DietWise

Funded by the European Union

DietWise

SYSTEMIC CHANGES | EMPOWERED CITIZENS

# Deliverable D5.3

## The pilot implementation plan for BE

Severijns, Rosaly (KUL)
Coeckelbergh, Caroline (VIGL)

https://www.dietwise.eu

Funded by the European Union

This project is funded by the European Union's *Horizon Europe program* under grant agreement No. 101181692

# Project information

| | |
|---|---|
| Program: | Horizon Europe |
| Topic: | HORIZON-CL6-2024-FARM2FORK-01-5 |
| Type of action: | HORIZON-RIA HORIZON Research and Innovation Actions |
| Grant Agreement #: | 101181692 |
| Project title: | Systemic Solutions to Enhance Healthy and Sustainable Food Provision and Cooking at Home |
| Project Name: | DietWise |
| Project Start Date: | 2024-11-01 |
| Project End Date: | 2027-10-31 |

# Document information

| | |
|---|---|
| Document name: | The pilot implementation plan for BE |
| Related Work Package: | WP5 |
| Related Task: | Task 5.3 “Setting up the pilots in Belgium” |
| Related Deliverables: | D5.3 |
| Author(s): | Severijns, Rosaly (KUL); Coeckelbergh, Caroline (VIGL) |
| Reviewer(s): | Botchway, Ebo (KUL); Bulotaite; Gabija (VU) |
| Submission date: | 2026-30-04 |
| Dissemination level: | Public |

# Document history

| Version | Date | Changes | Responsible partner |
|---|---|---|---|
| v0.1 | 2026-01-19 | First draft of deliverable template | VIGL |
| v0.2 | 2026-04-13 | First complete draft of pilot plan | VIGL, KUL |
| v0.3 | 2026-04-30 | Final version after internal review | VIGL, KUL |

*Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.*

# Abbreviations

| Abbreviation | Full Form |
|---|---|
| AI | Artificial Intelligence |
| EU | European Union |
| FOOD | Foodbag |
| GDPR | General Data Protection Regulation |
| KPIs | Key Performance Indicators |
| KUL | KU Leuven |
| RCT | Randomized Controlled Trial |
| WP | Work Package |
| VIGL | Vlaams Instituut Gezond Leven |

# Contents

The pilot implementation plan for BE ..... 1
1. Introduction & Background ..... 5
2. Scope & Pilot Descriptions ..... 5
2.1 MyRecipeWatch Pilot ..... 6
*2.1.1 Research questions and hypotheses* ..... 6
*2.1.2 Study Design* ..... 8
*2.1.3 SMART Goals* ..... 10
*2.1.4 Target sample* ..... 10
*2.1.5 Inclusion/Exclusion Criteria* ..... 12
*2.1.6 Data Collection, Types & Measures* ..... 12
*2.1.7 Evaluation & Analysis Plan* ..... 14
2.2 Foodbag Pilot ..... 16
*2.2.1 Research questions and hypotheses* ..... 16
*2.2.2 Study Design* ..... 16
*2.2.3 SMART Goals* ..... 18
*2.2.4 Target sample* ..... 18
*2.2.5 Inclusion/Exclusion Criteria* ..... 19
*2.2.6 Data Collection, Types & Measures* ..... 19
*2.2.7 Evaluation & Analysis Plan* ..... 21
3. Stakeholder engagement ..... 22
4. Data Management plan ..... 22
5. Risks & Limitations ..... 22
6. Ethics ..... 24
7. Reporting & Dissemination ..... 24
8. Conclusion ..... 24
9. References ..... 26
ANNEX ..... 27
A1. Pilot Questionnaire (Dutch) ..... 27

# 1. Introduction & Background

The DietWise project aims to improve food literacy and cooking practices in terms of health and sustainability through innovative tools and interventions. Two key components of the project are developing two AI-based tools and developing and testing behavioral interventions (i.e., message nudges). These tools and interventions will subsequently be tested in the field across three pilot countries: Belgium, Greece and Lithuania. This report provides the pilot plan for Belgium. The Belgian Food Consumption Survey indicates that the majority of the population does not consume the recommended amounts of healthy food products, such as water, wholegrain products, fruits and vegetables, legumes and nuts. Likewise, most people in Belgium exceed the recommended amounts for unhealthy products, such as red and processed meat and alcohol (Sciensano, 2023). Hence, there is a need to align eating practices with nutritional guidelines.

To address this gap, one of the key objectives of the DietWise-project is to develop a digital tool that supports healthier and more sustainable food choices in everyday contexts. One of the tools will be the MyRecipeWatch app. MyRecipeWatch allows users to enter recipe weblinks, extracts the ingredients, and then provides recipe recommendations (i.e., to replace ingredients) based on an algorithm guided by the Global Burden of Disease Dietary Risk factors (Afshin et al., 2019). It provides transparent, user-driven interactions, including acceptance statistics for substitutions and a score system that highlights the improvement potential of proposed changes. In terms of behavioral interventions, DietWise has done extensive online survey testing of nudge messages to encourage individuals to accept such ingredient substitutions. While most did not lead to a significant increase in acceptance (often acceptance rates were already high and depended on participants' characteristics), messages like 'Healthiness guaranteed', 'Join the 75% who accept this recommendation', 'Why not go for the better option?', and 'Recommended by MyRecipeWatch healthy diet experts', were effective for specific subgroups such as men and value-driven consumers. Both MyRecipeWatch and the behavioral interventions warrant field testing to account for real-life noise and daily practices.

In Belgium, KU Leuven (KUL) and Vlaams Instituut Gezond Leven (VIGL) will evaluate and disseminate MyRecipeWatch among the general population. In addition, one of DietWise's goals being 'empowering vulnerable populations', we will make additional efforts to deploy and evaluate the app among people at risk of vulnerability (e.g., migrants, low socio-economic status households). Next to the MyRecipeWatch app, Foodbag (FOOD) also aims to implement similar ingredient substitution suggestions in their mealbox ordering system, providing another opportunity for making diets healthier. This latter addition will be tested in a second field study. For each study, the team aims to adhere to the highest standards of experimental research (e.g., randomized design), while also taking into account practical relevance and limitations. To ensure usability and accuracy based on both citizens' perspectives and expert advice, we involve a range of stakeholders (e.g., citizens, dietitians) to test and improve the app.

This report outlines the pilot protocol for Belgium, including the design, goals, sample, data collection procedures, analysis plans, plans for stakeholder engagement, potential risks and mitigation strategies, and ethical considerations. It provides a guide to bringing the research into practice, fosters transparency, and prepares the team for ethical requests and pre-registration of the plans.

# 2. Scope & Pilot Descriptions

This deliverable focuses on establishing the methodology, timeline and pilot details of the two pilots in Belgium. The pilot implementation plans are based on SMART goals adapted for Belgium, testing two app-based tools (MyRecipeWatch and the Foodbag app) and behavioral interventions (the nudges that appeared to be most effective in online survey studies related to WP3-T3.1) and engaging relevant stakeholders. The Belgian pilots focus on improving ingredient choices to be more in line with predefined dietary guidelines, which are provided by the three field partners and find their base in the nutritional guidelines of the three countries respectively (Belgium, Lithuania, Greece). This will be done in two different ways, with varying tools and perspectives:

1. The MyRecipeWatch pilot, implemented by VIGL, aims to roll out the MyRecipeWatch-app in Flanders and test whether the app positively influences users' food literacy and healthy eating behavior, as well as whether there are heterogeneous effects for subgroups. The app provides recommendations for user-provided web-based recipes. In the main component of the study (Study 1a), participants are asked to test the app for (at least) a

month and fill in several surveys for evaluation. We also track their user behavior to analyze app retention and activity. The target group is the general population that VIGL can reach through their network, but with the specific aim of oversampling vulnerable populations. Moreover, in a sub-study (Study 1b), the team provides a physical onboarding session for populations at risk of vulnerability and tests whether this increases retention and hence behavioral effects.

2. The Foodbag pilot (Study 2), implemented by FOOD, will focus on providing healthy and meat-to-vegetarian ingredient swaps upon selection of meal kits to nudge people towards healthier ingredients. These swaps are tested in terms of effects on healthy meal choices (within the ordering system) and food literacy and healthy eating behavior in general. Examples would include swapping white pasta with its whole wheat alternative or switching a meat option for a plant-based one. This study targets Foodbag's clientele network.

The present deliverable also focuses on the development and management of stakeholder involvement, ranging from citizens and health professionals (e.g. dieticians) to public authorities.

The pilot activities will run during M19-M34 of the project, as part of WP7 activities.

## 2.1 MyRecipeWatch Pilot

### *2.1.1 Research questions and hypotheses*

The MyRecipeWatch pilot, implemented by VIGL (Gezond Leven), will roll out the MyRecipeWatch app in Flanders through two simultaneous studies:

- **Study 1a** is the main study that evaluates the effectiveness of the app among the general population and a larger subsample of vulnerable people. A clean design with waitlisted control group will be used, which allows singling out MyRecipeWatch's causal effect.
- **Study 1b** stimulates app usage among people with risks of vulnerability in practice, by providing physical onboarding sessions. Those who are willing also fill in the surveys so we can analyze changes over time, and if enough people participate, match them to a comparable external control group from Study 1a (as randomization is not possible in Study 1b) to test the effect of the onboarding session and the effectiveness of the app in this particular group.

Together, the studies evaluate:

1. **Effectiveness:** Does the app improve users' food literacy **and** positively influence healthy eating behaviors?
2. **Longer-term effectiveness**: Do the improvements due to the app last 1, 3 and 6 months after the initial intervention month?
3. **Heterogeneity:** Which groups experience smaller or larger effects from the app, and are more or less likely to keep using the app?
4. **In-app activity and behavioral interventions:** Does more app usage lead to larger effectiveness, and which in-app behavioral nudges increase the acceptance of recipe suggestions?
5. **Equity and accessibility:** Does the app differentially affect groups at risk of vulnerability, does an onboarding session help for longer-term engagement among those groups, and what improvements are needed for continued usage? This is integrated into the sampling strategy and heterogeneity testing of Study 1a and testing an onboarding session in Study 1b.

