$P_{j,iq}$ = the probability of person $i$from class q choosing alternative j
$\exp \beta_q X_{j,i}$ = the chance that alternative j is chosen by person i, based on the β of q
$\sum_{j=1}^{I_n}$ = the summation of chances of all alternatives j being chosen individually by person i, based on the β of class q

Based on the class specific model explained above, a certain number of classes are estimated. For all of these classes separate attribute weights and constants are estimated. Based on the class membership, the chance that a certain respondent belongs to a class can be determined. Based on the class a respondent is most likely to belong to, their preferences can be derived.

3.8. Model Interpretation

The ML model and LCCM model are estimated by means of the Apollo software to determine which model best fits the data and exercises more prediction power. The models obtain multiple parameter values for the attributes, covariates and constants included in the model. The attribute parameters should be interpreted as the level of utility gained or lost by a 1 unit increase of that attribute. Directly comparing the size and significance of parameters does not account for the differences between attributes in the utility scales associated with their levels. Therefore, relative importance is often used to say something meaningful about the model parameters. Relative importance denotes to what extent an attribute and the attribute level determine utility derived from a certain product. It uses the utility range of each attribute and divides that by the total attribute utility range of all attributes. Additionally, as discussed above, the models also allows for choice probability calculations. These choice probabilities can be used to predict expected market shares and - applied to health - can predict the uptake of the good or service(Lancsar & Louviere, 2008). In that case the value of the β denotes the relative impact of that attribute on the probability that a certain alternative is chosen.

3.9. Model Validation

It is important to understand the power and validity of the estimated model. The appropriateness of these models can be determined with Goodness-of-Fit statistics describing the fit of the model on the data. The statistics that were used to determine the best model fit are Log-Likelihood (LL), the McFadden's ρ^2, the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). The Log-Likelihood value quantifies how well the model can explain the data. It shows the probability of the observed data, given the model parameters. The higher the value, the higher the explanatory power of the model. The Likelihood Ratio Test (LRS) is used, which compares the LL of two models, to determine if the improvement of fit is significant. McFadden's ρ^2 uses the Log-Likelihood and the Null-Log-Likelihood (the Log-Likelihood of a model where parameters are set to zero, essentially equivalent to no model). It determines the relative increase in Log-Likelihood due to inclusion of the parameters in the model compared to a scenario where the parameters are not included. A higher value denotes an increase in model fit. Both AIC and BIC assess model fit with regards to its complexity. They use the Log-Likelihood and the number of parameters in the model, to determine the balance between goodness-of-fit and model complexity. With slightly different calculations, for both a lower value is preferred over a high value.

4 Survey Design

In this chapter it is explained how the selected attributes and levels are used to construct Discrete Choice Experiments. Afterwards the survey design is discussed, touching upon the included question on socio-demographic variables, perception and attitude statements, and questions on MTA use. Additionally, app characteristics included into the survey outside of the DCEs are discussed. Lastly, the validation and distribution of the survey among the target group is discussed.

4.1. Construction of the Experimental Design

As explained in chapter 3 different choice-sets have to be constructed. Within a choice-set two MTA alternatives are shown, these have the same attributes, but these attributes have different levels. The different combinations of variations of the attribute levels are constructed based on mathematical designs. For this experiment a mathematical design that can accommodate 7 attribute with 3 levels is needed. Basic plan 4 allowed for this amount of attributes and levels and resulted in 18 rows necessary. This means that preferably 18 choice-sets are constructed and included in the experiment. More detailed explanation of the use of basic plans and an image of basic plan 4 are shown in D.

Asking participants to make 18 choices is not realistic and would lead to respondents discontinuing the survey. As described in chapter 3, blocking can be applied to divide the choice-sets into two or more blocks. The more blocks are used, the more respondents are needed to obtain reliable results for each block. In this case two blocks were used to decrease the number of choice-sets per respondent but simultaneously not increase the number of respondents needed excessively. With this information as input, the software Ngene was used to obtain the eventual design. The syntax and the constructed experimental design can be found in appendix D, which also contains a description of the orthogonality check executed in SPSS.

Additionally, the experimental design was screened for any dominant alternatives, which are explained in 3.4. An elaborate explanation of discovering dominant alternatives can be found in appendix D. Eventually, no choice-sets had to be removed due to containing dominant alternatives. After these steps the experimental design was deemed suitable for the choice-set construction, which is shown in 15 in appendix D.

4.2. Presentation of choice-sets

With the final experimental design the choice-sets could be constructed. Each choice-set was designed separately. The alternatives consisted of the 7 attributes and the levels of these attributes corresponded with the values in the row corresponding to the choice-set of the experimental design. So row 2 depicted the values of choice-set 2 in 16. To increase the realistic component of the choice-sets, hypothetical apps that resembled the look of actual apps, were designed. See figure 2 for an example of a choice-set.

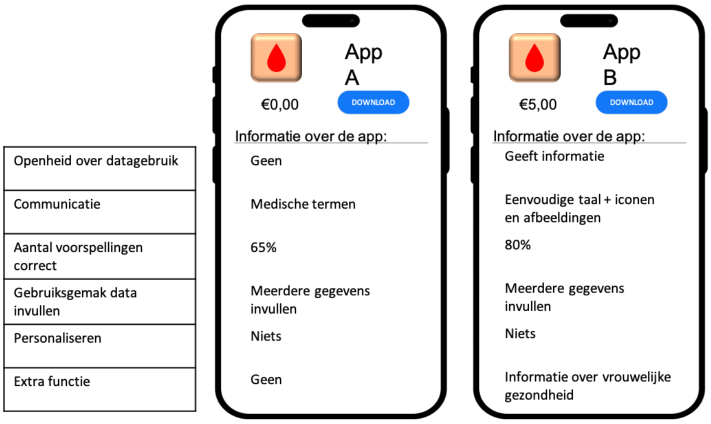

Figure 2: Example of a choice-set design

4.3. Structure of the Survey

The survey was made in Qualtrics and followed this structure:

- Opening statement
  This statement included a short introduction aiming to interest the respondent and address anonymity concerns.
- Selection questions
  This section included three questions to determine if the respondent is capable of using MTAs. The questions concerned menopause, menstruation and pregnancy. If the respondent was not fertile or did not menstruate, they were redirected to the end of the survey and thanked for their interest.
- Introduction + Current use questions
  In this part MTAs were explained with an image of a Dutch MTA. Furthermore, it contained questions, described in section 4.4.1 with regards to current or previous use of MTAs.
- Attribute introductions
  This section included a question per attribute, apart from costs which was explained shortly. The three options were the three levels, which required respondents to thoroughly read the levels.
- Choice experiment
  This section contained an explanation of the choice-sets and an example of a choice-set. Followed by 9 choice-sets from either block 1 or 2 in a randomized sequence.
- Factor questions
  These questions can be found in section 4.4.2 and questioned other factors not included in the DCE.
- Socio-demographic characteristics
  The socio-demographic characteristic definitions included in the survey, were based on CBS, the standardized approved survey questions in Qualtrics and the studied literature. The included socio-demographics are *age, gender, education level, ethnicity, work situation and income*. Appendix E contains more detail on the definitions.
- Perception questions
  These questions can be found in section 4.4.3 and asked respondents for their perceptions of MTAs.
- Closing statement
  Here, the respondents were thanked, given the opportunity to leave behind comments and fill in their e-mail for winning the coupon.

To reduce fatigue, the perception and factor questions were separated by the less cognitive heavy socio-demographic characteristic questions. The factor questions, socio-demographic characteristics and perception questions are discussed in the sections below.

4.4. Additional Questions

4.4.1. Current MTA Use

This part consisted of the following questions to obtain an idea of the current usage of MTAs. Only relevant questions were displayed to the respondent.

- 'Have you ever heard of MTAs?'
- 'How have you first heard of MTAs?'
- 'Have you ever used a MTA?'
- 'Which app have you used the most?'
- 'For what reason did you primarily use the MTA app?'
- 'What is the main reason you stopped using the MTA?'
- 'Do you pay for your app monthly?'
- 'Why did you pay for your app?'
- 'What is the main reason that you have never used a MTA?'

4.4.2. Importance of Additional App Characteristics

Due to the pragmatical and theoretical boundaries of DCE, not all important factors could be included in the experiments. Therefore, factors that were found in literature to be important for mHealth use by people with a low SEP, were separately included in the survey. They were asked to rate the factors from 1 to 5, indicating their importance in app selection. Among other papers, most of the questions were retrieved from the research by Starling et al. (2018) where users were asked for their preferences by rating app characteristics on importance. Additionally, all the questions corresponded to factors found by Hengst et al. (2023) that were not included in the DCE. The used Dutch translation of the scale and the questions can be found in appendix E.

Table 3 shows the factors respondents had to rank from not important to very important. In the right column the corresponding app factors identified by Hengst et al. (2023) can be found. To distinguish these app factors from the app-specific factors used in the DCEs, these factors will be denoted with app characteristics from now on.

Table 3: Additional App Factors

Statement in survey	Associated Literature Factor
'If the app is mindful of different languages, cultures and religions'	Tailoring
'If the app is made for my target group'	Tailoring & Patient-centered Design
'If the app is made in collaboration with women from diverse backgrounds'	Tailoring & Patient-centered Design
'If the app is reviewed well in the app store'	Recommendation
'If the app is recommended by someone I know'	Recommendation
'If the app is recommended by a doctor or a medical organisation (like a hospital)'	Recommendation
'If the app is recommended by social media, a news article or a blog post'	Recommendation
'If the app is validated in scientific research'	Evidence-based
'If the app is made in collaboration with a doctor or a medical organisation (like a hospital)'	Evidence-based
'If the app shares my personal data with others'	Privacy & security
'If the app allows me to come in contact with other users'	Community

4.4.3. Perceptions & Attitudes

Additionally, the respondents were asked to provide their opinions with regards to perception statements. These perceptions were included to obtain an idea of attitude towards and the willingness to use MTA. Furthermore, these attitudes are incorporated into the model to determine if they significantly influence app preferences.The questions for this part are inspired by multiple papers, but mostly on the papers by Gambier-Ross et al. (2018) and Nair et al. (2023).

Participants were asked to rate these statements on a 5-point Likert scale ranging from totally disagree to totally agree:

- 'I think MTAs are useful'
- 'I'm open to using a MTA'
- 'I would use a MTA to track my period'
- 'I would use a MTA to learn more about my fertility'
- 'MTAs help to learn more about your cycle and fertility'
- 'Tracking your cycle is important'
- 'Knowledge regarding your own cycle is important'
- 'I would like to know more about female health, such as the cycle'

See appendix E for the Dutch translation of the questions and the scales.

4.5. Survey Validation

The questions of the survey were validated by means of feedback from a DCE expert and an app diversity expert. Furthermore, the language used in the survey was iteratively improved with feedback from Pharos, an organisation improving accessibility of healthcare and reducing healthcare disparities. With use of a website detecting b1 Dutch language level, difficult words were explained or removed from the survey.

The survey was pilot tested with a small number of respondents due to time constraints. A few respondents from different backgrounds answered the survey in individual online meetings. It took the respondetns around 15-20 minutes to complete the survey. Incorporating that the respondents took time to explain certain choices to the researcher, this time was still longer than desired. Therefore, some questions were removed and explanatory texts were shortened. Afterwards, some respondents filled out the survey link and provided feedback. These respondents took somewhere between 10-15 minutes to complete the survey, which is a more acceptable time. Additionally, based on these pilot tests some introduction texts were adapted as well as the structure of the survey. The images of the apps was also slightly adapted to enhance user-friendliness.

4.6. Survey Distribution

After validating and improving the survey it was distributed via multiple social media platforms. Namely, Facebook, WhatsApp and LinkedIn. Also, via the organisation Voices for Women the survey was distributed among a larger care-related female audience. Additionally, the survey was printed out and distributed among general practitioner waiting rooms, pharmacies and houses in a neighbourhood with social housing. People with diverse networks were asked to spread the survey as well. This included female network organisations, individual people and social work organisations. A highschool was also approached to participate. Lastly, women were approached on an female healthcare information market hosted by the Amsterdam Universitair Medisch Centrum, which is the teaching hospital in Amsterdam. The survey was made accessible for 14 days. Furthermore, participation was encouraged by rewarding a 15 euro coupon for Bol.com and explicitly noting this in the distribution message.

5 Statistics and Choice Model Estimation

This chapter entails the data cleaning from the DCE data and the model estimation description. Firstly, the retrieved data is examined and cleaned. Secondly, the statistics of the sample are shown and discussed. Afterwards, the results of the additional questions in the survey are examined. Afterwards, the estimation method of the different models is explained.

5.1. Data Cleaning

The survey was open for 14 days from March 4th to March 18th. The survey was opened by 290 respondents and finished completely by 213 respondents. Of those 213 respondents, 45 respondents were redirected to the end of the survey due to not complying with the target group requirement of being able to menstruate. All of those responses were removed from the dataset, so in the end 168 responses were included into the data analysis and model.

Subsequently, it is important to check if block 1 and block 2 were completed the same amount of times. Block 1 was completed 87 times and block 2 81 times. Therefore, the blocks were shown in similar amounts and the distribution will not interfere with the estimates.

Lastly, the responses were screened in two ways to detect unreliable responses. Firstly, the answers were checked for non-trading behavior. This can be detected if a respondent consequently selects the same alternative throughout the choice-sets. Due to fatigue, indifference or misunderstanding the respondent could disregard the trade-off aspect of the questions and just select a random choice. Throughout all responses, there was only one respondent who selected the same alternative in every choice-set. That response was examined further and on the basis of the other answers there was no reason to suspect unreliable answers.

Since the third choice-set contained a dominant alternative, it could be used to check invalid responses (see 4.1 for an explanation of a dominant alternative). It must be noted that only the respondents that got appointed block 2, saw the third choice-set and can be checked. There were 6 respondents that did not choose the dominant alternative. All of these respondents preferred a different level for some of the attributes than assumed by the researcher, therefore the 'dominant' alternative was not dominant based on their preferences.

5.2. Statistics Respondents

An overview of the sample composition is supplied in table 4. Some socio-demographic variables were re-coded into categories. An explanation of how these socio-demographics were re-coded, can be found in G. As there were no participants that declared their gender other than female, this variable is left out. As table 4 shows, the sample composition of the survey is not equally distributed for all socio-demographic variables. Firstly, for the two socio-demographics education level and ethnicity, the variance between the two levels is low. Highly educated Dutch respondents are over-represented in comparison to low and middle educated non-Dutch respondents. As the underrepresented groups have been re-coded to both contain at least 20 respondents, analysis is possible. Furthermore, the age variable consist of 19-26 year old respondents for more than half of the sample. The older age group and teenagers are less present in this sample. With regards to income levels, the lowest level is most represented in this sample. However, it is important to note that the respondents in this group consist mostly of students. If these are left out of the sample, the lowest level contains 16 respondents, which would be 15% of the total sample of 105 respondents (leaving out 63 students). Lastly, the work situation variable shows that the sample consists of 52,8 % working respondents and 41,1% students. Respondents without a job or homemakers, are underrepresented in this sample.

As logit choice models are based on correlations, the lack of diversity in the sample does not pose a major problem. However, this only applies if 1) each category of the variable is sufficiently represented, 2) the dropout of a category underrepresented must be random.

As at least 30 observations for each category is normally considered sufficient, none of the variables comply with the first requirement. However, if it appears that the socio-demographics are not of influence on the choices made in the first place, this does not pose a problem. For example, if *age* has a high p- value, indicating insignificance, but a low β value, indicating that it does not affect utility a lot, it is less important if the finding can be generalized to the population as it has little impact anyway. This requires that while investigating the impact of socio-demographics, the focus should be on both the significance and the size of the impact a variable has on the utility. This will be considered when discussing the exclusion of socio-demographics for each model in each corresponding section in chapter 5.4.

The second requirement is hard to determine. As respondents were recruited to obtain a diverse sample group, but this turned out to be difficult. It is possible that the sample consists of respondents within proximity of the researcher. Additionally, it might be that respondents with certain characteristics are more likely to complete this survey. Respondents who are more open to MTAs, might be more likely to participate in research concerning MTA. If either one of these possibilities is the case, the dropout is not random. This is impossible to determine.

The sample composition is not ideal to determine preference differences with regards to socio-demographic variables. Especially the under-representation of non-Dutch and low educated respondents complicate that aim. As the researcher deployed various methods to reach this target group, it is another confirmation of the difficulty to include marginalized or low SEP individuals in research.

5.3. Results Additional Questions

This section contains the results of the questions posed alongside the DCEs in the survey. These consisted on questions concerning current MTA use, respondents' level preferences, additional app characteristics importance and perception & attitude questions.

5.3.1. Current MTA Use

For an overview of all the data of the current use questions see table 5 and 6 for the questions regarding discontinuation and non-use of MTAs. The most interesting findings are discussed shortly.

Firstly, most of the respondents were aware of MTAs (91,7%) and thus participated with prior knowledge of these apps. However, out of all the respondents 61 had never used a MTA. Adding the 14 respondents that were not aware of the existence of MTAs, this means that 75 respondents filled in the survey without experience of MTA use. This is 44,6% of the respondents, so the sample is equally distributed between users and non-users.

The most used apps were Flo (35,5%) by a big share and Apple Health (17,2%), however 19 different apps were mentioned. The main reason respondents used a MTA is to know when they will menstruate (46,2%) and to learn about their cycle (34,4%). Fertility awareness is often mentioned in literature as a reason for MTA use (see section 2.1.3), but less represented in this sample (9,7%). Which might be caused by the relatively young sample, which is not concerned with pregnancy yet. From the 62 current app users, 8 (12,9%) reported paying for a monthly subscription.

The main reason previous app users stopped using their apps was due to lack of a cycle to track (40%). This could be due to pregnancy, irregular cycles, illness or often due to a form of contraceptives that suppresses the cycle. This was confirmed by at least 5 respondents commenting that they stopped

Table 4: Overview Socio-demographic Statistics

Variable	**Category**	**Count**	**Column N %**
Ethnicity	Non-Dutch	20	11.9%
	Dutch	148	88.1%
Age	<19	11	6.5%
	19-26	89	53,0%
	27-35	47	28,0%
	36-45	11	6,5%
	>45	10	6,0%
Education Level	Low & middle education	26	15.5%
	High education	142	84.5%
Work situation	Fulltime	51	30.4%
	Parttime	36	21.4%
	No job	4	2.4%
	Homemaker	2	1.2%
	Student	69	41.1%
	Other	6	3.6%
Income	<20.000 euro per year	79 (63 students)	47.0% (79,7% students)
	20.000 - 30.000 euro per year	17	10.1%
	30.000 - 40.000 euro per year	31	18.5%
	40.000 - 50.000 euro per year	16	9.5%
	50.000 - 60.000 euro per year	13	7.7%
	>60.000 euro per year	12	7.1%

menstruating due to either the pill or IUD and therefore were not interested in MTAs. For 8 respondents the tracking was too much burden (22,9%). Out of the 75 respondents who never used a MTA, 61 were aware of these apps, but mostly not interested in tracking their menstruation or fertility (49,2%) or tracked their cycle in a different way 15 (24,6%).

Table 5: Overview Current MTA Use Questions

Question	Answer	Total respondents	Amount	Percentage
Have you ever heard of MTAs?	Yes	168	154	91,7%
	No		14	8,3%
How have you first heard of MTAs?	Through friends or family	154	58	35,1%
	Through social media		40	26%
	I don't remember		32	20,8%
	Through an influencer		7	4,5%
	It was available on my phone/smartwatch		6	3,9%
	Looked it up on internet		4	2,6%
	Through a doctor		3	1,9%
	Other		4	3%
Have you ever used a MTA?	Yes, I am currently using a MTA	154 [1]	62	40,3%
	Yes, I have in the past		31	20,1%
	No, I have never		61	39,6%
Which app did/do you use the most?	Flo	93	33	35,5%
	Apple Health		16	17,2%
	Clue		15	16,1%
	Fitbit or Garmin app		6	6,5%
	MijnKalender		5	5,4%
	Menstruatie Kalender		4	4,3%
	I don't remember		2	2,2%
	Other		12	12,9%
What did/do you use for MTA for the most?	To know when I will menstruate	93	43	46,2%
	To learn more about my cycle		32	34,4%
	To know when I am fertile		9	9,7%
	To learn about/track my cycle symptoms		6	6,5%
	All of the above		2	3,2%
Do you pay for your MTA monthly?	Yes	62	8	12,9%
	No		54	87,1%
Why do you pay for your MTA?	To gain more insights into my cycle/symptoms	8	4	50%
	Access to all the information		2	25%
	Other		2	25%

Table 6: Overview Discontinuation or Non-Usage MTA Questions

Question	Answer	Total respondents	Amount	Percentage
Why did you stop using your MTA?	I have no cycle to track anymore	35	14	40%
	Tracking was too exhausting		8	22,9%
	The app asked me to subscribe too often		3	8,6%
	Other		4	11,4%
Why have you never used a MTA?	I have no interest in tracking my fertility or menstruation	61	30	49,2%
	I track my fertility or menstruation in a different way		15	24,6%
	I don't think MTAs are useful		3	4,9%
	I am afraid it will increase my anxiety on fertility		2	3,3%
	I have not thought about it		2	3,3%
	MTAs are not accurate enough for me		2	3,3%
	Other		7	11,5%

5.3.2. Level preferences

The attributes and their levels were included in the survey as questions to explain them to respondents. The respondents were asked to select their preferred level. As these questions to not directly contribute to the research objective, but are mainly useful to provide insight into other results, all the results are shown and further discussed in appendix G.

The attribute that had a more surprising answer was the *ease of use* attribute. For that attribute the assumed preferred level was almost equally as preferred to the second level. With 44,1% choosing *entering one data point + reminder notification* as opposed to 37,5% choosing *entering one data point*. Since the only difference is a reminder notification, this could point out that some respondents do not prefer a reminder notification for data entry. For some respondents reminders are only beneficial if they can be adjusted to their needs, as discussed previously in section 2.2.1.

5.3.3. Additional App Characteristics Questions

This section entails the ranking of additional app characteristics. See 5.3.3 for an explanation on the choice for these characteristics. Table 7 shows the average values the respondents assigned to the app features and the associated standard deviations. Additionally, One-Way ANOVA and a t-test were performed to examine if respondents had significant differences in their answers based on their socio-demographic variables.

[1]This question was not displayed to the 14 respondents stating they have never heard of MTAs.

Table 7: App factor rankings

How important do you deem the following app feature? (1; not at all important - 5; very important)	**Mean**	**Std. Deviation**
If the app...		
Is approved in scientific research	3,91	1,17
Is made in cooperation with doctors or a medical organisation (like a hospital)	3,85	1,16
Shares my personal data with others	3,65	1,22
Is recommended by a doctor or medical institution (like a hospital)	3,63	1,23
Is made for the target group I belong to	3,53	1,25
Is made in cooperation with women from diverse backgrounds	3,24	1,42
Is recommended by someone I know	2,82	1,29
Has good ratings in the app store	2,71	1,25
Incorporates different languages, cultures and believes	2,55	1,39
Is recommended through social media, a news article or a blog post	1,60	0,84
Allows me to come in contact with other app users	1,29	0,68

The rankings are categorised based on the maximum length of the 5-point Likert scale. The lowest value they could appoint to a factor was 1 and the highest was 5. The range of the scale is divided by the greatest value of the scale to determine the size of each scale level. $(5-1)/5 = 0,80$ Applied to the first scale level, the values below represent the different levels:

- From 1 until 1.80 represents not at all important
- From 1.81 until 2.60 represents not important
- From 2.61 until 3.40 represents somewhat important
- From 3.41 until 4.20 represents important
- From 4.21 until 5.00 represents very important

As can be seen the highest ranked factors were *approved in scientific research* and *made in cooperation with doctors or a medical organisation*. Which indicates that the sample group highly values the medical grounding of a MTA. Furthermore, *sharing personal data*, *recommendation by a doctor or medical institution*, *designed for the target group* were all ranked 3,5 or higher. The two lowest scoring factors are *recommendation through social media, a news article or a blog post* and the *ability to come in contact with other app users*. With regards to those two factors, the standard deviation was lower than the other factors, indicating less variation in the sample group's rankings. For this sample group, those two factors were considered not at all important and not important.

