

Food & You: A Digital Cohort on Personalized Nutrition

Harris Héritier¹, Chloé Allémann¹, Oleksandr Balakiriev¹, Victor Boulanger¹, Sean F. Carroll¹, Noé Froidevaux¹, Germain Hugon¹, Yannis Jaquet¹, Djilani Kebaili¹, Sandra Riccardi¹, Geneviève Rousseau-Leupin¹, Rahel M. Salathé¹, Talia Salzmänn¹, Rohan Singh¹, Laura Symul^{1,2}, Elif Ugurlu-Baud¹, Peter de Verteuil¹, Marcel Salathé¹

¹Digital Epidemiology Lab, School of Life Sciences, School of Computer and Communication Sciences, EPFL, Switzerland

²Department of Statistics, Stanford University, USA

Nutrition is a key contributor to health. Recently, several studies have identified associations between factors such as microbiota composition and health-related responses to dietary intake, raising the potential of personalized nutritional recommendations. To further our understanding of personalized nutrition, detailed individual data must be collected from participants in their day-to-day lives. However, this is challenging in conventional studies that require clinical measurements and site visits. So-called digital or remote cohorts allow *in situ* data collection on a daily basis through mobile applications, online services, and wearable sensors, but they raise questions about study retention and data quality. “Food & You” is a personalized nutrition study implemented as a fully digital cohort in which participants track food intake, physical activity, gut microbiota, glycemia, and other data for two to four weeks. Here, we describe the study protocol, report on study completion rates, and describe the collected data, focusing on assessing their quality and reliability. Overall, the study collected data from over 1000 participants, including high-resolution data of nutritional intake of more than 46 million kcal collected from 315,126 dishes over 23,335 participant days, 1,470,030 blood glucose measurements, 49,110 survey responses, and 1,024 stool samples for gut microbiota analysis. Retention was high,

32 **with over 60% of the enrolled participants completing the study. Various data quality**
33 **assessment efforts suggest the captured high-resolution nutritional data accurately**
34 **reflect individual diet patterns, paving the way for digital cohorts as a typical study**
35 **design for personalized nutrition.**

36

37 **Keywords** personalized nutrition · digital cohort · gut microbiota · glycemia

38

39

40 Introduction

41 Nutrition plays a significant role in moderating the risk and/or severity of several
42 diseases, such as type 2 diabetes¹, cardiovascular diseases,^{2,3} or cancer⁴. Findings from
43 nutritional epidemiology studies have led to dietary guidelines and public health
44 campaigns designed to support healthy diets. However, while these recommendations
45 are generally based on results aggregated at the population level, a more individualized
46 approach to health⁵ has led to the concept of personalized nutrition. For example, a
47 randomized study showed that personalized recommendations improved diet quality as
48 measured by the Healthy Eating Index (HEI) compared to a control group⁶. Zeevi and
49 colleagues showed how personalized nutrition algorithms could be used to design diets
50 that lower postprandial glucose responses⁷. Further studies showed the importance of
51 personal features such as gut microbiota compositions on glycemic responses^{8,9}. Another
52 intervention study showed significant improvement in the food categories consumed
53 when receiving personalized diet advice, compared to generic or no advice¹⁰. These
54 findings highlight the need for a more holistic and individual approach to nutritional
55 epidemiology, encompassing diet, gut microbiota, physical activity, and lifestyle.

56

57 Blood glucose response is a particularly interesting outcome measure for nutritional
58 studies due to its association with insulin resistance, metabolic syndrome, and diseases
59 such as stroke, type 2 diabetes, and heart disease¹¹. Reducing blood glucose levels is thus
60 a recommendation of public health authorities around the globe. Identified risk factors
61 for elevated blood glucose levels include a carbohydrate-rich diet, lack of physical
62 activity, and poor sleep, among others. In addition, studies have begun to investigate the

63 role of the gut microbiome in modulating the blood glucose response to food intake.
64 Nutritional studies trying to understand postprandial glucose response (PPGR) are thus
65 faced with the challenge of obtaining data on relevant factors all at once, ideally
66 continuously and *in situ*, that is, in the regular environment in which participants' lives
67 unfold.

68

69 In digital health studies - also called remote or siteless studies - all interactions with
70 participants, as well as data collection, are digital or digitally coordinated. These studies
71 leverage an array of digital devices, wearable sensors, and online services. Digital cohorts
72 and trials have been heralded as a new major development for epidemiological and clinical
73 studies. However, since digital cohorts are a relatively new study approach, open
74 questions regarding selection bias, retention, and data quality remain. Indeed, access to
75 devices connected to the internet and digital literacy may lead to selection bias which in
76 turn may lead to a lack of representativity of the study population compared to the
77 general population. Furthermore, the time burden generated by following the study
78 protocol and collecting the data might create response fatigue, which could in turn
79 translate into lower study adherence, or data quality. Finally, novel data collection
80 methods may not have been thoroughly validated.

81

82 Here, we address some of these questions by evaluating the completion rates, adherence,
83 and data quality of the “Food and You” digital cohort on personalized nutrition. The “Food
84 & You” study started in early 2019 and consisted of two distinct digital sub-cohorts; the
85 sub-cohort “Basic” (cohort B) restricted to non-diabetic participants, and the sub-cohort
86 “Cycle” (cohort C) restricted to non-diabetic women of reproductive age who did not use
87 hormonal contraceptive or medication (see Methods for inclusion / exclusion criteria).
88 The study duration for cohort B was 14 days, whereas cohort C participants were enrolled
89 for 28 days, matching the length of a typical menstrual cycle. The study was performed
90 in Switzerland, and all participant were required to have a postal address in Switzerland.
91 Throughout the study, participants were requested to report i) food consumption using
92 an AI-assisted food tracking app (MyFoodRepo), ii) continuous blood sugar levels using a
93 continuous glucose monitor, and iii) physical activity and sleep using either activity
94 trackers or daily surveys. They were furthermore asked to follow a protocol that included

95 a one-time stool sample collection for gut microbiota analysis, and the consumption of
96 standardized breakfasts.

97

98 The present paper details the study protocol, and reports study engagement data by
99 looking at the individual characteristics of participants on their journey from enrollment
100 to completion. Further, we provide an overview of the data collected in the “Food & You”
101 cohort from October 2018 to March 2023, and describe our efforts to assess data quality,
102 including the comparison of nutritional and microbiota data collected in “Food & You”
103 with data collected in traditional (on-site) studies. We also discuss the challenges of
104 running a complex digital cohort, and how we addressed them. Overall, retention rates
105 were relatively high, with more than 60% of enrolled participants completing the study.
106 Despite certain fatigue over time, adherence very high, especially for glucose response
107 data, nutritional data, microbiota data, and data from daily surveys. While the study
108 population shows some demographic differences compared to the overall population, the
109 nutrition patterns are in very good agreement with data obtained in another study from
110 a representative sample of the general Swiss population.

111 Methods

112 **STUDY DESIGN AND SETTING**

113

114 The Swiss “Food & You” study is a digital cohort study collecting data on glycemia,
115 nutrition, gut microbiota, lifestyle, and physical activity as well as demographic data
116 (Figure 1). The cohort was open to anyone fulfilling the inclusion criteria listed in
117 Supplementary Table 1. The study consisted of four sequential phases: enrollment phase,
118 preparatory phase, tracking days phase, and follow-up phase (Figure 2). In the *enrollment*
119 *phase*, interested participants were first required to perform a self-check of their
120 eligibility, and fill out a consent form and a short screening questionnaire. Following this,
121 and upon acceptance by a member of the study team (based on the available capacity to
122 accommodate new participants), the participants were considered enrolled. During the
123 subsequent *preparatory phase*, enrolled participants were given instructions on the “Food
124 & You” website and asked to fill out a series of questionnaires. Next, they were instructed

125 to download the AI-assisted nutrition tracking app MyFoodRepo¹ (MFR) to track their
126 food intake for a trial period of at least three days. After the successful completion of the
127 trial period, participants would order the study material which included, among other
128 things, a continuous glucose monitoring (CGM) sensor for each period of 14 days and
129 stool-sample self-collection material. Upon receipt, they were asked to choose the
130 starting date of their tracking days phase. In the *tracking days phase*, participants were
131 required to log all their food and drink intake via the MyFoodRepo app, wear the CGM
132 sensor, and answer two daily surveys for 14 (cohort B) and 28 (cohort C) days. In addition,
133 participants were instructed to take standardized breakfasts in accordance with dietary
134 restrictions from the 2nd to the 7th day during the first week. They were asked to avoid
135 altering standardized meals, as well as to refrain from eating or engaging in physical
136 activity for the subsequent two hours. Cohort C participants repeated these instructions
137 on their third tracking week (Supplementary Table 2). On days 6 and 7 (additionally on
138 days 21 and 22 for cohort C) they were asked to perform an oral glucose tolerance test by
139 drinking 50g of glucose. In addition, study participants were asked to provide one stool
140 sample collected anytime during the tracking days phase. At the end of the tracking days
141 phase, participants were asked to upload their physical activity and CGM data on the
142 “Food & You” website. In the *follow-up phase*, they were requested to fill out a feedback
143 questionnaire regarding their experience. Participants were also provided with
144 interactive visualizations of their data (Supplementary Figure 1). Cohort C participants
145 were followed for two additional menstrual cycles during which they continued to track
146 their menstrual cycle and fertility-related body signs such as morning temperature and
147 cervical mucus characteristics.