Specifically, this leads to the main research question and hypotheses presented in **Table 1**.

**Table 1:** Study 1a and 1b goals, research questions and hypotheses.

| ***Goal*** | ***Hypothesis*** |
|---|---|
| **Study 1a** | |
| Main research question: *Does an app that provides personalized AI-driven ingredient recommendations (MyRecipeWatch) improve food literacy and healthy eating behavior?* | |
| **1. Effectiveness** | **H1**: Using MyRecipeWatch significantly improves food literacy compared to a waitlisted control group. |
| | **H2**: Using MyRecipeWatch significantly improves healthy eating behavior compared to a waitlisted control group. |
| **2. Longer-term effectiveness** | **H3**: One month after the intervention, the treatment group will still display a significantly higher level of food literacy and healthy eating behavior than their baseline values. |
| | **H4**: Three and six months after the intervention, the food literacy and healthy eating behavior of the treatment group will still be significantly higher than their baseline values. |
| **3. Heterogeneity** | **H5**: Motivation for healthy eating, motivation for sustainable eating, vulnerability dimensions and gender moderate the effectiveness of MyRecipeWatch. |
| | **H6**: Motivation for healthy eating, motivation for sustainable eating, vulnerability dimensions and gender moderate the continued usage (retention) of MyRecipeWatch. |
| **4. In-app activity and behavioral interventions** | **H7:** Individuals who use the app more often experience a larger improvement in food literacy and healthy eating behavior (app usage as a moderator). |
| | **H8**: Nudges have a differential effect on the likelihood of accepting a recipe suggestion. |
| | **H9**: Motivation for healthy eating, motivation for sustainable eating, vulnerability dimensions and gender moderate the effect of nudges on accepting recipe suggestions. |
| **Study 1b** | |
| Main research question: *Does a physical onboarding session explaining the MyRecipeWatch app help vulnerable groups to start and keep using the app and improve their food literacy and healthy eating behavior?* | |
| **1. Effectiveness** | **H1a:** MyRecipeWatch improves food literacy over time (improvement compared to baseline values) |
| | **H1b:** MyRecipeWatch improves food literacy compared to a matched sample from Study 1a's control condition.[1] |
| | **H2a:** MyRecipeWatch improves healthy eating behavior over time (improvement compared to baseline values) |
| | **H2b:** MyRecipeWatch improves healthy eating behavior compared to a matched sample from Study 1a's control condition.[1] |
| **5. Equity and accessibility** | **H3:** The onboarding session increases app engagement and retention compared to a matched sample from Study 1a's treatment condition (no onboarding).[1] |
| | **H4:** The onboarding session increases food literacy and healthy eating behavior compared to a matched sample from Study 1a's treatment condition (no onboarding). [1] |
| | **H5:** The onboarding session is most effective for people with a low educational level and high financial constraint. [1] |
| **3. Heterogeneity** | *H5-7 from Study 1a are also tested in Study 1b.* |

[1] Only under the condition that a sufficient number of people participates in the surveys of Study 1b. If matching is not possible, we will compare changes without matching (difference-in-differences) and ask for qualitative feedback from onboarding participants.

### 2.1.2 Study Design

MyRecipeWatch is a mobile application that provides ingredient swap suggestions upon entering a recipe weblink. The app relies on an algorithm that scores recipes based on 15 dietary components derived from the Global Burden of Disease (GBD) dietary risk framework[1] (Afshin et al., 2019). These components correspond to food groups or nutrients that are either encouraged (e.g., fruits, legumes, whole grains) or recommended to be limited (e.g., sodium, processed meat). The suggestions are accompanied by a recipe score comparing the new recipes to the old one, and non-punitive advice explaining the reasoning behind the suggestions.

In addition, if resources allow by the time the app is first launched, behavioural nudges will be displayed next to the recipe recommendations. These nudges represent the three most effective interventions among the 20 tested in the DietWise WP3 megastudy, a previous survey conducted within DietWise that evaluated the impact of different messages on recommendation acceptance. Each time someone in enters a recipe, they will randomly see no message, or one of the three nudges. We test the effectiveness of the app through a randomized, mixed (between-subject and within-subject) design with two groups and repeated measures. The mixed design arises from a waitlist approach, in which treatment participants are compared to waitlisted participants (between-subject), while for longer-term follow-ups within-person changes over time are analyzed due to the lack of a longer-term control group. The design proceeds as follows:

1. *Pre-test phase*. New apps, especially AI-driven ones, need a period of iterative development and testing to ensure usability and accuracy. From May to July 2026, MyRecipeWatch will be extended to the pilot country languages, pre-tested by citizens and dieticians and improved. Upon finalizing the app, we will conduct a pre-pilot of the study around August-October 2026, with only around 20 participants. The latter can lead to further improvements to the app and study design.

2. *Recruitment*. **Study 1a:** VIGL recruits participants through their network (17.700 followers on social media, 12.772 subscribers to the general newsletter, extensive partner network), interested in testing a new app for them. They are informed about the design and that some groups may have to wait as the app is tested in different 'phases'. As it is crucial, and generally more difficult to reach populations at risk of vulnerability with initiatives stimulating healthy eating knowledge and habits, we will put extra effort into recruiting people with a risk of vulnerability, to be able to analyze their experiences with the app. These efforts include spreading the call through partners (health, primary care, local health actors, social organizations, local governments) and intermediaries (community health workers, social services) within VIGL's network, physical advertisements in supermarkets, schools and other public spaces in lower socio-economic neighborhoods, and posting the call in specific Facebook and Reddit groups. **Study 1b:** To recruit people for the onboarding sessions, VIGL simultaneously identifies organizations and events that target people with a risk of vulnerability. The most feasible path will be to identify events we can be a part of, so it will cost the organization less effort. In case the event-based method does not yield much success, we organize stand-alone sessions in community organizations, companies, or public locations to explain the app.

3. *Onboarding session*. As part of **Study 1b**, attendees receive an explanation of the app, are instructed to download the app, and try out the app. They will also be able to ask questions. Attendees are invited to also participate in a study, enabling VIGL and KUL to improve the app.

4. *Baseline survey*. During recruitment, **Study 1a** participants receive the immediate suggestion to fill in the baseline survey. They can also opt to complete the baseline survey within the next week to prevent immediately pushing potential participants away. During the **Study 1b** onboarding sessions, interested attendees who have time can stay a bit longer to immediately fill in the baseline survey. Others are asked to do this within 2 days (to

[1] *Encouraged*: Fruits, vegetables, legumes, whole grains, nuts and seeds, milk, fiber, calcium, seafood omega-3 fatty acids, polyunsaturated fatty acids; *Discouraged*: red meat, processed meat, sugar-sweetened beverages, trans fatty acids, sodium.

prevent overlap with app usage). We will measure several socio-demographics, food literacy, healthy and sustainable eating motivations, healthy eating behavior, and potential other moderators such as a variety of vulnerability dimensions.

5. *Randomization and intervention*. Baseline participants are randomized into two groups, using simple randomization:
    i. Treatment group: This group is immediately asked to test the app for a month, receiving regular reminders via notifications and e-mail.
    ii. Waitlisted control group: This group is asked to test the app starting a month later ('phase 2'), receiving instructions at a later stage.
    iii. Meanwhile, the non-randomized **Study 1b** participants (people at risk of vulnerability, who received the onboarding session) continue life as usual while testing the app.
6. *Endline survey*. All participants receive a survey after the first month. Treatment participants receive additional multiple choice and open-text questions about the app.
7. *Waitlisted group participates*. The waitlisted control group receives instructions to test the app for a month.
8. *Follow-up survey*. After the test month of the waitlisted participants, everyone receives another survey to track the main outcomes. The waitlisted group now receives the same questions about the app.
9. *Short follow-ups*. Through very short follow-ups 3 and 6 months after the initial intervention with only the main outcomes, we track longer-term effects of the intervention.

