

## Article

# Estimating Dietary Intake from Grocery Purchase Data - A Comparative Validation of Relevant Indicators.

Jing Wu<sup>1</sup>, Klaus Fuchs<sup>2,\*</sup> , Jie Lian<sup>1</sup>, Mirella Haldimann<sup>3</sup>, Tanja Schneider<sup>1,\*</sup> , Simon Mayer<sup>1</sup>, Jaewook Byun<sup>4</sup>, Roland Gassmann<sup>5</sup>, Christine Brombach<sup>5</sup>, Elgar Fleisch<sup>2</sup>

<sup>1</sup> University of St. Gallen; jing.wu@unisg.ch, jie.lian@unisg.ch, tanja.schneider@unisg.ch, simon.mayer@unisg.ch

<sup>2</sup> ETH Zurich; fuchsk@ethz.ch, efleisch@ethz.ch

<sup>3</sup> D-One Solutions AG; mirella.haldimann@d-one.ai

<sup>4</sup> Sejong University; Jaewook Byun jwbyun@sejong.ac.kr

<sup>5</sup> Zurich University of Applied Science; garo@zhaw.ch, broc@zhaw.ch

\* Correspondence: fuchsk@ethz.ch; Tel.: +41-78-858-7037

† These authors contributed equally to this work.

**Abstract:** In light of the globally increasing prevalence of diet-related non-communicable diseases (NCDs), new scalable and non-invasive dietary monitoring techniques are urgently needed. Automatically collected digital receipts from loyalty cards have the potential to serve as an objective and automatically traceable digital biomarker for individual food choice behavior and do not require patients to manually log each individual meal item. Until recently, such electronic purchase records were hard to collect for researchers and were only validated in national empirical studies. Multiple quantitative indicators for purchase quality have been suggested, but so far no comparison has validated the potential of these alternative indicators to discriminate between health-beneficial and -detrimental food choices. With the introduction of the General Data Privacy Regulation in the European Union, millions of consumers gained the right to access their purchase data in a machine-readable form, representing a historic chance to leverage purchase data for scalable monitoring of food choices. This study hence is the first study comparing the calibration capacity and validating the discrimination potential of previously suggested purchase indicators for the nutritional quality of purchased groceries, incl. HEI-2015, HETI, GPQI, and FSA-NPS DI. To assess the indicators' potential, 464 study participants were asked to complete a validated food frequency questionnaire (FFQ) and to donate their digital receipts from the loyalty card programs of the two leading Swiss grocery retailers, representing 69% of the national grocery market. 89 participants fulfilled the eligibility criteria, i.e. completed the FFQ and were frequent users of the loyalty card systems. Compared to absolute food and nutrient intake, correlations between density-based relative food and nutrient intake and food purchase data are stronger. Counterintuitively, although the frameworks of the HETI and the GPQI are centered around food groups, both indicators do not capture food group intake such as vegetables or sweets very well. The FSA-NPS DI has the best calibration and discrimination performance in classifying participants' consumption of nutrients and food groups, and seems to be a superior indicator to estimate nutritional quality of a user's diet based on digital receipts from grocery purchases in Switzerland.

**Keywords:** dietary intake; diet monitoring; digital receipts; purchase quality indicators; FSA-NPS DI



**Citation:** Wu, Jing; Fuchs, Klaus; Lian, Jie; Haldimann, Mirella; Schneider, Tanja; Mayer, Simon; Byun, Jaewook; Gassmann, Roland; Brombach, Christine; Fleisch, Elgar  
Title. *Preprints* 2021, 1, 0.  
<https://doi.org/>

Received:  
Accepted:  
Published:

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## 1. Introduction

The globally increasing prevalence of diet-related non-communicable diseases (NCDs) including obesity, diabetes, certain types of cancer, and cardiovascular diseases represents a continuously growing burden for affected patients and health-care systems alike [1–4]. Due to societal trends such as urbanization and the transformation of food systems towards more processed, convenience, and fast food, dietary patterns around the world show an increasing consumption of (added) sugar, sodium, saturated fats, and calorific energy [5], thereby elevating the risks of diet-related NCDs [6–8]. Besides the uptake of food items of low nutritional quality, the global demand for meat, fish, exotic fruits throughout the

year, and the fact that one third of all food is wasted takes a significant toll on the planetary ecosystem as well [1]. Dealing with the burden of unhealthy dietary patterns, monitoring and eventually improving food choices via automatic and scalable monitoring tools are of key importance. Scalable monitoring tools enable the identification of at-risk populations and the support of mitigating health-related and/or environmentally-related concerns that may be harmful for the individual.