With regards to the recommendation questions, the most preferred recommendation is through doctors or medical institutions. The recommendation from someone you know and a recommendation through social media, a news article or a blog post are respectively valued as somewhat important and not important at all.

Interaction with Personal Characteristics

To determine if there was significant difference between the additional factor rankings of the respondents based on personal characteristics, a One-way ANOVA test and a t-test were executed. The detailed explanation of the tests and their results can be found in appendix G. The most important results are shortly discussed.

With regards to *age*, one statement was answered significantly different between groups. This was *if the app is recommended by someone I know*. The significant differences were found between the age category <19 and both the categories 19-26 and 27-35. With the lower age category assigning a lower ranking to this statement than the other two older age categories. Which indicates that teenagers on average appoint less importance to a recommendation from someone they know than young adults do. The *user* variable had significant differences for all medical grounding related statements. Which shows that users value scientifically based MTAs as less important than non-users. The reasoning behind this observation requires more research.

5.3.4. Perception Questions

In the survey respondents were asked to denote their level of agreement with regards to attitude and perception questions on MTAs. The answers were linked to the numbers 1 to 5 and averaged across all respondents to obtain the values displayed below in table 8. The same scale ranges as discussed in section 5.3.3 are applied to these answers, which comes down to:

- From 1 until 1.80 represents totally disagree
- From 1.81 until 2.60 represents disagree
- From 2.61 until 3.40 represents neither agree or disagree
- From 3.41 until 4.20 represents agree
- From 4.21 until 5.00 represents totally agree

Table 8: Average Scores on Perception & Attitude Statements

To what extent do you agree with the statement?	Mean	Std. Deviation
I find fertility & menstrual tracking apps useful	3,76	0,84
I am open to using fertility & menstrual tracking apps	3,81	1,08
I would use a fertility & menstrual tracking app to track my menstruation	3,75	1,14
I would use a fertility & menstrual tracking app to learn more about my fertility	3,59	1,12
Fertility & menstrual tracking apps can help to learn more about your cycle and fertility	4,07	0,86
Tracking your menstruation is important	3,65	0,99
Knowledge of your own cycle is important	4,04	0,89
I want to know more about female health such as the cycle	3,67	1,04

From these results it is evident that the sample overall scores relatively high on all statements. For all statements, the sample score falls into the agree category. Respondents agree the most with the statements that MTAs contribute to knowledge improvement on the cycle and fertility and that knowledge on female health is important in general. Both of these statements have relatively low standard deviation, indicating that the sample was reasonably united on their opinion. Since all statements obtained scores relatively close to each other, further comparison is less meaningful. Apparently, the sample of this research has a fairly positive attitude towards using MTAs, their usefulness and the importance of female health knowledge.

Interaction with Personal Characteristics

To determine if there were significant differences between the attitudes of the respondents based on their personal characteristics, a One-way ANOVA test was executed for variables with more than 2 categorical levels and a t-test for variables with 2 categorical levels. The execution and detailed results can be found in G. The most important findings will be discussed shortly.

It appeared that older respondents are less willing to use MTAs and less convinced of the usefulness of fertility & menstrual tracking apps than the younger respondents. As younger generations are known to be more open towards technology and apps, this is not unexpected (Gambier-Ross et al., 2018). It is also interesting to note, that with regards to the importance of female health knowledge, no significant differences were found. So the difference is strictly applicable to MTAs.

Additionally, the user variable also led to significant differences for all these statements. For every statement, except one, the user falls into a higher agreement category than the non-user. The influence of having used or currently using a MTA is thus substantial in determining an individuals' attitude towards MTAs.

Exploratory Factor Analysis

To include the perception factors into the model, they were reduced through Principal Axis Factoring (PAF). The PAF analysis determines if the measured perceptions can be aggregated based on the underlying factors. It looks for highly inter-correlated variables, which could be derived from the same underlying latent variable, which are called components.

The PAF can be executed using two types of factor rotations; orthogonal and oblique. Orthogonal rotation assumes no correlation between the factors, whereas oblique rotation does. To determine the best way to derive underlying components, both methods were tested and compared. A detailed description of the iterations of the analysis are described in F.

With oblique rotation, a simple structure was obtained without removing variables. Which meant that all variables could be incorporated into the two found components. The same steps were repeated with orthogonal rotation to determine if this also led to a simple structure without removing additional variables. This was not the case, as 3 variables had to be removed to obtain a simple structure. Therefore the pattern matrix obtained with oblique rotation was used to compose the two components. See 9 for the derived factors and the scoring of the variables. The two factors were named *'ProMTA'* and *'FemaleHealthKnowledge'*. The first encompassed the statements with regards to willingness to use MTAs and perceived usefulness of MTAs. Therefore, that factor denotes the attitude towards MTAs of the respondent. As table 9 shows, all factor loadings are positive. Therefore, the higher a respondent scores on *'ProMTA'*, the more positive they are towards MTAs. The second factor denotes how the respondent values possessing knowledge of female health. Again all factor loadings are positive, therefore the higher a respondent scores on this factor, the more they value having knowledge of female health. All respondents were given a factor score for the newly constructed perception factors by taking their average of numerical scores for the corresponding variables.

Table 9: PAF factors and the loading variables

Statement	**Factor Score**
Factor 1: Pro MTA attitude	
I find fertility & menstrual tracking apps useful	0.843
I am open to using fertility & menstrual tracking apps	0.871
I would use a fertility & menstrual tracking app to track my menstruation	0.846
I would use a fertility & menstrual tracking app to learn more about my fertility	0.659
Factor 2: Importance female health knowledge	
Tracking your menstruation is important	0.857
Knowledge of your own cycle is important	0.941
I want to know more about female health such as the cycle	0.593

5.4. Model Estimation

This part contains a description of how the DCE data was used to estimate the varying choice models. First, the choice data is discussed in general to obtain an overview of the experiment data. Afterwards, the process of the various model estimations are discussed.

5.4.1. DCE Data and Model Specification

Throughout all DCEs the opt out option was chosen 808 out of 1408 times (55%). 32 respondents - 16 across both blocks - choose no app throughout all choice-sets. 11 respondents choose an app throughout all of the choice-sets. This indicates that there some respondents were very open to MTAs regardless of their attributes and vice versa.

Based on the choices made in the surveys, the model can be estimated. Since all attributes were categorical they had to be re-coded to be included into the model. The attributes and personal characteristics were effect-coded, except for age and income, which were included as a continuous and interval variable respectively. The re-coded attributes and personal characteristics are displayed in appendix G.

After the attributes were re-coded into effect-coded variables, the systematic utility function had to be defined. This function contains an Alternative Specific Constant (ASC) and all the attributes. The βs are the parameters for the attributes that are estimated by the model. Next to the utility function, which is similar for both alternatives, the opt out alternative was included as a third alternative with its utility fixed to zero.

$$
\begin{aligned}
V_0 &= 0 \\
V_j &= ASC + \beta_{\text{low\_costs}} \cdot low\_costs_j + \beta_{\text{high\_costs}} \cdot high\_costs_j \\
&\quad + \beta_{\text{consent}} \cdot consent_j + \beta_{\text{information}} \cdot information_j \\
&\quad + \beta_{\text{simple\_language}} \cdot simple\_language_j + \beta_{\text{icons}} \cdot icons_j \\
&\quad + \beta_{\text{low\_accuracy}} \cdot acca_j + \beta_{\text{high\_accuracy}} \cdot accb_j \\
&\quad + \beta_{\text{single\_entry}} \cdot single\_entry_j + \beta_{\text{reminder}} \cdot reminder_j \\
&\quad + \beta_{\text{notifications}} \cdot notifications_j + \beta_{\text{features}} \cdot features_j \\
&\quad + \beta_{\text{health\_information}} \cdot health_information_j + \beta_{\text{chat}} \cdot chat_j
\end{aligned} \tag{8}
$$

Where:
V_0 = the systematic utility of alternative 0 (which is the opt-out alternative)
$V_j(j = 1, 2)$ = the systematic utility of app alternative j
ASC = the alternative specific constant for the app attributes

All the other function variables are the effect-coded attributes and their corresponding βs.

5.4.2. MNL model

Estimating the MNL model allows for comparison in prediction power improvement with the ML model and the LCCM. Firstly, the first MNL model only included the attribute parameters and no other covariates. The Log Likelihood of this first MNL model was -1236,78. Additionally, all socio-demographic variables were added to the model to observe if the fit improved. These were:

- Age
- Education
- Ethnicity
- Income
- Work situation (effect coded)

Simultaneously, the MNL model was extended by including interaction effects. The included interaction effects were between all the attributes and all the covariates discussed above; age, education, ethnicity, income, work situation.

As discussed in section 5.2, determining which socio-demographic variables will be included in the model was based on significance and impact on the utility of the variable. As income and work situation had a low t-ratio and a p value above 0,01, their impact was further examined. Both estimated βs were below 0,1 which implies a low impact on utility compared to the other socio-demographic variables. Therefore, they were left out of the model.

The only interaction effect that was significant for both attribute indicator variables, was between ethnicity and personalisation. This model had a LL of -1213,91 and the adjusted ρ^2 was 0,1805. Additionally, the likelihood ratio test was executed to determine if this model provided a better fit to the data. The model improvement turned out to be significant at the 99 % level (LRS=47,74; df=5).

Additionally, the model was extended by adding the *user* variable which depicted if the respondent has experience with using a MTA. Again, interaction effects of the *user* variable with all the attributes were added. The interaction effects were not significant and did not impact the model significantly, thus they were excluded. This model had an LL of -1179,15 and the adjusted ρ^2 was 0,2029. Furthermore, the model improvement was significant at the 99% level (LRS =69,52; df=1).

Lastly, the two constructed perception factors were added to the model. As only the first factor *ProMTA* was significant the second factor *FemaleHealthKnowledge* was further investigated to determine if it should be included. The factor had decent influence on the utility function with a β value of 0,134. However, since its t-value was 1,05 and the p value of 0,29 was significantly higher than 0,1, it was still excluded. Including the ProMTA factor, lead to the age variable becoming insignificant. However, as the variable was significant when just socio-demographics were involved, the socio-demographic age is deemed important in previous research (see 2.1.3) and the utility contribution range is relatively high, the variable remains included. The improvement of this model was also significant at the 99& level (LRS=60,62; df=1). The final LL was -1148,83.

5.4.3. Mixed Logit Model

Next a Mixed Logit model was estimated. This is done to determine possible nesting effects among the alternatives and acknowledge the panel characteristics of the data. This will tell if there are any unobserved utilities that are correlated between the app alternatives. Furthermore, the Mixed Logit can acknowledge the panel nature of the input data. Acknowledging the panel nature of the data ensures that the correlations between choices made by individuals over time are considered. For further explanation of both these features, see section 3.7.3.

Firstly, a basic Panel ML model with nesting effects was estimated. This model consisted of all the attributes, the same constant *ASC* as used in the MNL model and an additional Error Component. To establish the estimated σ of the error component (capturing the nesting effects), Halton draws were used. Additionally, the number of draws was increased until the results were stable and the estimates and LL hardly changed. To establish that, 700 draws were adequate. All model iterations were therefore estimated with 700 draws. The model had a LL of -1035. This is a substantial change in a LRS of 403,46 with the basic MNL. This shows that including nesting effects and acknowledging the panel nature of the data, resulted in a model that was able to significantly better predict choice preference estimates.

Next the background variables and their interactions were added. Just like the MNL model, these included the socio-demographic variables, the user variable, associated interaction effects and lastly the perception factors. Just like with the MNL model, the variables were excluded based on insignificance and impact on the utility. For the ML model the exact same variables and interaction effects were significant and thus influenced the utility derived from alternatives. These were; age, ethnicity, education, MTA experience and the ProMTA factor. Again adding MTA experience, resulted in a significantly better model. Again the last addition of the factor caused age to become insignificant. This was dealt with the same way as with the MNL model. After removing insignificant variables the improvement of fit was significant at the 99% level (LRS = 64 ; df=7). This model obtained the following Goodness-of-Fit statistics:

Table 10: Final Goodness-of-Fit Metrics ML model

Statistic Metric	Output
Log Likelihood	-1003,76
Adjusted ρ^2	0,317
AIC	2057,52
BIC	2190,55

The estimated coefficients are displayed in table 13 and discussed in section 6.1. Since the σ estimated for the Error Component of the ML model is significant, heterogeneity with regards to unobserved utility is present in this sample. This model is not further expanded to investigate taste heterogeneity because, the LCCM also captures taste heterogeneity, but additionally can explain the taste heterogeneity. Therefore, the utility function of the final ML model is as follows:

$$\begin{aligned}
V_0 &= 0 \\
V_j &= ASC + \upsilon_{\text{n,App}} \\
&\quad \beta_{\text{low_costs}} \cdot low_costs_j + \beta_{\text{high_costs}} \cdot high_costs_j \\
&\quad + \beta_{\text{consent}} \cdot consent_j + \beta_{\text{information}} \cdot information_j \\
&\quad + \beta_{\text{simple_language}} \cdot simple_language_j + \beta_{\text{icons}} \cdot icons_j \\
&\quad + \beta_{\text{medium_accuracy}} \cdot low_accuracy_j + \beta_{\text{high_accuracy}} \cdot high_accuracy_j \\
&\quad + \beta_{\text{single_entry}} \cdot single_entry_j + \beta_{\text{reminder}} \cdot reminder_j \\
&\quad + notifications_j \cdot (\beta_{\text{notifications}} + \beta_{\text{notifications,ethnicity}} \cdot ethnicity) \\
&\quad + features_j \cdot (\beta_{\text{features}} + \beta_{\text{features,ethnicity}} \cdot ethnicity) \\
&\quad + \beta_{\text{health_information}} \cdot health_information_j + \beta_{\text{chat}} \cdot chat_j \\
&\quad + \beta_{\text{age}} \cdot age + \beta_{\text{education}} \cdot education \\
&\quad + \beta_{\text{ethnicity}} \cdot ethnicity + \beta_{\text{user}} \cdot user \\
&\quad + \beta_{\text{ProMTA}} \cdot ProMTA
\end{aligned} \tag{9}$$

$$\upsilon_{\text{n,App}} \sim \mathcal{N}(0, \sigma_{\upsilon_{\text{App}}})$$

5.4.4. LCCM

After the MNL model was finished, a LCCM model was executed. This was done to determine if the sample consisted of consumer segments with taste heterogeneity. The LCCM model finds classes within the sample that have differing preferences between the classes but similar preferences within a singular class. The identified classes can be profiled by means of the personal characteristics, showing the probability that a respondent belongs to a certain class. Eventually, this clarifies to what extent respondents that possess certain characteristics associated with a segment prefer different app attributes.

Firstly, this model was executed without covariates to determine the right number of classes within the data as advised by Bertrand and Hafner (2014). The Goodness-of-Fit statistics can be seen in table 11, where the MNL model without covariates is also included for comparison.

Table 11: Goodness-of-Fit Statistics LCCM Classes

Classes	Log-Likelihood	Parameters	ρ^2	AIC	BIC	Smallest class
1 (MNL)	-1236,78	16	0,1686	2503,56	2583,37	100 %
2	-1053,12	32	0,2815	2168,25	2333,2	31 %
3	-1174,46	48	0,2008	2414,91	2590,51	0 %

As can be seen, the 2 class model scores better on all estimates than the MNL model without classes. Therefore, allowing for consumer heterogeneity by estimating a LCCM apparently leads to better estimates. Additionally, it is apparent that adding a third class leads to worse statistics. With a probability of 0 % for class 3 and very large standard errors for the estimated parameters, a third class with differing preferences could not be identified. Therefore, the 2 class model was used for subsequent steps.

It is important to note that for the second class only the ASC was measured and no attribute parameters. The first model run with 2 classes resulted in an ASC of -13,65 for the second class, this indicates that the second class would derive -13,65 utils from an app as opposed to the 'no app'-option. Additionally, the standard errors of the estimated βs were very large, with some values of s.e. above 400. This indicates that the probability that that class would choose an app alternative, either A or B, was so low, that there were too little observations for the influence of the parameters on the choice between A or B. Therefore, those estimated values had little prediction power. Consequently, those parameters were fixed to zero and only the ASC was estimated. This way of latent class modeling was also applied by (El Zarwi et al., 2017), who identified a 'non-adopting' class in their model for which only the ASC was estimated.

Subsequently, the covariates were added to the model. These allow for profiling of the segments identified. All socio-demographic variables, the user variable and the two perception factors were included in the model. The results of this model are displayed and discussed in section 6.3. This resulted in the following Goodness-of-Fit statistics, which turned out to be a significant fit improvement (LRS = 53,98; df =10):

Table 12: Statistics Extended MNL Model 2

Statistic Metric	Output
Log Likelihood	-1042,8
Adjusted ρ^2	0,356
AIC	2139,6
BIC	2283,27

6 Results

This chapter entails the results from the estimated models. Firstly, the ML and MNL models are discussed and compared. Secondly, the estimated alternative specific constant (ASC) is discussed and interpreted. Thereafter, based on the ML model the found estimated preference weights are discussed per attribute. Next, the personal characteristics and their influence on MTA uptake are discussed. Subsequently, the LCCM is touched upon.The LCCM discussion consists of the identified classes, the class specific preferences and the class membership estimation. Based on the latter, a posterior class allocation is executed to predict the probability that a respondent belongs to either of the classes based on their personal characteristics.

6.1. Model Selection & Interpretation

6.1.1. Discussion of MNL vs. ML model

To determine if the improvement in fit of the ML model compared to the MNL model is significant, a different measurement than the previously discussed Likelihood Ratio is used. The ML model is a different model using draws to estimate the error component, and thus the two models are non-nested. Therefore, the Ben-Akiva & Swait test is executed to determine if the model fit improvement is significant. The result of the test indicated that the chance that the MNL fits the data better even though it has a lower model fit is nearly 0,00 %. Therefore, the ML model will be used for interpretation from this paragraph forward. The MNL will be used for comparison with the LCCM model later on in section 6.3.

6.1.2. Final ML Model Fit

The final ML model had a Log Likelihood of -1006,42 and a ρ^2 of 0,317, which means that it can explain away 31,7% of the initial uncertainty. A big improvement was seen after adding the *user* variable. Which can indicate that the fact if a respondent was a user or not has a relatively big impact on predicting preferences. Furthermore, adding the perception factor *ProMTA* also increased the model fit significantly. In comparison, adding socio-demographic variables and interaction effects did not lead to a high improvement.

6.1.3. Interpretation of ML Results

All attributes are effect coded and therefore, the average utility of an app attribute is zero. The utility of a level from the attribute is thus the utility it adds or detracts in relation to the average utility of all levels from that attribute. The same goes for the covariates, except for age and the ProMTA factor, which are both continuous variables. The final results are displayed in table 13 along with the MNL model estimates for comparison. This table contains only the robust t-ratio values. Appendix H contains a more extensive table, displaying the estimated ML attributes along with the t-ratio, two sided p-value and the 95% confidence interval. It is interesting that for most attributes the first indicator variable is significant, except for the *use* and *extra function* attributes. A insignificant second variable normally indicates that the attribute parameter is linear, meaning that the utility increases or decreases similarly between levels. However, as these variables are all categorical, the concept of linearity in general is not applicable.

Table 13: Estimated Results ML & MNL models

	Mixed Logit Model			**Multinomial Logit Model**		
	Est.	**Rob. s.e.**	**Rob. t.rat. (0)**	**Est.**	**Rob. s.e.**	**Rob. t.rat. (0)**
BETA_high_costs	-0,936	0,102	-9,186	-0.693	0.090	-8.412
BETA_low_costs	-0,345	0,114	-3,039	-0.346	0.089	-3.428
BETA_consent	0,431	0,157	2,753	0.506	0.077	6.289
BETA_information	0,305	0,124	2,464	0.071	0.086	0.872
BETA_icons	0,348	0,078	4,452	0.341	0.073	4.684
BETA_simple_language	-0,034	0,090	-0,374	-0.026	0.082	-0.367
BETA_high_accuracy	0,943	0,109	8,611	0.810	0.077	9.065
BETA_low_accuracy	0,138	0,095	1,460	0.011	0.087	0.129
BETA_reminder	0,071	0,077	0,918	0.154	0.075	2.389
BETA_single_entry	-0,021	0,091	-0,227	-0.108	0.083	-1.362
BETA_features	0,466	0,139	3,365	0.256	0.116	1.960
BETA_notifications	-0,177	0,128	-1,386	-0.114	0.114	-1.156
BETA_chat	0,097	0,085	1,136	0.107	0.077	1.423
BETA_health_information	0,116	0,085	1,368	0.063	0.082	0.911
BETA_age	-0,031	0,028	-1,111	-0.019	0.009	-1.075
BETA_ethnicity	-0,493	0,246	-2,003	-0.339	0.082	-2.076
BETA_education	-0,509	0,272	-1,874	-0.322	0.097	-1.949
BETA_user	0,366	0,209	1,755	0.213	0.070	1.639
BETA_ProMTA	1,141	0,293	3,896	0.712	0.096	3.705
BETA_features_ethnicity	-0,369	0,135	-2,727	-0.249	0.106	-2.088
BETA_notifications_ethnicity	0,297	0,128	2,322	0.175	0.107	1.814
ASC	-4,396	1,587	-2,771	-3.008	0.458	-2.990
SIGMA_app	1,944	0,177	10,957	-	-	-

6.2. Estimation Results ML

6.2.1. ASC app alternative

In general the ASC measured in utility functions depicts the utility derived from an alternative compared to the other alternative, that is not measured by included features. Applied to the base ML model without background variables, it measures the average utility consumers derive from or associate with downloading an app (both alternatives) compared to no app, that is not explicitly measured through the included attributes.

However, the above stated interpretation only applies if the model only includes attributes and no other covariates. As the final ML model also includes background variables, the magnitude of the estimated constant changes, which complicates interpretation of the ASC.

The ASC estimated for the final model, has a value of -4,40 and is thus hard to interpret. However, in the first ML model that was estimated, only effect-coded attributes were included, and the ASC had a negative and significant value of -1,22. This indicates that for that estimated model the average utility derived from choosing an app over no app is negative. Meaning that on average the sample had a preference for downloading no app as compared to downloading either one of the two apps.

For the ASC of the final ML model a σ was estimated to examine possible heterogeneity of the constant among the sample. The σ of 1,94 depicts the heterogeneity of that variable. This shows that among unobserved preference for the no downloading option compared to the app alternatives, the utility association differs between respondents. The probability density function of this constant in presented in figure 3 down below. It shows that even though there is variance among the respondents for the constant, for most respondents the ASC is negative. There is however, also a small part within the range that has a positive value and thus not all respondents have a negative ASC value.