148

149 The Geneva ethics commission has reviewed and authorized the project (Ethical Approval
150 Number: 2017-02124). The study is registered on the website of the Federal Office of
151 Public Health (SNCTP000002833) and the platform clinicaltrials.gov (NCT03848299).

152

153

154

155

¹ <https://www.myfoodrepo.org/>

156 **DATA COLLECTION**

157

158

159 **Questionnaires:** During the enrollment phase, interested subjects were requested to fill
160 out a screening questionnaire with items regarding age, gender, height, weight, type of
161 mobile phone and dietary restrictions. During the preparatory phase, enrolled
162 participants had to fill out a lifestyle and health-related questionnaire. Participants were
163 asked about their general health (smoking, diet, food supplement intake, general hunger
164 levels, health state, past diagnoses, antibiotic intake, and menstrual health), physical
165 activity (exercise frequency, duration, and intensity), sleep (bedtime, wake-up time),
166 sociodemographic variables (nationality, socioeconomic status, job status, and household
167 description), and requested to provide self-measured anthropometric measurements
168 (waist and hip circumference, height, and weight). In the tracking days phase, participants
169 had to fill out a short form each evening to validate their adherence to protocol and
170 document medication intake. Cohort C participants had to answer additional questions
171 on menstrual blood, cervical mucus and provide self-reported temperature
172 measurements on a daily basis.

173

174 **Dietary Intake:** Participants of the “Food & You” study were asked to log any dietary
175 intake in real time on the MyFoodRepo app (MFR) using one the following options: taking
176 pictures of the food/drink, scanning the product’s barcode (if available), or describing
177 the food item with text. A logged entry is defined as a “dish”, and can contain multiple
178 food items. For example, tuna, steamed potatoes, and green beans are all single food
179 items, and together compose a dish. The pictures were automatically segmented and
180 classified by an image recognition algorithm¹². Portions, segmentations, and food classes
181 were subsequently verified or edited by a team of trained annotators. The MyFoodRepo
182 app also allows annotators to communicate with the participant for clarifications about
183 the food, and participants were able to leave comments through the app. This process
184 ensured that every single dish in the nutritional data was reviewed by a member of the
185 study team. Each food item was linked to a nutritional value database containing 2’129
186 items built on the Swiss Food Composition Database¹³, MenuCH data¹⁴, and Ciqua¹⁵. When
187 food intake was logged through barcode scanning, nutritional values of the food items

188 were fetched from the Open FoodRepo database API¹⁶. Manual entries were matched to
189 food items by the annotators.

190

191 As the aforementioned nutritional values data sources did not provide standard portion
192 sizes, these were manually extracted from the portion list of the WHO MONICA study¹⁷,
193 and the Mean Single Unit Weights of Fruit and Vegetables report¹⁸ by the German Federal
194 Office of Consumer Protection and Food Safety. When a standard portion was not
195 available for a particular food item, we assigned the standard portion of a similar food
196 item.

197

198 Each dish logged through the MyFoodRepo app carries a timestamp, which enables
199 dietary analysis at a high temporal resolution. Food items were classified into categories
200 based on the menuCH study¹⁴ (Chatelan, Marques-Vidal, et al. 2017). Barcoded food items
201 from the Open FoodRepo database were categorized based on the food product
202 description. When such a description was not available, we assigned the category
203 extracted from the Open Food Facts database². While “Food & You” is the first large
204 cohort to use MyFoodRepo, the app annotation quality had previously been validated¹⁹.

205

206 **Glycemia:** Glycemic data was collected by using the Flash Glucose Monitor Freestyle
207 Libre (Abbott Diabetes Care). The system, which has been validated in numerous
208 studies,²⁰⁻²² consists of a disposable sensor applied to the back of a participants’ upper
209 arm, and a reader device or a smartphone app allowing to collect data from the sensor
210 via NFC technology. It measures glycemia every 15 minutes via a subcutaneous filament
211 carrying enzyme glucose sensors²³. To encourage high adherence to protocol, we chose
212 a non-blinded glucose monitoring system to allow the participants to see their glycemia
213 in real time. Participants self-applied the sensor at home following explanations provided
214 in writing and video. Cohort B participants wore a single sensor for 14 days, whereas
215 cohort C participants wore two sensors consecutively for a period of 28 days to cover the
216 length of a typical menstrual cycle. Notably, when participants scan the sensor, the data
217 from the previous eight hours is collected. Thus, unless participants scan the sensor at
218 least every 8 hours, some data may remain unretrievable.

² <https://world.openfoodfacts.org/>

219

220 **Gut Microbiota:** Participants were requested to collect a stool sample following detailed
221 written and video instructions. They could collect and ship their sample anytime during
222 the tracking days phase. Samples were collected with stool nucleic acid collection and
223 preservation tubes from Norgen Biotek, stored at room temperature and shipped in
224 batches of 100 to 192 samples to Microsynth AG (Balgach, Switzerland) for sequencing
225 and bioinformatics analysis. V4 region of the bacterial 16S rRNA gene was sequenced via
226 creation of two-step Nextera PCR libraries using the primer pair 515F
227 (NNNNNGTGYCAGCMGCCGCGGTAA) and 806R
228 (NNNNNGGACTACNVGGGTWTCTAAT). The primers use 5 bases at their 5' end to
229 increase diversity of the bases during the first five sequencing cycles. Subsequently, the
230 Illumina MiSeq platform and a v2 500 cycles kit were used to sequence the PCR libraries.
231 The produced paired-end reads which passed Illumina's chastity filter were subject to
232 de-multiplexing and trimming of Illumina adaptor residuals using Illumina's real time
233 analysis software included in the MiSeq reporter software v2.6 (no further refinement or
234 selection). The quality of the reads was checked with the software FastQC version 0.11.8.
235 The locus specific V4 primers were trimmed from the sequencing reads with the software
236 cutadapt v2.8. Paired-end reads were discarded if the primer could not be trimmed.
237 Trimmed forward and reverse reads of each paired-end read were merged to in-silico
238 reform the sequenced molecule considering a minimum overlap of 15 bases using the
239 software USEARCH version 11.0.667. Merged sequences were then quality filtered
240 allowing a maximum of one expected error per merged read. Reads that contained
241 ambiguous bases or were considered outliers regarding the amplicon size distribution
242 were also discarded. Samples that resulted in less than 5000 merged reads were
243 discarded, to not distort the statistical analysis. The remaining reads were denoised using
244 the UNOISE algorithm²⁴ implemented in USEARCH to form amplicon sequence variants
245 (ASVs) discarding singletons and chimeras in the process. The resulting ASV abundance
246 table was then filtered for possible bleed-in contaminations using the UNCROSS²⁵
247 algorithm, and abundances were adjusted for 16S copy numbers using the UNBIAS²⁶
248 algorithm. ASVs were compared against the reference sequences of the RDP 16S database,
249 and taxonomies were predicted considering a minimum confidence threshold of 0.5 using
250 the SINTAX algorithm²⁷ implemented in USEARCH.

251

252 **Physical Activity and Sleep:** Study participant's physical activity (PA) and sleep data were
253 collected using one of two methods: objectively via Apple Health, Google Fit, or smart-
254 watches, or subjectively, i.e. self-reported on the study website via the morning and/or
255 evening questionnaire. The different formats of the objective PA and sleep data were
256 harmonized and stored in a single database, and comprised daily step count, daily calories
257 burned, bedtime, and wake-up time. In addition, for PA measured with smart devices, the
258 type of physical activity, the start and end times, the amount of burned calories, the
259 average heart rate, and the maximum heart rate were collected. Participants self-
260 reporting their sleep and PA on the website had to report the times at which they fell
261 asleep and woke up, if they did any physical activity, and if so, when they started and
262 finished, as well as the perceived intensity of their effort.