Unfortunately, contemporary monitoring solutions such as paper-based or digital recall diaries or food frequency questionnaires (FFQ) rely on manual logging of each single meal item, leading to high user attrition rates, under-reporting and self-selection of primarily healthy users. To analyze an individual's dietary pattern, dietitians and researchers typically rely on traditional dietary assessment approaches. Weighed food diaries or records and 24-hour recalls are widely used as the primary validation tools of food intake [9–11]. The high effort involved in collecting and transcribing such manual log leads to a substantial burden on the user as well as the supervisor side. For example, manual diet logging apps (e.g. MyFitnessPal) are only actively used by 8% of smartphone users [12]. Hence, the traditional approaches mainly remain short-lived and lead to relatively high costs, making them inappropriate for large-scale epidemiological studies or interventions. To make food logging more convenient and the sub-sequential transcription towards intake amounts easier, FFQs were suggested. In comparison to food diaries, FFQs contain a predefined, exhaustive list of relevant food items that respondents confirm or deny of having consumed in a given time period (e.g. a day or month). Hence, FFQs are relatively inexpensive compared to food diaries and do not require trained interviewers. Similar to food diaries, they are however also limited by the fact that specified portion sizes are difficult to estimate and that the correct capture of consumed food items remains moderated by memory bias. Consequently, results derived from FFQ could potentially be imprecise, e.g. underestimating the saturated fatty acid intake [13]. In addition, FFQs might not include all relevant food items that subjects have consumed. Finally, since FFQs feature exhaustive lists of food items, their purpose is usually designed towards specific nutrients, e.g. sodium, rather than capturing all health- or sustainability-related intakes. To allow for a more holistic food choice monitoring that can mitigate the ongoing obesity epidemic and nudge consumers towards more sustainable food choices, more scalable food choice monitoring tools are needed that do not suffer from the aforementioned biases [14–16].

With the proliferation of digital payments and loyalty cards, digital receipts from grocery purchases can serve as a novel scalable, automatically capturing and continuously self-updating monitoring tool for food choices. Digital receipts as machine-readable, electronic substitutes for their contemporary paper-based printed counterparts are expected to be adopted around the globe over the next decade as they promise significant advantages in regard to environmental footprint [17], mitigating tax evasion [18,19] and offering superior advantages and transparency for consumers [20]. Previous research shows that receipts from grocery purchases correlate with and predict individual diets [21–26]. For example, researchers were able to show that the dietary calorific intake of a person correlates with the amount of energy-dense food products purchased. Especially since in the case of salt, where 80% of sodium intake originates from packaged products [27], purchase data is promising for diet monitoring.

An additional key aspect is mapping electronic shopping history to study participants through loyalty cards for example (Ni Mhurchu et al., 2007, 2010; Taylor et al., 2015; Brewster et al., 2017 a). Validity of dietary intake assessed through purchase data is mostly tested with common epidemiological individual dietary intake methods, i.e. FFQ, 24HR, FR (as described in 2.1.1) and results show moderate to strong agreements depending on the method (Nelson, 1985; Ransley et al., 2001; Raynor et al., 2004; Fulkerson et al., 2008; Eyles et al., 2010 a). Relatively few studies have been able to analyze purchase data based on comparison(s) between groups of participants, e.g. overweight vs. lean, or high income vs. low income (Nelson, 1985; Ransley et al., 2003; Martin et al., 2006; Tang et al., 2016; Brewster

et al., 2017 a). In these studies, factors of interest were collected together with purchase and nutritional data. Yet the focus remained often limited since they did not involve more than one factor (e.g. biometrics, socioeconomic background, shopping behavior, household composition). Correlating factors of interest with purchase and nutritional data lead to the development of indexes for describing food purchase nutritional quality (Taylor et al., 2015; Tran et al., 2015 a; Brewster et al., 2017 a).

In the past, research in the domain under study has been strongly limited by restricted access to digital receipt data, considerable effort involved with collecting product data, and small sample sizes. Due to the introduction of the General Data Protection Regulation (GDPR) [28], millions of customers in the European Union recently gained the right to access their digital receipts from loyalty cards, and researchers may obtain this data with the subjects' consent. Therefore, product purchases – including food products – that are recorded on loyalty cards are now traceable, distinguishable, and analysable. By itself a digital receipt does not contain any nutritional details of food products, however. To this end, food product composition databases, - which were mandated by the regulation on mandatory declaration of nutrients on food items sold online (EU)1169 [29], - allow the data fusion of digital receipts and nutritional information on the purchased products. But to identify whether a purchase history of a user is healthy or not, quantitative indicators are needed to offer a normative, reliable, interpretable basis for comparison between users, retailers and regions.

Although multiple quantitative purchase indicators have been suggested to evaluate the nutritional quality of food and beverage purchase records, including the Healthy Trolley Index (HETI) [30], the Grocery Purchase Quality Index-2016 (GPQI) [31] and the Healthy Purchase Index (HPI) [32], there does not yet exist any published assessment that compared the discrimination potential of these indicators. All suggested indicators assign food items into different categories and thereby allow estimating how balanced a given user's purchase history, and hence the respective individual's diet might be. Taking local dietary guidelines as the golden standard, these indicators might not necessarily be directly 'transferable' to all regions. In addition, the expenditure share of food was used in these indices to calculate the compliance to corresponding dietary guidelines. In absence of food composition data, price-based indicators might make sense. But if available, purchased quantities in grams or milliliters would give a more accurate representation of the dietary impact from the evaluated purchase records rather than price-based comparisons. In addition to the three suggested purchase indices, traditional diet indices can be used to evaluate purchase behavior. It was suggested that the Healthy Eating Index-2010 (HEI-2010) scores derived from food purchases showed moderate agreement and minimal bias with HEI-2010 scores from 24-h recalls [33]. Thus, using purchase data to calculate HEI-2015 [34], the latest version of HEI, might be a feasible way of assessing shopping quality from digital receipts. Last, but not least, the British Food Standards Agency Nutrient Profiling System Dietary Index, FSA-NPS DI [35], from which the Nutri-score is calculated [36], can also be utilized to assess shopping quality. Instead of assigning food to different food groups, FSA-NPS DI takes all food items into account, and focuses on the overall nutritional quality of single products rather than the diet balance or basket composition. Despite the existence of multiple shopping indices, it remains unclear which of the proposed indicators has superior discrimination potential in distinguishing between healthy and unhealthy user behavior. The main goal of this study is hence to compare the calibration and discrimination ability of the aforementioned indices, namely HETI, GPQI, HEI-2015 and FSA-NPS DI. Because the HPI is also based on micro-nutrients, which are usually not available in food composition databases, this index is not included in our comparison. Still, the results should equip researchers, practitioners, and policy-makers with the insights required to select among the existing indices, and the conclusions might be relevant for dietary health researchers working on novel indicators as well.