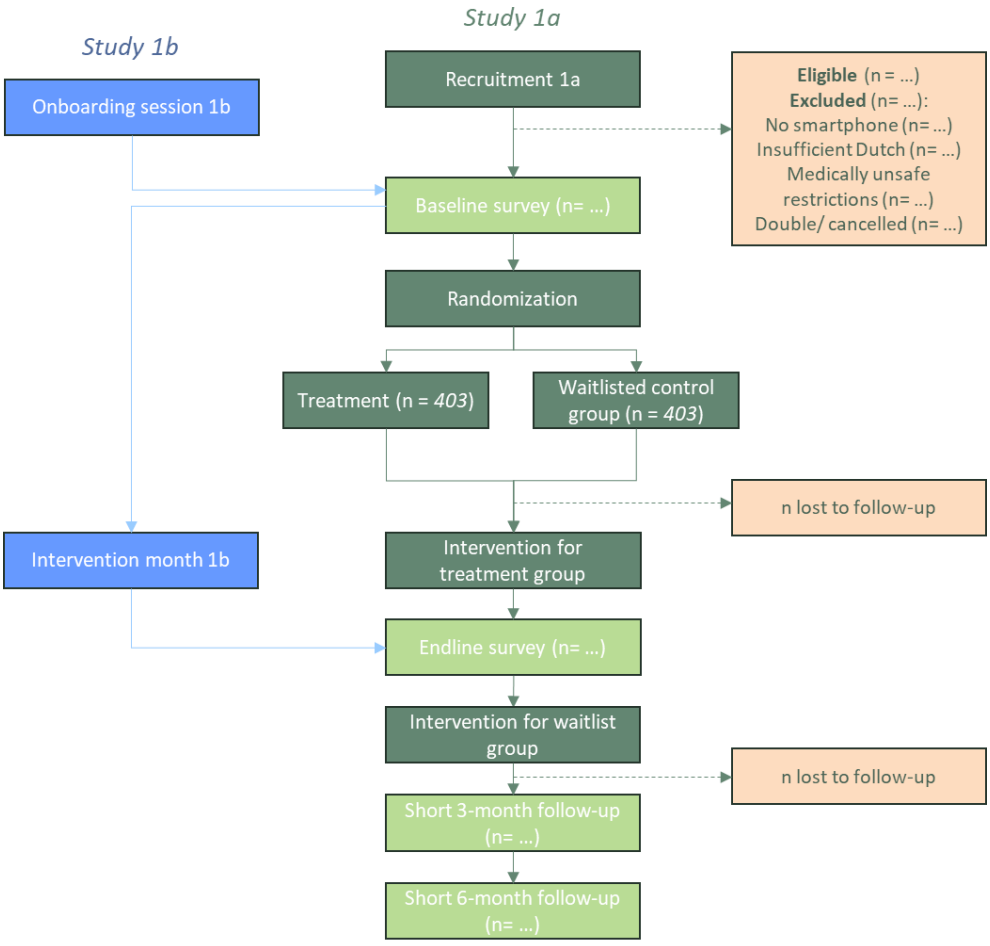


**Figure 1**. CONSORT flow diagram of design

### 2.1.3 SMART Goals

- **Specific:** Roll out MyRecipeWatch in Flanders; evaluate short and longer-term impact on food literacy and healthy eating behaviors; assess heterogeneity in effects; test nudges for recipe acceptance; assess and stimulate use among vulnerable groups; optimize dissemination via VIGL networks.
- **Measurable:**
    - *Key Performance Indicators (KPIs) outlined by the project:*
        - **Reach:** ≥ 5,000 users (downloads/registrations) (KPI #21) (1/3rd goal for Belgium: ~1700 users).
        - **Literacy improvement:** ≥ **15% improvement** in food/cooking literacy score of vulnerable people (KPI #32).
        - **Vulnerable group reach: At least 100** users from vulnerable groups recruited via intermediaries (KPI #33).
        - **Typical recipe corrections:** ≥ **45 corrected typical recipes** by citizens (number of typical meals) (KPI #23).
        - **Compliance with corrections: 15–30% of users** comply with suggested MyRecipeWatch corrections (recommendation acceptance) (KPI #22).
    - *Study impact measurement (hypotheses)*
        - Expected effect size of at least **0.20 standard deviations** for the app itself right after the intervention, and **0.15** at follow-ups (see also power calculation in Section 2.1.4).
        - 15% improvement in literacy and behavior (based on KPI #32).
- **Achievable:** Our goals will be achieved through VIGL's broad network and partnerships, the simple, non-invasive design of the app, nudges/reminders, and targeted support for vulnerable groups.
- **Relevant:** The pilot study aligns with DietWise objectives for healthy and sustainable food choices and digital tool adoption and encourages healthy eating and food literacy.
- **Time-bound:** Pre-testing and pre-pilot activities take place during **M19–M34** (May 2026 – September 2027) with measurement activities at baseline, endline, and 1 month, 3 months and 6 months post-intervention. After thorough pilot-testing and finetuning of the app and ethical procedures from May-September, we aim to start recruitment end September 2026, moving on to the intervention in October-November 2026 (for the treatment group) and November-December 2026 (for the waitlisted group). Follow-ups take place approximately in December 2026, March, and June 2027.

### 2.1.4 Target sample

The target sample differs depending on the perspective: (1) from the perspective of the studies, Study 1a targets the general population of adults, with an oversampling of individuals at risk of vulnerability, and Study 1b targets individuals at risk of vulnerability through VIGL's intermediaries and partners; (2) from the perspective of wider dissemination, VIGL will roll out the app to the wider population, without study recruitment, after Study 1a and 1b have taken place.

Participants will be recruited for Study 1a mainly via advertisements, starting from VIGL's large network in Flanders. The advertisements will directly ask people whether they want to test a new app for a month, with transparency about the number of surveys. To incorporate the waitlisted design, we alert participants that there will be several phases of testing, so they may only be asked to test the app in one month. Study 1b participants will be recruited through the physical onboarding session, organized via VIGL's intermediaries (e.g., community workers and NGOs). If the app turns out to be effective and acceptable by the study users, VIGL will use more general, non-study-focused, advertisements for wider dissemination after the study period.

**Table 2:** Target samples

| Population Group | Number |
|---|---|
| **Participants Study 1a** | At least 806 |
| **Vulnerability risk individuals Study 1a** | 300 (out of the 806 above) |
| **(Vulnerability risk) participants Study 1b** | At least 100 |
| **Subsample for qualitative follow-up** | 50 |

**Note on groups with a risk of vulnerability:** Specific target groups are to be finalized (e.g., low SES, low health literacy, migrants/newcomers, older adults, people with chronic conditions) and will be based on the Vulnerability Construct as defined by WP2-T2.4.

To calculate the sample size needed for the main study (1a), we conducted a power analysis. We use G*power to calculate the required sample size to compare group means, based on the primary hypotheses, expected effect sizes, expected explained variance by control variables, desired significance level, and desired power (0.80). Previous randomized controlled trials studying ingredient swap suggestions in an online supermarket setting found large effect sizes of around 0.50 standard deviations on nutritional quality (Jansen et al., 2021; Schruff-Lim et al., 2024). Moreover, meta-analyses of healthy eating interventions in general and digital healthy eating interventions find effect sizes of around 0.30 standard deviations (Michie et al., 2009; Vanwinkelen et al., 2026). Because MyRecipeWatch has a less controlled and more voluntary nature than many other intervention settings, i.e., people decide when and how to use the app themselves in their daily lives, we assume the smaller effect size of 0.20 standard deviations on healthy eating behavior for the main comparison between the treatment and waitlisted control group (**H1**). Our power analysis is based on healthy eating behavior, as the intervention is expected to have smaller effects on behavior than on cognitive characteristics like food literacy. To calculate the minimum sample size, only the smallest effect size is relevant as detecting smaller effect sizes requires larger samples. For the paired t-tests, we conservatively assume a correlation of 0.50 between timepoints, based on studies that measured food behavior (Parr et al., 2006).

Table 3 shows the required sample sizes for the different study scenarios, based on reaching a power of 0.80, and a significance level of 0.05 for each comparison.

**Table 3:** Power analysis and required sample size Study 1a

| Comparison | Test | Effect size | Sign. level | Required sample | Total per design (including 30% attrition barrier) |
|---|---|---|---|---|---|
| **Design: treatment and waitlisted control group (two arms)** | | | | | |
| **Treatment versus control (H1-2)** | T-test of independent means | 0.20 | 0.05 | 620 (310 per group) | **806**[2] (403 per group) |
| **Baseline to follow-ups within-person comparison of treatment group** | Paired t-test | 0.15[1] | 0.05 | 277 (in treatment group with two measurements) | |

[1] In the case of Paired t-tests between repeated measurements, we use Cohen's $d_z$ rather than Cohen's d, which incorporates the expected correlation between measurements (here r=0.50 is assumed (Parr et al., 2006)). Cohen's $d_z = d / \sqrt{(2(1-r))}$. [2] Includes a 30-percent attrition barrier (1.30 x 620).

App development projects face an inherent risk that development takes longer than expected. For example, the AI algorithm may not yet give accurate or appropriate recommendations in all cases. If pre-testing in May-July 2026 shows that the app needs more work before wide dissemination, we will refrain from testing the app among 800 individuals, and instead reshift the focus from behavioral effectiveness to app usability and user experience among a smaller sample (e.g., 200 people).

### 2.1.5 Inclusion/Exclusion Criteria

Upon registration, potential participants will fill in the baseline survey. The first questions will be about the eligibility requirements. If one of the requirements is not fulfilled, participants will be redirected to the end page with a message about ineligibility. Citizens will be recruited based on the following criteria (Table 4).

**Table 4:** Inclusion & exclusion criteria for citizens' recruitment

| Criteria | Inclusion Criteria | Exclusion Criteria |
|---|---|---|
| **Age** | ≥ 18 years old | < 18 years old |
| **Geography** | Living in Flanders (Belgium) | Living outside Flanders (Belgium) |
| **Language** | Able to use app and complete surveys in Dutch | Insufficient Dutch proficiency |
| **Technology** | Access to smartphone | No smartphone access |
| **Dietary characteristics** | No medical condition or dietary restriction that would make generic ingredient swap suggestions inappropriate or unsafe. | Medical condition or dietary restriction that would make generic ingredient swap suggestions inappropriate or unsafe. |

### 2.1.6 Data Collection, Types & Measures

The main data is collected through 2 main surveys and 3 follow-up surveys: a baseline survey at registration, and endline survey right after the intervention month, a follow-up survey 1 month after the intervention, and follow-ups 3 and 6 months after the intervention via shortened versions of the survey. The latest version of the questionnaire can be found in Annex A.1. The questionnaire is subject to change throughout the pre-testing and pre-piloting phases. Qualtrics links are sent to participants via email and WhatsApp in case a phone number is available. Participants can win small monetary incentives to encourage participation, based on a random selection of a registered email address in our pool of testers. We will send several reminders to respond to the survey. If response rates remain lower than 80%, we will use intensive tracking of a randomly selected non-responder sample (Millán & Macours, 2017).

App usage data is collected via the app software. Activities such as opening the app, entering a recipe, and accepting ingredient swap suggestions are recorded automatically and linked to user IDs. Email addresses will serve as common identifiers to merge survey and app usage datasets.