263

264 **STATISTICAL ANALYSIS**

265

266 **Data Preprocessing:** Sociodemographic questionnaires variables and anthropometric
267 measurements were re-coded as follows; Monthly household income was classified in 4
268 categories: <6000, 6000 to 8999, 9000 to 12999 and >=13000 Swiss Francs (CHF). Note
269 that the median income in Switzerland was 6,665 CHF per month in 2020. Household
270 type categories were defined as either "alone" for single-person households, "couple w.
271 children", "couple w.out children", and "other". Age at study start was transformed into
272 the following categories; 18 to 34, 35 to 49, 50 to 64, and older than 65 years. Citizenship
273 was categorized as "Swiss", "Binational", or "Foreigner". Education level was categorized
274 as "low" (mandatory, primary school), "intermediate" (high school and professional
275 diploma) or "high" (university). Smoking status was categorized as "non smoker" (never
276 smoked), "former smoker" or "current smoker" (occasionally or daily). The self-rated
277 health questions were re-coded as binary variables "Not good to average" (not good,
278 average, somewhat good) or "Good to very good" (good, very good). Residential addresses
279 were geocoded, and municipality number (Spatial Development ARE) was extracted for
280 each address. The municipality number was then joined with a urban/rural municipality
281 typology and linguistic region dataset from the Swiss Federal Statistical Office so that
282 each address was categorized as "urban" or "rural," whereas linguistic region was
283 categorized as "german" or "latin," the latter encompassing french, italian, and romansh-
284 speaking regions.

285

286 Missing values in income and education (<0.2% and less than 8%, respectively) were
287 imputed via multiple imputations with chained equations using the mice package²⁸. Self-
288 reported general physical activity levels were assessed based on the weekly frequency
289 and average duration in minutes. The data was then coded into “active” and “inactive”
290 based on the WHO cutoff of 75 min per week²⁹. Body-mass index (BMI) was calculated
291 based on self-reported height and weight and classified into “Underweight,” “Normal,”
292 “Overweight,” “Obese” based on the WHO cutoffs for BMI. For food intake, we removed
293 days for which energy intake was below 1000 kcal and aggregated by study subject to
294 calculate the individual weighted mean to account for day-of-week and seasonal
295 variations of intake over the observation period.

296

297 To assess the extent of glycemic excursions for each participant, we calculated the
298 proportion of readings below, in, and above the target range based on the cut-off values
299 of 3.9 and 10 mmol/L³⁰. We also computed the mean amplitude of glycemic excursions
300 (MAGE) for each subject using the R package gluvarpro (4.0).

301

302 **Completion rate and study population analysis:** Study completion rate was defined as
303 the ratio of the number of participants having completed the study over the total number
304 of participants enrolled in the study. The completion rate was then calculated for each
305 cohort (B and C) and each of the following strata: gender, age group, BMI, phone type,
306 and dietary restriction.

307

308 For the study population analysis, we only included the participants who have completed
309 the study, and calculated their proportion in each cohort for the following strata: gender,
310 household income, household type, age group, citizenship, education, smoking status,
311 health status, urbanity, linguistic region, physical activity, and BMI.

312

313 Analysis of adherence to study protocol was conducted by cohort given that their
314 tracking phase duration differed. For each participant, we reported the number of days
315 with (i) CGM measurements, (ii) reported food (distinguishing between days with a total
316 intake above 1000 kcal or days with any food logged), (iii) sleep data, (iv) PA data, and (v)

317 questionnaire data. We also reported the number of standardized breakfasts taken by
318 participants, of glucose oral tolerance tests performed, and stool samples collected.

319

320 **Data quality analysis:** Given the multimodal nature and multi-dimensionality of the
321 collected data, assessment of data quality may take several forms. Here we performed a
322 series of analyses to evaluate if each type of collected data followed expected patterns.

323

324 Specifically, we performed three analyses assessing the food data quality with respect to
325 timing, adherence, and composition. First, we assessed the distributions of food intake
326 times by calculating the count of the logged food intakes by day of the week and hour of
327 the day. Second, we hypothesized that most CGM peaks should be preceded by a food
328 intake. We thus conducted an experiment using CGM data as ground truth. For each CGM
329 peak (i.e., a local maximum above the 90th percentile of the participant's CGM
330 distribution) we reported whether a food intake of at least 50 kcal had been logged within
331 the 90 min interval preceding the peak, or within the 90 min preceding the peak minus 2
332 hours (control case). We then reported, for each participant, the proportions of CGM
333 matching a food intake in both experiment and control situations. To test our hypothesis
334 that the proportion of peaks with logged food intake within 90 minutes of the peak would
335 be higher than the same proportion for the control time-window, we used a non-
336 parametric Wilcoxon rank sum test. Third, to assess the quality of the composition of
337 reported food intake, we compared food intake data obtained in the "Food & You" cohort
338 to data reported by the Swiss nutritional survey menuCH which was obtained from a
339 demographically representative sample. We considered amounts of energy, meat, dairy,
340 water intake as well as the proportion of the study population that consumed more than
341 five portions of fruits and vegetables a day. For each of the following strata: sex, age group
342 and linguistic region, we calculated the weighted means of the intake to account for
343 weekday variations.

344

345 For sleep, the data were aggregated at the participant level to calculate the weekday and
346 weekend individual mean values for bedtime, wake-up time, and sleep duration. Sleep
347 durations of less than 4 hours or more than 12 hours were excluded.

348

349 For gut microbiota data comparison, we used relative abundances for the microbiome
350 samples of other studies available from the R package CuratedMetagenomicsData
351 (version 3.7). These samples were selected based on the filtration criteria that all samples
352 were of gut origin and from healthy adults. Four studies with most of such samples were
353 selected for comparison with the “Food & You” microbiome^{7,31,32,33}. Relative abundances
354 from these studies and “Food & You” were aggregated at the genus taxonomic level. Bray-
355 Curtis distances were calculated between the samples using the vegdist function of the
356 vegan package in R, which were then transformed into two-dimensional principal
357 coordinates using the pcoa tool of ape package in R.

358

359

360 Results

361

362 STUDY COMPLETION RATE

363

364 Overall study completion rate - defined as the ratio of the number of participants having
365 completed the study over the total number of participants enrolled in the study - was
366 relatively stable over the years, with 69.5%, 56.4%, 65.7%, and 57.3% in 2019, 2020, 2021,
367 and 2022, respectively. Detailed completion rates are shown in Table 1. Within the
368 investigated groups - gender, BMI, age, phone type, and diet - we did not see any major
369 differences, with the exception of diet. Participants who reported dietary restrictions had
370 higher completion rates (82.5% and 81.5% in cohorts B and C, respectively) than their
371 counterparts without restrictions (62.7% in cohort B and 56.7% in cohort C).

372

373 COHORT CHARACTERISTICS

374

375 1,014 study subjects have completed the “Food & You” study, of which 870 in cohort B and
376 144 in cohort C (Table 2). In cohort B, both sexes were equally distributed. The proportion
377 of healthy and educated subjects was found to be higher in “Food & You” as compared to
378 the Swiss population. A direct comparison is difficult, because representative statistics
379 for the Swiss subpopulation with study inclusion and exclusion criteria applied are not

380 available. For example, in cohort B, the proportions of overweight and obese cohort
381 participants was 24% and 6%, while these proportions are slightly higher in the general
382 Swiss population (31% and 11% respectively). However, the combination of diabetes as an
383 exclusion criterion and the well-known association between diabetes and overweight /
384 obesity³⁴ may explain the difference. Because representative data from the Swiss
385 population filtered by our exclusion and inclusion criteria is not available, we did not test
386 whether the observed differences could be due to chance.. Some of the large differences
387 are unlikely to be explained by exclusion and inclusion criteria alone. Most strikingly, the
388 “Food & You” study population is much more highly educated, with almost three quarters
389 of participants having a university degree (compared to less than one third in
390 Switzerland). In addition, the 18-34 and 45- to 49 age classes are substantially over-
391 represented, while the elderly (ages 65 and over) are severely underrepresented (less than
392 3% of participants).