## 2. Materials and Methods

This manuscript describes the first study comparing the calibration capacity and validating the discrimination potential of multiple previously suggested purchase indicators for the nutritional quality of purchased groceries, including HEI-2015, HETI, GPQI, and FSA-NPS DI. In the following, the digital receipt integration, food composition database and study design are introduced.

### 2.1. Digital receipt integration

The digital receipt infrastructure was implemented in Switzerland, due to the availability of digital receipts from the loyalty card systems of the two leading supermarket chains. To support the comparative analysis of the suggested quantitative purchase indicators, a technical setup was designed and eventually implemented to allow the collection of digital receipts from users who consented to participate in the study. The study was deployed on the Bitsaboutme (BAM) online platform<sup>1</sup>. BAM is a GDPR-compliant data marketplace service located in Switzerland, which allows users to request their own personal user data from data controllers and store their data in an encrypted data vault. Just as data sources like social media or messaging services process personal data, also financial transactions from bank accounts and digital receipts are considered personal data by the GDPR. Hence, users of the BAM service are able to retrieve their digital receipts from the two leading Swiss loyalty card providers, namely Migros Cumulus and/or their Coop Supercard data. In this regard, Switzerland can be considered as an ideal region to validate quantitative purchase indicators, as just these two leading Swiss grocery retailers, i.e. Migros and Coop respectively, represent 76% of the national B2C grocery market [37]. More concretely, the Swiss consumers can be considered frequent users of loyalty cards in general, as loyalty cards are used for a total of 80% of Swiss retailers' revenue [37,38], thus representing an promising study location for the comparative analysis of the suggested quantitative purchase indicators. In this respect, Switzerland is certainly a few years or even decades ahead of many countries in terms of adopting digital receipts. In regions other than Switzerland, however, digital receipts are likely to also be implemented and adopted in the future due to the steadily increasing acceptance of digital payment methods such as credit cards or mobile payment, and the expected transition from paper to digital receipts. Hence, we believe that the results and implications from carrying out this study in Switzerland can be generalized towards other regions as well.

Once users decide to donate their digital receipts to the study via the BAM service, they need to agree to multiple opt-in consent forms before their historic and future receipts can be integrated into the study (See Figure 4). To join the study, users had to opt-in at least four times before their digital receipts would become part of the sample analyzed in this study. First, prospective study participants needed to be enrolled into at least one of the two Swiss loyalty card systems. Consequentially, users also needed to accept the terms and conditions of at least one or even both of the loyalty card providers and opt-in towards collecting digital receipts in digital form. Second, prospective participants needed to join the BAM service before they could participate in the study. Hence, they agreed to the terms & conditions of the BAM service to allow the service to retrieve their personal data from data controllers on their behalf. Third, users who already collected digital receipts from their loyalty cards need to consent the BAM service to retrieve their digital receipts on their behalf from the loyalty card system provider directly. In one case, i.e. Migros Cumulus, this is done by linking the corresponding online account, similar to using a Facebook connect to hand over user data. In the other case, i.e. Coop Supercard, a user needs to share email-based digital receipts with the BAM service. Only then, a user's historic (i.e. up to two year history) and new digital receipts - i.e. those that are created every time a user buys groceries and uses the own loyalty card at the supermarket checkout from now on - are then automatically imported into the BAM service and stored in a standardized

<sup>1</sup> See <https://bitsabout.me/>

form in the personal BAM data vault. Fourth and finally, a prospective user then had to join our study which was displayed towards eligible users on the BAM service platform. A BAM service user was considered eligible for the study, once a user has successfully linked at least one Swiss loyalty card to the BAM service and at least 40 product items were stored in the personal data vault. The call to join the study presented in this manuscript was then displayed to eligible BAM service users who could then decide to consent to the study protocol and to donate their digital receipts to the study.