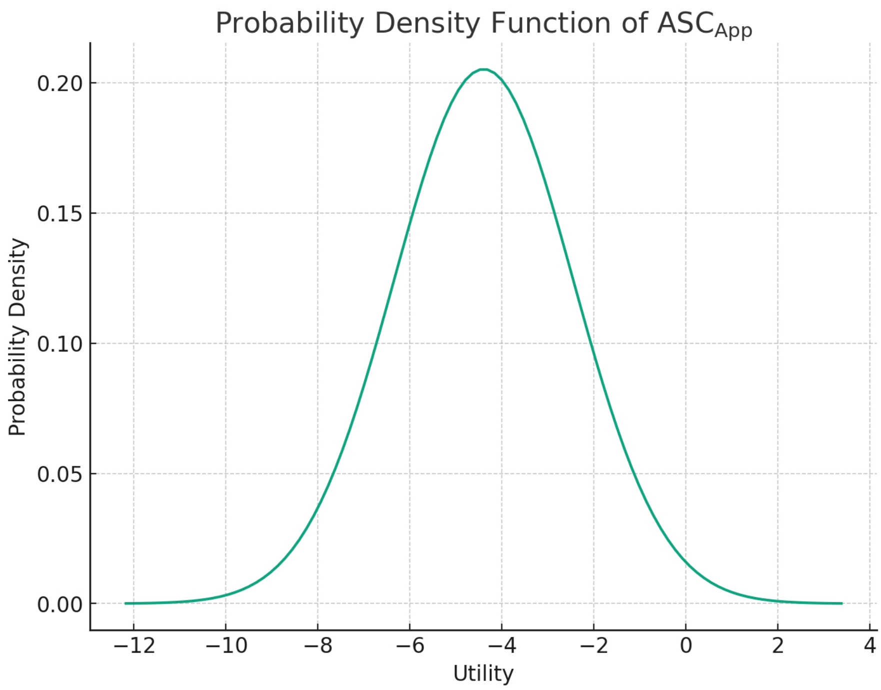

Figure 3: Probability Density Function of ASC_{App}

6.2.2. Identification of App Preferences

Firstly, the relative utility contributions of all attributes are compared. Afterwards, all level contributions are discussed per app attribute. Additionally, the interaction effect between ethnicity and personalisation is discussed. Thereafter, the influence of personal characteristics will be touched upon.

Comparison of Attribute Influence

To compare the influence the attributes have on the utility function and thereby obtain an idea of which attributes determine app preferences most, the utility range of all attributes were compared. Due to the average utility of all attributes being zero and their utility range is equal to the most highest estimated β value level of the attribute. Even though all attributes have the same levels and thus can be compared directly, to obtain more clarity the relative importance will be measured. This is done by dividing the utility range of an attribute by the total utility range from all attributes together. The obtained results in 4 show that the lowest impact on the utility function is made by the attributes *use* and *extra function*. The highest contributor is the *cost* attribute, followed by *accuracy* and *privacy*. The attributes personalisation and communication are of moderate influence. This sample thus bases their choice for a MTA mostly on its price and the level of accuracy it provides as these two are responsible for 31% and 26% of the utility respectively.

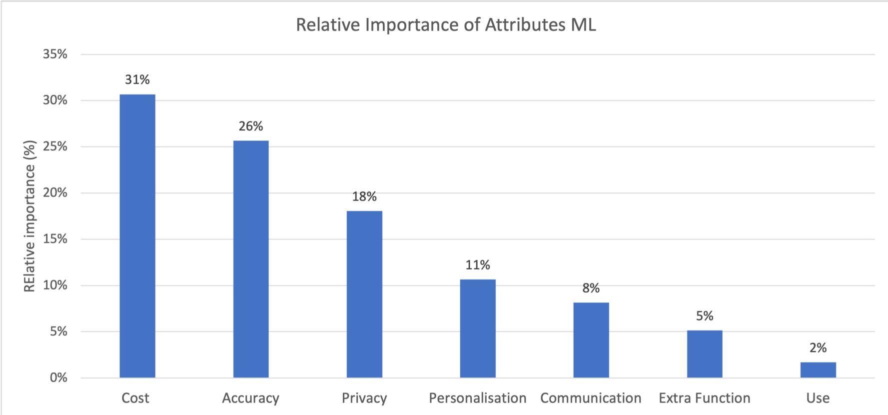

Figure 4: Relative Importance Attributes

Cost

The utility contributions of all the *cost* levels can be seen in figure 5. The signs are as expected; a free app contributes the most to the utility and the most expensive app derives the most from the utility. The amount in which the free app level contributes to the utility compared to the average utility of *cost*, is relatively high with 1,28 utils. The utility difference with the second level €2,50 is 1,63 as opposed to the difference between the €2,50 and €5,00 level being 0,59. The utility contribution of this attribute is thus not linear. Meaning that the biggest decline in utility contribution compared to average happens when a MTA is raised from €0,00 to €2,50. If the app price is increased with €2,50 again the difference in utility contribution is significantly smaller. The first increase in price is thus more influential.

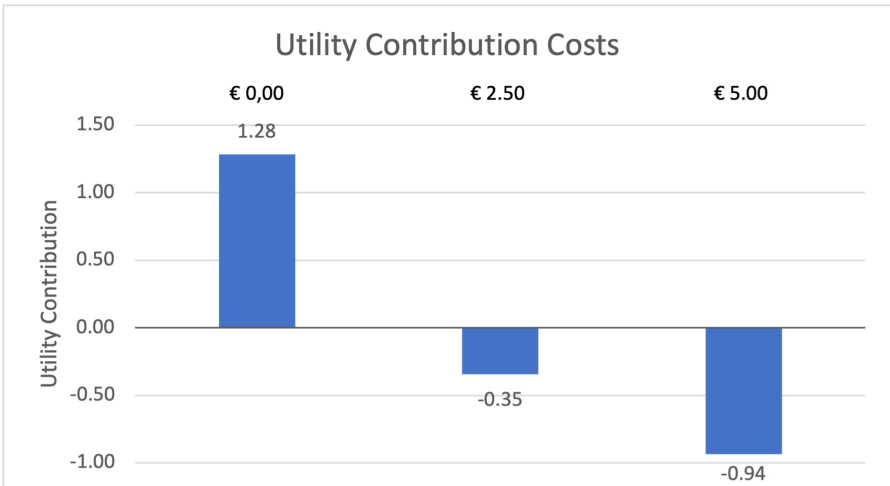

Figure 5: Utility Contribution Cost

Privacy

The utility contributions of all the *privacy* levels can be seen in figure 6. As can be seen, an app offering information about the way it handles data, contributes to the utility of an app. Whereas an app that supplies no information detracts from the utility of the app. Interestingly, the additional utility offered by asking for permission next to supplying information is only 0,13.

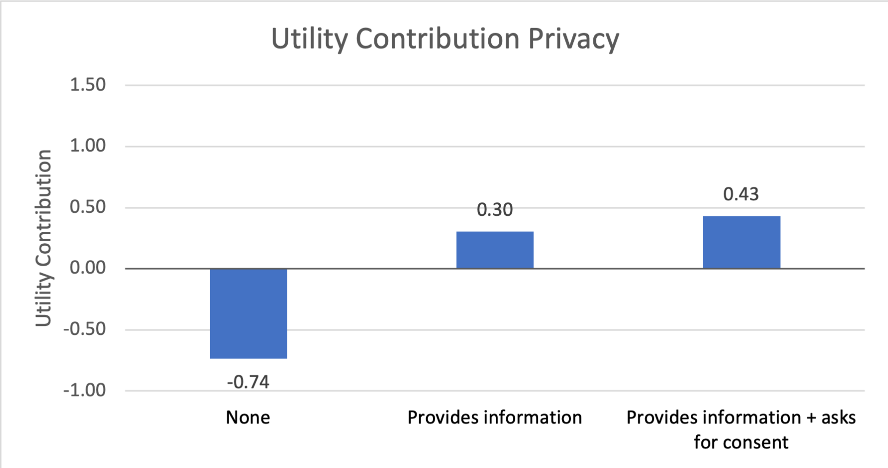

Figure 6: Utility Contribution Privacy

Communication

The utility contributions of all the *communication* levels can be seen in 7. Again here the expected preferred level has the highest utility contribution compared to the average contribution. Noticeably, is that an app communicating with simple language detracts utility compared to the average utility contribution of the communication app, however it is still preferred to communication with medical terms. The added value of using icons and images for communication within this sample is apparent.

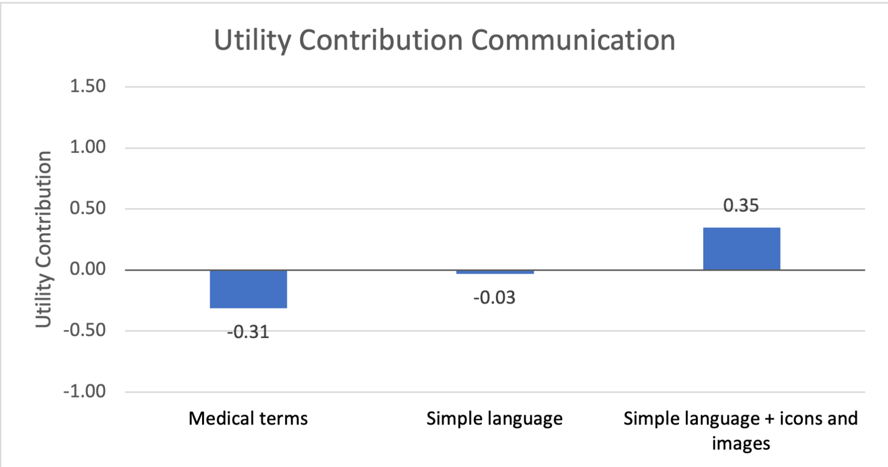

Figure 7: Utility Contribution Communication

Accuracy

The utility contributions of all the *accuracy* levels can be seen in 8. As can be expected the highest accuracy level is the most preferred. The highest accuracy level adds significantly more to the utility contribution as compared to the middle accuracy level. The lowest accuracy level has a significantly lower utility contribution than average, with -1,08 utils.

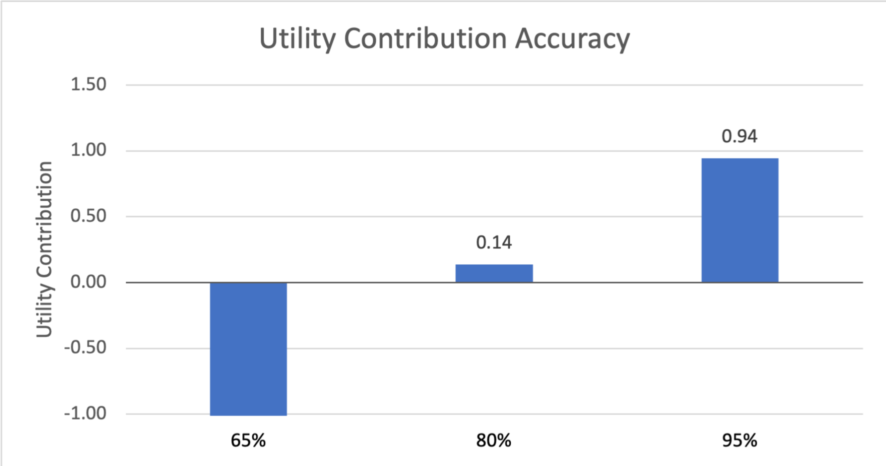

Figure 8: Utility Contribution Accuracy

Use

The utility contributions of all the *use* levels can be seen in 9. The first thing that is noticeable, is the very low influence of this attribute of all levels. The most influence this attribute can exercise on the utility is -0,07, which in comparison to all the other attributes is very little. This indicates that this attribute is less important for respondents compared to the other included attributes. Which is interesting as ease of use is often mentioned as very important. This is discussed extensively in chapter 8. Consequently, interpreting the utility contributions is harder due to little to no difference between the levels.

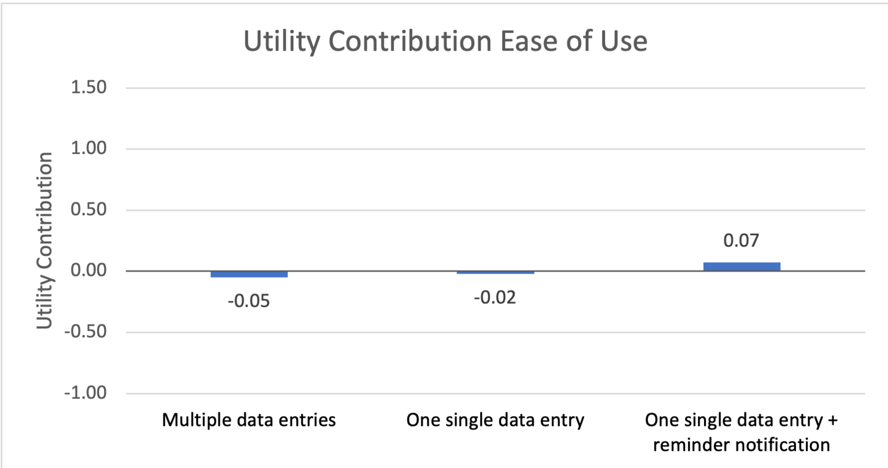

Figure 9: Utility Contribution Use

Personalisation

The utility contributions of all the *personalisation* levels can be seen in 5. The ability to personalise both reminders and functions adds the most utility compared to average. The added value of the ability to turn off functions is apparent compared to the utility contribution of just personalising the notifications. Which is not that much more preferred than being able to personalise nothing.

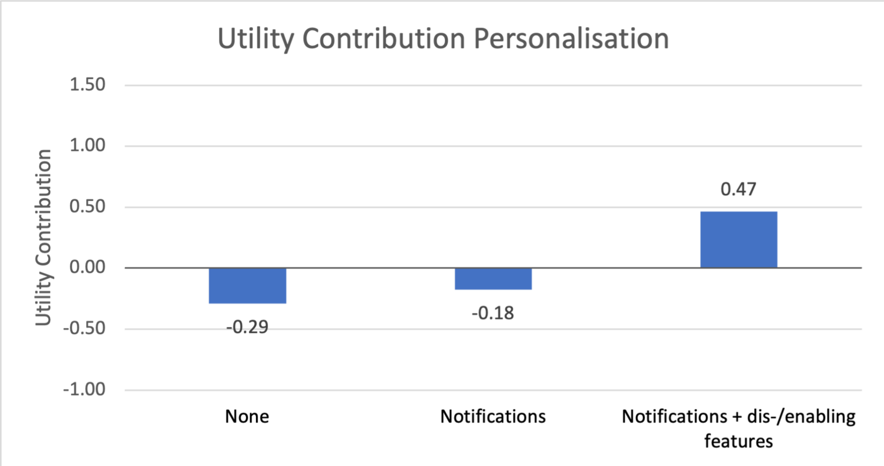

Figure 10: Utility Contribution Personalisation

Interaction Effect

The attribute personalisation turned out to interact with the personal characteristic ethnicity. This shows that in this sample Dutch and non-Dutch respondents valued the levels of personalisation differently. As shown in table 13, the estimated interaction βs are -0,369 for the features component of personalisation and 0,297 for the notifications component. To illustrate the meaning of an interaction effect clearly, the part of the utility function that includes the *personalisation* attribute will be filled in. The function looks as follows:

$$\begin{aligned} U_{personalisation} =& features \cdot (\beta_{features} + \beta_{features,ethnicity} \cdot ethnicity) + \\ & notifications \cdot (\beta_{notifications} + \beta_{notifications,ethnicity} \cdot ethnicity) \end{aligned} \tag{10}$$

Now substituting the found β values into this function, we obtain:

$$U_{personalisation} = features \cdot (0,466 + -0,369 \cdot ethnicity) + notifications \cdot (-0,177 + 0,297 \cdot ethnicity) \tag{11}$$

Now let ethnicity be 1 (i.e. a Dutch respondent), the equation takes on the following form:

$$\begin{aligned} U_{personalisation} =& features \cdot (0,466 + -0,369) + notifications \cdot (-0,177 + 0,297) \\ =& \ features \cdot 0,097 + notifications \cdot 0,120 \end{aligned} \tag{12}$$

If a respondent is non-Dutch (i.e., ethnicity = -1), the equation takes on the following form:

$$\begin{aligned} U_{personalisation} =& features \cdot (0,466 + 0,369) + notifications \cdot (-0,177 + -0,297) \\ =& \ features \cdot 0,835 + notifications \cdot -0,474 \end{aligned} \tag{13}$$

This shows that the impact of this is more influential for non-Dutch respondents. If the personalisation level is at the highest level (i.e., $features = 1$ and $notifications = 0$), the non-Dutch respondents derives more utility from the personalisation attribute of that MTA than a Dutch respondent would. However, as soon as the personalisation level decreases, the utility drops at a higher rate for non-Dutch respondents than for Dutch respondents. Which makes the chance that a respondent downloads a MTA with a lower personalisation level decrease more if the respondent is non-Dutch opposed to Dutch. Applied to the level definitions, this means that the addition of personalising functions on top of notifications is thus valued more highly by non-Dutch respondents compared to Dutch respondents.

Extra Function

The utility contributions of all the *extra function* levels can be seen in 11. Again, this attribute appears to be relatively unimportant, as the most influence it can exercise on the utility of an app is -0,17. What is interesting to note is that these values do indicate a slight preference for a chat function as an extra feature over information on female health.

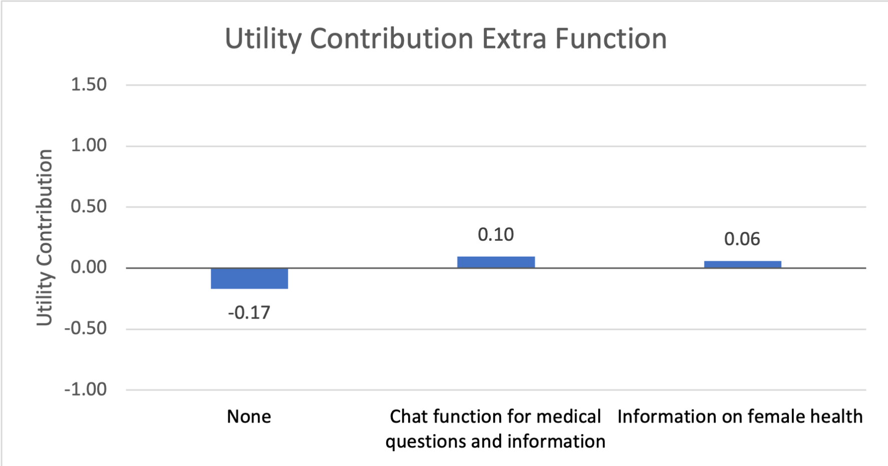

Figure 11: Utility Contribution Extra Function

6.2.3. Influence of Personal Characteristics

To determine if different consumers make different choices with regards to MTAs, personal characteristics were added to the utility function. These consisted of socio-demographics, user experience and attitudes towards MTAs. Apart from the influence on MTA uptake in general, it was also examined if different consumers had different attribute preferences. This turned out to be true only for ethnicity and the personalisation attribute, which is discussed in section 6.2.2.

Socio-Demographics

The socio-demographics *age*, *income*, *education*, *ethnicity*, *work situation* were all added to the model. Only age, education and ethnicity were significant. As discussed earlier, age became insignificant when attitudes were added to the model, but is still taken into consideration when discussing the influence of socio-demographics.

For the continuous age variable figure 12 shows the influence it has on utility. As expected with a linear estimate, the older the respondents are the less likely they are to download an app. With the utility decreasing with -1,64 for the oldest respondent of 53 years.

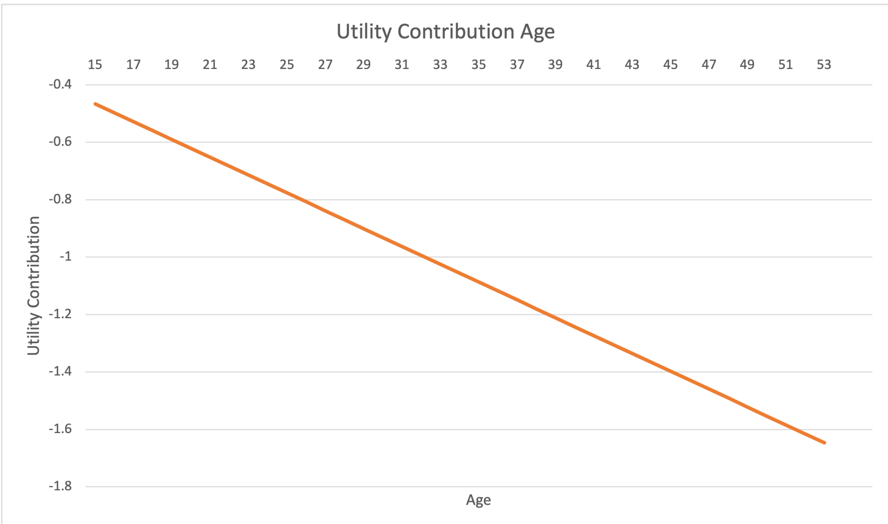

Figure 12: Utility Contribution Age

Respondents are more likely to download an app than the average respondent when they are not Dutch or have a mixed ethnicity or if they have a low or middle educational level.

As the reviewed literature states the opposite, it is interesting to further discuss these findings. A possible explanation for this finding is that firstly, the variance in these variables was very slight. Most of the sample consisted of highly educated Dutch respondents. Often at least 30 respondents is preferred for subgroups in samples, and both these groups do not contain that number of respondents, therefore this could have influenced the estimations.

Furthermore, the marginalized groups found in literature are not represented in this sample composition. So the ethnic group defined as non-Dutch and the low & middle income group as defined in this research do not represent the marginalized or low SEP groups identified in research for which the differing outcomes were found. The ethnic group defined as non-Dutch consists of 11 respondents identifying as a different ethnicity than Dutch and 9 respondents identifying as Dutch mixed with another ethnicity. These were grouped together to allow for a more significant non-Dutch respondent group, which results in a different ethnic subgroup than often targeted in research. The lower education group in this sample consists of mostly middle education and two VMBO high-schoolers. Therefore, the definition of low education in this sample does not overlap with the target group that is identified as low educated in research.

The other socio-demographic variables *work situation* and *income* were not of influence on the utility derived from a MTA. Therefore, students, full-time or part-time workers and respondents without a job or homemakers are not more likely to download a MTA as opposed to the average respondent. However it must be noted that the latter two were grouped into one category for the estimation and that group still consisted of only 12 respondents, which may have influenced the lack of effect. Furthermore, respondents are not more or less likely to download a MTA than the average respondent based on their income range.

User vs. Non-user

Additionally, the fact if a respondent was a current MTA user or had ever used a MTA was included into the model. This appeared to be a very insightful predictor for MTA uptake. A respondent was more likely to download an app if they ever used a MTA than if they had not. This effect is logical as being familiar with technology is an important factor in mHealth uptake (Rajak & Shaw, 2021). The influence of the *user* variable is 0,208 for users and -0,208 for non-users.

Attitudes

Lastly, the measured perception factors were added to the model. Looking at the first attitude factor *ProMTA*, which measures respondents' positive attitude towards MTAs, there is a significant influence present. Respondents that score higher on that perception factor and thereby are more open to using MTAs and perceive MTAs as more useful, are more likely to download a MTA. This attitude had a significant utility contribution with its utility contribution ranging from 0,741 to 3,705.

The second factor was not significant. Indicating that for this sample group, respondents that scored higher on the second factor *FemaleHealthKnowledge* were not more likely to download a MTA compared to the average respondent. As this factor measured the perceived importance of knowledge on female health, this means that women who find knowledge on female health important, are not more likely to download a MTA than women who are neutral towards the importance of female health knowledge.

6.3. Estimation Results LCCM

To identify heterogeneity among the sample apart from the heterogeneity estimated by incorporating the error component as discussed in 6.2.1, a LCCM model was estimated. The results of that model will be discussed in this chapter. The final estimates can be seen in table 14

6.3.1. Identification of Classes

The estimated LCCM identified two classes. However, as discussed in chapter 5, the second class contains only the estimated ASC due to high standard errors for the attributes. This class can be defined as the 'non-adopters'. Indicating that a substantial part of the respondents prefers to download no app regardless of the attributes. This is in line with the sigma found for the distribution of the error component in 6.2.1, that found that for most respondents the unobserved utility for either of the apps compared to no app, was negative.

The estimated parameters and the class membership values can be seen in table 14. The estimated parameters indicate the weight of the attribute in determining the utility. The estimated γ depicts how important a personal characteristic is in determining if a respondent is likely to belong that class.

6.3.2. Class Specific Preferences and Covariates

Since two classes are identified, but parameters are estimated for only one class, no comparison can be made between the preferences of the classes. However, the ASCs of both classes can be compared. Additionally, the relative importance of the attributes of the adopters segment will be compared against those of the MNL model.