393

394 **ADHERENCE**

395

396 Of the participants who completed the study, adherence to protocol was generally high,
397 with a large majority of participants able to collect the requested amount of data for most
398 modalities. In cohort B, 93.9% provided at least 13 days of food tracking data with daily
399 energy intake above 1000 kcal, whereas this figure was slightly lower (84.7%) for cohort
400 C participants, albeit for at least 27 days. Cohort C participants were asked to provide
401 self-reported mucus quality and temperature measurements on a daily basis. 97% and
402 76% reported bleeding data on any day, and during the last week, respectively. The 1,014
403 participants who completed the study logged a total of 297,626 dishes amounting to 43.6
404 million kcal. Participants mostly logged their food intake by taking pictures (76.1%),
405 whereas a smaller proportion of entries were logged by barcode scanning (13.3%) or
406 manually (10.6%). Sleep data was available for 64.8% of the participants who completed
407 the study. In total, participants ate 6,944 standardized breakfasts including 2,158 glucose
408 drinks.

409

410 In total, 6,460 subjective (i.e. self-reported) physical activities were reported. A large
411 proportion (33% in cohort B, 42.4% in cohort C) of users in both cohorts did not report
412 any activity at all, but the proportion that did report activities was higher in cohort C

413 (Figure 3). Around 4/5 of the participants who provided physical activity data reported
414 only subjective activity via website questionnaire. In terms of objective activities collected
415 through activity trackers such as smartwatches, the most reported activity types (66.5%)
416 were walking, running, and cycling.

417

418 In total, 997 participants who completed the study provided a stool sample for microbiota
419 composition quantification. The distribution of relative abundances at phylum and
420 species levels are depicted in supplementary figures 2 and 3.

421

422 **DATA QUALITY**

423

424 Breakfasts were observed from 05:00 onwards, whereas the hourly distribution of this
425 meal was shifted later during the weekend as compared to weekdays (Figure 4a). Logged
426 lunches were found to be consistent with work schedules during the week, with a peak
427 at noon, whereas the number of entries at those hours was much lower during weekends.
428 The number of entries after 20:00 tended to be higher Fridays and Saturdays, likely
429 reflecting social dinners. In general, and as expected, a shift in the later hours was
430 observed for all meals taken during the weekend as compared to weekdays (difference in
431 average meal times of 33 minutes, Wilcoxon rank sum test p -value < 0.001). The exception
432 to this pattern was Sunday dinners: the timing of those was more similar to Monday
433 dinners' than to Saturday dinners'.

434

435 The glucose data displayed a similar pattern as the food data with a clear shift between
436 weekdays and weekend, reflecting the later wake-up and eating times (Figure 4b, maximal
437 cross-correlation between weekend and weekdays mean glucose levels is found with a
438 30-minute shift). Consistently, wake up and bedtimes were also shifted toward later
439 hours during weekends. Participants slept 7.9 hours a night on average (SD: 1.19, Figure
440 4f). On average, participants slept 16 minutes longer on weekends than on weekdays ($n =$
441 591, paired t -test p -value < 0.001). On average, participants' bedtime was 23:41 (SD:
442 1.37, Figure 4e), 23:37 on weekdays, and 23:51 on weekends. Participants woke up on
443 average at 07:35 (SD: 1.43, Figure 4e), and woke up significantly later on weekends, at 07:57
444 vs 07:27 (average difference of 30 minutes, paired t -test p -value < 0.001).

445

446 The proportion of glucose peaks matching a food item was significantly higher (median
447 proportion: 68%) within the experimental group compared to the control group (median
448 proportion: 27%, one-sided Wilcoxon rank sum test p -value < 0.001), indicating a good
449 agreement between the food intake and CGM data. However, as we have no objective
450 ground truth data on nutrition, we do not know what the expected proportions would be.

451

452 In general, participants had a healthy glycemic profile with mean amplitude of glucose
453 excursion (MAGE) values of 1.41 mmol/L and 1.21 mmol/L for cohorts B and C, while the
454 proportion of readings above 10 mmol/L (hyperglycemia) was below 0.3%.

455

456 **COMPARISON WITH OTHER STUDIES**

457

458 For participants who completed the study, daily mean energy, meat, dairy and water
459 intake were 2,205.1 kcal, 92 g, 124.9 g, and 964.9 ml, respectively. The proportion of
460 participants more than 5 portions of fruits and vegetables per day was 6.71%. These values
461 were in good agreement with the mean values reported by the Swiss nutritional survey
462 “menuCH” were obtained from a demographically representative sample. The only
463 exception was for water, for which intake was much lower in menuCH than in our study.
464 Means and standard deviations for energy, meat, dairy, water intake as well as proportion
465 of participants eating more than 5 portions of fruits and vegetables per day are reported
466 in supplementary table 3 for the whole population and for various strata (sex, language,
467 age group, and combination of sex and age group).

468

469 The PCoA plot of the gut microbiota profile (Figure 5) shows the Bray-Curtis distance
470 dissimilarities of gut microbial community samples found in healthy individuals of “Food
471 & You” and four other gut-related studies. To allow for comparability, bacterial
472 taxonomies from these microbiomes were aggregated at the genus level. The plot shows
473 that most participants from the “Food & You” cohort and the LifeLinesDeep study tend
474 to form their own cluster distinct from other studies.

475 Discussion

476 The “Food & You” study is a fully digital nutrition cohort collecting diverse multimodal
477 data remotely, without any physical contact between the participants and any member
478 of the study team. This study has gathered a wealth of nutritional intake data, blood
479 glucose measurements, survey responses, and gut microbiota samples from 1,014
480 participants. Here, we described the study protocol, reported on study completion rates,
481 and described the collected data, focusing on assessing their quality and reliability.

482
483 The overall completion rate was high, with over 60% of enrolled participants completing
484 the study. In comparison with other digital health studies³⁵, the retention rate for 14 and
485 28 days (cohort B and C, respectively) was rather high. Several factors may have
486 contributed to this outcome. Perhaps most importantly, we approached the study's
487 design from a participant's perspective, which led to the decision to develop a new food-
488 tracking app (MyFoodRepo) from scratch, emphasizing ease of use. We also developed
489 the study website and data collection system from scratch in order to directly integrate
490 data collected from sensors or apps on the study website. For example, we combined
491 nutritional data from the app and glucose data from the CGM system to generate
492 interactive charts on the study website. Completion rates in younger and older age
493 groups were comparable, indicating no major hurdle related to the use of digital tools for
494 data collection for older subjects. Subjects with dietary restrictions were particularly
495 committed to the study in both cohorts (completion rates over 80%), in line with previous
496 findings from disease-based digital health studies showing that participants affected by
497 the disease showed higher study retention³⁵.

498
499 With respect to adherence, data availability was high for most indicators in both cohorts,
500 with the exception of physical activity and sleep. The limited provision of physical activity
501 and sleep data by participants possibly indicates that these aspects, unlike diet, were not
502 viewed as central to the study, despite encouragement to include this information. The
503 minimal study duration was 14 days (cohort B), which may represent a significant time
504 investment for participants. Besides participant's fatigue, other parameters may have
505 impacted adherence. For example, technical issues such as faulty glucose sensors or
506 omission to scan the sensor and temporary technical issues in the early versions of the

507 food tracking app or the “Food & You” website may at times have contributed to a lower
508 amount of data delivered. However, across both cohorts, data availability was high for
509 most indicators.

510

511 Because the “Food & You” study obtained high-resolution dietary data from a mobile app
512 (MyFoodRepo) that was originally created specifically for the study, care needed to be
513 taken to assess the quality of the nutritional data. As other studies have also begun to use
514 the app, the first independent validation study indicated strong data quality¹⁹. In addition,
515 the fact that every submitted data point on nutrition was reviewed by a study annotator
516 provides additional confidence in the data quality. We observed expected patterns
517 related to weekdays and weekends in terms of timing of food intake, glucose curves, and
518 wake-up and bedtimes at the population level. The high proportion of CGM peaks
519 matching a food intake further suggests that participants logged their food intake
520 appropriately and that the number of missing intakes is expected to be low. In addition,
521 overall intake, and intake of main food groups is in agreement with results reported from
522 a representative sample of the Swiss population.

523

524 “Food & You” is the first study that collected food intake via the AI-assisted nutrition
525 tracking app MyFoodRepo, generating an unparalleled dietary high temporal resolution
526 dataset of over 300'000 food dishes with a total of over 45 million kcal. We therefore have
527 a very precise picture of dietary patterns over at least two weeks from over 1000
528 participants. The data provided by the MyFoodRepo app is primarily based on time-
529 stamped pictures, and thus allows for objective temporal assessment of food intake. The
530 high study completion rates after participants completed the test of the app, combined
531 with the overall positive feedback from post-study surveys indicate that MyFoodRepo
532 was well accepted by participants.