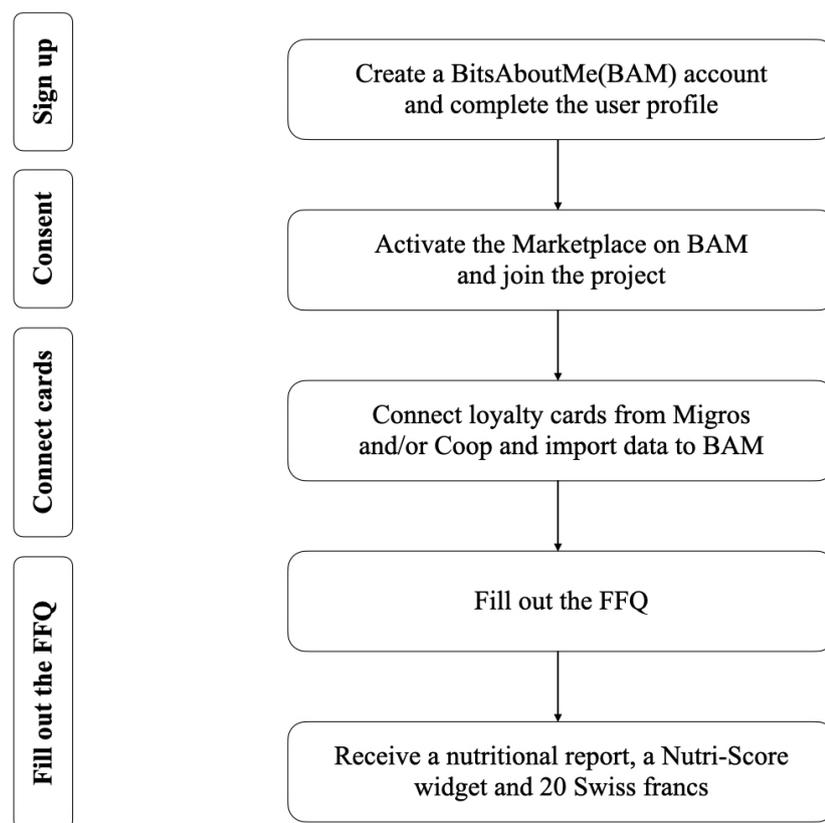
The study and its consent form were approved by the Ethics Committee of ETH Zurich (protocol code 2019-N-134, approved on October 15th, 2019) prior to the launch of the study. In particular, the study protocol and the consent form on the BAM service required the anonymization of the donated digital receipt data. Concretely, no directly identifying personal data such as names, email addresses, phone numbers or loyalty card identifiers were shared by the BAM service with the study. To ensure the anonymity of the data donors, even the shopping locations and time of day were removed from the digital receipt dataset before the analyses conducted in this manuscript. Hence, each receipt in the final digital receipt dataset donated by the N=464 users of the BAM service who participated in the study (see Figure 4) only contained a randomized study identifier of the respective user, the day of the year of the purchase, the purchased quantity amount, the identifier of the food item that was bought and the price and potentially applied discounts. The necessary effort to potentially re-identify individual consumers from a potentially maliciously acquired copy of the anonymized dataset was communicated to prospective study participants in the consent form and considered an acceptable study risk, given the study's contribution.

## 2.2. Food composition database

To conduct the comparison of purchase indicators, the anonymized digital receipt dataset was enriched with data on the nutritional composition of purchased food items, as digital receipts itself usually do not contain such information. When using receipts as proxies for food choices, a common challenge is mapping food products captured via printed or digital receipt to nutrient information [22,39]. Hence, this study's authors leveraged an existing food composition database containing detailed information on over 50'000 grocery products frequently sold and consumed in Switzerland [40,41]. Driven by the recent mandates for online food nutrition databases [29], there are now trusted, curated databases (such as GS1 trustbox) as well as crowdsourced databases (e.g. OpenFoodFacts) available to retrieve detailed nutritional information on products sold in a retail environment. This information becomes particularly useful when combined with a consumer's shopping history.

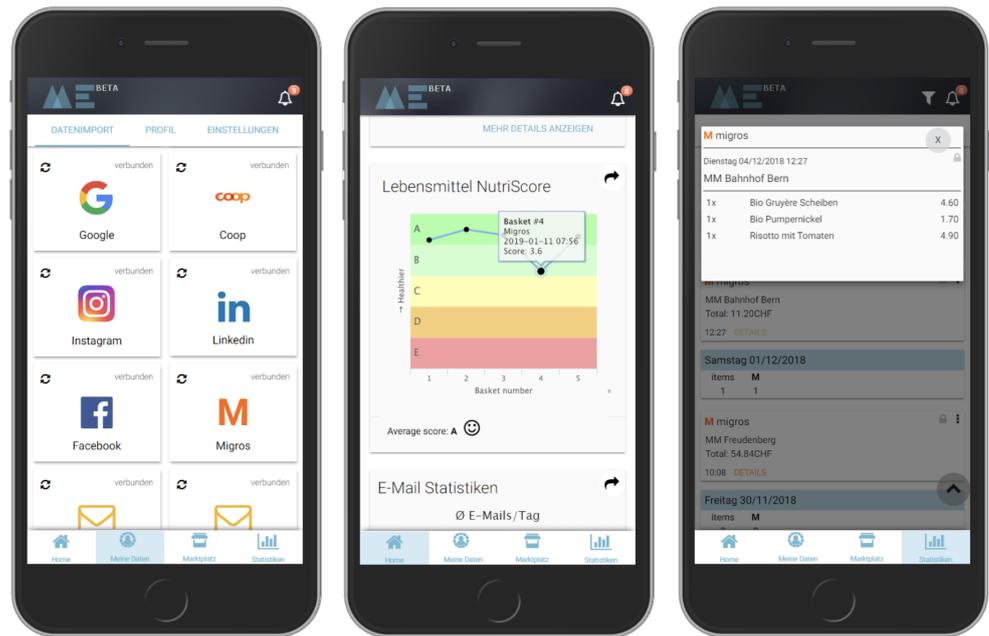
The required mapping between digital receipt based product identifiers and Global Trade Item Number (GTIN) was done manually and enabled a detection of ca. 69% of the products present in the anonymized digital receipt dataset. In food composition databases, products sold in retail environments are usually identified via their GTIN, a globally unique product identifier distributed by GS1, a globally operating non-profit standards organization. Unfortunately, paper or digital receipts usually only contain a product's name as identifiers. Identifying a product's GTIN from a digital receipt, represents one of the key challenges in digital receipt based monitoring of food choices. Future regulatory mandates, e.g. GDPR 2.0, could enforce the use of available industry standards to ease the barriers for data portability of personal datasets. Until then, manual mapping of article identifiers on digital receipts (e.g. product names) to GTINs needs to be done manually. Via collaboration and crowd-sourcing, such manual efforts could be shared, in order to allow other researchers and practitioners to re-use and contribute to the identifier mappings already done by fellow researchers. The product matching was achieved using the data attributes in the digital receipt, namely the retailer at which a specific product was bought, its product name and its price. As some products' names alone were ambiguous, this approach proved to be viable process for the study. To maximise the frequency with