Table 5 presents a list of key variables and other data collected. The outcome variable *Healthy eating behavior* is measured through the Rapid Prime Diet Quality Score Screener (rPDQS), a validated 13-question food-frequency screener (Kronsteiner-Gicevic et al., 2023). The screener measures the frequency of 13 consumed product types in the past month on a 5-point Likert scale: 'Less than once per week', 'Once per week', '2-4 times per week', 'Nearly daily or daily', and 'Twice per day or more'. Examples were adapted to the Belgian context and the local nutritional guidelines. The 13 food groups largely align with the Global Burden of Disease Dietary Risk factors (Afshin et al., 2019) used as a guideline for developing MyRecipeWatch: *discouraged*- processed meats, red meats, full-fat dairy products, fast food or takeout, sweets and desserts, sugar-sweetened beverages, and white-grain products; *encouraged*- fish, whole-grain products, legumes, vegetables, fruits, and nuts. As in Belgium and according to the GBD, full-fat dairy is not discouraged, but dairy in general is rather encouraged, we changed the full-fat dairy item to 'milk products' and reversed the scoring. We use the same two scoring systems as Kronsteiner-Gicevic et al. (2023). One is a traffic light scoring system that accounts for the recommended frequencies of consuming a product. For example, fruits and vegetables only get the most optimal score if consumed twice or more per day, while fish already receives the most optimal score from 2-4 times per week. As

a secondary check, we also use the other system, which is a numeric count of the frequencies from 0 to 4 for each food group and reversed for the discouraged items.

The second key outcome measure, *Food literacy,* is assessed using a self-developed food knowledge questionnaire, inspired by Section 1 of the General Nutrition Knowledge Questionnaire (Kliemann et al., 2016), and based on the Global Burden of Disease dietary risk factors and Belgian nutritional guidelines. Participants will be asked whether national nutritional guidelines recommend to eat more, the same or less of 14 food products, for example wholegrain, vegetables, salt, legumes and sugar-sweetened beverages.

**Table 5**: List of data types, measurements and methods - MyRecipeWatch pilot

| Data Category | Data Type/Variables | Collection mode |
|---|---|---|
| **Key outcome variables** | • Healthy eating behavior<br>• Food literacy | Questionnaire |
| **Socio-demographics and vulnerability dimensions** | • Age<br>• Gender<br>• Country of birth<br>• Household situation<br>• Education<br>• Occupational status<br>• Ethnic background<br>• Living area (urban/rural)<br>• Region<br>• Household income<br>• Food security | Questionnaire |
| **Other food-related questions** | • Healthy eating motivation<br>• Sustainable eating motivation<br>• Ambivalence towards sustainable/ healthy dietary change<br>• Intention to use recipe improvement tools<br>• Frequency of home-cooking<br>• Frequency of using recipes | Questionnaire |
| **User experience & satisfaction** | • Frequency of using MyRecipeWatch<br>• Perceived usefulness, necessity and informativeness<br>• Ease of use<br>• Accuracy/ appropriateness of suggestions<br>• Satisfaction<br>• Likelihood to recommend<br>• Qualitative feedback | Questionnaire |
| **Usage Data** | • App downloads<br>• Unique users<br>• Recipe entries<br>• Acceptance occurrences<br>• User time and activity (e.g., clicks on the info button) | App data |

### 2.1.7 Evaluation & Analysis Plan

The evaluation consists of three components: evaluating the KPIs, analyzing the effectiveness (testing the hypotheses) of the intervention compared to the control group and over time, and evaluating the implementation and app itself.

Pilot outcomes will be evaluated against the following list of relevant KPIs (Table 6).

**Table 6:** List of relevant KPIs and evaluation details - MyRecipeWatch pilot

| # KPI | KPI Description | Evaluation Measure | Data source |
|---|---|---|---|
| **21*** | ≥ 5000 users testing MyRecipeWatch | # registered users or #downloads or #unique logins | MyRecipeWatch in-app metrics |
| **22** | 15-30% of users comply with the suggested MyRecipeWatch-corrections | # acceptance rate | MyRecipeWatch in-app metrics |
| **23*** | ≥45 corrected recipes by citizens | # entered meals | MyRecipeWatch in-app metrics |
| **32** | ≥ 15% improvement in food/cooking literacy score of vulnerable people | % change in food literacy score | Survey question |
| **33** | Reach at least 100 vulnerable users | # participants with a risk of vulnerability in the onboarding sessions | Survey question |

**KPI target shared across pilots in GR, BE, LT*

In terms of hypothesis-testing, we adhere to an Intention-to-treat (ITT) approach, meaning that all participants are analyzed regardless of whether they complied with our request to use the app or not. This leads to a more policy-relevant conclusion, as the effect will reflect that of a real-life campaign to advertise such an app. The analysis will proceed through the following steps.

1. *Balance check.* Directly after randomization, the experimental groups are compared in terms of key outcomes (food literacy and healthy eating behavior) using ANOVA. If randomization was successful, the groups should not significantly differ.
2. *Selective attrition checks.* Experiments with multiple measurements commonly lose participants over time (attrition). This can influence the comparability of the groups if attrition is non-random and the group characteristics are no longer balanced. This is checked after each measurement by regressing a binary responder versus non-responder variable on the main characteristics. If responders differ significantly from non-responders on certain characteristics, we will further analyze heterogeneity based on those characteristics. Furthermore, several robustness checks will be conducted: reweighting the results using an intensely tracked random sample of non-responders and calculating Lee bounds.
3. *Propensity score matching.* If enough people participate in the onboarding sessions (e.g., >200), we can ensure comparability between the treated group in Study 1b and the external control group from the concurrent Study 1a, by applying propensity score matching (PSM). Propensity scores are estimated using a logistic regression model including all relevant baseline covariates. Each treated participant is matched to a control participant using nearest-neighbor matching. After matching, covariate balance between groups will be assessed using standardized mean differences. We repeat this process to identify a matched sample in both the control group and the treatment group of Study 1a.
4. *Main analysis at endline.* To compare the treatment group to the control group we regress the two primary outcome variables on experimental group and relevant control variables. The baseline value of the outcome, as well as socio-demographic characteristics and healthy and sustainable eating motivations are included to boost power (the more variance is explained, the less participants are needed to detect smaller effects). Both standardized effect sizes and percentage change are calculated to more easily interpret the results. For Study

1b, the group that received the onboarding session will be compared to the matched control group through regression analysis, or analyzed through within-person change if a matched control group is not feasible.

5. *Longer-term analysis of follow-ups.* To evaluate within-person change in the outcomes from endline to follow-up, we will estimate linear mixed-effects models with Time as a fixed effect and a random intercept for participants.
6. *Heterogeneity.* To test heterogeneity based on specific characteristics, we add interaction effects with experimental group to the regression analyses mentioned in point 4. Significant interactions are also visualized to further examine differences.
7. *Behavioral nudges.* To compare the effect of the different nudges on recommendation acceptance, we use a mixed-effects logistic regression with a random intercept for each participant to account for the nested data structure (recommendations within individuals) using all recipe recommendation data (i.e., one row per recommendation across both treatment and waitlisted control group app users). The binary acceptance variable is regressed on message condition and control variables.

Finally, the implementation and app will be evaluated through several metrics and qualitative feedback:

- Reach: # downloads, registrations, demographic representativeness.
- Adoption: % users started using the app in the week after registration.
- Retention: % users still active after 2 weeks, a month, 2 months, 3 months and 6 months
- Acceptability: user satisfaction, qualitative feedback.
- Reach and experience of vulnerable groups: # individuals with risk of vulnerability reached, qualitative feedback.

## 2.2 Foodbag Pilot

Foodbag (FOOD) is a company that delivers food boxes with recipes. Customers choose a weekly menu out of more than 35 recipes. Their ordering system is not one of DietWise's digital innovations but provides a useful platform to improve individuals' cooking behavior, for instance through ingredient swap suggestions, contributing to DietWise's objectives.

### *2.2.1 Research questions and hypotheses*

The Foodbag pilot, implemented by FOOD, will test the implementation of healthy and meat-to-vegetarian ingredient swap recommendations in their mobile and web-based ordering application. The main research question will be: *Do healthy and meat-to-vegetarian swap suggestions in a meal box ordering system (app/ web-based) improve healthy and meatless meal choice, food literacy and healthy eating behavior?*

The study will focus on the effectiveness of the ingredient swap intervention, heterogeneity in terms of this effectiveness, and Foodbag-specific hypotheses on ordering system usage and retention. Table 7 displays the hypotheses we plan to test.

**Table 7**: Study 2 research questions and hypotheses

| *Goal* | *Hypothesis* |
|---|---|
| **Study 1a** | |
| **Main research question:** ***Do healthy and meat-to-vegetarian swap suggestions in a meal box ordering system (app/ web-based) improve healthy and meatless meal choice, food literacy and healthy eating behavior?*** | |
| **1. Effectiveness** | **H1:** Receiving ingredient swap suggestions in the Foodbag ordering system significantly improves healthy meal choices compared to a control group. |
| | **H2:** Receiving ingredient swap suggestions in the Foodbag ordering system significantly improves meatless meal choices compared to a control group. |
| | **H3:** Receiving ingredient swap suggestions in the Foodbag ordering system significantly improves food literacy compared to a control group. |
| | **H4:** Receiving ingredient swap suggestions in the Foodbag ordering system significantly improves healthy eating behavior compared to a control group. |
| **2. Heterogeneity** | **H5:** Motivation for healthy eating, motivation for sustainable eating, vulnerability dimensions, gender and order frequency moderate the effectiveness of the ingredient swap suggestions. |
| | **H6:** Motivation for healthy eating, motivation for sustainable eating, vulnerability dimensions, gender and order frequency moderate the rate of acceptance of the ingredient swap suggestions. |
| **3. Order application usage** | **H7:** Receiving ingredient swap suggestions does not influence the retention rate (% that keeps using Foodbag) of participants. |
| | **H8:** There is a difference in effect between web application and mobile application users. |

### *2.2.2 Study Design*

Foodbag (FOOD) is a company that delivers food boxes with recipes. Customers choose a weekly menu out of more than 35 recipes. They can hold a weekly or biweekly membership, or place one-time orders. While they generally aim to offer healthy meals and already have some health-directed nudges (e.g., labeling meals with '50% vegetables'), some ingredients could be improved to be more in line with nutritional guidelines (e.g., less red meat, more whole grain

products). Therefore, they plan to introduce 'ingredient swaps' in their order applications (web and mobile). This means that when people select a certain meal, they receive suggestions to substitute specific ingredients in two ways:

1. Healthy ingredient substitutions (e.g., refined grains → whole grains)
2. Meat-to-vegetarian protein swaps (e.g., minced meat → lentils or plant-based alternatives)

Customers can either accept or ignore the swap suggestion.