533

534 Self-selection bias may have led to a non-representative study population with regard to
535 some particular sociodemographic variables. Based on the exclusion criteria targeting the
536 non pre-diabetic and diabetic population, we expected our study population to lead a
537 healthy lifestyle. Indeed, the high proportion of physically active people who self-report
538 a good health status among the participants points in that direction. Similarly, blood
539 glucose-related indicators such as MAGE and proportion of hypo- and hyperglycemic

540 excursion show no indication of subjects affected by pre-diabetes. Recruitment via social
541 media may have increased the proportion of younger participants with scientific affinity.
542 Further, the digital nature of the project could have increased the selection of digitally
543 savvy participants. The lack of socio-demographic representation of the “Food & You”
544 study can be overcome through appropriate weighting of the imbalanced strata as
545 conducted in other studies³⁶. This adjustment procedure will be crucial for further
546 publications, since it has been shown that socio-demographic factors greatly influence
547 dietary factors³⁷. Finally, by excluding pre-diabetic people from the study, we may have
548 filtered out participants with high BMI, since the former is known to be associated with
549 the latter. A Swiss study³⁸ on self-reported anthropometric measurements has reported
550 that BMI measurements based on self-reported height and weight are underestimated by
551 a factor of 1.6. The true distribution of BMI observed in the “Food & You” project may
552 therefore be shifted towards higher BMI ranges.

553

554 The use of a non-blinded CGM system may be considered a limitation of this study as
555 participants may have adjusted their dietary habits with the intention to control their
556 glucose levels. The decision to use a non-blinded CGM system was motivated by two
557 factors. First, we believed that participants would remain more engaged than with a
558 blinded system. Second, as self-tracking health sensors become more commonly
559 available, future personalized nutrition systems deployed at scale must be designed in
560 the context of full data visibility to participants. Whether such potential self-adjustments
561 in digital cohorts are limited to the short term and initial use is an important question for
562 further research.

563

564 Microbiome samples in "Food & You" were sequenced using 16S rRNA, whereas the
565 samples from the other studies to which the microbiome data was compared to used
566 shotgun sequencing. The distinct clustering of "Food & You" and LifeLinesDeep samples,
567 originating from Switzerland and the Netherlands respectively, suggests a potential
568 geographical influence. These geographical differences are likely contributing factors to
569 the observed variations in gut microbiota between the two cohorts. More detailed
570 analyses of the microbiome and its associations with other data collected in the study are
571 ongoing.

572

573 Taken together, our results show that collecting a large amount of high-quality data with
574 high study protocol adherence is feasible in the framework of a digital nutrition cohort,
575 opening the path toward large-scale and detailed personalized nutrition studies.

576 Acknowledgments

577 We are deeply grateful to the participants of the “Food & You” study. This work was
578 supported by the Kristian Gerhard Jebsen Foundation, the Seerave Foundation, and the
579 Fondation Leenaards. The funders had no role in the design or execution of this study
580 and will have no role in the analyses, interpretation of the data, or decision to submit
581 results. We thank Stéphanie Bonnewyn, Sandra Rüeeggsegger-Tanner, and Bennan Tong
582 for help with the annotations, as well as Dr Jardena Puder for her clinical guidance. We
583 further thank everyone who helped with various parts of the project, especially Chloé
584 Greub, Leila Munaretto, and Jil Zanolin.

585 Author Contributions

586 Study Design: CA, MS

587

588 Technological Development: AB, SC, NF, GH, YJ, DK, PdV

589

590 Nutritional Analysis: CA, GR-L, RMS, EU-B

591

592 Project Management: CA, SR, TS, HH, MS

593

594 Data analysis: HH, VB, DK, RS, LS, MS

595

596 Writing: HH, RS, LS, MS

597 Competing Interests

598 The authors declare that there are no competing interests.

599 Data Availability

600 The data to reproduce the results presented will be made available in a public repository
601 upon publication. We're happy to share the repository with reviewers during the review
602 process.

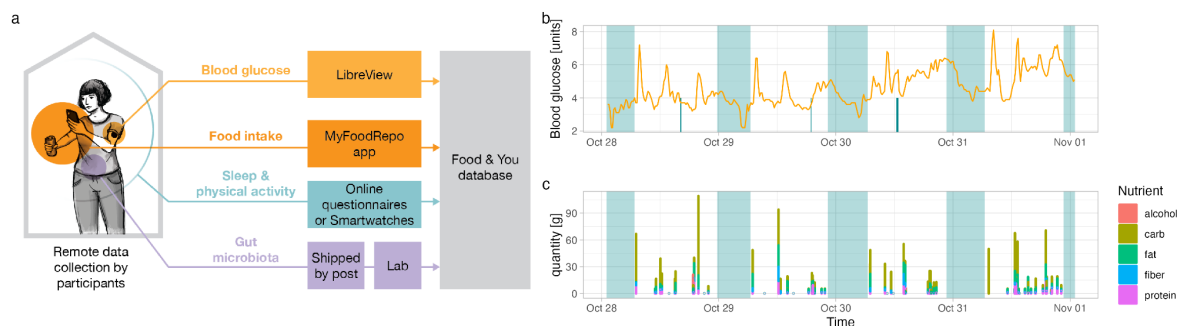
603

604

605 Figures

606

Figure 1



607

608 **Data collection.** a) Schematic illustrating the data collection process. (Left) Participants
609 track the study variables *in situ* (from home, work, etc.). (Center) Data or samples are
610 collected via web platforms and apps, or shipped to the lab by mail. (Right) Data is
611 processed in the Food & You database. b and c) Example of data collected by one
612 participant over 5 days. Top panel shows blood glucose levels (orange line), physical
613 activity (turquoise spikes), and sleep (translucent turquoise rectangles). Bottom panel
614 shows time and micronutrient composition (colors) of reported food intake. Like in the
615 top panels, translucent turquoise rectangles show when the participant is asleep.