which purchased items are identified, the manuscript's authors chose a heuristic strategy of mapping the most frequently occurring articles within the anonymized digital receipt dataset. In total, 5'950 products' article identifiers from the digital receipts were mapped to corresponding GTINs and in turn to their nutritional composition. Overall, this manual mapping which took significant effort led to the detection of circa 69% of the products present in the anonymized digital receipt dataset consisting of purchases from N=464 users from two loyalty card systems. For each of the matched products, its attributes such as nutrition details (e.g. calorific energy and macro-nutrients such as protein, carbohydrate, sugar, fat, saturated fat, dietary fiber, and minerals such as sodium, all per 100 g or ml of product; 1 g = 1 ml), its logistical data (e.g. product size in l, kg, g or ml), product images, allergens, ingredients were available for the analysis.



**Figure 1.** Participant onboarding flow

The study setup described in this manuscript not only allowed for the post-hoc analysis of nutritional quality of purchased food items from digital receipts within this study, but also allowed study participants to analyze the nutritional quality of their purchases in near-time after joining the study (See Figure 2). After joining the study on the BAM service website successfully, participants would gain access to a new widget that would show users the aggregated Nutri-Score of their recent purchases. The analysis was provided via an API and demonstrated the potential of assessing the nutritional quality of digital receipts for tailored interventions to consumers, aimed at supporting healthy food choices.



**Figure 2.** Enriching digital receipts with food composition data and displaying the weight-averaged Nutri-Scores of aggregated baskets.

### 2.3. Food frequency questionnaire

After having donated their digital receipts, prospective study participants were encouraged to also complete a previously validated FFQ in order to collect ground truth data on individual diets for the validation of the purchase indicators. Multiple, alternative FFQs were identified and evaluated for the study. A systematic literature research and the assessment of a meta platform that compares available FFQs<sup>2</sup> has led to the comparison of five FFQs and one web-mediated 24h recall tool. Concretely, the VioCare FFQ<sup>3</sup>, the Ernahrungserhebung FFQ provided by Zurich University of Applied Science<sup>4</sup> [42], the DHQ3 FFQ<sup>5</sup>, the Block FFQ<sup>6</sup>, the MiniMeal-Q FFQ<sup>7</sup> were compared. In the comparison, aspects such as cost, required time, Swiss aptitude, accuracy, a published validation and setup effort were structurally assessed. Finally, the web-mediated FFQ<sup>8</sup> provided by Zurich University of Applied Sciences (ZHAW) was selected (See Figure 3), primarily due to its previous validation in Switzerland, which also was the focus region of this study [42], as well as its easy-to-use web-based administration. The chosen FFQ took an average of around 30 minutes for each of the N=181 users who filled out the FFQ completely (i.e. at least 70% which included all diet-related questions). The technical link between the BAM service and the web-based FFQ was realized via personalized links. Concretely, participants who agreed to the study consent on the BAM service received an email from BAM and where thus invited via a personalized link which included an pseudonymous user identifier that identified a user's digital receipts as well as the user's FFQ responses. Table 1 shows the daily food and nutrient intake of participants, based on the self-reported FFQ results.

An overview of the FFQ's questions and potential answers to can be found in the link provided in the appendix as well as online (See <https://r2n.ernaehrungserhebung.ch/>). All donated data, i.e. digital receipts as well as the FFQ responses, were anonymized

<sup>2</sup> See <https://www.nutritools.org/>

<sup>3</sup> See <https://vioscreen.com/>

<sup>4</sup> See <https://r2n.ernaehrungserhebung.ch/>

<sup>5</sup> See <https://www.dhq3.org/login/>

<sup>6</sup> See <https://nutritionquest.com/login/>

<sup>7</sup> See <https://www.nutritools.org/tools/199>

<sup>8</sup> See <https://r2n.ernaehrungserhebung.ch/>

Table 1: *Absolute* individual daily nutritional intake, N = 89.

Category	Mean	Standard deviation
Meat, portions/day	1.17	1.20
Vegetables, portions/day	2.55	1.71
Fruits, portions/day	1.38	1.16
Whole grains, portions/day	0.32	0.33
Sweets, portions/day	2.93	1.95
Sodium, g/day	2.1	1.5
Dietary fibers, g/day	27.1	14.3
Saturated fatty acids, g/day	37.5	26.0
Added sugar, g/day	10.4	8.52