As explained in section 6.2.1, the ASC interpretation depends on the variables included in the estimation. For the LCCM model a MNL model is estimated without incorporating any covariates. Therefore, the utility is determined by attributes only and the constant is not influenced by background variables. Consequently, the constant can be interpreted as the average utility difference from all app alternatives shown to respondents as compared to the utility of the no app option. As can be seen the second class *non-adopters* has a constant of -4,050. Indicating that no matter how the alternatives were presented, on average the segment obtained a negative utility for choosing an app. As expected their constant is much lower than that of the first class *adopters*. The adopters segment also experiences a slight negative ASC value. Thus the utility difference between the no app option and the average of all alternatives shown is slightly negative. However, the constant is significantly lower than for the non-adopters segment.

For the adopters segment the utility range of each attribute is retrieved and subsequently the relative importance is calculated. These values are compared to the MNL model, shown in 13. The estimated values for the MNL model can be found in 13.

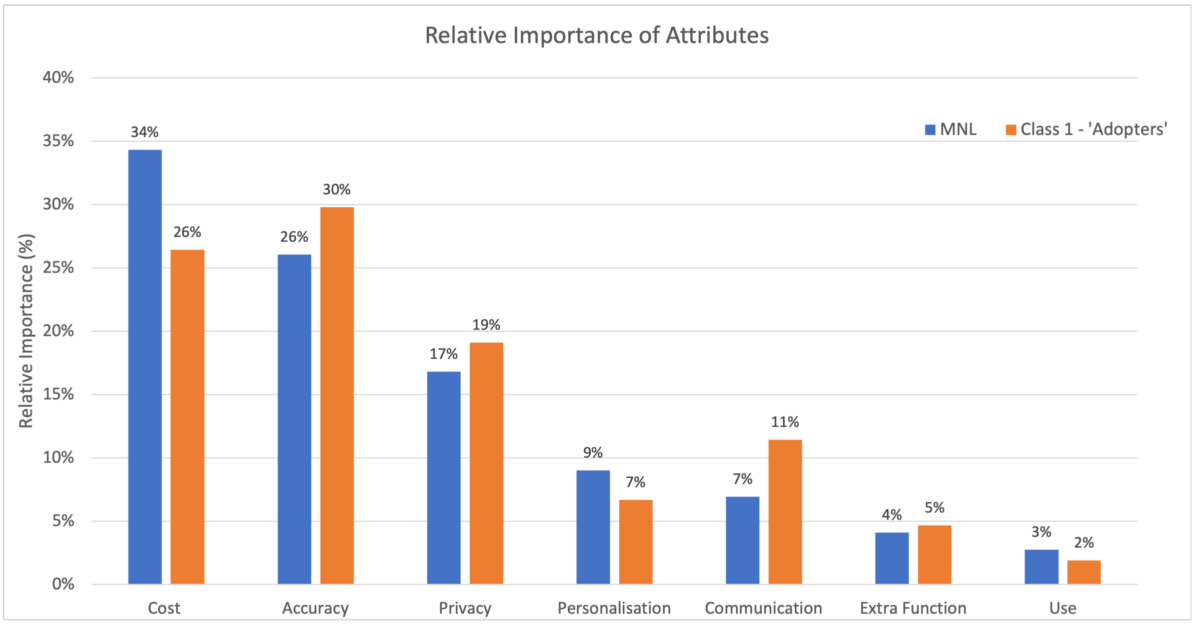

Figure 13: Comparison of Relative Importance Attributes

The figure shows that the sequence of important attributes changes for this class as opposed to the MNL model. Firstly, the most important attribute is no longer *cost*, but *accuracy*. This may be due to this segment having more preference for a MTA than the other segment and therefore is more open to app use. Being more open to app use might make the segment more willing to pay for an app and address more importance to the *accuracy* of a MTA. *Privacy* is still the third most important attribute. Interestingly, this segment appoints more importance to *communication* as opposed to *personalisation* compared to the MNL model. As to why this effect occurs, more research is required. The *extra function* and *use* attribute are still relatively unimportant.

At the bottom half of the table the class membership estimates are displayed. The estimated constant shows the probability of belonging to that class regardless of the respondents' characteristics. The higher that constant is, the higher the probability of belonging to that segment. As the estimated delta for the class membership of the *adopters* segment has a low value but a high probability of 74%, this does not comply with that statement. However, there is no delta to compare this delta with to determine if it is relatively high. Additionally, the delta parameter is not significant. Taking this into account, it is hard to base conclusions on this estimate for delta.

The other membership estimates are estimated compared to the reference class. In this case the *'non-adopters'* class was the reference class. The estimates should be interpreted as follows: a positive and significant estimate, such as γ_{user}, indicates that being a MTA user improves the probability of belonging to that segment compared to belonging to the second segment. Generally a significant estimate is likely to be a predictor of class membership.

6.3.3. Class Memberships

Looking at the estimated gamma values estimated for the personal characteristics, it is evident that there are two coefficients that exercise a larger influence than the others. These are the only two significant coefficients $gamma_{\text{user}}$ and $gamma_{ProMTA}$. Both of these have relatively high positive estimations. The first estimated coefficient indicates that if a respondent is an user the probability that the respondent belongs to the adopter segment is higher. This an expected relation due to previous or current users having adopted a MTA at least once. Therefore, they would logically be a part of the *adopter* segment. The second significant estimated parameter is the first perception factor *ProMTA*. The higher a respondent scores on the pro MTA attitude factor, the more likely this respondent is to be a part of the *adopter* segment. Again, this effect can be expected as being more open towards MTA use and perceiving MTAs as useful, can both contribute to being more likely to adopt a MTA.

Even though both of these factors are quite significant and influential, their added contribution in uptake recommendations is less obvious. For the application of this research to provide recommendations to app developers, characterizing the adopter segment based on previous use and the attitude towards MTAs do not provide the most tangible recommendations to app developers. As this information is not always easily available to app developers, the absence of these factors was examined. It must be noted that this is not common practice within the DCM field. However, this was done to explore if more relevant results could be found to ensure useful application of the method. Consequently, firstly the user variable and the *ProMTA* factor were removed from the model. This resulted in the second perception factor *FemaleHealthKnowledge* being significant. The value estimated for this factor implied that the more a respondent assigned importance to knowledge on female health, the more the respondent was likely to belong to the *adopters* segment. As this factor also entails knowledge of attitudes of possible consumers, which can be hard to determine, further analysis was done without incorporating this factor. This resulted in *age* becoming a significant estimate. The estimated gamma value of *age* was -0,07. Thereby indicating that the older the respondent is, the higher the chance they do not belong to the first segment of *adopters*.

6.3.4. Posterior Class Allocation

The estimated LCCM model offers the opportunity to determine the probability for each individual that they belong to the estimated classes. This is done as follows: The posterior class allocation of each individual for each class is given by the model calculation. For each level of the personal characteristics the sum of the probabilities of the individuals that obtain that level of the characteristic is summed and divided by the total probability of that class. The calculated values and the amounts of respondents corresponding to the characteristic level are displayed in 15.

The four personal characteristics named in the section above are included in the posterior class allocation. Furthermore the factors scores were divided into four categories. Lastly, the sample composition was included.This is necessary to compare the class compositions to and determine if specific characteristics influence the probability that an individual belongs to a certain class.

Profile of the Adopter Segment

As table 15 shows, the individuals belonging to this segment are way more likely to be a (previous) MTA user than the second segment and even more likely as compared to the sample. Furthermore, the age distribution in this segment is similar to that of the sample. Comparing this distribution to the distribution of the second segment, it contains relatively less older individuals, seeing the biggest difference in the 19-25, the 41-45 and the 45+ age categories. Furthermore, compared to the sample, this segment consists of a lot of individuals with a *ProMTA* score above 4,00. Compared to the other segment it contains way more individuals within the highest score level, but significantly less within the second highest score level. The other score levels are represented relatively less than in the sample and significantly less than in segment 2. Lastly, for the second factor *FemaleHealthKnowledge* the last level is also significantly more present in the first class than in the sample and the first segment. However, the second and third score levels are less represented than in segment 2, but relatively similar to the sample composition. In conclusion, using or having ever used a MTA, being younger than 40, being very pro MTAs and valuing female health knowledge as very important, improves the probability of belonging to the *non-adopter* segment.

Profile of the Non-adopter Segment

The most noticeable determinant is MTA experience. If an individual has not used a MTA before it is highly likely that they belong to the non-adopter segment compared to the other segment, sample composition and average. Furthermore, the older two age categories are significantly more represent in this segment. The individuals scoring up until 3,99 on the first factor *ProMTA* are relatively more present in this segment than in the sample and significantly more than in the other segment. Lastly, examining the second factor *FemaleHealthKnowledge*, shows that individuals that the third scoring level is more represent in this segment. The highest level is significantly less present compared to the sample and the first segment. In conclusion, being a non-user, that is older than 40, is not pro MTAs or somewhat pro MTAs and values female health knowledge as not important to somewhat important, increases the probability of belonging to the *non-adopter* segment.

Table 14: LCCM Model Results

	Adopters					Non-adopters				
Variable	**Est.**	**Rob. s.e.**	**Rob. t-ratio (0)**	**CI 95%**		**Est.**	**Rob. s.e.**	**Rob. t-ratio (0)**	**CI 95%**	
ASC	-0,553	0,132	-4,200	-0,711	-0,395	-4,050	0,362	-11,199	-4,681	-3,423
β_high_costs	-0,786	0,093	-8,425	-0,974	-0,598	0,00	-	-	-	-
β_low_costs	-0,409	0,113	-3,618	-0,596	-0,222	0,00	-	-	-	-
$\beta\_consent$	0,568	0,087	6,523	0,406	0,729	0,00	-	-	-	-
$\beta\_information$	0,117	0,084	1,389	-0,056	0,290	0,00	-	-	-	-
β_icons	0,340	0,074	4,595	0,189	0,490	0,00	-	-	-	-
$\beta\_simple\_language$	-0,011	0,086	-0,124	-0,192	0,171	0,00	-	-	-	-
$\beta\_high\_accuracy$	0,886	0,102	8,700	0,721	1,050	0,00	-	-	-	-
$\beta\_medium\_accuracy$	0,053	0,088	0,600	-0,128	0,234	0,00	-	-	-	-
$\beta\_reminder$	0,049	0,077	0,640	-0,112	0,210	0,00	-	-	-	-
β_single_entry	-0,057	0,094	-0,603	-0,240	0,127	0,00	-	-	-	-
$\beta\_features$	0,199	0,099	2,003	0,015	0,382	0,00	-	-	-	-
$\beta\_notifications$	-0,022	0,078	-0,286	-0,192	0,147	0,00	-	-	-	-
β_chat	0,068	0,082	0,835	-0,089	0,226	0,00	-	-	-	-
$\beta\_health\_information$	0,139	0,080	1,727	-0,035	0,312	0,00	-	-	-	-
Class Membership										
Probability	74%					26%				
δ	-2,126	1,936	-1,098	-5,575	1,322	0,00	Reference			
γ_age	-0,040	0,035	-1,117	-0,119	0,040	0,00				
γ_income	-0,113	0,212	-0,531	-0,535	0,309	0,00				
$\gamma\_education$	-0,366	0,329	-1,113	-1,053	0,321	0,00				
$\gamma\_ethnicity$	-0,457	0,473	-0,966	-1,387	0,474	0,00				
γ_other	-0,634	0,633	-1,001	-2,048	0,780	0,00				
$\gamma\_student$	-0,611	0,479	-1,273	-1,604	0,383	0,00				
γ_user	0,818	0,297	2,754	0,280	1,356	0,00				
$\gamma\_parttime$	0,096	0,448	0,214	-0,823	1,014	0,00				
γ_ProMTA	1,008	0,438	2,299	0,203	1,813	0,00				
$\gamma\_FemaleHealth.$	0,452	0,344	1,314	-0,230	1,135	0,00				

Table 15: Posterior Class Allocation Characteristics

	Adopters		Non-adopters		Sample Composition		Sample Average
	Amount	Percentage	Amount	Percentage	Amount	Percentage	Percentage
User							
Yes	84	68%	9	20%	93	55%	44%
No	39	32%	36	80%	75	45%	56%
Age							
tot 18	7	6%	4	9%	11	7%	7%
19-25	59	48%	19	42%	78	46%	45%
26-30	29	24%	10	22%	39	23%	23%
31-35	16	13%	3	7%	19	11%	10%
36-40	6	5%	2	4%	8	5%	5%
41-45	2	2%	3	7%	5	3%	4%
45+	4	3%	4	9%	8	5%	6%
ProMTA factor ProMTA (score)							
1-1,99	1	1%	5	11%	6	4%	6%
2-2,99	7	6%	8	18%	15	9%	12%
3-3,99	36	29%	25	56%	61	36%	42%
4-4,99	79	64%	7	16%	86	51%	40%
Factor2 FemaleHealthKnowledge (score)							
1-1,99	2	2%	2	4%	4	2%	3%
2-2,99	7	6%	7	16%	14	8%	11%
3-3,99	44	36%	23	51%	67	40%	43%
4-4,99	70	57%	13	29%	83	49%	43%

7 Model Application

This chapter contains the application of the estimated model. Different scenarios are constructed and the associated app uptake is predicted. The different scenarios are created by varying the values of the attributes and providing these as input to the choice model. The model then computes the probability of the choice for downloading an app or no app. Based on those probabilities, recommendations to app developers can be made with regards to the influence of app factors on app uptake.

7.1. LCCM Application

As the LCCM is able to separately predict uptake for *adopters* and *non-adopters*, it is the model used for application. The values used for scenario's and prediction are equal to the final LCCM model estimates, which can be seen in 14.

The predictions of the *adopters* segment and the entire model were included into the result tables. The prediction of the *non-adopters* segment is only dependent on the ASC and thus did not vary between scenarios. The predictions for the *non-adopters* segment throughout all scenarios were 97% for downloading no app, and 3% for downloading an app. The predictions for both segments together are calculated as follows: It takes the probabilities of both segments and multiplies them by the chance of the respondent belonging to that segment, which is 74% for segment one and 26% for segment two. As the adopters segment is relatively large as compared to the non-adopters segment, the probability uptake associated with that segment weighs more in determining the overall probability uptake, which results in a higher uptake probability.

7.2. Base Scenario

Firstly, the personal characteristics of the respondent were set to the most average (e.g., age) or most occurring (e.g., ethnicity) values. Secondly, the base scenario represented a plain app with the lowest levels as can be seen in 16. This base scenario was chosen to expose if adding or improving the levels of the attributes would result in potential uptake improvement, enhancing MTA recommendations. Another potential base scenario could have resembled the current average app available on the market. However, the research aim is not to predict uptake for existing MTAs, but to investigate possible apps that would have high uptake among a diverse consumer group. Therefore, an 'average' MTA was not used as the base scenario.

The base scenario levels and the uptake probability for the different segments are shown in table 16. As shown, the uptake probability for this scenario is relatively low for all segments. Indicating that a free app is not that desireable if it is very basic and simple. As expected the lowest uptake is expected by segment non-adopters of the LCCM model. This expected uptake for this segment applies to all scenarios.

Table 16: Base Scenario Values

App attribute	Level Value
Costs	€0,00
Privacy	None
Communication	Medical terms
Accuracy	65 %
Ease of Use (Data Input)	Multiple data points input
Personalisation	None
Extra Function	None
Uptake Probability LCCM Adopters Segment	28 %
Uptake Probability LCCM Non-adopters Segment	3%
Uptake Probability LCCM Both Segments	18 %

7.3. Scenario Results

To establish the trade-offs a developer can make with MTAs, a high quality free app is tested, then paid apps are explored and lastly a regulation scenario is executed. For all scenarios the average of the choice probabilities of the respondents choosing either no app or either one of the apps are displayed. Also, these estimates are compared to the base scenario. The results are displayed per scenario, and discussed and compared in section 7.4.

7.3.1. Free High Quality Apps

Secondly, the uptake probability for the best app were estimated, with the results displayed in table 17. The best app was free and contained all the highest levels. Since it is not clear which extra function of the app is preferred most, both are tested in the scenario. This app is downloaded by 97%. As the class of the LCCM that is incorporated consists of (previous) users and individuals with a pro-MTA attitude, the high uptake is to be expected. A free app with all the best features and an extra function, results in almost all the class members choosing the app. Between the extra functions available, no considerate differences were observed. Apparently, the presence of one function does not convince possible users to download a MTA more than the other does.

Table 17: Free High Quality App Scenario Uptake

App attribute	Best App	
Costs	€0,00	
Privacy	Information + asks consent	
Communication	Simple language + icons and images	
Accuracy	95%	
Ease of Use (Data Input)	Single data point + reminder notification	
Personalisation	Reminders + dis-/enabling features	
Extra Function	Chat function	Information on female health
Uptake Probability Adopters Segment	97 %	98 %
Difference Base Scenario	+69%	+70%
Uptake Probability Both Segments	61%	62%
Difference Base Scenario	43%	+44%

7.3.2. Paying for Functions

The third scenario considers apps that require payment but contain all the features and best levels. It will tell how many consumers value extensive apps and would pay for these apps. Again both extra features - chat function and information on female health - were examined, which again did not lead to any differences in uptake. The LCCM model predicts that a substantial part of the respondents would pay a high amount for an app with high accuracy, usage of images and icons, full transparency on data use, easy data input, a high amount of personalisation and an extra function. Additionally, it is interesting to note that doubling the price has an effect on the uptake but it does not decrease significantly. This means that for a high quality app, app developers can charge a high price and still ensure 82% uptake among adopters and 52% among a normal consumer group.

Table 18: Paid High Quality App Scenario Uptake

App attribute	Moderately Expensive App		Expensive App	
Costs	€2,50		€5,00	
Privacy	Information + asks consent		Information + asks consent	
Communication	Simple language + icons and images		Simple language + icons and images	
Accuracy	95%		95%	
Ease of Use	Single data point + reminder notification		Single data point + reminder notification	
Personalisation	Reminders + dis-/enabling features		Reminders + dis-/enabling features	
Extra Function	Chat function	Information on female health	Chat function	Information on female health
Uptake Probability LCCM	88 %	88%	82 %	82 %
Difference Base Scenario	+60%	+60%	+54%	+54%
Uptake Probability LCCM Both Segments	60%	60%	52%	52%
Difference Base Scenario	+42%	+42%	+34%	+34%

7.3.3. Regulation Scenario

The regulation scenario examines how external influences that force app developers to develop certain apps, will persist in the uptake probability. A likely development in the mHealth sector is regulation, as discussed in 2.1.2. Currently, regulatory institutions, such as the European Commission, are in the process of updating or developing guidelines for mHealth development Essén et al. (2022). According to papers, these regulatory guidelines should be focused on privacy & security issues and the validity or efficacy of apps and their content (see 2.1.2). Therefore, this scenario simulates stricter guidelines to be followed by app developers with regards to the privacy and accuracy of the apps. 'Strict regulations' require app developers to hold their apps to the highest standards, in the case of this experiment; information and consent for privacy and 95% for accuracy. 'Moderate regulations' require 80 % accuracy and apps to provide information on their data use. Furthermore, if app developers are obliged to improve their apps, this could persist in app prices for the consumers. Therefore, the scenarios also include increasing the costs attribute. The other attributes are again similar to the base scenario.

The results in table 19 show that strict regulations with regards to privacy and accuracy of apps have a positive impact on the uptake of apps. As privacy and accuracy are the first and third most important attributes for the adopters segment, ensuring that these are up to the highest standards improves app attractiveness for consumers. It even appears that consumers are willing to pay for these improvements, as the uptake is still above 50% for the highest app price. However, for the mediocre regulations the substantial improvement only upholds if the app is free of charge. As soon as the consumer is expected to pay, the increased uptake of adding information on data use and improving accuracy to 80% is compensated by the utility decrease of price. In the case of an app price increasing to €5,00, it even decreases uptake probability. The difference in uptake between the app with consent and 15% more accuracy and the other app, is at its lowest for the free app (20% difference) and is higher when the app costs €2,50 (32% difference). When an app becomes more expensive, a higher privacy and accuracy level results in relatively more improvement in uptake probability than when an app is free. Furthermore, for both the scenarios going from a free app to an app of €2,50 has more negative impact on the uptake than raising the price from €2,50 to €5,00.

Table 19: Strict Regulation Scenario Values

App attribute	Strict regulations			Moderate regulations		
Costs	€0,00	€2,50	€5,00	€0,00	€2,50	€5,00
Privacy	Information + asks consent			Information		
Communication	Medical terms			Medical terms		
Accuracy	95%			80 %		
Ease of Use (Data Input)	Multiple data points input			Multiple data points input		
Personalisation	None			None		
Extra Function	None			None		
Uptake Probability LCCM	88%	62%	52%	68%	30%	24%
Difference Base Scenario	+60%	+34%	+24%	+40%	+2%	−4%
Uptake Probability LCCM Both Segments	56%	40%	34%	44%	20%	16%
Difference Base Scenario	+38%	+22%	+16%	+26%	+2%	

7.4. Discussion Scenarios

This section will synthesize the findings discussed above. Before the scenarios are discussed, it is important to note that most of the remarks are made with regards to the adopters segment probabilities, as the non-adopter segment probabilities are fixed. Therefore, the non-adopters segment probabilities and the combined probability do not provide additional insights. The four scenarios that were executed in this part revealed interesting insights, which will be discussed per scenario.

Firstly, it was investigated how much uptake a free app of high quality would yield. This showed that almost all (96%-98%) individuals from the adopters segment would download the best possible app and the majority of both segments combined would as well (60%-62%). Indicating that the highest uptake of apps is obtained when the app is free, provides a lot of accuracy, is transparent on data use, allows for easy data input, communicates with icons and images, allows for personalisation of notifications and features and contains an extra feature.

However, if an app developer would decide to supply that same app but increase the price, the uptake will decrease. With 82% of the adopter segment still downloading an app that is priced at €5,00, the price of an app is less relevant if all the other attributes are of the highest quality. Comparing the uptake to the previous scenario, which contained the same app but free of charge, the probability uptake only decreased with $\sim -14\%$. Increasing the price to €5,00 from the base scenario to the worst app scenario lead to a decrease of 23%. In that scenario all the attributes were kept at the lowest levels. Thereby this finding indicates that increasing the price but increasing the level of privacy, accuracy, communication, ease of use and adding an extra function, lessens the decrease in uptake. Additionally, the height of the price is influential, but app developers do not lose too much consumers if they request €5,00 as opposed to €2,50

Lastly, the regulation scenarios showed that improving just accuracy and privacy to the highest levels also slowed down decrease in uptake due to costs. However, increasing accuracy and privacy with one level was only beneficial for uptake if the apps were free of charge. Indicating that the adopters segments accepts a costly app only if it returns high quality of privacy and accuracy. Additionally, just improving accuracy and privacy lead to 88% probability uptake, which was a +60% increase from the base scenario. Again, indicating that mainly these app attributes influence the uptake probability of this segment.

8

Conclusion and Discussion

This thesis investigated the uptake of menstrual tracking applications (MTAs) among women from diverse backgrounds, emphasizing on the knowledge gap on different preferences among women.

This was done by firstly obtaining an overview of identified factors influencing MTA uptake and general mHealth uptake, which was synthesized to identify overlapping and differing factors. The most important factors were included in a DCE, studying the trade-offs respondents made between them. Based on that the influence of the attributes and of personal characteristics on app uptake was determined.

8.1. Conclusions Sub-research Questions

In the introduction four sub-research questions and one main research question were posed in order to fulfill the research goal. In this section the answers to these sub-research questions will be discussed individually and finally synthesized in order to answer the main research question. However, before these questions are discussed and answered, perhaps the most important conclusion with regards to including people with a low SEP in research is discussed.