616

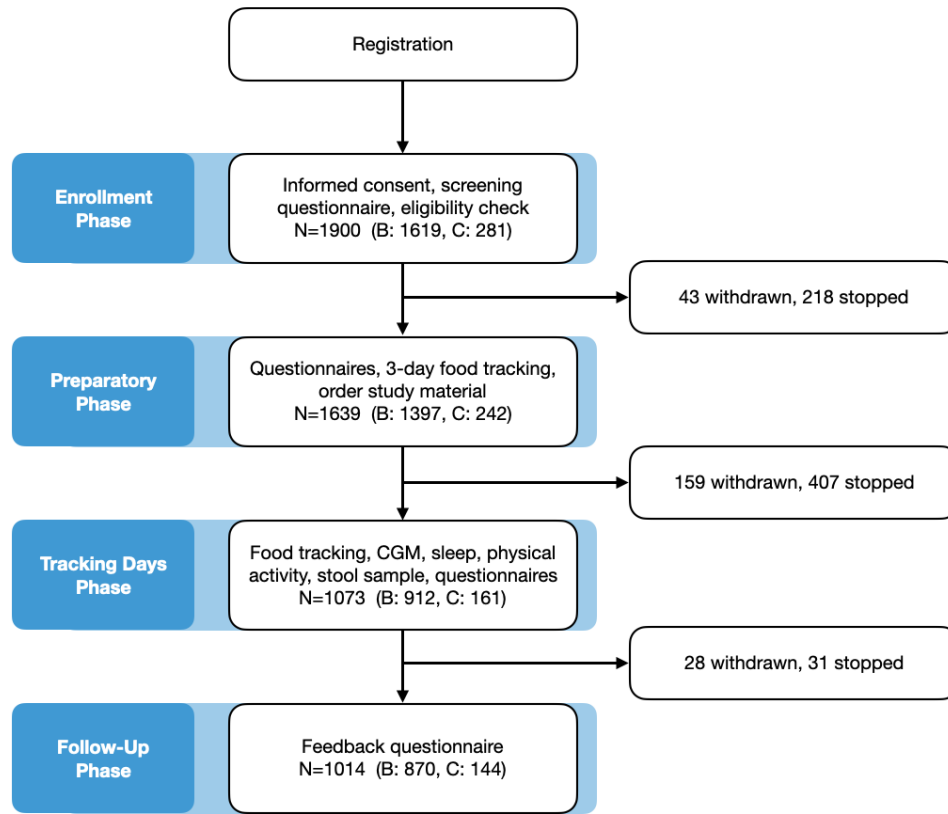
617

618

619

Figure 2

620



621

622

623 Study phases with participants per phase, and exit numbers.

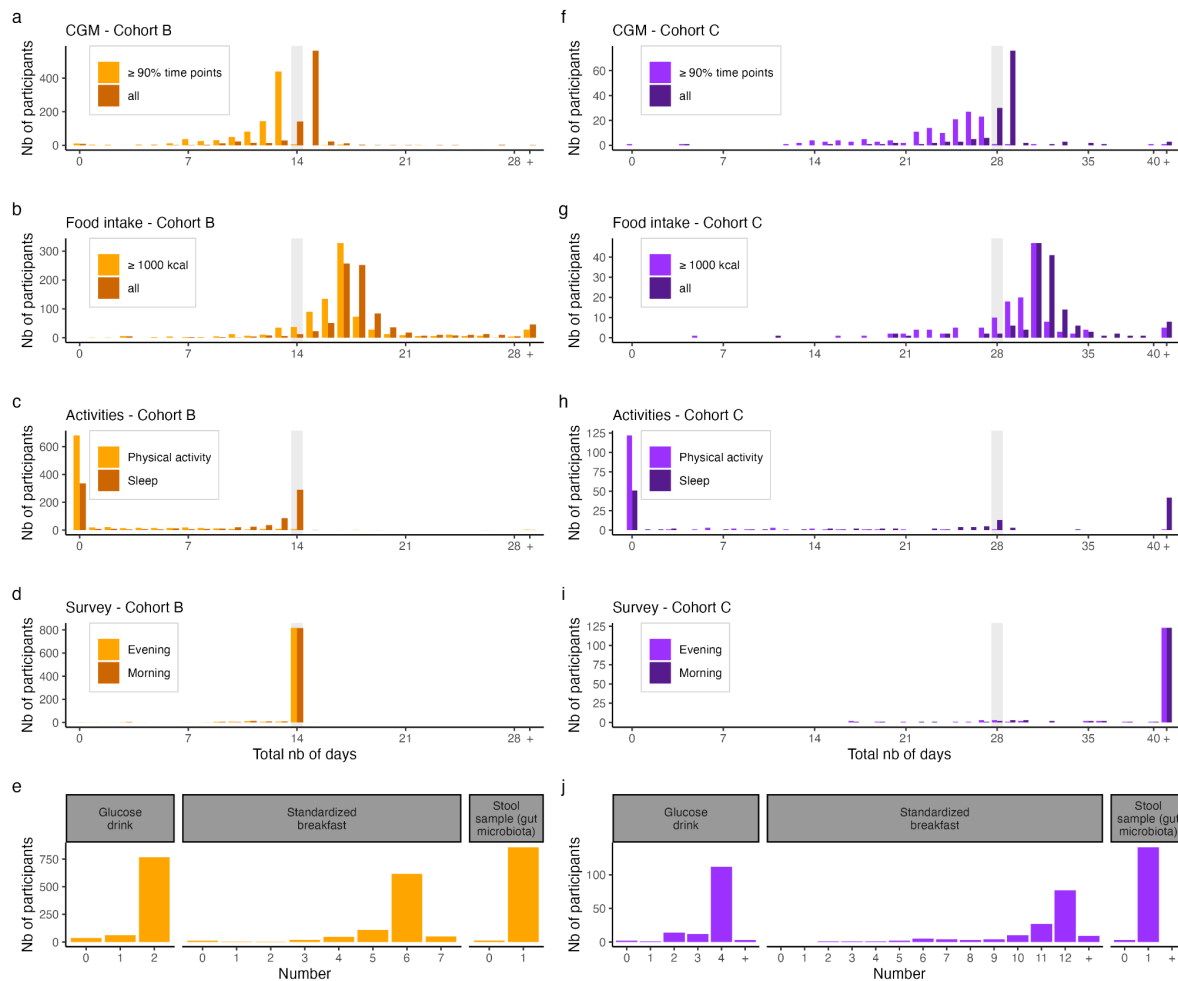
624

625

626

Figure 3

627



628

629

630 Distribution of collected data.

631 (a-d) and (f-i): Distribution of the total number of days (x-axis) with collected data by
 632 participants (y-axis) in cohort B (a-d) and cohort C (f-i) for each of the study data
 633 modality. Light gray vertical bar indicates the duration of the study for both cohorts (i.e.,
 634 the number of days for which participants were instructed to report data). Top panels
 635 (a,f) show that distribution for the blood glucose data (CGM: continuous glucose
 636 monitoring). Darker shades show the distribution in the case where days are included if
 637 there is at least 1 data point collected that day, while lighter shades show the distribution
 638 in the case that days are included only if there are data points for at least 90% of the day.
 639 2nd row panels (b, g) show the distribution for the food intake data. In a similar fashion
 640 to the top panels, darker shades show the distribution for days with any food intake, while
 641 lighter shades only include days during which total caloric intake was above 1000 kcal.

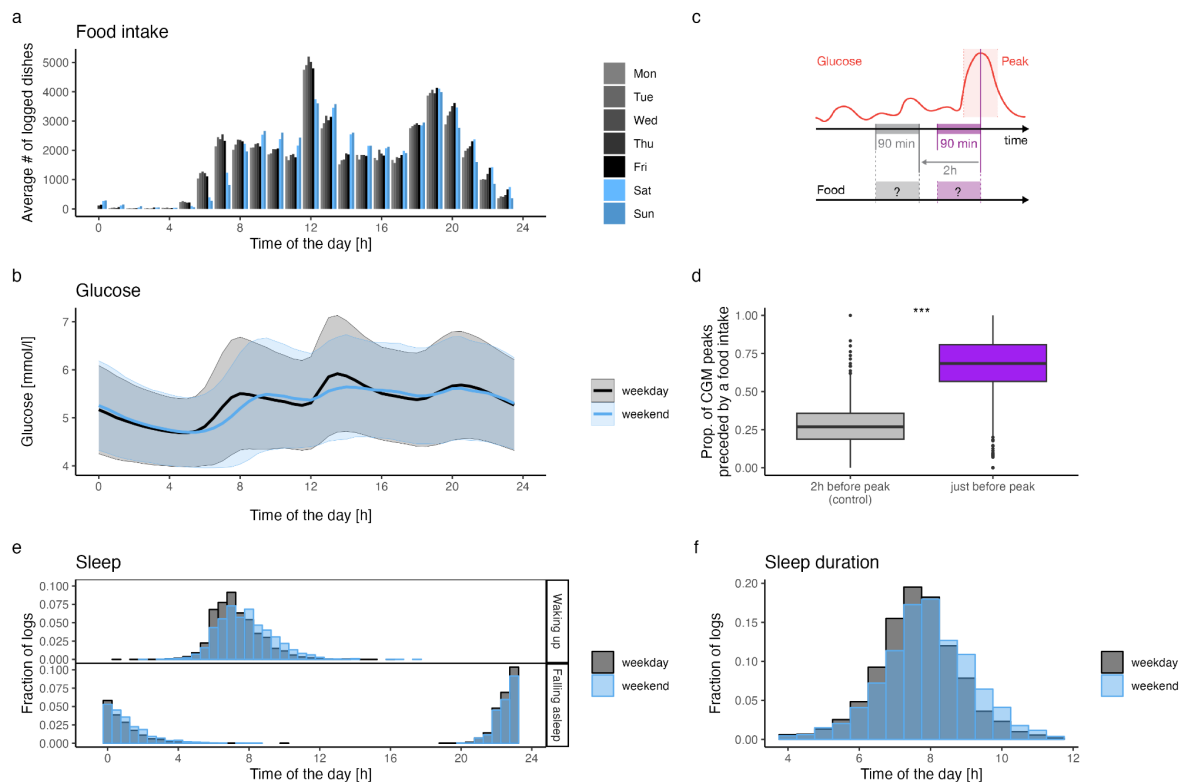
642 3rd row panels (c, h) show the distribution for the sleep and physical activity data. 4th
643 row panels (d, i) show the distribution for daily questionnaires. Days were included if
644 participants filled the morning (lighter shade) or evening (darker shade) surveys. 5th row
645 panels (e, j) show the number of glucose drink (left), standardized breakfast (center), and
646 stool samples (right) reported or sent by participants. Cohort B (resp. C) participants were
647 instructed to eat two (four) glucose drinks and 6 (12) standardized breakfasts. For panel
648 (i) and (h), note that participants in Cohort C were allowed to keep answering survey for
649 150 days after start.

650

651

652

Figure 4



653

654 **Temporal patterns of measured or self-reported data**

655 a. Total number across all participants of logged dishes (any food or drink intake) per
656 hour (x-axis) and weekday.