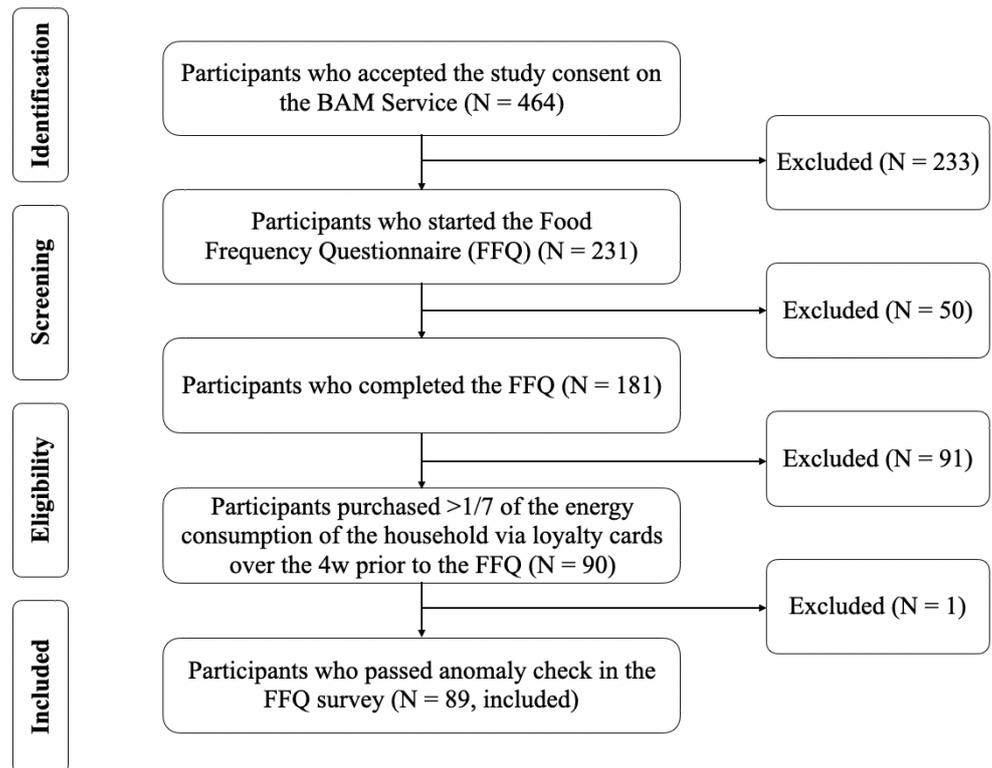
prior to the analyses conducted in this study. After successfully finishing these tasks, all participants received a nutritional assessment report based on their FFQ (See Figure 3) and an automatically self-updating Nutri-Score widget based on their recently purchased grocery baskets, displayed within the BAM service website (See 2). In addition, participants received a financial compensation of 20 Swiss Francs (CHF) (i.e. 21.80 USD, June 28th, 2021) for joining the study and filling out the FFQ.

Figure 3. Web-mediated Food Frequency Questionnaire (FFQ)

#### 2.4. Study participants

The presented study was deployed on the BAM service online platform<sup>9</sup> and was advertised together with an invitation link through a variety of channels, including BAM's marketplace, BAM's newsletter, social media and billboards on the local university campus. This approach of using BAM's marketplace and the university network ensured to address younger as well as more mature households, as the socio-demographic characteristics on the BAM website represent more mature consumers (See Table 2).

<sup>9</sup> See <https://bitsabout.me/>



**Figure 4.** The procedure of excluding ineligible participants.

Participants were recruited from December 22nd, 2018, to June 10th, 2021, and used their own devices (e.g. laptops and mobile phones) to enroll in the study. As explained in the previous chapter, the enrollment required having and actively using at least one of the two supported loyalty cards, as well as at least four opt-ins towards the BAM service, at least one of the loyalty card services, connecting the card to the BAM service, and finally opting into the study consent form online. These relatively high barriers to join the study led to a slow uptake in terms of participants. Hence, the participant recruiting was conducted on a rolling basis of over two years in order to collect the required sample size needed to conduct a robust comparison of the suggested purchase quality indicators. A detailed overview of the enrollment process is displayed in Figure 1.

Table 2: Demographic summary of participants.

<b>Sample</b>	<b>Count (%)</b>
<b>Gender</b>	
Male	68 (76.4)
Female	21 (23.6)
Other	0 (0.0)
<b>Age [yrs]</b>	
18-29	29 (32.6)
30-39	29 (32.6)
40-49	18 (20.2)
>50	13 (14.6)
<b>BMI [kg/m<sup>2</sup>]</b>	
Underweight (<18.5)	2 (2.3)
Normal (i.e. $\geq 18.5$ and $< 25.0$ )	55 (61.8)
Overweight (i.e. $\geq 25.0$ and $< 30.0$ )	22 (24.7)
Obese ( $\geq 30$ )	10 (11.2)
<b>Total</b>	<b>89 (100.0)</b>

To ensure completeness and comparability between the FFQ data and the digital receipt data, participants are required to meet several eligibility criteria, which are illustrated in Figure 1). In total, N=464 participants joined the study via accepting the study consent form that was granted by the ETH Zurich Ethics Commission and displayed on the BAM service website. Out of those, N=231 followed the email-based invitation to also start the web-based FFQ (i.e. email response rate of 50.2%). Fifty participants were excluded because they did not complete the FFQ, which took over 30 minutes to complete (i.e. drop out rate 21.6%). To support the assessment of the suggested purchase indicator, eligible users should have at least a certain minimum of their expected calorific energy intake represented in the purchased groceries within their digital receipts over the course of the four weeks preceding their FFQ. The four-week time window was defined by the FFQ, which was validated to estimate a patient's typical diet on a monthly basis [42]. The minimum for the energy supply to be represented in the digital receipt dataset was defined to be at least 1/7 of the estimated energy intake from the household estimated from each participant's responses in the FFQ. Consequently, 91 participants were excluded for the validation because their digital receipts represented less than 1/7 of their estimated calorific energy over the four-weeks prior to the FFQ (exclusion rate 50.2%, i.e. 91 from 181 users who completed their FFQ and donated digital receipts). Finally, one participant self-reported the own BMI to be 109.8 kg/m<sup>2</sup> and was therefore excluded. The final dataset hence included 89 users with a completed FFQ and an acceptable amount of purchased products captured by digital receipts in the four weeks prior to the FFQ.