Minority Groups in Research

A serendipitous but significant result of this research was the confirmation of difficulty to reach the target group. This research aimed to include a broad range of respondents to determine if their preferences would significantly differ. More importantly, the inclusion of respondents with a lower SEP was important. Therefore, it employed multiple methods to reach the demographic groups that are harder to reach, while designing the survey as discussed in 4.5 and while distributing the survey as discussed in 4.6. Among the other methods women's charities and female health organisations were contacted. A large amount of time within the distribution phase was dedicated to reaching these women. However, as discussed in 5.2 the respondent sample composition consisted of mostly Dutch, highly educated, young women. As this research did have time and resource constraints, these factors could have influenced the lack of diversity within the sample group. However, contacted professional agencies supplying panel participants for research, stated that even with their resources these socio-demographic groups were hard to reach. Which underlines that including females from minority groups or low SEP in research is difficult. Thereby, this research has underscored the importance of specifically targeting minority groups throughout the entire research, as they do not participate in surveys voluntarily even when these surveys are distributed in surroundings familiar to these groups. Within sections 8.5 and 8.6 more in-depth reasoning behind this and possible mitigations are proposed.

1. What barriers and facilitators for uptake of health apps in general and MTAs in particular have been identified?

According to the papers examined, there is an abundance of app-specific features that influence uptake. The app-specific themes that were derived from these features are: Costs, ease of use, accuracy, scientific grounding, privacy, design, discretion, the available functions, personalisation, reminders and the reputation of an app. For almost all of the factors, their absence or presence determines if its a barrier or facilitator. The papers focused on mHealth uptake by consumers with a low SEP, found the following additional features to be important: Tailoring, language, visuals and patient-centered design.

Furthermore, with regards to personal characteristics, papers established the influence of age and the associated stage-of-life a consumer is in. With motivations for using a MTA changing according to women's needs in their stage-of-life, e.g. wanting children or entering menopause. Other personal characteristics contain attitudes, perceptions or app/technology/phone user habits. Their influence on MTA uptake are discussed in the answer to sub-research question 2.

2. How do women from diverse backgrounds currently utilize, perceive and approach MTAs?
According to the reviewed papers that conducted research on MTA with inclusion of non-users, the awareness of MTAs was relatively high among the researched group. Which was underscored by the survey executed in this research. Additionally, the literature mentions that there is an overall positive attitude towards MTAs and in general the majority of the participants was willing to use these apps sometime in the future. This was found in all papers examined, which included diverse target groups. In this research this was underpinned, as the sample on average agreed with all the statements depicting willingness to use MTAs. On average respondents above 45 and non-users were less willing to use MTAs than respectively younger respondents and respondents that (previously) used MTAs. With regards to use, the reviewed literature found that MTAs were used mainly for these reasons: Trying to conceive, trying not to conceive, and tracking and understanding the cycle. Additionally, these reasons changed according to the age and stage-of-life of consumers.

With regards to the perceptions on efficacy and usefulness of MTAs, self-reported health knowledge improvement due to MTA use was evident in multiple reviewed papers. This underlines the usefulness that users ascribed to MTAs. This research confirmed this, as the sample was on average positive on the usefulness of MTAs. The perceived usefulness statements were answered even more positively by younger respondents and users than respondents above 45 and non-users.

3. To what extent do certain MTA attributes influence app uptake?
This question will be answered mainly based on the results of the adopters segment of the LCCM model estimated and discussed in respectively section 5.4 and chapter 6. The question will be complemented by the survey questions regarding additional app features as discussed in section 5.3.3.

The extent to which attributes influence uptake, was estimated for costs, privacy, communication, accuracy, ease of use, personalisation and extra features. Based on the estimated models, the extent to which privacy (19%), costs (26%) and accuracy (30%) are relatively important in determining app preferences is fairly large, whereas the influence of communication (11%) is moderate, and lastly the influence of personalisation, use and an extra function can be considered small.

Estimating scenarios of possible apps showed that a free app with low quality and no additional features, still acquired a moderate uptake. A free app of which all the attributes were at their highest level, had an uptake of 96%. When the app was no longer free, but required €5,00, the uptake was 82%.

Furthermore, the significant decrease in uptake due to increasing the price to €5,00, was mitigated by ensuring at least the highest levels of accuracy and privacy (uptake of 52%), but preferably by ensuring the highest levels of all attributes (uptake of 82%). Thus more than half of the sample is still willing to download an expensive app, if its accuracy is 95% and it provides information and asks for consent on data use.

Additionally, with regards to factors not specific to the features available in an app, medical grounding was noted as important by the sample. With apps approved in scientific research and made in cooperation with doctors or medical institutions having the highest average score. Both the sharing of data with third parties and the patient-centered design features were deemed important as well. Tailoring and recommendations (apart from doctors) were not deemed as important.

4. To what extent do fertility app attribute preferences and uptake vary based on personal characteristics?
This question is answered for the 7 app-specific features by means of the two class LCCM model, discussing the preference of the two segments. This is complemented by the ML model, indicating the influence of personal characteristics on attribute preferences and overall utility. Lastly, the question is also answered for the additional app characteristics, by discussing the interaction effects of personal characteristics estimated.

Firstly, with regards to overall MTA preferences, heterogeneity was present in the sample. Two consumer segments could be identified in the sample: the adopters and the non-adopters. The non-adopters segment on average choose the no app alternative over either of the app alternatives very often. Thereby indicating that there was a clear group of respondents that on average had a strong preference for downloading no app, no matter the features of the MTA. A respondent was more likely to belong to the adopters segment than to the non-adopters segment, when the respondent was a MTA user, between 19 and 40 years old, very positively towards MTAs and ascribed great importance to female health knowledge.

If the distinction between the adopter and non-adopter segment is not made, costs is more influential than accuracy in determining the influence of the attributes on uptake of a MTA. Consequently, if a consumer is likely to belong to the adopter segment, they might sooner choose the most accurate app with a higher price.

Adding personal characteristics to the model, showed that ethnicity, education, age, the attitude with regards to MTAs and if an individual is a (previous) MTA user, all influence the probability that a consumer decides to download a MTA. The older a respondent is, the lower the chance they would download a MTA as opposed to a younger respondent. Non-dutch respondents are more likely to download the same app than Dutch respondents. Lastly, low & middle educated respondents are more likely to download a similar MTA than highly educated respondents. Furthermore, only one of attribute preferences differed significantly based on personal characteristics. For non-Dutch respondents an improvement in personalisation abilities of a MTA, leads to a higher increase in the uptake chance than it does for Dutch respondents. Which means that non-Dutch respondents ascribe more value to the personalisation attribute of a MTA than Dutch respondents.

Lastly, the personal characteristics that influenced additional app characteristic preferences, were age, ethnicity and users. If a respondent is younger than 19, they ascribed less importance to an app recommendation from someone they know than if they were between 19-35 years old. Furthermore, Dutch respondents found an app's rating in the app store less important than non-Dutch respondents. Finally, respondents familiar with MTAs valued a scientifically based MTA as significantly less important than non-users.

8.2. Conclusion Main Research Question

With regards to the main research question: 'To what extent do various factors influence fertility & menstrual tracking app uptake by women from diverse backgrounds?', the study concludes the following:

For consumers more open to adoption of MTAs, accuracy is the most important followed by costs and privacy. Furthermore, if privacy and accuracy are of high quality, these consumers are still willing to download an app even if it is expensive. These consumers are likely to be between 19-35 years old, familiar with and positive towards MTAs, and convinced of female health knowledge importance. When no distinction is made between consumers based on their willingness to adopt, costs is more important than accuracy. Furthermore, age, ethnicity, education level, familiarity with and attitude towards MTAs influence a consumers choice to download a MTA. Additionally, the importance of app personalisation is dependent on the consumer's ethnicity. Lastly, the medical grounding characteristic of a MTA is perceived to be important by all respondents, but more so by respondents unfamiliar with MTAs.

8.3. Comparison with Existing Literature

The research by (Starling et al., 2018) harbors the most similarity with this research. As it differs from most MTA research papers by requiring participants to rank the importance of app attributes on a 4-point scale from not important to very important. This study does note the lack of diversity within their sample as a limitation. However, as in this research the sample group was not diverse either the comparison is still valid. Similar to this research they found that scientific grounding, recommendations by healthcare providers, an app that was tested in published research, costs and privacy were considered somewhat to very important. However, an app that was easy to navigate was also considered somewhat to very important, which is contrary to the finding in this research. Reasoning for this is extensively discussed later on in section 8.5. Personalisation, extra functions and communication were not considered in this research.

Papers conducting DCEs on mHealth find similar results, again apart from the ease of use importance. It must be noted that these papers do not have diverse sample groups with regards to ethnicity, income and educational level. However, as this research tried to achieve this but obtained a mostly homogeneous sample group, the comparison is still relevant. Simblett et al. (2023) found that a higher level of privacy and increased accuracy were important, with accuracy being the most important. Relative high importance of costs and privacy were also important in the research by Folkvord et al. (2022). Cost was also found important by Nittas, Mütsch, and Puhan (2020), however they found that privacy was not an important driver. The differences with this research with regards to privacy might be due to the nature of the data used by their app, which focused on sun protection monitoring. Which requires significantly different data than the sensitive data that MTAs require. Their research also found data input/ease of use to be of importance, contrary to this research. Which is underscored by most papers reviewed for this research, with some papers considering ease of use a necessity for app uptake no matter if the consumer group is regular or consisting of minority groups (Bretschneider, 2015; Epstein et al., 2017). The explanation to the differing findings is discussed in section 8.5

With regards to the found influence of age, ethnicity, education level, a positive attitude towards MTAs and experience with MTA use, some but not all findings are similar to literature. Firstly, the negative effect of age on app uptake, is confirmed in multiple papers (Epstein et al., 2017; Gambier-Ross et al., 2018; Haile et al., 2018). With these three papers respondents from different countries, ages, sexual orientations are considered. However, the effect of low SEP on app preferences are not investigated specifically. Gambier-Ross et al. (2018) also found that respondents that use other health apps are more likely to download MTAs, which is not equal to but similar to the finding in this research that experience with MTA leads to higher MTA uptake. Furthermore, multiple papers found the positive effect of positive technology/app/mHealth perceptions and attitudes on mHealth uptake among both general sample groups and sample groups specified to low SEP (Aamir et al., 2018; Hengst et al., 2023; Rai et al., 2013; Szinay et al., 2020)

Interestingly, this research found that being non-Dutch and educated on a low to middle level, influenced utility from apps positively. This is contrary to the literature indicating that app uptake of mHealth is low among marginalized and low SEP groups, who often are non-Dutch and possess a lower education level (Figueroa et al., 2023; Hengst et al., 2023; Hughson et al., 2018). As mentioned before in section 6.2.3 this probably due to the small size and composition of the sample size and sub-group definitions. Additionally, the mentioned papers focused on a sample group that consisted of low SEP, whereas this was attempted but not fully achieved in this research. Which could explain the difference. Lastly, this group might exercise a willingness to use thi

Additionally, the impact of ethnicity on the relatively higher importance of personalisation for non Dutch respondents aligns with the finding by (Hengst et al., 2023) who identified that personalisation is an important facilitator specifically for people with a low SEP, who according to the same research encompass ethnic minorities. Similarly Mora et al. (2020) found that Latina women valued an app personalised to their needs highly. It must be noted that in these researches the preference for personalisation is not compared between ethnic groups, which is the case in this research.

8.4. Recommendations and Implications

Before the recommendations are given, it should be noted that further research with a larger and more diverse sample group should be executed to confirm if the found results of this research are also present in the population.

8.4.1. Recommendations

In general, the found influences underscore the importance of a segmented approach in app development, tailored to meet the diverse needs and expectations of various user groups. Which could enhance uptake by diverse female consumers.

Non-adopters or Low SEP Consumer Recommendations

According to Hengst et al. (2023) low SEP or marginalized population groups often have less familiarity with mHealth and are less open to technology and mHealth in general. Therefore, they are more likely to belong to the non-adopters segment.

If a consumer group is targeted that does not necessarily consist of consumers likely to adopt an app, developers should focus on costs before all the other attributes are considered. If costs are ensured to be low, then the developer should focus on accuracy and privacy to increase uptake, lastly personalisation should be focused on to moderately increase uptake. When targeting non-Dutch respondents, paying more attention to the level of personalisation should be more prioritised than when targeting Dutch respondents.

Additionally, a negative attitude towards and low familiarity with MTAs did influence uptake regardless of app specific features, indicating that improving uptake among such a group requires consideration of features outside of the apps. Likewise, an increase in familiarity and attitude towards MTAs leads to significant uptake increase. Consequently, app developers with not just a commercial motive targeting marginalized groups, should focus not just on the development of the app but also on the societal context factors that refrain these groups from using the apps in the first place. These encompass a broad range from low digital literacy to policies in place. Depending on the app developer, some of these factors are outside of the possible means. However, collaboration with different stakeholders, such as the RIVM to establish MTAs that are approved by governmental guidelines on for example privacy and accuracy, offer opportunities. One important factor is the manner of introduction of MTAs. As this research showed that recommendations from healthcare professionals are valued as the most important, developers should collaborate with healthcare professionals to introduce MTAs to patients.

Additionally, further research should point out if other interaction effects with personal characteristics and app attributes can be identified.

Adopter Segment Recommendations

MTA developers should target 19-35 year old consumers with a positive attitude towards MTAs, the perception that female health knowledge is important and familiarity with MTAs. These consumers are likely to belong to the adopt prone segment. Targeting these consumers requires knowledge of consumers' perceptions. As this is not straightforward, these consumers should be targeted through latent indicators. One way to do this, is through targeted social media advertising, where advertisements can be specifically targeted towards consumers with a certain profile that indicates interests similar to MTAs. The uptake among this group is the highest (96%) if the developed MTA is a free app, has a high level of privacy, accuracy, provides reminders of easy data input, communicates simply and with icons, allows for personalisation of features and reminders and additional features. However, if these features are ensured, the developer could also charge these apps for €2,50 and even €5,00 and still the uptake would be significantly higher than a free basic app. A large uptake for a high price is also achieved by just ensuring the highest levels of data transparency and accuracy. Therefore, those attributes should be prioritised. If the app is very basic and of low quality, developers should provide the app for free to ensure moderate uptake levels.

8.4.2. Scientific Contribution

With regards to the scientific contribution, the most important finding may be that this research underscores how hard it is to reach consumers from low SEP. Simultaneously, this research also pointed out the importance of including different respondents, as even with low diversity some different preferences were found. Therefore, inclusion of diverse target groups in MTA research should be prioritised.

Furthermore, this research aimed to close the knowledge gap on the factors influencing MTA preferences among women from diverse backgrounds. With consideration of the sample size and composition issue this research contributed to knowledge on this issue. More specifically, this research added an overview of important factors in uptake for MTAs and mHealth in general for people with a low SEP. Furthermore, it quantified the influence of the most important app attributes on app preferences, while including and accounting for personal characteristic influences. Which is often disregarded in current research. The research established differences in app preferences present among different socio-demographic groups. Additionally, it executed a method to predict app uptake based on different app compositions. All of these findings were translated into recommendations for app developers, establishing scientifically based and consumer-specific focus points for app development.

8.5. Limitations

As with any research conducted, this research also contains its limitations. The most important limitations will be discussed in this section.

Firstly, as mentioned in 5.2 and 6.2.3 , the composition and size of the sample group complicated answering the main research question to a certain extent. In general if a sample group is not representative of the target group, its findings may be more difficult to apply to the real life target group. For this research specifically, the sample did not include a diverse representation of women, which complicated deriving conclusions for the target group of low SEP women. As low representation of subgroups of respondents may compromise finding possible interaction effects of these subgroups on utility or attribute preference. Therefore, interaction effects that were not found in this research, could still be present in the targeted population. As the researcher tried to reach out to diverse populations in the distribution process, an attempt to prevent this was made. Additionally, it is possible that the sample is over-represented by respondents with an interest in the subject matter and respondents similar to and within proximity of the researcher.

Secondly, related to the first limitation, the issue of another research method should be raised. This method was selected to investigate trade-offs and predict quantified uptake, to determine different preferences among consumer segments and thereby including various respondents within the entire range of possible app consumers. However, as this research underscored, ensuring participation of minority groups and people with low SEP, perhaps requires a method adapted to these subgroups. As surveys require a certain level of digital literacy and DCEs in particular have a large cognitive demand, this might have compromised participation of women with a low SEP.

Thirdly, the definition used in this research to measure the impact of the attribute *ease of use* might have compromised the validity of results with regards to this attribute. The found extent to which this attribute exercised influence on the choice for an app, was substantially lower than expected based on reviewed literature. When the respondents were asked to select their preferred value for the *ease of use* attribute, the *entering one data point* and *entering one data point + reminder notification* were selected almost an equal amount of times. Thereby indicating that there was no obvious preferred value for this attribute and the levels did not have considerate differences. As considerable differences between levels is a requirement to make sure that respondents do not disregard that attribute (see section 3.2.2 for elaborate explanation), it is likely that this influenced the low relative importance of the ease of use attributes. To prevent these kinds of errors a validation round was executed with participants answering a pilot survey. In the future this could be prevented by more extensive pilot testing and qualitative pre-research.

Fourth, executing DCEs might lead to hypothetical bias. Within this research the models were estimated based on the data retrieved from the choice-sets. Those choice-sets required respondents to make a choice between hypothetical apps and subsequently declare if they would download that app or would prefer to download no app at all. Seeing as the choices were based on a hypothetical situation, hypothetical bias could influence the results. This means that there were no real-life consequences for the respondents based on their choice. When selecting an expensive app as their preferred app, respondents did not actually have to pay €5,00. Respondents might *state* that they would prefer this apps, while in real-life they could possible *choose* differently. However, there is no way of knowing if and to what extent this influenced the results.

Lastly, included attributes did not cover all important app aspects. Since DCEs should remain simple, only a small number of attributes were included. Whereas the scope of this research was specified to app attributes, it is important to note that the reported influence of attributes was thus compared to only the included attributes. Therefore, these app attributes (e.g. the accuracy level) were not traded against app characteristics (e.g. who recommended the app) or societal or policy influences (e.g. privacy policies or regulations for mHealth). However, the selection of included app attributes encompassed the attributes that literature deemed as the most important.

8.6. Future Research

This section contains future research recommendations. These recommendations are build on the limitations mentioned in the previous section 8.5 and complemented by additional future research recommendations.

Firstly, with regards to the sample group, as this research emphasized both the difficulty of reaching marginalized groups for research and the importance of consideration of diverse respondents, diversity should be prioritised in further research on MTA uptake. As this research pointed out, it is not sufficient to employ a research method aimed at reaching all possible consumers and focusing the distribution on consumers of low SEP. Contrary, the research methods should be aimed at the groups that are hard to reach specifically. Possible research opportunities could still employ DCEs, aiming to keep the benefits of that method, but adjusting the circumstances of the experiment to the minority group. Gatlin and Johnson (2017) state that developing adapted surveys should be done by conducting cognitive interview studies. Therefore, within the qualitative phase of DCM, the target group should be included into multiple focus groups to ensure well designed experiments. Additionally, the experiments should then be conducted face-to-face. Furthermore, research investigating protocols to set-up DCEs to include minorities in general promises an interesting direction. As multiple DCE protocols for mHealth have been established, but the possibilities for DCE to propel research with regards to people with a low SEP remain insufficiently examined. The research by Gatlin and Johnson (2017) also states that the preferred research method is face-to-face. Consequently, research conducting qualitative research methods such as focus groups and interviews, could more easily acquire the target group instead of DCEs conducted through surveys. Establishing trust is mentioned as an important prerequisite for participation by minority groups (Gatlin & Johnson, 2017). Therefore, for both research directions mentioned above, it is essential that the target group is included throughout the entire research and that trust between the researchers and the target group is established by investing time and effort.

Secondly, to determine the true impact of ease of use on MTA uptake, similar research should be conducted with a more suited definition. To ensure considerate level differences, extensive pilot testing is required. Where the focus should be on respondents indicating if the levels are significantly different. The ease of use level of a MTA entails many varying aspects, i.e. navigation, data entry, setting changes, etc. Therefore, next to simply defining the attribute levels better, future research could perhaps test the influence of this attribute by means of a Hierarchical Information Integration (HII) approach. This approach suggest that possible consumers process choices with many attributes in a hierarchical manner. Where a separate experiment for each decision construct (e.g. the navigation or data entry) is conducted to estimate its contribution on the evaluation of the corresponding higher order decision construct (e.g. ease of use) (Bos et al., 2004). This allows for estimation of the separate preferences of for example data entry as well as the on the aggregated preferences for ease of use.

Thirdly, similar experiments could be conducted with differing factor scopes to determine the attributes. These could be focused on possible policy implementations to increase MTA uptake among certain population groups. An interesting approach could be to use context-dependent experiments. This entails regular choice-sets being presented to respondents under different contexts, which allows for examination of the influence the context has on choices (Arentze & Molin, 2013).

Likewise, experiments focused on more extensive background variable investigation could be insightful. It could be interesting to further examine the influence of perceptions and attitudes with regards to smartphones, apps, mHealth and technology. Additionally, it could be very interesting to include smarthpone or app use habits. Likewise, including other socio-demographics such as religion, sexuality, contraceptive use, regularity of menstruation or health status might also shed light onto the differing needs of diverse MTA users.

Lastly, an important future research opportunity is examining the reasoning behind the choices by conducting focus groups or in-depth interviews with respondents. As shown in the literature review (see 2.1.3) an abundance of research on consumers' use of, attitudes towards, and perception on MTAs is available. Therefore, some assumptions of the choices made by respondents in this research could be substantiated based on those findings. However, complementing and explaining the found quantitative data of DCEs by in-depth interviews or focus groups within one research scope, could lead to very meaningful and well-substantiated interpretation of DCEs.

References

Aamir, J., Ali, S., Boulos, M., & Anjum, N. (2018). Enablers and inhibitors: a review of the situation regarding mHealth adoption in low-and middle-income countries. *Health Policy and Technology*. https://www.sciencedirect.com/science/article/pii/S2211883717300874?casa_token=I36KTjsc5sAAAAAA:RgpvG5urEKG2uPB1r0VdmwMcAmopKMhnd_fV7b3qxkGS74rkj218l-zJUQy669enxsfOD-ieUw

Abelson, J. S., Kaufman, E., Symer, M., Peters, A., Charlson, M., & Yeo, H. (2017). Barriers and benefits to using mobile health technology after operation: A qualitative study. *Surgery*, *162*(3), 605–611. https://doi.org/10.1016/j.surg.2017.05.007

Alfawzan, N., Christen, M., Spitale, G., & Biller-Andorno, N. (2022). Privacy, Data Sharing, and Data Security Policies of Women's mHealth Apps: Scoping Review and Content Analysis. *JMIR mHealth and uHealth*, *10*(5). https://doi.org/10.2196/33735

Almeida, T., Shipp, L., Mehrnezhad, M., & Toreini, E. (2022). Bodies Like Yours: Enquiring Data Privacy in FemTech. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3547522.3547674

Al-Rshoud, F., Qudsi, A., Naffa, F. W., Al Omari, B., & AlFalah, A. G. (2021). The Use and Efficacy of Mobile Fertility-tracking Applications as a Method of Contraception: a Survey. *Current Obstetrics and Gynecology Reports*, *10*(2), 25–29. https://doi.org/10.1007/S13669-021-00305-4

Angum, F., Khan, T., Kaler, J., Siddiqui, L., & Hussain, A. (2020). The Prevalence of Autoimmune Disorders in Women: A Narrative Review. *Cureus*, *12*(5). https://doi.org/10.7759/CUREUS.8094

Arentze, T. A., & Molin, E. J. (2013). Travelers' preferences in multimodal networks: Design and results of a comprehensive series of choice experiments. *Transportation Research Part A: Policy and Practice*, *58*, 15–28. https://doi.org/10.1016/j.tra.2013.10.005

Becker, C. M., Gattrell, W. T., Gude, K., & Singh, S. S. (2017). Reevaluating response and failure of medical treatment of endometriosis: a systematic review. *Fertility and Sterility*, *108*(1), 125–136. https://doi.org/10.1016/j.fertnstert.2017.05.004

Ben-Akiva, M., McFadden, D., & Train, K. (2019). Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-based Conjoint Analysis. *Foundations and Trends® in Econometrics*, *10*(1-2), 1–144. https://doi.org/10.1561/0800000036

Bernasco, W., & Block, R. (2013). Discrete Choice Modelling. Ned Levine & Associates /National Institute of Justice.