657 b. Mean (solid line) and 50% interquartile range (shaded areas) of the blood glucose
658 levels measured during weekdays (black) or weekends (blue).

659 c. Schematic illustrating the experiment to assess if glucose peaks are more likely to
660 be directly preceded by food intake. The time-window of interest (*i.e.*, in which
661 food intake is expected) is displayed in purple, while the control time-window is
662 displayed in gray.

663 d. Distribution of the percentage of glucose peaks directly (purple) or distantly
664 (control - gray) preceded by food intake. Percentages are computed per
665 participant: for each participant, glucose peaks are identified, food intake in the
666 two time-window is reported as a binary variable, and percentages are computed
667 as the fraction of time-windows with reported food intake.

668 e. Distribution of the times at which participants woke up (upper panel) or fell asleep
669 (lower panel) during weekdays (gray) or weekend (blue). These times are either

670 self-reported by participants when filling the morning questionnaires or obtained
671 from participants' connected devices (smartwatches) data.

672 f. Distribution of average sleep durations (x-axis, in hours) during weekdays (gray)
673 or weekends (blue).

674

675

676

677

678

679

680

681

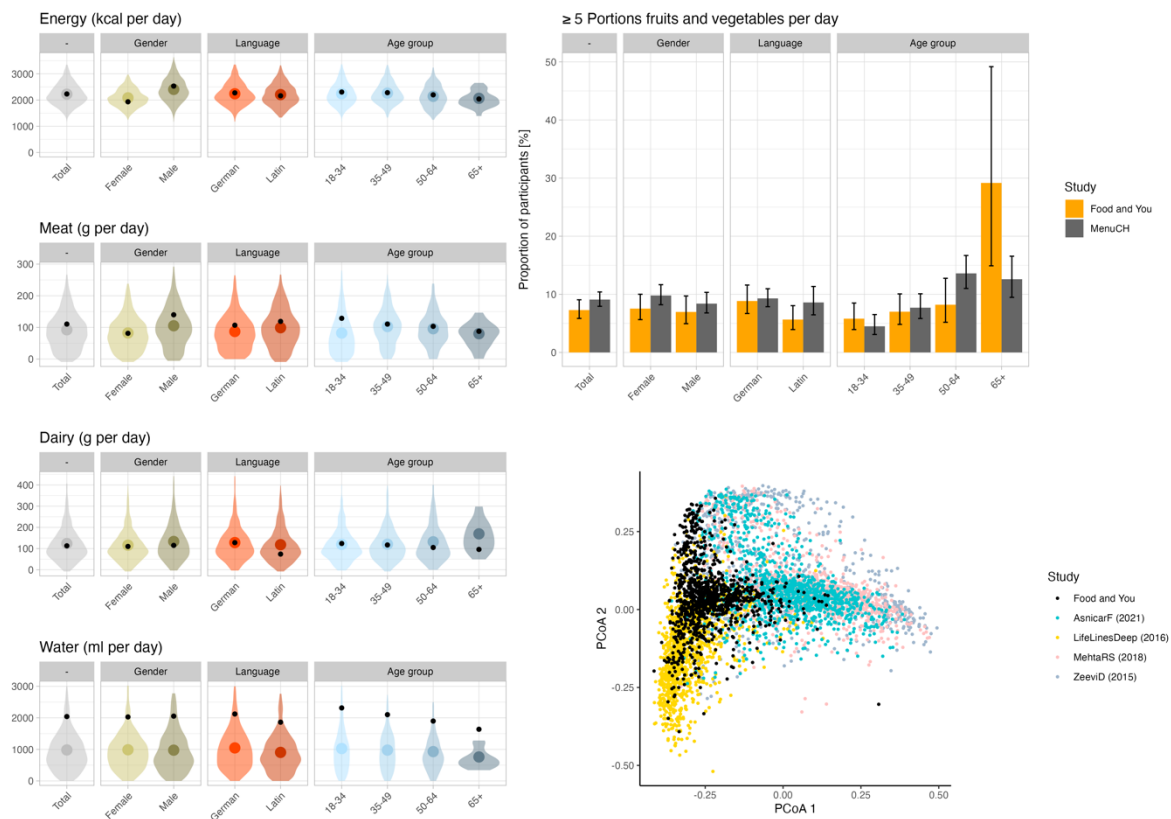
682

683

684

685

Figure 5



686

687 Comparison with other studies

688 (Left panels): Comparison with the menuCH results. Distribution of the daily intake in
689 energy (top row), meat (2nd row), dairy (third row), and water (bottom row). In each panel,
690 the colored violins and dots represent the distribution and mean of the Food & You data
691 while the smaller black dot is the mean for the corresponding population stratum
692 reported by the menuCH study. (Top right panel): Proportion of participants that, on
693 average, eat more than 5 portions of fruits and vegetables daily. For menuCH, the
694 averages are over 2 days of data collection. For “Food & You”, the average is over 14
695 (cohort B) or 28 (cohort C) days of data collection. Bar heights show the proportion of
696 participants, black whiskers show the 95% CI. (Bottom right panel): Microbiota
697 composition of F&Y participants compared to that of other cohorts. Each dot is a sample,
698 and samples are colored by cohort. The coordinates of each sample are the first two
699 coordinates resulting from a principal coordinate analysis on the Bray-Curtis distances
700 between samples (Methods).

701

702

703

704 Tables

705

706

Table 1

		Cohort B	Cohort C
Age Group	18-34	63.27 (348)	61.76 (84)
	35-49	64.23 (316)	59.18 (58)
	50-65	60.4 (183)	50.0 (2)
	65+	57.5 (23)	0.0 (0)
BMI	Normal	63.64 (574)	62.15 (110)
	Underweight	61.54 (16)	63.64 (7)
	Overweight	63.4 (220)	47.62 (14)
	Obese	57.14 (60)	58.33 (7)
Gender	female	63.24 (425)	59.5 (144)
	male	61.38 (445)	0.0 (0)
Dietary Restriction	No	62.73 (771)	56.74 (122)
	Yes	82.5 (99)	81.48 (22)
Phone Type	Android	66.44 (396)	56.14 (64)
	iPhone	59.18 (474)	62.5 (80)

707

708

709

710 **Cohort completion rates.** For each section, the table shows percentages per section and
711 cohort, and absolute numbers in parentheses. Percentages are calculated as the number
712 of participants having completed the study over the number of participants enrolled in
713 the study.

714

715

Table 2

		Cohort B (n)	Cohort C (n)
Gender	Female	48.85 (425)	100 (144)
	Male	51.15 (445)	0 (0)
Household Income	<6000	17.82 (155)	22.92 (33)
	6000-8999	20.11 (175)	24.31 (35)
	9000-12999	29.31 (255)	28.47 (41)
	>=13000	32.76 (285)	24.31 (35)
Household Type	Alone	18.28 (159)	21.53 (31)
	Couple w. Children	30.69 (267)	33.33 (48)
	Couple w.out Children	35.06 (305)	29.17 (42)
	Other	15.98 (139)	15.97 (23)
Age Group	18-34	37.82 (329)	57.64 (83)
	35-49	35.86 (312)	40.97 (59)
	50-64	23.56 (205)	1.39 (2)
	65+	2.76 (24)	0 (0)
Citizenship	Swiss	52.64 (458)	45.14 (65)
	Binational	10.69 (93)	13.19 (19)
	Foreigner	36.67 (319)	41.67 (60)
Education	Low	1.72 (15)	2.78 (4)
	Intermediate	24.71 (215)	15.28 (22)
	High	73.56 (640)	81.94 (118)
Smoking	Non Smoker	48.05 (418)	51.39 (74)
	Former Smoker	42.18 (367)	34.03 (49)
	Smoker	9.77 (85)	14.58 (21)
Health Status	Not Good to Average	8.85 (77)	6.25 (9)
	Good to very Good	91.15 (793)	93.75 (135)
Urbanity	Urban	76.55 (666)	78.47 (113)
	Rural	23.45 (204)	21.53 (31)
Linguistic Region	German	47.93 (417)	53.47 (77)
	Latin	52.07 (453)	46.53 (67)
Physical Activity	Active	66.55 (579)	65.97 (95)
	Not active	33.45 (291)	34.03 (49)
BMI	Underweighted	1.84 (16)	4.86 (7)
	Normal	65.98 (574)	76.39 (110)
	Overweighted	25.29 (220)	13.89 (20)
	Obese	6.9 (60)	4.86 (7)

716

717 **Cohort characteristics.** For each section, the table shows percentages per section and
 718 cohort, and absolute numbers in parentheses. Household income is monthly in Swiss
 719 Francs (CHF), which is roughly in parity with USD.