Table 3: Purchase characteristics of the study population.

Characteristics of observed purchase behavior <sup>a</sup>	Mean (SD)
<b>Household</b>	
Adults sharing the loyalty card(s)	1.7 (1.0)
Children sharing the loyalty card(s)	0.5 (0.9)
<b>Purchased quantity identified via digital receipts</b>	
Amount spent in Swiss Francs (CHF)	230.30 (175.60)
Amount spent in US Dollars (USD) <sup>b</sup>	250.28 (190.83)
Weight <sup>c</sup> of purchased grocery products in kg	39.9 (32.1)

<sup>a</sup> of the N=89 participants in the four weeks before their FFQ

<sup>b</sup> conversion CHF/USD on date: June 30th, 2021

<sup>c</sup> conversion for liquids: 1g = 1ml

The final N=89 participants completed their FFQ successfully and had purchased sufficient grocery products via their loyalty cards in the four weeks preceding the FFQ. On average a participant was 36.2 years of age (standard deviation (SD): 10.3) and had an average BMI of 24.4 (SD: 3.8). Hence, compared to Switzerland, the recruited population (See Table 2) is more male (76.4% vs. 49.2% in Switzerland), lower in BMI (BMI of 24.4 vs 25.3 in Switzerland), and younger (36.2 years vs. 43.1 in Switzerland). With regard to the observed shopping behavior, each participant in the final dataset spent on average a total of 230.30 CHF (SD: 175.60 CHF) for grocery products that amounted to an average of 39.9 kg (standard deviation: 32.1 kg) over the four weeks before finishing the FFQ. The average number of adults and children who are sharing their loyalty cards are 1.7 (SD: 1.0) and 0.5 (SD: 0.9) respectively (See Table 3). In terms of individual diet as determined by the FFQ, the N=89 users on average ate a mean of 1.17 (SD: 1.2) portions of meat per day (See Table 1). Similarly, they consumed on average 2.55 (SD: 1.71) portions of vegetables and 1.38 (SD: 1.16) portions of fruit per day. Hence, on average, the average participant did not reach the publicly recommended three portions of vegetables and two portions of fruit per day. The users consumed on average 0.32 (0.33) portions of whole grains per day. Finally, the users consumed an average of 2.93 (1.95) portions of sweets per day. The fact, that the Swiss population consumes too many sweets has been observed in the annual food intake study MenuCH for many years [43,44]. In terms of nutritional intake, the N=89 participants consumed 2.1 grams (SD: 1.5) of sodium per day on average, 27.1 (SD: 14.3) grams of dietary fibers, 37.5 (SD: 26.0) grams of saturated fatty acids and an estimated amount of 10.4 (SD: 8.5) grams of added sugar per day. Compared to the annual food intake study MenuCH, these values seem realistic and in line with the observations for the dietary behavior of the Swiss population [43,44].

### 2.5. Validation and comparison of purchase indicators

To assess the calibration and discrimination capacity of the alternative purchase quality indicators, four different indicators were selected, calculated and their respective discrimination performances for individual dietary behavior computed. For each of the N=89 participants that fulfilled the eligibility criteria a set of purchase quality indicators was calculated. We selected the HEI-2015, HETI, GPQI and FSA-NPS DI for the assessment for multiple reasons. First, in contrast to the HPI [32], which relies on micro-nutrients, these four indicators can all be calculated using the digital receipts mandated by GDPR [28], as well as the declaration of macro-nutrients for products sold online, as mandated by EU1169 [29]. Hence, these indicators have potential to support millions of consumers in the European Union alone in their monitoring of the nutritional quality using their digital receipts. Second, although these indicators have been referenced and discussed in the literature, we lack comparative assessments of their calibration and discrimination potential. Third, the four indicators have been validated separately in different geographic regions, but not yet within a neutral region, thereby Switzerland represents an interesting

study context for this validation study. While the HEI-2015 was defined in North America, HETI and GPQI were defined in Australia. Further, the FSA-NPS-DI was defined in Great Britain and later adopted within the Nutri-Score framework within France. Finally, all four indicators were defined to represent a qualitative indication of an individual's general dietary habitual patterns. Hence, all four indicators can be calculated on a monthly basis, serving as an ideal validation context for the FFQ, which was also validated over the course of a four-week time period. Consequentially, all four indices were calculated based on timestamps within the mentioned period prior to the completion of the FFQ. The original definitions of the HETI, GPQI, and FSA-NPS DI were used and not adapted in terms of calculation. As discussed in the previous chapter, the HEI-2015 [34] was adopted to digital receipts, similar to how the Healthy Eating Index-2010 (HEI-2010) was adapted to food purchases [33]. As defined in their publications, the HEI-2015, HETI and GPQI are primarily based on health-relevant food groups and not on nutrients present in all food items purchased during the observation period. Hence, for their calculation only items belonging to the relevant food groups used in their respective definitions were included in their calculation. When calculating the FSA-NPS DI, all purchased food items were included, as the FSA-NPS DI framework is centered around purchased nutrients primarily. To make the four indicators directly comparable, they were all normalized to zero mean and unit variance.