Bertrand, A. M. E., & Hafner, C. M. (2014). On heterogeneous latent class models with applications to the analysis of rating scores. *Computational Statistics*, *29*(1-2), 307–330. https://doi.org/10.1007/s00180-013-0450-5

Bidmon, S., Terlutter, R., & Röttl, J. (2014). What Explains Usage of Mobile Physician-Rating Apps? Results From a Web-Based Questionnaire. *Journal of Medical Internet Research*, *16*(6), e148. https://doi.org/10.2196/jmir.3122

Blair, D. L., Morgan, H. M., & McLernon, D. J. (2021). Women's perspectives on smartphone apps for fertility tracking and predicting conception: a mixed methods study. *The European Journal of Contraception & Reproductive Health Care*, *26*(2), 119–127. https://doi.org/10.1080/13625187.2021.1874336

Bos, I. D. M., Van der Heijden, R. E. C. M., Molin, E. J. E., & Timmermans, H. J. P. (2004). The Choice of Park and Ride Facilities: An Analysis Using a Context-Dependent Hierarchical Choice Experiment. *Environment and Planning A: Economy and Space*, *36*(9), 1673–1686. https://doi.org/10.1068/a36138

Bretschneider, R. A. (2015). A goal- and context-driven approach in mobile period tracking applications. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9177*, 279–287. https://doi.org/10.1007/978-3-319-20684-4{_}27/FIGURES/3

Broad, A., Biswakarma, R., & Harper, J. C. (2022). A survey of women's experiences of using period tracker applications: Attitudes, ovulation prediction and how the accuracy of the app in predicting period start dates affects their feelings and behaviours. *Journal of Women's Health*, *18*, 174550572210952. https://doi.org/10.1177/17455057221095246

Centraal Bureau voor de Statistiek(2016). Bevolking naar migratieachtergrond. https://www.cbs.nl/nl-nl/achtergrond/2016/47/bevolking-naar-migratieachtergrond.

Centraal Bureau voor de Statistiek. (2022). *Standaard Onderwijsindeling 2021* (tech. rep.). Centraal Bureau voor de Statistiek. Den Haag/Heerlen. https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/onderwijs-en-beroepen/standaard-onderwijsindeling--soi--/standaard-onderwijsindeling-2021#:~:text=De%20publicatieindeling%20naar%20niveau%20van,het%203e%20aggregatieniveau%208%20categorie%C3%ABn.

Centraal Planbureau(2023). Augustusraming 2023 (cMEV 2024). https://www.cpb.nl/augustusraming-2023.

Chorus, C.(2022). Choice behaviour modeling and the Logit-model. What and How? https://brightspace.tudelft.nl/d2l/le/content/401506/viewContent/2239904/View.

Chorus, C.(2022). Mixed Logit model. https://brightspace.tudelft.nl/d2l/le/content/401506/viewContent/2239905/View.

Clue(2024). About Clue. https://helloclue.com/about-clue.

Corbin, B. A. (2020). Digital Micro-Agressions and Discrimination: FemTech and the 'Othering' of Women. *Nova Law Review*, *44*. http://www.vox.com/the-goods/2018/11/13/18079458/menstrual-tracking-surveillance-

Dudouet, L. (2022). Digitised fertility: The use of fertility awareness apps as a form of contraception in the United Kingdom. *Social Sciences & Humanities Open*, *5*(1), 100261. https://doi.org/10.1016/j.ssaho.2022.100261

Earle, S., Marston, H. R., Hadley, R., & Banks, D. (2021). Use of menstruation and fertility app trackers: a scoping review of the evidence. *BMJ Sexual & Reproductive Health*, *47*(2), 90–101. https://doi.org/10.1136/BMJSRH-2019-200488

Eijkholt, S.(2021). Groei femtech door aandacht voor onderschatte zorg - Smarthealth. SmarthHealth. https://smarthealth.live/2021/06/04/groei-femtech-door-aandacht-voor-onderschatte-zorg/.

El Zarwi, F., Vij, A., & Walker, J. L. (2017). A discrete choice framework for modeling and forecasting the adoption and diffusion of new transportation services. *Transportation Research Part C: Emerging Technologies*, *79*, 207–223. https://doi.org/10.1016/j.trc.2017.03.004

Epstein, D. A., Lee, N. B., Kang, J. H., Agapie, E., Schroeder, J., Pina, L. R., Fogarty, J., Kientz, J. A., & Munson, S. A. (2017). Examining Menstrual Tracking to Inform the Design of Personal Informatics Tools. *Proceedings of the SIGCHI conference on human factors in computing systems. CHI Conference*, *2017*, 6876–6888. https://doi.org/10.1145/3025453.3025635

Erickson, J., Yuzon, J. Y., & Bonaci, T. (2022). What You Do Not Expect When You Are Expecting: Privacy Analysis of Femtech. *IEEE TRANSACTIONS ON TECHNOLOGY AND SOCIETY*, *3*(2), 121. https://doi.org/10.1109/TTS.2022.3160928

Eschler, J., Menking, A., Fox, S., & Backonja, U. (2019). Defining Menstrual Literacy with the Aim of Evaluating Mobile Menstrual Tracking Applications. *CIN - Computers Informatics Nursing*, *37*(12), 638–646. https://doi.org/10.1097/CIN.0000000000000559

Essén, A., Stern, A. D., Haase, C. B., Car, J., Greaves, F., Paparova, D., Vandeput, S., Wehrens, R., & Bates, D. W. (2022). Health app policy: international comparison of nine countries' approaches. *npj Digital Medicine*, *5*(1), 31. https://doi.org/10.1038/s41746-022-00573-1

Faubion, S. S. (2021). Femtech and midlife women's health: good, bad, or ugly? *Menopause (New York, N.Y.)*, *28*(4), 347–348. https://doi.org/10.1097/GME.0000000000001742

Figueroa, C. A., Luo, T., Aguilera, A., & Lyles, C. R. (2021). The need for feminist intersectionality in digital health. *The Lancet Digital Health*, *3*(8), e526–e533. https://doi.org/10.1016/S2589-7500(21)00118-7

Figueroa, C. A., Sundqvist, J., Mathieu, S., Farrokhnia, N., Nevin, D., & Andersson, S. W. (2023). The opportunities and challenges of women's digital health: A research agenda. *Health Policy and Technology*, *12*(4), 100814. https://doi.org/10.1016/j.hlpt.2023.100814

Folkvord, F., Bol, N., Stazi, G., & Lupiáñez-Villanueva, F. (2022). Preferences in the willingness to download a mobile health app: A discrete choice experiment in Spain, Germany and the Netherlands. https://doi.org/10.21203/RS.3.RS-1859416/V1

Freis, A., Freundl-Schütt, T., Wallwiener, L. M., Baur, S., Strowitzki, T., Freundl, G., & Frank-Herrmann, P. (2018). Plausibility of Menstrual Cycle Apps Claiming to Support Conception. *Frontiers in Public Health*, *6*, 347358. https://doi.org/10.3389/FPUBH.2018.00098

Gambier-Ross, K., McLernon, D. J., & Morgan, H. M. (2018). A mixed methods exploratory study of women's relationships with and uses of fertility tracking apps. *DIGITAL HEALTH*, *4*, 205520761878507. https://doi.org/10.1177/2055207618785077

Gambon, E., Stotz, C., & Sandhu, N.(2020). Femtech is expansive - it's time to start treating it as such. Rockhealth. https://rockhealth.com/femtech-is-expansive-its-time-to-start-treating-it-as-such/.

Gatlin, T. K., & Johnson, M. J. (2017). Two Case Examples of Reaching the Hard-to-Reach: Low Income Minority and LGBT Individuals. *Journal of Health Disparities Research and Practice*, *10*(3). https://digitalscholarship.unlv.edu/jhdrp/vol10/iss3/11/

Gonçalves, A. S. S., Prado, D. S., & Silva, L. M. (2021). Frequency and experience in the use of menstrual cycle monitoring applications by Brazilian women. *European Journal of Contraception and Reproductive Health Care*, *26*(4), 291–295. https://doi.org/10.1080/13625187.2021.1884222

Goodman, A., Mahoney, R., Spurling, G., & Lawler, S. (2023). Influencing Factors to mHealth Uptake With Indigenous Populations: Qualitative Systematic Review. *JMIR mHealth and uHealth*, *11*, e45162. https://doi.org/10.2196/45162

Gross, M. S., Hood, A., & Corbin, B. (2021). Pay No Attention to That Man behind the Curtain: An Ethical Analysis of the Monetization of Menstruation App Data. *https://doi.org/10.3138/ijfab-2021-03-22*, *14*(2), 144–156. https://doi.org/10.3138/IJFAB-2021-03-22

Gutterman, A. S. (2023). FemTech and Older Women. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4372007

Haile, L. T., Fultz, H. M., Simmons, R. G., & Shelus, V. (2018). Market-testing a smartphone application for family planning: assessing potential of the CycleBeads app in seven countries through digital monitoring. *mHealth*, *4*, 27–27. https://doi.org/10.21037/MHEALTH.2018.06.07

Hankivsky, O., Reid, C., Cormier, R., Varcoe, C., Clark, N., Benoit, C., & Brotman, S. (2010). Exploring the promises of intersectionality for advancing women's health research. *International Journal for Equity in Health*, *9*(1), 5. https://doi.org/10.1186/1475-9276-9-5

Hendl, T., & Jansky, B. (2022). Tales of self-empowerment through digital health technologies: a closer look at 'Femtech'. *Review of Social Economy*, *80*(1), 29–57. https://doi.org/10.1080/00346764.2021.2018027

Hengst, T. M., Lechner, L., Dohmen, D., & Bolman, C. A. (2023). The facilitators and barriers of mHealth adoption and use among people with a low socio-economic position: A scoping review. *Digital Health*, *9*. https://doi.org/10.1177/20552076231198702

Hess, S., & Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, *32*, 100170. https://doi.org/10.1016/j.jocm.2019.100170

Hughson, J.-a. P., Daly, J. O., Woodward-Kron, R., Hajek, J., & Story, D. (2018). The Rise of Pregnancy Apps and the Implications for Culturally and Linguistically Diverse Women: Narrative Review. *JMIR mHealth and uHealth*, *6*(11), e189. https://doi.org/10.2196/mhealth.9119

Jacobs, N., & Evers, J. (2023). Ethical perspectives on femtech: Moving from concerns to capability-sensitive designs. *Bioethics*. https://doi.org/10.1111/bioe.13148

Johnson, S., Marriott, L., & Zinaman, M. (2018). Can apps and calendar methods predict ovulation with accuracy? *Current Medical Research and Opinion*, *34*(9), 1587–1594. https://doi.org/10.1080/03007995.2018.1475348

Karasneh, R. A., Al-Azzam, S. I., Alzoubi, K. H., Muflih, S. M., & Hawamdeh, S. S. (2020). Smartphone Applications for Period Tracking: Rating and Behavioral Change among Women Users. *Obstetrics & Gynecology*, *2020*. https://doi.org/10.1155/2020/2192387

Kløjgaard, M. E., Bech, M., & Søgaard, R. (2012). Designing a Stated Choice Experiment: The Value of a Qualitative Process. *Journal of Choice Modelling*, *5*(2), 1–18. https://doi.org/10.1016/S1755-5345(13)70050-2

Ko, S., Lee, J., An, D., & Woo, H. (2023). Menstrual Tracking Mobile App Review by Consumers and Health Care Providers: Quality Evaluations Study. *JMIR mHealth and uHealth*, *11*. https://doi.org/10.2196/40921

Krebs, P., & Duncan, D. T. (2015). Health App Use Among US Mobile Phone Owners: A National Survey. *JMIR mHealth and uHealth*, *3*(4), e101. https://doi.org/10.2196/mhealth.4924

Lancsar, E., & Louviere, J. (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. *PharmacoEconomics*, *26*(8), 661–677. https://doi.org/10.2165/00019053-200826080-00004

Levy, J., & Romo-Avilés, N. (2019). "a good little tool to get to know yourself a bit better": A qualitative study on users' experiences of app-supported menstrual tracking in Europe. *BMC Public Health*, *19*(1). https://doi.org/10.1186/S12889-019-7549-8

Lupton, D. (2016). The use and value of digital media for information about pregnancy and early motherhood: A focus group study. *BMC Pregnancy and Childbirth*, *16*(1), 1–10. https://doi.org/10.1186/S12884-016-0971-3/PEER-REVIEW

Maas, A. H., & Appelman, Y. E. (2010). Gender differences in coronary heart disease. *Netherlands heart journal : monthly journal of the Netherlands Society of Cardiology and the Netherlands Heart Foundation*, *18*(12), 598–603. https://doi.org/10.1007/S12471-010-0841-Y

Mandeville, K. L., Lagarde, M., & Hanson, K. (2014). The use of discrete choice experiments to inform health workforce policy: a systematic review. *BMC health services research*, *14*, 367. https://doi.org/10.1186/1472-6963-14-367

Moglia, M. L., Nguyen, H. V., Chyjek, K., Chen, K. T., & Castaño, P. M. (2016). Evaluation of smartphone menstrual cycle tracking applications using an adapted applications scoring system. *Obstetrics & Gynecology*, *127*(6), 1153–1160. https://doi.org/10.1097/AOG.0000000000001444

Mohamed, A. H. H. M., Tawfik, H., Al-Jumeily, D., & Norton, L. (2011). MoHTAM: A Technology Acceptance Model for Mobile Health Applications. *2011 Developments in E-systems Engineering*, 13–18. https://doi.org/10.1109/DeSE.2011.79

Molin, E.(2023). Constructing choice sets: Orthogonal designs.

Molin, E.(2023). Introduction to experimental designs. https://brightspace.tudelft.nl/d2l/le/content/597386/viewContent/3392341/View.

Mora, A. C., Krishnamurti, T., Davis, A., & Simhan, H. (2020). A culturally appropriate mobile health application for pregnancy risk communication to latino women. *American Journal of Obstetrics and Gynecology*, *222*(1), S575–S576. https://doi.org/10.1016/j.ajog.2019.11.941

Nair, A., Padmakumar, K., Marketing, S. K. .-. I. J. o., & 2023, u. (2023). Menstrual Tracking Apps in India: User Perceptions, Attitudes, and Implications. *researchgate.net*. https://doi.org/10.17010/ijom/2023/v53/i2/172630

Nijland, N., van Gemert-Pijnen, J. E., Kelders, S. M., Brandenburg, B. J., & Seydel, E. R. (2011). Factors Influencing the Use of a Web-Based Application for Supporting the Self-Care of Patients with Type 2 Diabetes: A Longitudinal Study. *Journal of Medical Internet Research*, *13*(3), e71. https://doi.org/10.2196/jmir.1603

Nittas, V., Mütsch, M., Braun, J., & Puhan, M. A. (2020). Self-Monitoring App Preferences for Sun Protection: Discrete Choice Experiment Survey Analysis. *Journal of Medical Internet Research*, *22*(11), e18889. https://doi.org/10.2196/18889

Nittas, V., Mütsch, M., & Puhan, M. A. (2020). Preferences for Sun Protection With a Self-Monitoring App: Protocol of a Discrete Choice Experiment Study. *JMIR Research Protocols*, *9*(2), e16087. https://doi.org/10.2196/16087

Or, C. K. L., & Karsh, B.-T. (2009). A Systematic Review of Patient Acceptance of Consumer Health Information Technology. *Journal of the American Medical Informatics Association*, *16*(4), 550–560. https://doi.org/10.1197/jamia.M2888

Ovia Health(2021). Ovia Health. https://www.oviahealth.com/.

Patel, U., Broad, A., Biswakarma, R., & Harper, J. C. (2023). Experiences of users of period tracking apps: which app, frequency of use, data input and output and attitudes. *Reproductive BioMedicine Online*, 103599. https://doi.org/10.1016/J.RBMO.2023.103599

Peterlin, B. L., Gupta, S., Ward, T. N., & MacGregor, A. (2011). Sex Matters: Evaluating Sex and Gender in Migraine and Headache Research. *Headache: The Journal of Head and Face Pain*, *51*(6), 839–842. https://doi.org/10.1111/J.1526-4610.2011.01900.X

Pratap, A., Neto, E. C., Snyder, P., Stepnowsky, C., Elhadad, N., Grant, D., Mohebbi, M. H., Mooney, S., Suver, C., Wilbanks, J., Mangravite, L., Heagerty, P. J., Areán, P., & Omberg, L. (2020). Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. *npj Digital Medicine 2020 3:1*, *3*(1), 1–10. https://doi.org/10.1038/s41746-020-0224-8

Rai, A., Chen, L., Pye, J., & Baird, A. (2013). Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. *Journal of Medical Internet Research*, *15*(8). https://doi.org/10.2196/JMIR.2635

Rajak, M., & Shaw, K. (2021). An extension of technology acceptance model for mHealth user adoption. *Technology in Society*, *67*. https://doi.org/10.1016/J.TECHSOC.2021.101800

Ramos, G., Ponting, C., Labao, J. P., & Sobowale, K. (2021). Considerations of diversity, equity, and inclusion in mental health apps: A scoping review of evaluation frameworks. *Behaviour Research and Therapy*, *147*, 103990. https://doi.org/10.1016/J.BRAT.2021.103990

Rijksinstituut voor Volksgezondheid en Milieu(2018). Gezondheidsverschillen | Volksgezondheid Toekomst Verkenning.

Ryan, M. (2000). Using conjoint analysis to elicit preferences for health care. *BMJ*, *320*(7248), 1530–1533. https://doi.org/10.1136/bmj.320.7248.1530

Sarkar, U., Gourley, G. I., Lyles, C. R., Tieu, L., Clarity, C., Newmark, L., Singh, K., & Bates, D. W. (2016). Usability of Commercially Available Mobile Applications for Diverse Patients. *Journal of General Internal Medicine*, *31*(12), 1417–1426. https://doi.org/10.1007/s11606-016-3771-6

Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*, *99*(9), 1013–23.

Simblett, S., Pennington, M., Quaife, M., Siddi, S., Lombardini, F., Haro, J., Peñarrubia-Maria, M., Bruce, S., Nica, R., Zorbas, S., Polhemus, A., Novak, J., Dawe-Lane, E., Morris, D., Mutepua, M., Odoi, C., Wilson, E., Matcham, F., White, K., . . . Wykes, T. (2023). Patient preferences for key drivers and facilitators of adoption of mHealth technology to manage depression: A discrete choice experiment. *Journal of Affective Disorders*, *331*, 334–341. https://doi.org/10.1016/j.jad.2023.03.030

Smith, B., & Magnani, J. W. (2019). New technologies, new disparities: The intersection of electronic health and digital health literacy. *International Journal of Cardiology*, *292*, 280–282. https://doi.org/10.1016/j.ijcard.2019.05.066

Starling, M. S., Kandel, Z., Haile, L., & Simmons, R. G. (2018). User profile and preferences in fertility apps for preventing pregnancy: an exploratory pilot study. *mHealth*, *4*, 21–21. https://doi.org/10.21037/MHEALTH.2018.06.02

Stowell, E., Lyson, M. C., Saksono, H., Wurth, R. C., Jimison, H., Pavel, M., & Parker, A. G. (2018). Designing and Evaluating mHealth Interventions for Vulnerable Populations. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3173574.3173589

Szinay, D., Cameron, R., Naughton, F., Whitty, J. A., Brown, J., & Jones, A. (2021). Understanding Uptake of Digital Health Products: Methodology Tutorial for a Discrete Choice Experiment Using the Bayesian Efficient Design. *Journal of Medical Internet Research*, *23*(10), e32365. https://doi.org/10.2196/32365

Szinay, D., Jones, A., Chadborn, T., Brown, J., & Naughton, F. (2020). Influences on the uptake of and engagement with health and well-being smartphone apps: Systematic review. *Journal of Medical Internet Research*, *22*(5). https://doi.org/10.2196/17572

Taylor, A. M. (2021). Fertile Ground: Rethinking Regulatory Standards for Femtech. https://www.naturalcycles.com/other/legal/

Tiffany, K.(2018). Period-tracking apps are not for women. Vox. https://www.vox.com/the-goods/2018/11/13/18079458/menstrual-tracking-surveillance-glow-clue-apple-health.

Train, K. (2009). Behavioral models. In *Discrete choice methods with simulation* (2nd ed., pp. 11–33). Cambridge University Press.

Trapero-Bertran, M., Rodríguez-Martín, B., & López-Bastida, J. (2019). What attributes should be included in a discrete choice experiment related to health technologies? A systematic literature review. *PLOS ONE*, *14*(7), e0219905. https://doi.org/10.1371/journal.pone.0219905

van Rede, Y. (2021). *Gendergelijkheidsplan* (tech. rep.). Rijksinstituut voor Volksgezondheid en Milieu (RIVM).

Vicente, M. R. (2022). ICT for healthy and active aging: The elderly as first and last movers. *Telecommunications Policy*, *46*(3), 102262. https://doi.org/10.1016/j.telpol.2021.102262

Volksgezondheid en Zorg(). Sociaaleconomische status | verantwoording | Definities. https://www.vzinfo.nl/sociaaleconomische-status/verantwoording/definities#.

Worsfold, L., Marriott, L., Johnson, S., & Harper, J. C. (2021). Period tracker applications: What menstrual cycle information are they giving women? *Journal of Women's Health*, *17*. https://doi.org/10.1177/17455065211049905

Xie, Z., Liu, H., & Or, C. (2023). A discrete choice experiment to examine the factors influencing consumers' willingness to purchase health apps. *mHealth*, *9*. https://doi.org/10.21037/MHEALTH-22-39/COIF

Xie, Z., & Or, C. K. (2023). Consumers' Preferences for Purchasing mHealth Apps: Discrete Choice Experiment. *JMIR mHealth and uHealth*, *11*, e25908–e25908. https://doi.org/10.2196/25908

Zhaunova, L., Bamford, R., Radovic, T., Wickham, A., Peven, K., Croft, J., Klepchukova, A., & Ponzo, S. (2023). Characterization of Self-reported Improvements in Knowledge and Health Among Users of Flo Period Tracking App: Cross-sectional Survey. *JMIR mHealth and uHealth*, *11*, e40427. https://doi.org/10.2196/40427

Zwingerman, R., Chaikof, M., & Jones, C. (2019). A Critical Appraisal of Fertility and Menstrual Tracking Apps for the iPhone. *Journal of Women's Health*. https://doi.org/10.1016/j.jogc.2019.09.023

A

Appendix: MTA Functionalities

The following table depicts FTA functionalities examined by Zwingerman et al. (2019) They reviewed and appraised 140 menstrual cycle tracking apps available in the Apple App Store. Their appraisal is not important for this research, but their research also provides a feature overview. However, it must be noted that this study was published in 2019. In the mean time multiple apps have been launched, that might contain different features. Therefore, the paper by (Ko et al., 2023) was examined to indicate newly launched features, since it was published in 2023. They selected 34 apps from the Korean and English Google Play Store and iOS App Store. They excluded apps with less than 10,000 reviews, paid apps and apps whose last update was more than 180 days ago. No differing features or significant changes with regards to app functionalities were found, so therefore the overview by Zwingerman et al. (2019) is used to examine app functionalities. Table 20 contains the features identified by Zwingerman et al. (2019).