720 References

- 721 1. Papamichou, D., Panagiotakos, D. B. & Itsiopoulos, C. Dietary patterns and
722 management of type 2 diabetes: A systematic review of randomised clinical trials.
723 *Nutrition Metabolism Cardiovasc Dis* **29**, 531–543 (2019).
- 724 2. Rodríguez-Monforte, M., Flores-Mateo, G. & Sánchez, E. Dietary patterns and CVD: a
725 systematic review and meta-analysis of observational studies. *Brit J Nutr* **114**, 1341–1359
726 (2015).
- 727 3. Grosso, G. et al. Mediterranean Diet and Cardiovascular Risk Factors: A Systematic
728 Review. *Crit Rev Food Sci* **54**, 593–610 (2014).
- 729 4. Key, T. J. et al. Diet, nutrition, and cancer risk: what do we know and what is the way
730 forward? *Bmj* **368**, m511 (2020).
- 731 5. Hood, L. & Friend, S. H. Predictive, personalized, preventive, participatory (P4) cancer
732 medicine. *Nat Rev Clin Oncol* **8**, 184–187 (2011).
- 733 6. Celis-Morales, C. et al. Effect of personalized nutrition on health-related behaviour
734 change: evidence from the Food4Me European randomized controlled trial. *Int J*
735 *Epidemiol* 578–588 (2016) doi:10.1093/ije/dyw186.
- 736 7. Zeevi, D. et al. Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163**,
737 1079–1094 (2015).
- 738 8. Mendes-Soares, H. et al. Assessment of a Personalized Approach to Predicting
739 Postprandial Glycemic Responses to Food Among Individuals Without Diabetes. *Jama*
740 *Netw Open* **2**, e188102 (2019).
- 741 9. Berry, S. E. et al. Human postprandial responses to food and potential for precision
742 nutrition. *Nat Med* **26**, 964–973 (2020).
- 743 10. Hoevenaars, F. P. M. et al. Evaluation of Food-Intake Behavior in a Healthy
744 Population: Personalized vs. One-Size-Fits-All. *Nutrients* **12**, 2819 (2020).
- 745 11. Collaboration, E. R. F. et al. Diabetes mellitus, fasting blood glucose concentration,
746 and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies.
747 *Lancet* **375**, 2215–2222 (2010).
- 748 12. Mohanty, S. P. et al. The Food Recognition Benchmark: Using Deep Learning to
749 Recognize Food in Images. *Frontiers Nutrition* **9**, 875143 (2022).
- 750 13. Federal Food Safety and Veterinary Office. Swiss Food Composition Database.
751 <https://valeursnutritives.ch/en>.
- 752 14. Chatelan, A. et al. Lessons Learnt About Conducting a Multilingual Nutrition Survey
753 in Switzerland: Results from menuCH Pilot Survey. *Int J Vitam Nutr Res* **87**, 25–36 (2017).
- 754 15. French Agency for Food Environmental and Occupational Health and Safety. ANSES-
755 CIQUAL food composition table. <https://ciqual.anses.fr/>.
- 756 16. Lazzari, G., Jaquet, Y., Kebaili, D. J., Symul, L. & Salathe, M. FoodRepo: An Open Food
757 Repository of Barcoded Food Products. *Frontiers in nutrition* **5**, 57 (2018).
- 758 17. Dinauer, M., Winkler, A. & Winkler, G. Moncia Mengenliste. [https://www.ble-](https://www.ble-medienservice.de/0654/monica-mengenliste)
759 [medienservice.de/0654/monica-mengenliste](https://www.ble-medienservice.de/0654/monica-mengenliste) (1991).

- 760 18. Prüße, U., Hüther, L. & Hohgardt, K. Mittlere Gewichte einzelner Obst- und
761 Gemüseerzeugnisse.
762 [https://www.bvl.bund.de/SharedDocs/Downloads/04_Pflanzenschutzmittel/rueckst](https://www.bvl.bund.de/SharedDocs/Downloads/04_Pflanzenschutzmittel/rueckst_gew_obst_gem%C3%BCde_pdf.pdf?__blob=publicationFile)
763 [_gew_obst_gem%C3%BCde_pdf.pdf?__blob=publicationFile](https://www.bvl.bund.de/SharedDocs/Downloads/04_Pflanzenschutzmittel/rueckst_gew_obst_gem%C3%BCde_pdf.pdf?__blob=publicationFile) (2002).
- 764 19. Zuppinger, C. et al. Performance of the Digital Dietary Assessment Tool
765 MyFoodRepo. *Nutrients* **14**, 635 (2022).
- 766 20. Ólafsdóttir, A. F. et al. A Clinical Trial of the Accuracy and Treatment Experience of
767 the Flash Glucose Monitor FreeStyle Libre in Adults with Type 1 Diabetes. *Diabetes*
768 *Technol The* **19**, 164–172 (2017).
- 769 21. Scott, E. M., Bilous, R. W. & Kautzky-Willer, A. Accuracy, User Acceptability, and
770 Safety Evaluation for the FreeStyle Libre Flash Glucose Monitoring System When Used
771 by Pregnant Women with Diabetes. *Diabetes Technol The* **20**, 180–188 (2018).
- 772 22. Tsoukas, M. et al. Accuracy of FreeStyle Libre in Adults with Type 1 Diabetes: The
773 Effect of Sensor Age. *Diabetes Technol The* **22**, 203–207 (2020).
- 774 23. Blum, A. Freestyle Libre Glucose Monitoring System. *Clin Diabetes Publ Am Diabetes*
775 *Assoc* **36**, 203–204 (2018).
- 776 24. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for
777 next-generation sequencing reads. *Bioinformatics* **31**, 3476–3482 (2015).
- 778 25. Edgar, R. C. Accuracy of microbial community diversity estimated by closed- and
779 open-reference OTUs. *Peerj* **5**, e3889 (2017).
- 780 26. Edgar, R. C. UNBIAS: An attempt to correct abundance bias in 16S sequencing, with
781 limited success. *Biorxiv* 124149 (2017) doi:10.1101/124149.
- 782 27. Edgar, R. C. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS
783 sequences. *Biorxiv* 074161 (2016) doi:10.1101/074161.
- 784 28. Buuren, S. V. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained
785 equations in R. *Journal of statistical software* **45**, 1–67 (2011).
- 786 29. Bull, F. C. et al. World Health Organization 2020 guidelines on physical activity and
787 sedentary behaviour. *Brit J Sport Med* **54**, 1451–1462 (2020).
- 788 30. Battelino, T. et al. Clinical Targets for Continuous Glucose Monitoring Data
789 Interpretation: Recommendations From the International Consensus on Time in Range.
790 *Diabetes Care* **42**, 1593–1603 (2019).
- 791 31. Asnicar, F. et al. Microbiome connections with host metabolism and habitual diet
792 from 1,098 deeply phenotyped individuals. *Nat Med* **27**, 321–332 (2021).
- 793 32. Zhernakova, A. et al. Population-based metagenomics analysis reveals markers for
794 gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
- 795 33. Mehta, R. S. et al. Stability of the human faecal microbiome in a cohort of adult men.
796 *Nat Microbiol* **3**, 347–355 (2018).
- 797 34. Abdullah, A., Peeters, A., Courten, M. de & Stoelwinder, J. The magnitude of
798 association between overweight and obesity and the risk of diabetes: A meta-analysis of
799 prospective cohort studies. *Diabetes Res Clin Pr* **89**, 309–319 (2010).
- 800 35. Pratap, A. et al. Indicators of retention in remote digital health studies: a cross-study
801 evaluation of 100,000 participants. *Npj Digital Medicine* **3**, 21 (2020).

- 802 36. Chatelan, A. et al. Major Differences in Diet across Three Linguistic Regions of
803 Switzerland: Results from the First National Nutrition Survey menuCH. *Nutrients* 9, 1163
804 (2017).
- 805 37. Krieger, J.-P. et al. Dietary Patterns and Their Sociodemographic and Lifestyle
806 Determinants in Switzerland: Results from the National Nutrition Survey menuCH.
807 *Nutrients* 11, 62 (2018).
- 808 38. Faeh, D., Marques-Vidal, P., Chiolero, A. & Bopp, M. Obesity in Switzerland: do
809 estimates depend on how body mass index has been assessed? *Swiss Med Wkly* **138**,
810 204–210 (2008).