We test the effectiveness of the ingredient swap suggestions through a randomized, mixed (between-subject and within-subject) design with two groups and repeated measures. The mixed design arises from comparing a treatment to control group (between-subject), as well as accounting for both pre- and post-intervention values (repeated measures). The procedure is as follows:

a. *Pre-test phase:* The implementation of the ingredient swaps will be pre-tested among project partners and nutritional experts, for example in terms of appropriateness. In addition, the entire study will be piloted with a small group of Foodbag customers (~ 20 people) to smooth out potential issues.
b. *Recruitment*: A random sample from Foodbag's clients is invited to participate in the study. They are told that they will receive two questionnaires about eating behavior and perceptions with the goal of getting to know Foodbag's clientele as well as evaluate their meal box approach.
a. *Baseline survey*: The invitation asks participants to fill in the baseline survey. They can also opt to complete the baseline survey within the next week to prevent immediately pushing potential participants away. We measure socio-demographics, food literacy, healthy and sustainable eating motivations, healthy eating behavior, dimensions of vulnerability, and perceptions of Foodbag.
c. *Randomization and intervention*: Participants are randomized into a control or treatment group. The treatment group starts receiving ingredient swap suggestions in the order system. The control group keeps seeing the same order application as before.
d. *Follow-up survey*: After 2 months, both control and treatment participants are asked to fill in another survey. The questions are approximately the same as in the baseline survey, but treatment participants receive additional questions assessing their awareness and perceptions of the ingredient swap suggestions.

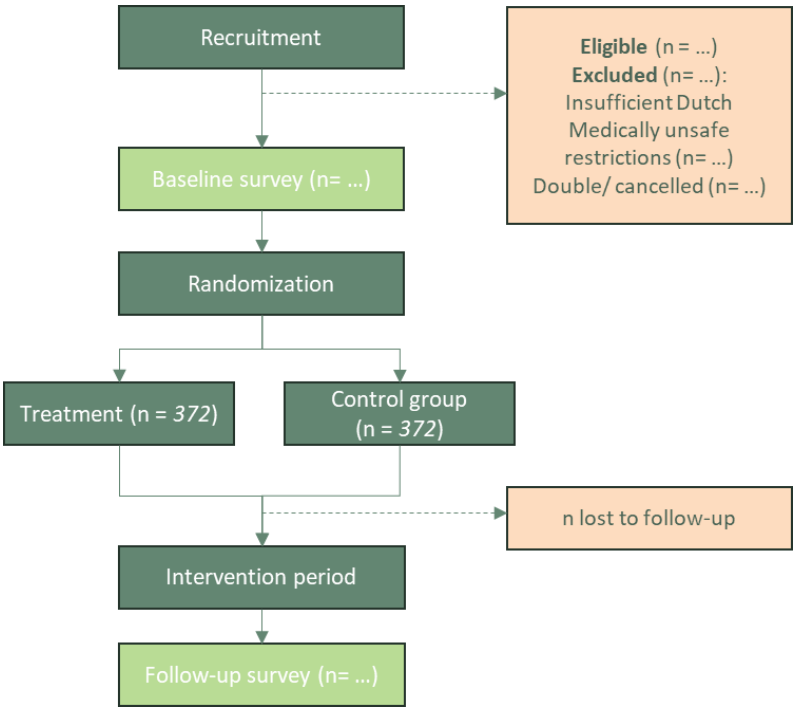


**Figure 2**. CONSORT flow diagram of Study 2: Foodbag pilot

### 2.2.3 SMART Goals

- **Specific:** Implement and test the effects of healthy ingredient swap recommendations in Foodbag meal kit order systems; evaluate heterogeneity in effects; evaluate retention, assess feasibility in routine operations.
- **Measurable:**
    - *Key Performance Indicators (KPIs) (as most project KPIs are about MyRecipeWatch, here we also devised our own additional KPIs):*
        - Study implementation: **>300** subscribed users exposed to swap suggestions
        - Further dissemination upon success: **>9000** subscribed users exposed to swap suggestions
        - Swap uptake rate: **15-30%** of swap suggestions accepted
        - Retention: The number of subscribers remains stable throughout the intervention
        - **> 15%** improvement in food literacy of vulnerable people (KPI #32)
    - *Study impact measurement (hypotheses)*
        - Expected effect size of at least **0.30 standard deviations** for healthy and meatless meal choices
        - Expected effect size of at least **0.20 standard deviations** for food literacy and healthy eating behavior
        - **15%** improvement in literacy and behavior (based on KPI #32)
- **Achievable:** Our goals are achievable through leveraging Foodbag's existing tools and customer base, and the straightforward design of providing suggestions to people that are already making a choice. It is based on a limited set of high-feasibility swaps and simple communication integration.
- **Relevant:** The intervention supports healthier cooking at home via scalable market channels.
- **Time-bound:** Activities will take place from M19-M34 (May 2026 - August 2027). Implementation depends on the speed of integrating the swaps in Foodbag's ordering system. Developing the swap option in the ordering system and planning the logistics will take place from May 2026 to September 2026. After testing the intervention with a small group of people in October and November 2026, recruitment will start in January 2027. The main study takes place between February and May 2027.

### 2.2.4 Target sample

Study participants will be recruited from Foodbag's customer pool, which currently consists of 9.500 subscribed customers. Through emails, subscribed customers (who have previously given permission to contact them for studies) will be invited to the study and informed that the goal is to get to know the customers, and evaluate the order applications, which is done through two surveys. We will also highlight any compensation given for participation. If the intervention turns out to be successful and feasible, all Foodbag customers will be exposed to the swap suggestions.

**Table 8:** Target sample by population group – Foodbag pilot

| Population Group | Number |
|---|---|
| **Study participants (subscribed customers)** | **>744**<br><br>If not feasible, >334 participants with a focus on H1-H2 |
| **Foodbag customers exposed to swap suggestions if effective and feasible** | **>9.000** |

G*power was used to calculate the required sample size to compare group means, based on the primary hypotheses, expected effect sizes, desired significance level, and desired power (0.80). Previous randomized controlled trials studying ingredient swap suggestions in an online supermarket setting found large effect sizes of around 0.50 standard deviations on nutritional quality (Jansen et al., 2021; Schruff-Lim et al., 2024). Because our Foodbag pilot will take place in a less controlled setting than those interventions, we assume the smaller effect size of 0.30 standard deviations on *healthiness of the selected meals* and *proportion of meatless meals* for the main comparison between the treatment and control group (**H1-2**). For overall *healthy eating behavior* and *food literacy* we expect a smaller effect size of 0.20, as those are less directly linked to the intervention (**H3-4**). To calculate the minimum sample size, only the smallest effect size (0.20) is relevant as detecting smaller effect sizes requires larger samples.

Based on reaching a power of 0.80, an effect size of 0.20, and a significance level of 0.05, and an attrition barrier of 20%, we need **744 participants** (372 per group). If we cannot reach this in the designated recruitment period within Foodbag's customer pool, we focus on the sample we need to test hypotheses 1 and 2 (effect size 0.30), and aim for at least **334 participants** (167 per group).

### 2.2.5 Inclusion/Exclusion Criteria

**Target participants** are Foodbag customers and will be recruited based on the following criteria (Table 9).

**Table 9:** Inclusion and exclusion criteria for recruitment

| Criteria | Inclusion | Exclusion |
|---|---|---|
| **Customer status** | Subscribed weekly or biweekly Foodbag customer | Not a Foodbag customer, or a customer placing one-time orders |
| **Language** | Able to complete surveys in Dutch | Insufficient Dutch proficiency |
| **Dietary characteristics** | No medical condition or dietary restriction that would make generic ingredient swap suggestions inappropriate or unsafe. | Medical condition or dietary restriction that would make generic ingredient swap suggestions inappropriate or unsafe. |

### 2.2.6 Data Collection, Types & Measures

The data is collected in two ways. First, Foodbag-app and order data show which meals customers ordered and whether they accepted swap suggestions or not. These activities are recorded automatically and linked to user IDs/ email addresses. Second, survey data will be collected to measure the other outcome variables (food literacy and healthy eating behavior), as well as moderator and control variables (socio-demographics, motivation for healthy eating, vulnerability dimensions). Data collection will take place online through Qualtrics. Qualtrics links are sent to participants via email and WhatsApp in case a phone number is available. Email addresses will serve as common identifiers to merge survey and app usage datasets.

Randomly selected participants will win small incentives (e.g., a Foodbag coupon) to encourage participation. We will send several reminders to respond to the survey. If response rates remain lower than 80%, we will use intensive tracking of a randomly selected non-responder sample.

The primary outcome variables are 'healthiness of selected meals', 'proportion of meatless meals', 'food literacy', and 'healthy eating behavior'.