Table 20: Identified FTA Functionalities by Zwingerman et al. (2019)

	No.(%)	
Feature	Free (n=90)	Paid (n=50)
Privacy policy	40(44,4)	23(46)
No advertisements	70(77,8)	49(98)
Clinical disclaimer	33(36,7)	28(56)
References or attributions for content	9(10)	6(12)
Can use without registering	70(77,8)	43(86)
Password protection option	51(56,7)	25(50)
Data export function	28(31,1)	22(44)
Backup function	31(34,4)	18(36)
Social media or message board component	17(18,9)	7(14)
Syncs with external devices	22(24,4)	7(14,0)
Generates export	40(44,9)	30(60)
Generates reminders	53(58,9)	27(54)
Provides general information about menstrual cycles or fertility	32(35,6)	19(38)
Menstrual cycle tracker	76(84,4)	39(78)
Basal body temperature tracker	50(55,6)	26(52)
Intercourse tracker	52(57,8)	32(64)
Cervical mucous tracker	30(33,3)	23(46)
Ovulation stick tracker	24(26,7)	17(34)
Medical appointment tracker	7(7,8)	5(10)
General medication tracker	35(38,9)	16(32)
Ovulation prediction calculator	54(60)	33(66)
Fertile window calculator	77(85,6)	36(72)
Expected next menstrual cycle start calculator	70(77,8)	35(70)
Estimated due date calculator	14(15,6)	11(22)
Provides information about infertility or infertility treatment	8(8,9)	7(14)
Infertility treatment specific menstrual cycle calendar	3(3,3)	2(4)
Infertility treatment specific appointment tracker	5(5,6)	3(6)
Fertility medication tracker	6(6,7)	5(10)
Real-time interaction with other fertility parents	6(6,7)	4(8)

B

Identified Facilitators and Barriers for mHealth Uptake

In this chapter a table containing the sourced reviewed to obtain identified facilitators and barriers can be seen in table 21. The eventually identified facilitators and barriers and their derived themes, can be found in 22. This table uses the numbers assigned to the sources in table 21, to indicate in which paper a factor was found. Additionally, this chapter also contains the explanation of the factors identified to influence mHealth uptake by people with a low SEP as discovered by (Hengst et al., 2023).

B.1. Identified Facilitators and Barriers for MTA Uptake

Table 21: Overview of Reviewed Sources for Influencing Factors

Number	Source
MTA/Female Health apps	
1	Bretschneider (2015)
2	Levy & Romo-Avilés (2019)
3	Gambier-Ross et al. (2018)
4	Zhaunova et al. (2023)
5	Eschier et al. (2019)
6	Dudouet (2022)
7	Epstein et al. (2017)
8	Patel et al. (2023)
9	Haile et al. (2018)
10	Ko et al. (2023)
11	Karasneh et al. (2020)
12	Lupton (2016)
13	Earle et al. (2021)
14	Starling et al. (2018)
15	Blair et al. (2021)
General mHealth	
1	Aamir et al. (2018)
2	Alaiad et al. (2019)
3	Hengst et al. (2023)
4	Alam et al. (2020)
5	Anaya et al. (2021)
6	Hughson et al. (2018)
7	Mora et al. (2020)

Table 22: Assessment of Factors Impacting App Uptake Across Various Studies

	Source		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Factors	**Theme**	**Occurrence**															
Barriers																	
Too complex/many features	Ease of use	3	x						x		x				x		
Too confusing	Ease of use	2	x							x	x						
No trust in calculated predictions/inaccurate	Accuracy	4	x		x				x	x							
List of possible moods/symptoms too confusing	Ease of use	4	x				x		x						x		
Missing pregnancy mode/pregnancy ruins algorithm	Accuracy	3	x				x		x								
No reminder for the pill	Function	1	x														
Too few gradations to enter menstruation/no normative cycle	Personalisation	2	x								x						
GUI lacks clarity	Ease of use	1	x														
Process to enter data too long	Ease of use	2	x							x							
Ads in the way of data entry	Ease of use	2	x							x							
Unnecessary features are annoying	Ease of use	2	x				x										
How is the data used	Privacy & Security	2	x				x										
Privacy issues	Privacy & Security	2		x								x					
Burden of tracking	Ease of use	2		x						x							
Changes in tracking behavior in different phases	Personalisation	3		x				x	x								

Continued on next page

Table 22 continued from previous page

	Source		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Factors	**Theme**	**Occurrence**															
Reminders are annoying	Notifications & Reminders, Personalisation	1			x												
Pinkification	Design	1			x												
Commercialization	Design, Costs	2			x					x							
Too stressful/too much pressure	Ease of use	1			x												
Stereotypical feminine assumptions, family planning + sexual activity	Personalisation	1							x								
(Hidden) cost	Costs	2								x						x	
Missing medical grounding	Accuracy, Evidence Based	1											x				
Facilitators																	
Tracking symptoms (e.g., headache, acne) (see if correlates to cycle)	Function	5	x		x	x			x				x				
Tracking mood	Function	2	x						x								
Notification and reminders	Notifications & Reminders	3			x								x	x			
Informative/education	Information Communication	4			x	x	x						x	x			x
History of cycles	Information Communication	1			x												
Easy to store and share data	Information Communication, function	2			x				x								
Prediction of fertile dates + ovulation prediction (plan sex)	Function	3			x	x							x				

Continued on next page

Table 22 continued from previous page

	Source		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Factors	**Theme**	**Occurrence**															
Community connection	Function	2			x									x			
Medically grounded	Accuracy, Evidence Based	6			x		x					x	x	x		x	
Accuracy	Accuracy	5			x					x				x	x	x	
Discrete design of app, not obviously menstruation app	Design, Discretion	2			x				x								
Adding notes	Informative, Function	2			x				x								
Additional tracking options other than fertility	Informative, Function	2			x				x								
Prediction of period dates	Function	1											x				
Pause function on tracking cycle	Function	2					x		x								
Customize + update symptom list	Personalisation	1					x					x					
Freedom to enable or disable elements	Personalisation	1					x										
Trend reporting related to symptoms	Information Communication	1					x										
Provide evidence-based action items for menstrual health	Function, Evidence based	1					x										
Interoperability with other platforms	Function	2							x					x			
Easy sharing with healthcare providers	Function, Ease of use	1							x								
Ease of use	Ease of use	3								x			x				x

Continued on next page

Table 22 continued from previous page

	Source		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Factors	**Theme**	**Occurrence**															
Good visuals for better navigation	Design	2								x			x				
Personalisation	Personalisation	3										x	x	x	x		
Ensured security	Privacy & Security	2											x			x	
No internet connection needed	Ease of use	1											x				
Tested app	Accuracy, Evidence Based	2											x			x	
User ratings and reviews	Reputation	1														x	
Healthcare provider recommended	Reputation	1														x	
Info without having to talk to someone	Discretion	1															x

B.2. Identified Facilitators and Barriers for mHealth Uptake by Low SEP Consumers

Table 23: App Specific Factors Influencing Adoption by People with a Low SEP (Hengst et al., 2023)

Facilitator or Barrier	Explanation
Tailoring	The extent to which an app and its' information are adapted to the needs, preferences and characteristics of the target group
Usability	Simple, easy to use, readily available and reliable
Visuals	Using audio, images and video's is a facilitator, because they're more appealing and easier to recall
Evidence-based	Incorporate evidence-based guidelines into the design in all development stages
Language	Important due to language and low literacy being important barriers
Personalisation	The extent to which individuals can set up the application to her/his own liking
Patient-centered design	Involvement of the target group in development
Privacy and secure data sharing	Clear and easily interpretable data security statements
Frequency of reminders	Frequent but adequate delivery and ability to personalise the timing of reminders

C

Focus Group

C.1. Focus Group Set Up

The focus group was held online due to the higher chance of getting participants from different backgrounds together. The focus group was recorded and transcribed. Initially, the focus group consisted of 7 participants, that were within the fertile age range (12-45) and reported being fertile. One participant was not fertile at the moment of the focus group due to being pregnant. However, she reported having used MTA recently to support getting pregnant, so her opinion was very valuable and therefore she was invited to the focus group. Due to personal reasons, one participant was not available last minute. Therefore, composition of the focus group was the following:

Table 24: Overview of Focus Group Participants' Socio-demographics

Respondent	Age	Highest level of education	Employment status	Current Employment	Parents' Nationality	Income Range
1	25	Bachelor's Degree	Consultant	N/A	Dutch	€40.000
2	22	Higher Professional Education (HBO)	Student	N/A	Dutch + Belgian	Student
3	20	High School (MAVO)	Student	N/A	Dutch + Gambian	Student
4	39	Master's Degree	Employed	Radiologist	Dutch	> €100.000
5	30	MBO Degree	Employed	Massage Therapist	Iranian	€30.000
6	16	High School (VWO)	Scholar	N/A	Dutch	Scholar

As can be seen, the focus group representation consists mostly of young participants. This is due to the cancellation of one of the participants, which was a 40 year old woman with a Higher Professional Education (HBO) level and a income between €30,000-40,000. Consequently, 2 out of 6 participants (33,33%) were between the age range of 30-40, 1 out of 6 participants (16,67%) was younger than 20 years, and the other 3 participants (50%) were between 20-25 years old. This was kept in mind while processing the focus group data.

The focus group consisted of the following schedule:

1. Thesis and focus group objective introduction
2. Open discussion about perception on MTAs
3. Selecting relevant attributes

4. Discussion of selected attributes and levels
5. Closing

The focus group started with an introduction round and assuring all the participants had access to and understood the Miro board. During the introduction into the topic and focus group objective, it was made clear that the participants' opinions would be processed anonymously and that all opinions were equally valuable and accepted. Then, the first part of the focus group consisted of an open discussion among the participants to obtain an overview of their perception towards MTA usage. The results are discussed in C.2. The researcher refrained from participation as much as possible and acted as a discussion facilitator. However, it appeared some terms or features with regards to the apps were unclear. So, when needed the researcher provided explanation. The following questions were used as discussion starters:

- Do/did you use a MTA?
- For what purpose did you use a MTA?
- Why have you not used a MTA?
- Would you be willing to use a MTA in the future?
- Why would/wouldn't you?
- If you would download a MTA, what features would you pay attention to when selecting the right one?
- If you could design your own MTA, what would it look like and what features would it harbor?

For the second and third part, the attributes and levels that were considered were presented. Table 25 contains the attributes and levels proposed to the focus group in Dutch. The second part consisted of the participants using figures in the Miro board to declare if they deemed the proposed factors relevant, irrelevant or unclear. The results of this part are discussed in C.3. The third part consisted of collectively going through the attributes and levels together. The results of that part will be discussed in C.4. It is important to note that due to a small respondents group, decisions with regards to factors can not be made based on the focus group alone.

Table 25: Proposed Attributes and Levels Focus Group in Dutch

Factor	Niveaus
Kosten	
De app kost:	Maandelijks 0 euro
	Maandelijks 5 euro
	Maandelijks 10 euro
Personalisatie	
De volgende dingen zijn te personaliseren:	Niets
	De hoeveelheid meldingen
	De hoeveelheid meldingen + wat ik wil tracken
Informatief	
De app:	Bevat geen extra informatie
	Bevat informatie over vrouwelijke gezondheid

continued on next page

Table 25 – continued from previous page

Factor	Niveaus
	Bevat informatie over vrouwelijke gezondheid + geeft aan wanneer mijn cyclus afwijkt
Privacy	
De app:	Geeft niets aan over hun privacybeleid
	Heeft een privacybeleid
	Heeft een privacybeleid + vraagt om mijn toestemming voor het delen van data
Discreet	
De app heeft:	Een menstruatie/ovulatie gerelateerd(e) logo en naam
	Een neutra(a)l(e) logo en naam
Gebruiksgemak	
De app is:	Niet makkelijk in gebruik
	Relatief makkelijk in gebruik
	Erg makkelijk in gebruik
Wetenschappelijk gebaseerd	
De menstruatie/ovulatie voorspellingen en informatie worden:	Niet onderbouwd
	Onderbouwd door bronnen
	Onderbouwd door bronnen + onderbouwd door aan de app verbonden artsen
Aangeraden	
De app is aan mij aangeraden door:	Niemand
	Mijn vrienden/familie
	Mijn zorgverlener
Doelgroep gericht	
Hoeveel content van de app is op een bepaalde doelgroep gericht:	Geen
	De app is algemeen, maar bevat ook extra content voor jouw doelgroep
	De content van de app is helemaal toegespitst op jouw doelgroep
Communicatie	
De app communiceert naar de gebruiker als volgt:	Met veel medische termen
	Met simpel taalgebruik
	Met simpel taalgebruik + iconen en afbeeldingen

continued on next page

Table 25 – continued from previous page

Factor	Niveaus
Accuraat	
De app voorspelt de eerste dag van mijn menstruatie:	Bijna niet correct
	Regelmatig correct
	Bijna altijd correct
Reviews en ratings	
De app wordt met dit aantal sterren beoordeeld:	3.2
	4
	4.8
App ontwikkelaar	
De app is ontwikkeld door:	Een commerciële app ontwikkelaar
	Een overheidsinstantie (Bijvoorbeeld: Het ministerie van Volksgezondheid en Welzijn)
	Een medische instantie (Bijvoorbeeld: Nederlandse Vereniging voor Verloskunde en Gynaecologie)
Data input proces	
Input data in de app:	Moet ik handmatig invullen
	Moet ik handmatig invullen nadat ik een melding heb ontvangen
	Wordt automatisch ingevuld en hoef ik alleen maar te controleren

C.2. Results Focus Group Part 1.

- Two out of the six participants disclosed having utilized a MTA, with FLO being the application mentioned. One participant initially employed the app to ascertain whether her menstrual cycle was irregular and subsequently to aid in ovulation tracking for conception purposes, an endeavor that proved successful. She expressed satisfaction with the app's functionality and user experience. The other participant began using the app to monitor her menstrual cycle onset and to gain insights into her cycle's patterns.
- The remaining participants indicated a preference for traditional methods of tracking their menstrual cycle, including the use of a calendar, mental calculations, or not tracking at all.
- Despite varying degrees of familiarity and utilization of MTAs, there was a collective interest among the participants in exploring the use of such apps in the future, albeit with reservations about their added value compared to existing tracking methodologies.
- Key considerations for the adoption of an app or the design of an ideal MTA included:
  - A user-friendly interface without unnecessary complexity.
  - Aesthetic appeal in design was highlighted by one participant as a desirable attribute.
  - An ad-free experience to enhance usability and focus.
  - The inclusion of reminders or notifications to prompt data entry, underscoring the importance of engagement facilitation in app design.

C.3. Results Focus Group Part 2.

- The majority of factors were deemed significant by at least three of the six participants.
- Reviews and ratings were not considered important by half of the participants, with one individual expressing a desire to "form my own opinion about an app." Another participant noted, "What might be important to me, might not be important to someone else and vice versa. We just might have different app needs," highlighting the subjective nature of app evaluation.
- Privacy concerns were minimal for two participants. One simply stated "I just use a fake name when they ask for my personal information," while another was indifferent to data sharing: "I don't really mind that a third party might know my cycle data."
- The value of recommendations was dismissed by one participant, with the same reasoning as for the reviews and ratings.
- The provision of informative content on female health within MTAs was not a priority for one participant, who believed alternative sources could fulfill this need. After clarification of the type of information provided she stated: "Oh, that might be interesting. But, still not that important for me when selecting a MTA."
- Discretion in app design was unanimously deemed non-essential. One participant shared, "I don't care if the name of the app shows that it is a period tracking app," but she still preferred subtlety in notifications: "However, I would prefer not to receive notifications on my phone saying 'you will get your period in 2 days'."
- The identity of the app developer was irrelevant to all participants, with the consensus being that the credibility and evidence-based nature of the content were more important. "If the content is evidence-based, I don't care who developed it, because the content is reliable anyway," and "I care more about the content of the app than who produced the content," were sentiments echoed by the group. All later concurred that this factor was less important compared to the app being evidence-based.

C.4. Results Focus Group Part 3.

Apart from *ease of use*, *app developer*, *Data input process*, *communication*, *recommendations* and *accuracy* all the other attributes and level wording received remarks and comments from the participants.

- The *costs* levels were too big in their ranges. Paying 10 euro per month was highly unrealistic according to the participants. Even 5 euro per month was considered a very high amount. Additionally, they stated that they would consider a one-time payment earlier than a monthly payment. However, it is important to notice here that the bigger part of the focus group consisted of younger people and students, which can often be associated with a lower income (Centraal Bureau voor de Statistiek, 2022). This might have influenced this remark. However, the participants with middle and high income also found the 5 and 10 euro per month subscriptions unfavorable and also stated they would much sooner subscribe with a one-time payment.
- The factor *personalisation* was unclear to the participants. They interpreted the factor as being able to indicate your age or medical conditions, which are needed for the prediction of your cycle, instead of the level to which the app can be customized.
- The factor *informative*, should be defined more precisely. They wanted to know which information was meant specifically.
- With regards to *privacy*, the participants reported hardly ever reading the privacy policies. So therefor privacy policies were not of big influence. However, the sharing of data was deemed important. The participants would prefer to be able to determine which part of their data is shared with who.
- With regards to the *evidence-based* factor, they stated that the level 'doctors associated with the app' was vague. They doubted how objective and science based doctors affiliated with an app would actually be. Also the sequence of the levels was illogical.
- Considering the factor *tailored towards target group*, it was unclear to what extent the app then would be different. What would be the noticeable features of the tailoring for the users? Especially the third level was considered unclear.

- For *reviews and ratings*, they deemed the starting point of 3,2 stars for ratings too high. In their opinion there are a lot of apps with lower ratings, and therefore the difference between the levels was insignificant. Also they found the ratings too random.

D

Experimental Design

Firstly, a basic plan that could accommodate 7 attributes with 3 levels had to be found. Based on that basic plan the number of choice-sets necessary was obtained. This turned out to be basic plan four, which is shown in figure 14.

BASIC PLAN 4: 3^7; 18 trials

1	2	3	4	5	6	7
0	0	0	0	0	0	0
0	1	1	2	1	1	1
0	2	2	1	2	2	2
1	0	1	1	1	2	0
1	1	2	0	2	0	1
1	2	0	2	0	1	2
2	0	2	2	1	0	2
2	1	0	1	2	1	0
2	2	1	0	0	2	1
0	0	2	1	0	1	1
0	1	0	0	1	2	2
0	2	1	2	2	0	0
1	0	0	2	2	2	1
1	1	1	1	0	0	2
1	2	2	0	1	1	0
2	0	1	0	2	1	2
2	1	2	2	0	2	0
2	2	0	1	1	0	1

Figure 14: Basic plan 4

The basic plan works as follows: Each row is a choice-set and each column represents an attribute. The values 0,1 and 2 are used to depict which level each attribute should have in that choice-set. Therefore, the first row of the basic plan shows that all attributes should be at their first level in the first choice-set. As basic plan 4 has 18 rows, this was used as input into the Ngene syntax as the number of choice-sets that was probably necessary to construct the orthogonal design. As discussed in section 4.1, to reduce cognitive load on participants blocking was used. Therefor, the final Ngene syntax looked as follows:

design
;alts = app1, app2
;rows = 18
;block = 2
;orth = seq
;model:

$U(app1) = beta_{priv} * priv[0, 1, 2] + beta_{costs} * costs[0.00, 2.50, 5.00] + beta_{com} * com[0, 1, 2] + beta_{acc} * acc[0.65, 0.80, 0.95] + beta_{use} * use[0, 1, 2] + beta_{pers} * pers[0, 1, 2] + beta_{extra} * extra[0, 1, 2]/$

$U(app2) = beta_{priv} * priv + beta_{costs} * costs + beta_{com} * com + beta_{acc} * acc + beta_{use} * use + beta_{pers} * pers + beta_{extra} * extra\$$

This resulted in the experimental design shown in figure 15. All 18 choice-sets and the corresponding values for both alternatives are shown. Furthermore, the choice-sets are divided into blocks, which can be seen in the last column.

Design															
Choice situation	app1.priv	app1.costs	app1.com	app1.acc	app1.use	app1.pers	app1.extra	app2.priv	app2.costs	app2.com	app2.acc	app2.use	app2.pers	app2.extra	Block
1	0	0	0	0.65	0	0	0	1	5	2	0.8	0	0	1	2
2	0	2.5	1	0.8	1	1	1	1	0	1	0.8	2	0	2	2
3	0	5	2	0.95	2	2	2	0	5	1	0.95	1	0	0	2
4	0	5	1	0.95	1	0	0	2	5	1	0.65	0	1	2	2
5	0	0	2	0.65	2	1	1	1	2.5	1	0.65	2	2	0	1
6	0	2.5	0	0.8	0	2	2	2	2.5	0	0.95	2	0	1	1
7	1	2.5	2	0.95	0	1	0	0	2.5	1	0.8	1	1	1	1
8	1	5	0	0.65	1	2	1	2	0	2	0.8	1	2	0	2
9	1	0	1	0.8	2	0	2	0	2.5	0	0.8	0	2	2	1
10	1	2.5	1	0.65	2	2	0	0	5	2	0.95	2	2	2	1
11	1	5	2	0.8	0	0	1	2	0	1	0.95	0	2	1	1
12	1	0	0	0.95	1	1	2	2	5	0	0.8	2	1	0	1
13	2	5	0	0.8	2	1	0	2	2.5	2	0.65	1	0	2	1
14	2	0	1	0.95	0	2	1	1	2.5	2	0.95	0	1	0	2
15	2	2.5	2	0.65	1	0	2	0	0	2	0.65	2	1	1	2
16	2	0	2	0.8	1	2	0	1	0	0	0.95	1	1	2	2
17	2	2.5	0	0.95	2	0	1	1	5	0	0.65	1	2	1	2
18	2	5	1	0.65	0	1	2	0	0	0	0.65	0	0	0	1

Figure 15: Experimental Design Results

The designs were tested in SPSS to control for correlation between the attributes and thus ensure orthogonality. The correlations between the attributes within an alternative were 0. In between the alternatives low correlations were present. However, this is not an issue that interferes with the results, due to the choice-sets consisting of two MTA alternatives with equal attributes. Therefore, each attribute appears twice in a choice-set. This makes correlations less problematic.

Additionally, the design was be screened for so-called dominant alternatives. These are alternatives where no attribute scores lower than that of the other alternatives and at least one attribute scores higher than the other alternatives. For almost all attributes a preferred level could be assumed, except for the last attribute, which is the extra function. There is no literature available and it is impossible to determine the preferred level based on common sense, due to some respondents preferring a chat function over healthcare information or vice versa. Therefore, that attribute was not considered when looking for dominant alternatives. It appeared that choice-set 3 contained a dominant alternative, namely alternative 1. All attribute levels of alternative 1 were higher or similar to those of alternative 2. Therefor, this alternative was examined further to determine if it had to be removed. Looking at the actual levels, the choice-set was not excessively dominant. As for example, personalising features additionally to personalising reminders, is not proven to be much more preferred by MTA consumers. Consequently, it was included into the design.

Figure 16 displays the final experimental design.