To measure *Healthiness of selected meals* we use the UK Food Standard Agency (FSA) Nutrient Profiling System (also used as a basis for the Nutri-Score system) (FSA scores) (UK Department of Health & Social Care, 2026). It uses a scoring system based on the nutrient content of foods, including "negative" nutrients such as energy, saturated fat, sugar, and sodium, and "positive" nutrients such as protein, fibre, and percentages of fruits, vegetables, and nuts. The FSA score can range

from −15 (best) to +40 (worst nutritional quality) and is commonly used in academic studies to measure intervention effects on nutritional quality of food choices (e.g., Dubois et al., 2021; Schruff-Lim et al., 2024). With their point system, we assign points to each of Foodbag's meals' ingredients and average the scores across the periods of interest (2 months before the study and 2 months during the study). Accepting healthy and meatless ingredient swaps will lead to more beneficial scores.

*Proportion of meatless meals* is measured by the percentage of meatless meals ordered throughout the 2 months of the intervention period, and the 2 months before the intervention (as a baseline value).

*Healthy eating behavior* and *Food literacy* are measured through the surveys in the same way as in the MyRecipeWatch study (see Section 2.1.6). The same counts for data on socio-demographic information (only baseline survey), the potential moderators and control variables. Customer feedback and evaluation questions related to the ingredient swap suggestions will be adapted to the specific Foodbag context.

**Table 10:** List of data types, measurements and methods – Foodbag pilot

| Data Category | Data Type/Variables | Collection mode |
|---|---|---|
| **Outcome variables** | • Meal choice (healthiness / meatless)<br>• Food literacy<br>• Healthy eating behavior | Ordering system data (meal choice), questionnaire |
| **Socio-demographics and vulnerability dimensions** | • Age<br>• Gender<br>• Country of birth<br>• Household situation<br>• Education<br>• Occupational status<br>• Ethnic background<br>• Living area (urban/rural)<br>• Region<br>• Household income<br>• Food security | Foodbag customer data, questionnaire |
| **Other food-related questions** | • Healthy eating motivation<br>• Sustainable eating motivation<br>• Ambivalence towards sustainable/ healthy dietary change<br>• Intention to use recipe improvement tools<br>• Frequency of home-cooking<br>• Frequency of using recipes<br>• Frequency of ordering Foodbag meals<br>• Motivators of Foodbag meal choices (health, environment, less meat, taste,…) | Questionnaire |
| **Order and swap data** | • Number of orders during pilot period<br>• Order frequency<br>• Subscription<br>• Number of swap offers presented<br>• Number accepted<br>• Swap category (grain/protein)<br>• Clicks/ app usage | Ordering system data |
| **Customer feedback** | • Evaluation of swap suggestions (clarity, usefulness, tasteful, ease of application)<br>• Qualitative feedback | Questionnaire |

### 2.2.7 Evaluation & Analysis Plan

The evaluation consists of three components: evaluating the KPIs, analyzing the effectiveness (testing the hypotheses) of the intervention compared to the control group and over time, and evaluating the feasibility and customer perceptions of the intervention got potential future dissemination.

Descriptive statistics will be used to evaluate the following KPIs, also mentioned in Section 3.2.3:

- Study implementation: **>300** subscribed users exposed to swap suggestions
- Further dissemination upon success: **>9000** subscribed users exposed to swap suggestions
- Swap uptake rate: **15-30%** of swap suggestions accepted
- Retention: The number of subscribers remains stable throughout the intervention
- **> 15%** improvement in food literacy of vulnerable people (KPI #32)

In terms of hypothesis testing, we adhere to an Intention-to-treat (ITT) approach, meaning that all participants are analyzed regardless of how actively they kept ordering Foodbag meals. Participants who did not order anything throughout the study can unfortunately not be included in the final analysis of H1 and 2 as the outcome variables are calculated from order data (see also 'Missing data'). The analysis will proceed as follows.

1. *Balance check.* Directly after randomization, the experimental groups are compared in terms of key outcomes (healthiness of meals, proportion of plant-based meals, food literacy, and healthy eating behavior) using ANOVA. If randomization was successful, the groups should not significantly differ.
2. *Selective attrition checks.* Experiments with multiple measurements commonly lose participants over time (attrition). This can influence the comparability of the groups if attrition is non-random and the group characteristics are no longer balanced. This is checked after the follow-up measurement by regressing a binary responder versus non-responder variable on the main characteristics. If responders differ significantly from non-responders on certain characteristics, we will further analyze heterogeneity based on those characteristics. Furthermore, several robustness checks will be conducted: reweighting the results using an intensely tracked random sample of non-responders and calculating Lee bounds.
3. *Main analysis.* To compare the treatment group to the control group (**H1-4**) we regress the two primary outcome variables on experimental group and relevant control variables. The baseline value of the outcome, as well as socio-demographic characteristics and healthy eating motivations, are included to boost power (the more variance is explained, the less participants are needed to detect smaller effects).
4. *Heterogeneity and order application activity.* To test heterogeneity based on specific characteristics (**H5-6-8**), we add interaction effects with experimental group to the regression analyses mentioned in point 3. Significant interactions are also visualized to further examine differences. **H7** is tested by comparing retention rates between the control and treatment group.

Finally, customer perceptions and feasibility will be evaluated through several metrics and qualitative feedback:

- Retention: # of retained subscribers in treatment group.
- Acceptability: customer satisfaction, qualitative feedback.
- Feasibility: qualitative (experience) and quantitative (time, cost) feedback from Foodbag staff on the difficulty of implementation.

# 3. Stakeholder engagement

To maximize reach and recruitment efforts of both pilots, VIGL and FOOD will engage relevant stakeholders of their target populations respectively. Examples of such stakeholders include but are not limited to the list in Table 11.

**Table 11:** List of relevant stakeholders per pilot component in Belgium

| Pilot | Stakeholders | Role | Network/Source |
|---|---|---|---|
| **MyRecipeWatch** | Dieticians | Testing and evaluating the app, spreading study and/or app among clients | VIGL network |
| **MyRecipeWatch** | Welfare organisations | Promoting the app among clients, co-organizing onboarding sessions | VIGL network |
| **MyRecipeWatch** | Citizens | Testing and evaluating the app | VIGL network |
| **Foodbag** | Citizens | Providing feedback to the ingredient swap feature | Foodbag customers |

# 4. Data Management plan

DietWise's current data management plan is outlined in D1.2.

# 5. Risks & Limitations

Potential risks that could overhaul pilot efforts are relevant for recruitment quotas, intervention design and are presented in (Table 1).

**Table 1:** Risks & limitations and relevant pilot impact

| # | Category | Detailed risk | Pilot Impact | Mitigation Measure | Owner | Risk |
|---|---|---|---|---|---|---|
| **1** | **Technical** | MyRecipeWatch functionality and readiness / implementation of swaps in Foodbag system | Delay of studies; usability. | Early planning and a long period for pilot implementation (May '26 – August '27) which allows for some delays. Back-up pilots in case of delays: Testing MyRecipeWatch with a smaller sample, focusing more on usability and relevance; Testing Foodbag swaps in a hypothetical study. | ICCS/ Foodbag ICT department | medium |
| **2** | **Tool recruitment** | Insufficient reach/downloads for MyRecipeWatch | Study 1a underpowered; KPI #21 not met | Broad multi-channel dissemination; partnerships; (paid) promotion | VIGL | medium |
| **3** | **Study recruitment and attrition** | Low participation and retention in research | Limited pre-post data, underpowered | Incentivize survey completion, minimize survey burden, recruit | VIGL, KUL, FOOD | high |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | for statistical conclusions | for a test of the app so people know what they sign up for, leveraging VIGL network, reminders. | | |
| **4** | **Retention/ engagement** | High app dropout | Insufficient usage data; limited behavior change | Use reminders/ prompts; optimize onboarding; user support; qualitative feedback for improvements | VIGL, KUL, ICCS | high |
| **5** | **Vulnerability risk groups** | Vulnerability risk groups are underrepresented or face barriers (language, access, digital) | Limited equity impact of RW; findings not generalizable to vulnerable populations | Targeted recruitment via intermediaries; assisted onboarding; simplified instructions | VIGL | medium |
| **6** | **Operational feasibility** | Foodbag unable to implement swaps or control group as planned (logistical, time, cost, complexity) | Without control group: confounding; without implementation: study impossible. | Incorporating enough time for technical implementation. Back-up option for lack of control group is pre-post measurement, controlling for confounders. Back-up option for failing to implement entirely is a hypothetical study. | FOOD, KUL, VIGL | medium |
| **7** | **Data sharing & GDPR** | GDPR-compliance | Privacy breach | Obtaining ethical approval, obtaining informed consent, and establishing a data sharing agreement. | FOOD, KUL, VIGL | low |
| **8** | **External factors** | Seasonal effects, competing campaigns, economic factors | Confounding | Randomize participants and have a control group to minimize risk. Measure & adjust for known confounders; qualitative data to contextualize findings. | KUL, VIGL | low |
| **9** | **Algorithm bias / inappropriateness** | MyRecipeWatch AI might suggest swaps that are culturally insensitive, not feasible or relevant for the recipe | Reduced trust and high dropout (among specific groups) | Pre-test AI-generated swaps before full rollout. Flagging system after rollout. | ICCS | medium |
| **10** | **Timeline** | Pilots not finalized in time | Not reaching deliverable/ KPI deadlines | Long period/ buffer time to implement pilots. Early start by defining protocol plans. | KUL | low |

# 6. Ethics

A separate ethics submission will be filed for each study to the Social and Societal Ethics Committee of KU Leuven. All pilot activities will commence once ethical approvals have been acquired.

We will ensure that all participants are properly informed about the purpose of the study and apps, what data will be collected, how that data will be used, and who will have access to it. Participation will be entirely voluntary, and participants must be given the opportunity to provide explicit written consent before engaging with the study or app. Such consent for the research will be embedded in the online questionnaire, and for the apps it is embedded in registration (i.e., making an account).