Design														
Choice situation	app1.priv	app1.costs	app1.com	app1.acc	app1.use	app1.pers	app1.extra	app2.priv	app2.costs	app2.com	app2.acc	app2.use	app2.pers	app2.extra	Block
1	0	0	0	0.65	0	0	0	1	5	2	0.8	0	0	1	2
2	0	2.5	1	0.8	1	1	1	1	0	1	0.8	2	0	2	2
3	0	5	2	0.95	2	2	2	0	5	1	0.95	1	0	0	2
4	0	5	1	0.95	1	0	0	2	5	1	0.65	0	1	2	2
5	0	0	2	0.65	2	1	1	1	2.5	1	0.65	2	2	0	1
6	0	2.5	0	0.8	0	2	2	2	2.5	0	0.95	2	0	1	1
7	1	2.5	2	0.95	0	1	0	0	2.5	1	0.8	1	1	1	1
8	1	5	0	0.65	1	2	1	2	0	2	0.8	1	2	0	2
9	1	0	1	0.8	2	0	2	0	2.5	0	0.8	0	2	2	1
10	1	2.5	1	0.65	2	2	0	0	5	2	0.95	2	2	2	1
11	1	5	2	0.8	0	0	1	2	0	1	0.95	0	2	1	1
12	1	0	0	0.95	1	1	2	2	5	0	0.8	2	1	0	1
13	2	5	0	0.8	2	1	0	2	2.5	2	0.65	1	0	2	1
14	2	0	1	0.95	0	2	1	1	2.5	2	0.95	0	1	0	2
15	2	2.5	2	0.65	1	0	2	0	0	2	0.65	2	1	1	2
16	2	0	2	0.8	1	2	0	1	0	0	0.95	1	1	2	2
17	2	2.5	0	0.95	2	0	1	1	5	0	0.65	1	2	1	2
18	2	5	1	0.65	0	1	2	0	0	0	0.65	0	0	0	1

Figure 16: Experimental Design Results

E

Survey Design

E.1. Dutch Translations of Survey Questions

This appendix shows the Dutch translations of the additional app characteristics and perception measuring statements. For the additional app characteristics, respondents had to denote how important the app characteristic was to them. They were asked to ascribe a value of 1-5 to the app characteristic, with the following scales:

- 1 - Helemaal niet belangrijk
- 2
- 3 - Belangrijk
- 4
- 5 - Heel belangrijk

The perception questions were answered by denoting the extent to which the respondents agreed with the statement. The used scale looked as follows:

- Helemaal mee oneens
- Mee oneens
- Niet mee eens of oneens
- Mee eens
- Heel erg mee eens

This lists shows the included additional app characteristics:

- Of de app rekening houdt met verschillende talen, culturen en religies
- Of de app is gemaakt voor mijn doelgroep
- Of de app is gemaakt in samenwerking met vrouwen uit diverse achtergronden
- Of de app goed beoordeeld is in de app store
- Of de app wordt aanbevolen door iemand die ik ken
- Of de app wordt aanbevolen door een dokter of een medische organisatie (zoals een ziekenhuis)
- Of de app wordt aanbevolen door sociale media, een nieuwsartikel of een blogpost
- Of de app gevalideerd is in wetenschappelijk onderzoek
- Of de app is gemaakt in samenwerking met een dokter of een medische organisatie (zoals een ziekenhuis)
- Of de app mijn persoonlijke gegevens deelt met anderen

This list shows the included perception statements:

- Ik denk dat FTAs nuttig zijn

- Ik sta open voor het gebruik van een FTA
- Ik zou een FTA gebruiken om mijn menstruatie bij te houden
- Ik zou een FTA gebruiken om meer te leren over mijn vruchtbaarheid
- Ik zou een FTA gebruiken als ik zwanger wilde worden
- FTAs helpen om meer te leren over je cyclus en vruchtbaarheid
- Je cyclus bijhouden is belangrijk
- Kennis over je eigen cyclus is belangrijk
- Ik zou meer willen weten over vrouwelijke gezondheid, zoals de cyclus

E.2. Included Socio-demographics

The included socio-demographics were defined based on CBS, approved questions from Qualtrics and the reviewed literature. *Gender* is included to account for respondents with female biological reproductive organs that identify as male, non-binary or otherwise. Additionally the term 'female' is avoided where possible. Except in the case of 'female health'. The options for gender were adapted from Qualtrics approved questions and the questions used by Zhaunova et al. (2023). The *ethnicity* question was based on the Qualtrics questions and the biggest migration population groups in the Netherlands as described by CBS (Centraal Bureau voor de Statistiek, 2016). For *education* the distribution used by CBS to define low, middle and high income was used (Centraal Bureau voor de Statistiek, 2022). Lastly, *income level* is based on CBS and the standard Qualtrics questions, *work situation* on the Qualtrics questions and neighbourhood was determined based on the zip-code. Neighbourhood data is combined with the neighbourhood data that CBS has available on neighbourhoods with low socio-economic positions.

F
Principal Axis Factoring

This appendix contains a detailed explanation of the Principal Axis Factoring executed to determine any possible latent factors underlying the perception scores of respondents. This was done to derive variables that could be included into the ML and LCCM models depicting respondents' attitudes. The analysis was executed in SPSS through a few sequential steps.

When executing this method, the aim should be to derive a simple structure and simultaneously interpretable and useful perception factors. A simple structure means that each indicator loads high (>0,50) on one factor and low (<0,30) on all other factors. The factor loading depicts the correlation an indicator has with the constructed factor. Factor loadings lower than 0,30 will be suppressed, as these are considered negligible. Preferably, a minimum of two indicators score high on each factor.

Firstly, all steps were executed with oblique rotation and subsequently all steps were repeated using orthogonal rotation. Oblique rotation assumes that the underlying factors are correlated, where orthogonal does not. Both methods will be tested to examine if either one of the rotation methods requires less perception variables to be removed. To clarify the shown tables in this section, in SPSS the orthogonal rotation is denoted with Varimax and oblique rotation with Oblimin.

The first patterns matrix that included all perception variables looked as follows:

Pattern Matrix[a]

	Component 1	Component 2
Usefulness fertility apps	0.843	
Willingness to use a fertility app	0.871	
Would use a fertility app to track menstruation	0.657	
Would use a fertility app to learn about my fertility	0.846	
Fertility apps can help to learn more about your cycle and fertility	0.659	
Tracking your menstruation is important		0.857
Knowledge of your own cycle is important		0.941
I would want to know more about female health such as the cycle		0.593

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Figure 17: Principal Axis Factoring Oblique Rotation 1

As figure 17 shows, all variables score high on either one of the factors and no variable scores high on more than one factor. Therefor, a simple structure was achieved in one rotation. Furthermore, the established factors are interpretable and therefore suited for use in the choice model estimation. Additionally, a component correlation matrix was estimated, which can be seen in figure 18. This shows that the correlation between factor 1 and factor 2 is 0,473.

Component Correlation Matrix

Component	1	2
1	1.000	0.473
2	0.473	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Figure 18: Component Correlation Final Oblique Rotation

Subsequently, the same analysis was executed with orthogonal rotation. This analysis required removal of several perception variables and thus was less preferred than oblique rotation. Therefor, only the final iteration of this analysis was shown. As figure 19 shows, achieving a simple structure required the removal of three perception variables.

Rotated Component Matrix[a]

	Component	
	1	2
Usefulness fertility apps	0.832	
Willingness to use a fertility app	0.855	
Would use a fertility app to learn about my fertility	0.799	
Tracking your menstruation is important		0.873
Knowledge of your own cycle is important		0.909

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Figure 19: Final Iteration Orthogonal Rotation

The factors obtained with oblique rotation, that are showed in 17 are thus used to construct the new perception factor scores for each respondent. This is done by taking the average of all numerical values for the variables that score on the factor.

G

Statistics and Choice Model Estimation

This appendix contains details of the coding of the personal characteristics and all One-Way ANOVA and T-tests executed to determine interaction effects on perceptions, additional app characteristics and the constructed factors.

G.1. Coding of the Personal Characteristic Variables

This section explains how the retrieved socio-demographic data was re-coded into categories displayed in table 4 in section 5.2.

Firstly, age was coded into 4 categories. For income the highest income ranges were grouped together due to few respondents belonging to those categories. Additionally, respondents that answered 'I would rather not say' were assumed to earn the modal income in the Netherlands. This was 41.500 euro per year for 2023, which is level 4 (Centraal Planbureau, 2023). Therefor, the respondents were added to level 4. Education was re-coded into low & middle education and high education. This was done as the low education, only contained 2 (1,2%) respondents. Work situation was re-coded as well, the option 'other' and 'I would rather not say' were grouped together. The ethnicity questions were merged into one variable depicting Dutch and non-Dutch respondents. Respondents that were Dutch and had another ethnicity were coded as non-Dutch. In this research the ethnicity is elicited to determine if cultural or ethnic differences influence app preferences. Respondents that reported identifying as partly Dutch and partly another ethnicity, might have different cultural or ethnic rituals influencing their preferences and therefore they were coded into the variable. Additionally, the user variable was re-coded as follows:

Table 26: Re-coded User Variable

Variable	Original value	Recoded value	Amount
Have you ever used a MTA?	Yes, I am currently using a MTA	User	62
	Yes, I have in the past	User	31
Total			93
	No, I have never	Non-user	61
	Have never heard of MTA	Non-user	14
Total			75

G.2. Level Preferences

As table 27 shows, all though preferences were distributed among respondents, in most cases the respondents' preference aligned with the assumed preference. For *privacy*, *language* and *personalisation* the majority of the respondents preferred the assumed preferred attribute, with respectively 82,4%, 80,4% and 78,6% of the respondents preferring that level.

The attribute that had a more surprising answer was the *data input* attribute. For that attribute the assumed preferred level was almost equally as preferred to the second level. With 44,1% choosing *entering one data point + reminder notification* as opposed to 37,5% choosing *entering one data point*. Since the only difference is a reminder notification, this could point out that some respondents do not prefer a reminder notification for data entry. Or perhaps they would prefer to optionally receive reminders, instead of reminders being a given. For some respondents reminders are only beneficial if they can be adjusted to their needs, as discussed previously in section 2.2.1.

Two of the attributes were framed differently. These were the *accuracy* and *extra feature* attributes. The accuracy question was phrased as 'what accuracy percentage do you expect current apps to have?', due to the assumption that everyone prefers the highest accurate app. Half of the respondents expected apps to have 80% accuracy, 39,9% expected 65% accuracy and 10,1% expected 95% accuracy. The other attribute *extra feature* was phrased as follows: 'Which additional feature would you pay 2 euro for once?'. 45,2% of the respondents would pay 2 euro to be able to access *the chat for medical questions and information*, which makes it significantly more preferred to *information on female health* (15,5%). Whereas, 39,3% selected *none* and thus was not willing to pay 2 euro's for any of the two proposed extra functions. In existing apps the respondent always receives access to multiple functions when paying the premium subscription fee. Therefore, the choice in real life to pay for extra features is different due to this question only proposing one additional feature.

Table 27: Overview Attribute Level Questions

Attribute	Levels	Amount	Percentage
Privacy	None	16	9,5%
	Provides information	14	8,3%
	Provides information + asks for consent	138	82,1%
Communication	Medical terms	10	6,0%
	Simple language	23	23,7%
	Simple language + icons and images	135	80,4%
Accuracy | How often do you think MTAs predict correctly?	65%	67	39,9%
	80%	84	50%
	90%	17	10,1%
Ease of Use	Entering multiple data points	31	18,5%
	Entering a single data point	63	37,5%
	Entering a single data point + reminder notification	74	44%
Personalisation	None	12	7,1%
	Notifications	24	14,3%
	Notifications + dis-/enabling functions	132	78,6%
Extra Feature | For which function would you pay 2 euro?	None	66	39,3%
	Information on female health	26	15,5%
	Chat function for medical questions and information	76	45,2%

G.3. Socio-demographic Variable Interaction with App Characteristic Ranking

The socio-demographic variables with more than 2 categorical levels were *age*, *work situation* and *income*. All of these were included like they were displayed in table 4. Furthermore, the t-test variables were *ethnicity*, *education level* and lastly, the *user* characteristic. The first two were included like in table 4 as well. The user characteristic was included as it is displayed in table 26.

G.3.1. One-Way ANOVA Results

The One-Way ANOVA test revealed that with regards to *income*, no significant ranking differences between the groups were present. For the *work situation* variable the statement *if the app has good ratings in the app store* was answered significantly different between the groups (p = 0,03). Looking closer at the mean differences of the groups, the only p value that could be significant if a confidence interval of 90% is used, is the difference between respondents with a full-time job and the respondents falling under the other category. However, as the 'other' category is very difficult to profile, with respondents that did not answer, respondents that worked side-jobs and respondents that were part-time students, no relevant relation can be derived from this finding.

With regards to *age*, also one statement was answered differently between groups. This was *if the app is recommended by someone I know* (p = 0,028). The significant differences were found between the age category <19 and both the categories 19-26 and 27-35. With the lower age category assigning a lower ranking to this statement than the other two older age categories (p = 0,041; p=0,049). Which indicates that teenagers on average appoint less importance to a recommendation from someone they know than young adults do. As to why this relation is found, more research is necessary as the first intuition points to the opposite effect as teenagers are often known to value the approval of people they know highly. Perhaps, this does not uphold for mHealth and they seek approval from more validated sources.

G.3.2. T-test Results

For the variable *education* no significant differences between the lower education levels and higher educational levels between app factor ranking was found. The *ethnicity* variable did influence the ranking of *rating in app store* (p =0,04). With Dutch respondents (*M*= 2,64; *SD*= 1,22) rating this lower than non-Dutch respondents (*M*= 3,35; *SD* = 1,37). Indicating that for Dutch respondents the ratings made by other consumers are less important. As to why this is the case, more research is necessary. The variable *user* revealed interaction with multiple app factors. Firstly, the users (*M*= 3,3; *SD* = 1,3) valued *if the app is recommended by a doctor or medical institution* lower than non-users (*M*= 4,0; *SD* = 1,02)(p < 0,001). Additionally, users also differed on the value of *if an app is made in cooperation with doctors or a medical organisation* (p = 0,018). Users (*M*= 3,66; *SD* = 1,21)ascribed a lower value to this app factor than non-users (*M*= 4,08; *SD* = 1,05). Lastly, users (*M*= 3,69; *SD* = 1,26) also valued *if an app is approved in scientific research* significantly (p = 0,004) lower than non-users(*M*= 4,19; *SD* = 0,97). All these mean differences found for the user variable, show that users value scientifically based MTAs as less important than non-users. This covers recommendation, development and testing. Perhaps, users are more convinced of the additional values of MTAs that scientific grounding is less important. However, the reasoning behind this observation requires more research.

G.4. Socio-demographic Variable Interaction with Attitude Statement Scores

To determine if there were significant difference between the attitudes of the respondents based on their personal characteristics, a One-way ANOVA test was executed for variables with more than 2 categorical levels and a t-test for variables with 2 categorical levels. The execution was precisely similar to the execution of the tests for the factor questions, described in G.3.

G.4.1. One-Way ANOVA Results

The One-Way ANOVA pointed out that based on respondent's income and work situation no significant differences in the measured attitudes and perceptions were present. For age significant differences were found with regards to four statements. All of the differences were between the >45 age category and one or two of the younger categories. These are presented per statement down below.

- 'I find fertility & menstrual tracking apps useful' was answered significantly (p = 0,02) lower for respondents older than 45 compared to the 19-26 age category (mean difference of -0,91; p = 0,009) and to the 27-35 age category (mean difference of -0,95; p = 0,009).
- 'I am open to using fertility & menstrual tracking apps' was agreed on less (p = 0,007) by the >45 age category than the 19-26 category with a -1,09 difference (p = 0,018).
- 'I would use a fertility & menstrual tracking app to learn more about my fertility' was answered significantly (p = 0,009) lower for respondents older than 45 compared to the 19-26 age category (mean difference of -1,10; p = 0,02) and to the 27-35 age category (mean difference of -1,18; p = 0,02).
- 'Fertility & menstrual tracking apps can help to learn more about your cycle and fertility' was agreed on significantly less (p = 0,007) by the >45 age category than the 27-35 category with a -0,98 difference (p = 0,009).

Apparently, the older respondents are less willing to use MTAs (the second and third statement above) and less convinced of the usefulness of fertility & menstrual tracking apps (the first, second and fourth statement) than the younger respondents. Perhaps, they are less convinced of the usefulness of apps in the first place, since the younger categories use their phones more. It is also interesting to note, that with regards to the importance of female health knowledge, no significant differences were found.

So the difference is strictly applicable to MTAs. Also, no significant difference was found between the oldest age category and the youngest category. So the older respondents' attitudes are not significantly different from teenagers in this sample.

G.4.2. T-test Results

The t-test showed that education did not have a significant impact on how respondents answered the perception questions. For the ethnicity, one significant difference was found ($p = 0{,}04$) which was that non-Dutch respondents agreed more with the statement 'I would want to know more about female health such as the cycle' ($M=4{,}05$; $SD=0{,}83$) than Dutch respondents ($M=3{,}61$; $SD=1{,}05$). However, both still fall into the agree category. The user variable revealed significant differences for almost all statements. For all statements p had a value of $<0{,}001$, except for 'Knowledge of your own cycle is important' ($p = 0{,}01$) and 'Fertility & menstrual tracking apps can help to learn more about your cycle and fertility' ($p = 0{,}002$). For all statements the users had a higher mean agreement score than the non-users, which are shown in table 28 below:

Table 28: Mean Differences for the User Variable on Perception Questions

Statement	MTA User	Mean	Std. Deviation
I find fertility & menstrual tracking apps useful	Yes	4.16	0.680
	No	3.27	0.741
Difference		0.89	
I am open to using fertility & menstrual tracking apps	Yes	4.32	0.710
	No	3.17	1.129
Difference		1,15	
I would use fertility & menstrual tracking apps to track menstruation	Yes	4.29	0.788
	No	3.08	1.148
Difference		1,21	
I would use fertility & menstrual tracking apps to learn about my fertility	Yes	3.84	1.014
	No	3.29	1.175
Difference		0,55	
Fertility & menstrual tracking apps can help to learn more about your cycle and fertility	Yes	4.25	0.803
	No	3.84	0.886
Difference		0,41	
Tracking your menstruation is important	Yes	3.94	0.918
	No	3.31	0.972
Difference		0,63	
Knowledge of your own cycle is important	Yes	4.19	0.824
	No	3.84	0.931
Difference		0,35	
I want to know more about female health such as the cycle	Yes	3.91	0.940
	No	3.36	1.074
Difference		0,55	

The results show that the differences between the means are the greatest for the statements with regards to the willing to use MTAs. As the user has previously or is currently using these apps, it is expected that they score highly on both these statements. Where the user on average totally agrees with both the statements, the non-user on average is neither disagreeing or agreeing with both of them. It is interesting to note that the difference is significantly smaller if the willingness to use is specified to using the apps for fertility reasons. This could indicate that the (previous) users in the sample mostly use(d) their MTAs for menstrual tracking and less for fertility tracking and are therefor less interested in that feature of MTAs specifically. The differences are also less big for the statements with regards to usefulness (the first and fifth statement) and with regards to the importance of female health knowledge in general (the last three statements). However, for all these statements, except for the statement 'knowledge of your own cycle is important', the user falls into a higher agreement category than the non-user. The influence of having used or currently using a MTA is thus substantial in determining an individuals' attitude towards MTAs.

G.5. Socio-demographic Variable Interaction with Attitude Statement Scores

To determine if there were significant difference between the two constructed factors based on respondents' personal characteristics, a One-way ANOVA test was executed for variables with more than 2 categorical levels and a t-test for variables with 2 categorical levels. The execution was precisely similar to the execution of the tests for the factor questions, described in G.3.

G.5.1. One-Way Anova Test

For the One-Way Anova test, income, age and work situation were included. For both the first factor and the second factor, there were no significant mean differences found at the significance level of $p < 0{,}05$. Which means that based on socio-demographics the respondents did not have substantial different scores on the two factors.

G.5.2. T-Test

As ethnicity and user are two-categorical variables, a t-test was more fit for testing mean differences. The same holds for education, which was re-coded into a two categorical value, due to the lowest initial third level only containing 2 cases. The t-test performed for these variables lead to no significant differences at $p < 0{,}05$ for ethnicity and education. However, the user characteristic was significant for the ProMTA factor ($t(136{,}02) = -7{,}61$; $p < .001$) and the FemaleHealth factor($t(166) = -4{,}26$; $p < .001$). The mean for non-users was 3,5 and 3,3 for ProMTA and FemaleHealth respectively, compared to 4,2 and 4,0 for (previous) users. This indicates that the respondents that had ever used or are currently using a MTA score value the importance female health knowledge higher and are more open to MTAs than non-users.

G.6. Recoding of Variables for Choice Model Input

The two tables in this section show the re-coding of incorporated variables for the choice model. The personal characteristics were also effect-coded, except for income, which is included as an interval variable. These are shown in 29. The attributes were effect-coded like in table 30. The names of the two indicator variables of an attribute depict the level they represent.

Table 29: Effect Coded Personal Characteristics

Attributes	**Levels**	**Re-coded**		
Education Level	High	1		
	Low & middle	-1		
Ethnicity	Dutch	1		
	None-Dutch	-1		
User	User	1		
	Non-user	-1		
Work Situation		Part-time	Student	Other
	Full-time	-1	-1	-1
	Part-time	1	0	0
	Student	0	1	0
	Other	0	0	1
Income	<20.000 euro per year	1		
	20.000 - 30.000 euro per year	2		
	30.000 - 40.000 euro per year	3		
	40.000 - 50.000 euro per year	4		
	50.000 - 60.000 euro per year	5		
	>60.000 euro per year	6		

Table 30: Effect Coded Attributes

Attributes	Levels	Effect coded	
		High_costs	Low_costs
Costs	€ 0.00	-1	-1
	€ 2.50	0	1
	€ 5.00	1	0
		Consent	Information
Privacy	None	-1	-1
	Provides information	0	1
	Provides information + asks for consent	1	0
		Icons	Simple_language
Communication	Medical terms	-1	-1
	Simple language	0	1
	Simple language + icons and images	1	0
		High_accuracy	Medium_accuracy
Accuracy	65%	-1	-1
	80%	0	1
	95%	1	0
		Reminder	Single_entry
Ease of use	Input multiple data points	-1	-1
	Input a single data point	0	1
	Input a single data point + reminder notification	1	0
		Features	Notifications
Personalisation	None	-1	-1
	Notifications	0	1
	Notifciations + dis-/enabling features	1	0
		Chat	Health_information
Extra function	None	-1	-1
	Information on female health	0	1
	Chat function for medical questions and information	1	0

H

Results

Table 31: Overview ML Results - Extended Statistic Metrics

	Est.	Rob. s.e.	Rob. t.rat. (0)	p (2-sided)	CI 95%	
BETA_high_costs	-0,936	0,102	-9,186	0,00	-1,136	-0,737
BETA_high_costs	-0,345	0,114	-3,039	0,00	-0,568	-0,123
BETA_consent	0,431	0,157	2,753	0,01	0,124	0,738
BETA_information	0,305	0,124	2,464	0,01	0,062	0,547
BETA_icons	0,348	0,078	4,452	0,00	0,195	0,502
BETA_simple_language	-0,034	0,090	-0,374	0,71	-0,211	0,143
BETA_high_accuracy	0,943	0,109	8,611	0,00	0,728	1,157
BETA_medium_accuracy	0,138	0,095	1,460	0,14	-0,047	0,324
BETA_reminder	0,071	0,077	0,918	0,36	-0,081	0,223
BETA_single_entry	-0,021	0,091	-0,227	0,82	-0,198	0,157
BETA_features	0,466	0,139	3,365	0,00	0,195	0,738
BETA_notifications	-0,177	0,128	-1,386	0,17	-0,427	0,073
BETA_chat	0,097	0,085	1,136	0,26	-0,070	0,264
BETA_health_information	0,116	0,085	1,368	0,17	-0,050	0,282
BETA_age	-0,031	0,028	-1,111	0,27	-0,086	0,024
BETA_ethnicity	-0,493	0,246	-2,003	0,05	-0,975	-0,011
BETA_education	-0,509	0,272	-1,874	0,06	-1,041	0,023
BETA_user	0,366	0,209	1,755	0,08	-0,043	0,775
BETA_ProMTA	1,141	0,293	3,896	0,00	0,567	1,716
BETA_features_ethnicity	-0,369	0,135	-2,727	0,01	-0,634	-0,104
BETA_notifications_ethnicity	0,297	0,128	2,322	0,02	0,046	0,547
ASC	-4,396	1,587	-2,771	0,01	-7,506	-1,286
SIGMA_app	1,944	0,177	10,957	0,00	1,597	2,292

Table 31 shows the results of the ML model. The grey coloured estimates indicate that the parameters are insignificant. Their CI contains a negative and a positive value and their t-ratio values are less than

$|1{,}96|$.