The research includes vulnerable research participants, so paying close attention to ethical considerations is crucial. As the questionnaires may ask some sensitive questions about topics people may need help with (i.e., food insecurity), the questionnaire will include referral information to relevant local services and organizations. Moreover, participants have the option to answer sensitive questions with 'I'd rather not answer this question'.

Participants retain the right to withdraw from the pilot or app at any time and without consequence, and any data will be deleted upon request. All participant data will be stored securely, anonymized wherever possible, and used solely for the purpose of evaluating the pilot. Raw data will not be shared with third parties, unless it has a specific purpose and is agreed upon by the participant. We will have a clear process in place for participants to raise concerns or report negative experiences during the pilot.

For a more thorough explanation of data management and sharing practices, we refer to DietWise's Data Management Plan. In addition to these standards, DietWise's coordination team is obtaining advise from an external ethical advisor, who will give specific ethics feedback to the pilot plans.

# 7. Reporting & Dissemination

Results and outputs will be summarized and presented as part of the public Deliverable D7.2 – "The pilot operationalization in Belgium".

In terms of academic dissemination, we aim to publish each study in a Q1-journal in the behavioral science or food consumption field, such as Appetite or the Journal of Environmental Psychology. Furthermore, the results will be presented at international conferences through oral or poster presentations.

VIGL, FOOD and KUL will also spread the results through other channels, such as press releases and posts on social media and websites.

# 8. Conclusion

This report outlined the plans for pilot studies in Belgium within the DietWise project, which will be implemented by KU Leuven, Vlaams Instituut Gezond Leven and Foodbag from May 2026 until August 2027. The overarching goal is fostering food literacy and knowledge of nutritional guidelines, as well as improving healthy cooking and eating behavior. Two pilot studies will help people improve their cooking behavior through ingredient swap suggestions:

- A stand-alone app developed in the DietWise project that provides ingredient swap recommendations to recipe weblinks entered by citizens (MyRecipeWatch) (Study 1).
- Ingredient swap recommendations within a meal kit ordering system (Foodbag) (Study 2).

Both will be tested against a control group through a randomized controlled trial design. In addition, a sub-study of Study 1 implements an onboarding session for people at risk of vulnerability and analyzes whether this helps increasing the reach and reducing the barriers for vulnerability risk groups.

All studies make use of questionnaires implemented at least before and directly after the intervention, measuring socio-demographics, food perceptions, food literacy, and food behavior. Directly after the intervention, the intervention itself will also be evaluated through evaluation questions and open-text boxes.

Throughout the process, stakeholders will be involved. In particular, dietitians will be engaged in pre-testing the app and recruiting their clients for Study 1 and app dissemination in general. Non-profit organizations will also be involved to co-organize the onboarding sessions and reach more vulnerability risk groups.

All studies will be analyzed using an Intention-to-Treat approach and regression analyses comparing control to treatment groups, supplemented by pre-post analysis where control-treatment comparisons are not possible. Heterogeneity will be analyzed to identify more specific target groups, in terms of motivation, socio-demographics, and vulnerability dimensions. In addition, KPIs and interventions will be descriptively and qualitatively analyzed.

Next steps are obtaining ethics approval and starting the pre-testing phase in May-September 2026, followed by the actual pilot studies in October 2026-March 2027. Final reporting and paper submissions will take place from April to August 2027.

# 9. References

Afshin, A., Sur, P. J., Fay, K. A., Cornaby, L., Ferrara, G., Salama, J. S., Mullany, E. C., Abate, K. H., Abbafati, C., & Abebe, Z. (2019). Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *393*(10184), 1958–1972.

Jansen, L., van Kleef, E., & Van Loo, E. J. (2021). The use of food swaps to encourage healthier online food choices: a randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity*, *18*(1), 156.

Kliemann, N., Wardle, J., Johnson, F., & Croker, H. (2016). Reliability and validity of a revised version of the General Nutrition Knowledge Questionnaire. *European Journal of Clinical Nutrition*, *70*(10), 1174–1180.

Kronsteiner-Gicevic, S., Tello, M., Lincoln, L. E., Kondo, J. K., Naidoo, U., Fung, T. T., Willett, W. C., & Thorndike, A. N. (2023). Validation of the rapid prime diet quality score screener (rPDQS), a brief dietary assessment tool with simple traffic light scoring. *Journal of the Academy of Nutrition and Dietetics*, *123*(11), 1541–1554.

Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychology*, *28*(6), 690.

Millán, T. M., & Macours, K. (2017). *Attrition in randomized control trials: Using tracking information to correct bias*. SSRN.

Parr, C. L., Veierød, M. B., Laake, P., Lund, E., & Hjartåker, A. (2006). Test-retest reproducibility of a food frequency questionnaire (FFQ) and estimated effects on disease risk in the Norwegian Women and Cancer Study (NOWAC). *Nutrition Journal*, *5*(1), 4.

Schruff-Lim, E.-M., Van Loo, E. J., van der Lans, I. A., & van Trijp, H. C. M. (2024). Impact of food swap recommendations on dietary choices in an online supermarket: A randomized controlled trial. *Appetite*, *194*, 107158.

Sciensano. (2023). *Food consumption survey 2022-2023: Food consumption and adherence to dietary guidelines in the Belgian population*.

Vanwinkelen, K., Spruyt, B., & Smits, T. (2026). Digital Interventions Targeting Healthy and Sustainable Eating Behavior: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, *28*, e80821.

# ANNEX

## A1. Pilot Questionnaire (Dutch)

**Codes:**

**[ALL]** = MyRecipeWatch-pilot & Foodbag pilot

**[RW]** = enkel MyRecipeWatch-pilot

**[FOOD]** = enkel Foodbag-pilot

*[A more complete information sheet and consent form will be added at a later stage]*

Beste respondent,

Fijn dat je even de tijd neemt om deze enquête in te vullen! Met jouw hulp willen we MyRecipeWatch verder verbeteren.

In deze vragenlijst vragen we naar jouw mening en ervaringen met de app. Jouw feedback is ontzettend waardevol en helpt ons om MyRecipeWatch gebruiksvriendelijker en beter afgestemd op de noden van gebruikers te maken.

De enquête is volledig anoniem en je antwoorden worden uitsluitend gebruikt voor onderzoeksdoeleinden. Niets is terug te leiden naar jou.

Alvast heel erg bedankt voor je tijd en voor je bijdrage aan dit onderzoek!

**Sociodemografische kenmerken**

1. Gender **[ALL]**
- ☐ Man
- ☐ Vrouw
- ☐ Non-binair
- ☐ Anders
- ☐ Dit zeg ik liever niet

2. Leefijd **[ALL]**

3. Geboorteland: **[ALL]**
- ☐ België
- ☐ Andere (specifieer)

4. Huidige woonsituatie: **[ALL]**
- ☐ Alleenwonend
- ☐ Met partner zonder kinderen
- ☐ Met partner met kinderen
- ☐ Zonder partner met kinderen
- ☐ Met huisgenoten

- ☐ Andere
- ☐ Dit zeg ik liever niet

5. Hoeveel kinderen wonen er in je huishouden? **[ALL]**
- ☐ 0
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 of meer

6. Hoe zou je je woonomgeving omschrijven?
   - ☐ Stedelijk gebied
   - ☐ Randstedelijk gebied
   - ☐ Landelijk gebied

7. In welke provincie woon je?
   *Dropdown menu van Vlaamse provincies (Antwerpen, Limburg, Oost-Vlaanderen, West-Vlaanderen, Vlaams-Brabant)*

8. Hoogst behaalde opleidingsniveau: **[ALL]**
   - ☐ Minder dan basisonderwijs
   - ☐ Basisonderwijs
   - ☐ Secundair onderwijs
   - ☐ Bachelor
   - ☐ Master
   - ☐ Doctoraat

9. Welk beroepsstatuut heb je? **[ALL]**
   - ☐ Student
   - ☐ Zelfstandige
   - ☐ Bediende / ambtenaar
   - ☐ Arbeider
   - ☐ Gepensioneerd
   - ☐ Werkloos
   - ☐ Huisman/huisvrouw
   - ☐ Andere

10. Wat is je etnische achtergrond? **[ALL]**
    - ☐ Belgisch
    - ☐ Italiaans
    - ☐ Marokkaans
    - ☐ Frans
    - ☐ Turks
    - ☐ Nederlands
    - ☐ Ander Europees
    - ☐ Ander niet-Europees
    - ☐ Dit zeg ik liever niet

11. Wat is het gezamenlijk gemiddeld maandelijks netto inkomen van je huishouden? (Studenten: ga uit van je maandelijks beschikbaar budget). **[ALL]**

- ☐ Minder dan 500 EUR
- ☐ 501 EUR tot 1500 EUR
- ☐ 1501 EUR tot 2500 EUR
- ☐ 2501 EUR tot 3500 EUR
- ☐ 3501 EUR tot 4500 EUR
- ☐ 4501 EUR tot 5500 EUR
- ☐ 5501 EUR of meer
- ☐ Dit zeg ik liever niet

**Voedselgerelateerde vragen**

12. **Eetgewoontes** - Denk aan de afgelopen maand: hoe vaak at of dronk je volgende voedingsmiddelen? Kies één antwoord uit de volgende opties. **[ALL]**

| **rPDQS - België** | | | | | | |
|---|---|---|---|---|---|---|
| **Vragen** | | **Antwoordopties** | | | | |
| **Nr.** | | **Minder dan één keer per week** | **Eén keer per week** | **2–4 keer per week** | **Bijna dagelijks of dagelijks** | **Twee keer per dag of meer** |
| V1 | Bewerkt vlees (gehakt, worst, spek, charcuterie, …) | | | | | |
| V2 | Rood vlees (lam, rund, varken) | | | | | |
| V3 | Vis (kabeljauw, tonijn, zalm, …) | | | | | |
| V4 | Zuivel (melk, yoghurt, kaas, …) | | | | | |
| V5 | Fastfood (pizza, gefrituurde gerechten, kant-en-klaarmaaltijden, …) | | | | | |
| V6 | Suikerhoudende dranken (frisdrank, energiedranken, gezoete dranken, …) | | | | | |
| V7 | Witte graanproducten (witbrood, witte pasta, witte rijst, ...) | | | | | |
| V8 | Volle graanproducten (volkorenbrood, volkoren pasta, volle rijst, ...) | | | | | |
| V9 | Zoete snacks en desserten (koeken, gebak, snoep, ijs, …) | | | | | |
| V10 | Peulvruchten (bonen, kikkererwten, linzen, …) | | | | | |
| V11 | Groenten | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| V12 | Fruit | | | | | |
| V13 | Noten en zaden | | | | | |

13. Hoe gemotiveerd ben je om gezond te koken? [1 = helemaal niet gemotiveerd; 7 = heel gemotiveerd] **[ALL]**
14. Hoe gemotiveerd ben je om milieuverantwoord te koken? [1 = helemaal niet gemotiveerd; 7 = heel gemotiveerd] **[ALL]**

15. In de laatste 12 maanden, was er een moment dat je, door gebrek aan geld of andere middelen: **(Schaal: ja, nee, ik weet het niet, ik zeg dit liever niet) [ALL]**

a) Bezorgd was dat je niet voldoende zou hebben om te eten?
b) Niet in de mogelijkheid was om gezond en voedzaam te eten?
c) Je maar een select aantal voedingsmiddelen kon eten?
d) Je een maaltijd moest overslaan?
e) Je minder had gegeten dan je eigenlijk zou moeten?
f) Je huishouden geen eten meer had?
g) Je honger had maar niet hebt gegeten?
h) Je een hele dag niet hebt gegeten?

16. **Voedselvaardigheden (kennis) [ALL]** - Volgens de nationale voedingsrichtlijnen, wat is de ideale inname voor volgende voedingsmiddelen? Kies een van volgende antwoorden:

- Aanmoedigen: Dit voedingsmiddel is goed voor je gezondheid, de richtlijnen raden aan om hier meer van te eten.
- Matigen: Dit voedingsmiddel heeft een neutraal of onvoldoende bewezen effect; het past in een gezond eetpatroon, als het niet in overvloed wordt gegeten.
- Beperken: dit voedingsmiddelen is gelinkt aan het risico op ziekte. De voedingsrichtlijnen raden aan om hier zo weinig mogelijk van te eten.
- Ik weet het niet: ik weet de aanbeveling van dit voedingsmiddel niet.

| Voedingsmiddel | Aanmoedigen (meer eten) | Matigen | Beperken (minder eten) | Ik weet het niet |
|---|---|---|---|---|
| **Volle granen** | | | | |
| **Groenten** | | | | |
| **Vette vis** | | | | |
| **Plantaardige oliën** | | | | |
| **Noten en zaden** | | | | |
| **Melk(producten)** | | | | |
| **Peulvruchten** | | | | |
| **Fruit** | | | | |
| **Vezels** | | | | |
| **Gefrituurd eten en gebak** | | | | |
| **Suikerhoudende dranken** | | | | |
| **Zout** | | | | |
| **Rood vlees** | | | | |

| | | | | |
|---|---|---|---|---|
| **Bewerkt vlees** | | | | |

**Gevoelens rond duurzaam koken [ALL]**

17. Geef aan in hoeverre de volgende woorden je gevoelens beschrijven ten opzichte van het veranderen van je kookgewoonten om ze **gezonder** te maken (schaal van 0-5, waarbij '0' staat voor "ik heb dit gevoel niet" en '5' staat voor "ik heb dit gevoel heel erg":
    - Tegenstrijdige gevoelens
    - Gemengde gevoelens
    - Besluiteloosheid
18. Geef aan in hoeverre de volgende woorden je gevoelens beschrijven ten opzichte van het veranderen van je kookgewoonten om ze **milieuverantwoorder** te maken (schaal van 0-5, waarbij '0' staat voor "ik heb dit gevoel niet" en '5' staat voor "ik heb dit gevoel heel erg":
    - Tegenstrijdige gevoelens
    - Gemengde gevoelens
    - Besluiteloosheid

**Intentie om kooktools te gebruiken [RW]**

Beeld je in dat je wil gaan koken, en je een recept zoekt op het internet.

19. Zou je een nieuwe online tool gebruiken die je suggesties kan geven om een recept gezonder te maken? (1 = totaal niet; 5 = zeker wel)
20. Zou je een nieuwe online tool gebruiken die je suggesties kan geven om een recept meer milieuverantwoord te maken? (1 = totaal niet; 5 = zeker wel)

**Voedingsgewoonten [ALL]**

21. Hoe vaak eet je gemiddeld huisgemaakte maaltijden per week (zelf gekookt of door een huisgenoot, ouder, partner, ...)?
    - ☐ Dagelijks
    - ☐ 3–6 keer per week
    - ☐ 1–2 keer per week
    - ☐ (Bijna) nooit

22. Hoe vaak gebruik je recepten als je thuis maaltijden kookt? **[ALL]**

- ☐ Dagelijks
- ☐ Bijna dagelijks
- ☐ Enkele keren per week
- ☐ Een keer per week
- ☐ Een paar keer per maand
- ☐ Minder dan 1 keer per maand
- ☐ Nooit

**Evaluatie van MyRecipeWatch** (enkel voor de post-intervention questionnaire) **[RW]**

1. Hoe vaak heb je MyRecipeWatch gebruikt in de afgelopen 4 weken?
    - ☐ Dagelijks
    - ☐ 3–6 keer per week
    - ☐ 1–2 keer per week
    - ☐ Minder dan 1 keer per week
    - ☐ Nooit

2. Over het algemeen, in welke mate vind je de suggesties van MyRecipeWatch (1 = helemaal niet en 7 = heel erg): **[RW]**

a) Handig
b) Nodig
c) Informatief

3. Hoe gemakkelijk was MyRecipeWatch om te gebruiken (schaal van 1-7 waarbij 1 = extreem moeilijk en 7 = extreem gemakkelijk)? **[RW]**
4. Hoe handig was MyRecipeWatch bij het helpen van het identificeren van gezondere recepten (schaal van 1-7 waarbij 1 = helemaal niet handig en 7 = extreem handig)? **[RW]**
5. Heeft MyRecipeWatch je geholpen om gezondere voedselkeuzes te maken? **[RW]**
6. Hoe gemakkelijk was het om de suggesties van MyRecipeWatch in je dagelijks koken te integreren (schaal van 1-7 waarbij 1 = extreem moeilijk en 7 = extreem gemakkelijk)? **[RW]**
7. Hoe logisch vond je de suggesties die MyRecipeWatch voorstelde (schaal van 1-7 waarbij 1 = helemaal niet logisch en 7 = heel erg logisch)? **[RW]**

8. Over het algemeen, hoe tevreden ben je over MyRecipeWatch (schaal van 1-7 waarbij 1 = extreem ontevreden en 7 = extreem tevreden)? **[RW]**
9. Hoe waarschijnlijk is het dat je MyRecipeWatch in de toekomst zal blijven gebruiken? (1 = helemaal niet waarschijnlijk en 7 = heel waarschijnlijk) **[RW]**
10. Hoe waarschijnlijk is het dat je MyRecipeWatch zal aanraden bij je vrienden? (1 = helemaal niet waarschijnlijk en 7 = heel waarschijnlijk) **[RW]**

**Evaluatie van Foodbag-swaps** (enkel voor de post-intervention questionnaire)

1. **Hoeveel vaak bestel je gemiddeld maaltijden bij Foodbag? [FOOD]**

☐ Elke week
☐ Elke twee weken
☐ Elke maand

2. **Motivaties en houding [FOOD]**

Hoe belangrijk zijn de volgende factoren voor jou bij het kiezen van je Foodbag-maaltijden? (Schaal 1–7 waarbij '1' staat voor "helemaal niet belangrijk" en '7' staat voor "heel erg belangrijk")

☐ Gezond eten
☐ Duurzaamheid en milieu-impact
☐ Minder vlees eten
☐ Smaak en voorkeur van het gezin
☐ Prijs/budget
☐ Gemak en bereidingstijd
☐ Nieuwe smaken/ingrediënten ontdekken

3. **Evaluatie van de swaps [FOOD]**

Wat is jouw mening over de voorgestelde ingrediëntenswaps? (Schaal 1–7: Helemaal niet mee eens → Helemaal mee eens)

- ☐ De suggesties waren duidelijk en begrijpelijk.
- ☐ De suggesties waren nuttig voor mij.
- ☐ De voorgestelde alternatieven pasten goed bij mijn smaakvoorkeur.
- ☐ De swaps waren gemakkelijk toe te passen tijdens het koken.
- ☐ De maaltijden met de swaps smaakten even goed als het origineel.

4. **Drempels (waarom werd er niet geswapt?) [FOOD]**

Als je een suggestie niet accepteerde, wat was daarvoor de belangrijkste reden? (Meerdere opties mogelijk)

- ☐ Ik was bang dat het minder lekker zou zijn.
- ☐ Mijn gezinsleden wilden het originele ingrediënt.
- ☐ Ik kende het alternatieve ingrediënt niet.
- ☐ Het alternatief leek me moeilijker te bereiden.
- ☐ Ik vond het originele recept al gezond genoeg.
- ☐ Gewoonte / Ik heb er niet echt over nagedacht.
- ☐ Andere:

5. **Impact [FOOD]**

- ☐ Door de suggesties van Foodbag ben ik bewuster gaan nadenken over mijn ingrediëntkeuze.
- ☐ Ik heb nieuwe, gezonde ingrediënten leren kennen door de swaps.