To validate the indicators with the nutritional intake as determined by the FFQ, certain conditions were followed to ensure a coherent process. First, digital receipts that could not be identified (e.g. rare products) and therefore were missing nutritional data, were discarded. Second, to also validate the potential to discriminate the intake of added sugar whose declaration is not mandated in Europe [29], the added sugar amounts of products were calculated retrospectively [45]. The content of added sugar in purchased food products was hence estimated based on the declared sugar content and the respective product category, using an established approach by Louie et. al. [45]. To assess the number of portions eaten by each individual user as determined by the FFQ, or purchased as determined by the digital receipts, the following considerations were agreed upon. First, to calculate the number of consumed meat portions, dried meat portions were defined at 30 grams per portion, while non-dried meat at 120 grams per portion. Similarly, to assess the number of fruit portions purchased or eaten, 30 grams of dried fruits were defined as one portion, while 120 grams of fresh or frozen fruit or 200 ml of fruit juice were defined as one portion. Third, and similarly, 30 grams of dried vegetables account for one portion, while 120 grams of fresh or frozen vegetable or 200 ml of vegetable juice were defined as one portion. To address the number of whole grains portions eaten or purchased, the portions were defined as follows: 100g for bread, 30g for cereals or flakes, 60g for crisp bread and 30g for cereals. In this study, whole grains were defined as grain products that include at least 5 grams of dietary fiber per 100 grams of product [46–48]. Finally, for sweets, the definitions of one portion were set to: 17g for chocolate, 20g for cocoa products and jams, 30g for bonbons, cereal bars, or cookies, 50g for pudding, ice creams, 120g for sweet cakes and similar. A detailed overview of the portion sizes used in the data processing for the FFQ and digital receipts can be found in the datasets that were published together with this manuscript.

Based on the obtained data from the participants' FFQs and digital receipts, we assessed the *calibration capacity* and the *discrimination capacity* of the considered purchase indicators. First, the calibration capacity reflects how agreeable the models are in terms of prediction and outcomes. Second, the discrimination capacity evaluates how well the models can perform in separating cases with and without outcomes [49,50]. To evaluate the *calibration capacity* of the models, Pearson correlation coefficients between normalized purchase indicators and absolute as well as relative nutritional intake were calculated. In this context, higher correlation coefficients represent higher calibration capacity. More specifically, a high correlation of a (normalized) purchase indicator with health relevant nutrient or food group underlines a valid calibration of the purchase indicator for the

nutrient or food group respectively. For example, the HETI indicator is well calibrated for estimating absolute daily fruit intake (See Table 4. The corresponding results for the calibration capacity are displayed in Table 4 for absolute dietary intakes (in portions/day or grams/day) as defined in the FFQ, as well as in Table 5 for relative dietary intakes (in portions/1000 kcal or grams/1000 kcal) as defined in the FFQ.

To evaluate the *discrimination capacity* of the proposed purchase quality indicator models, nutritional intake of three compliance tertiles were compared. Similar to validation studies of single purchase quality indicators that divided their sample population into three segments, tertile T1 maps to the lowest compliance to a given indicator and T3 maps to the highest compliance. For the assessment of the discrimination capacity of the indicators, their potential to discriminate between all three tertiles T1-T3, as well as their ability to distinguish pairwise differences between each two combinations of the tertiles were addressed. Hence, in total, four statistical tests, namely one Kruskal-Wallis-Test for the 3-tertile comparison (T1, T2, T3) and 3 Mann-Whitney U tests for the pairwise comparisons (T1 and T2; T2 and T3; T1 and T3, respectively) were conducted for each combination of an indicator with a nutritional intake as determined by the FFQ. For each significant difference that we encountered, we then aggregated the results by assigning one point to an indicator, while no points were assigned for non-significant differences. A significantly different result is counted as one point, while a non-significant result is counted as zero points. The total points of each indicator were obtained by summing up the points across all comparisons where this indicator was involved, where more points represent better discrimination capacity (the maximal achievable number of points is 36). The corresponding results for the discrimination capacity are displayed in Table 6 for absolute dietary intakes (in portions/day or grams/day) as defined in the FFQ, as well as in Table 7 for relative dietary intakes (in portions/1000 kcal or grams/1000 kcal) as defined in the FFQ.

### 3. Results

We obtained the following results on the calibration and discrimination capacities of the investigated indicators.

#### 3.1. Calibration Capacity: Correlations between normalized purchase indicators and nutritional facts

As shown in Table 4, normalized purchase indicators and individual absolute daily nutritional intake were in general moderately correlated. Nevertheless, correlations between normalized purchase indicators and individual absolute daily sweets intake and sodium intake were relatively weak. The correlations were significantly stronger when using individual daily density-based relative nutritional intake, as shown in Table 5. Compared to the other three indicators, the FSA-NPS DI was correlated higher with the nutritional facts on both, absolute and relative scales. Specifically, the FSA-NPS DI demonstrated the highest calibration capacity, as it demonstrated the highest correlation coefficient for four (six) out of the nine dimensions for absolute (relative) dietary intake respectively. The FSA-NPS DI therefore has higher calibration capacity than the HEI-2015, HETI and GPQI.

#### 3.2. Discrimination Capacity: Comparisons of nutritional facts across compliance tertiles

Table 6 and Table 7 show the results of the indicators' assessment of their discrimination capacity regarding the differentiation of *absolute* and *relative* nutritional intake respectively. Regarding the *absolute* food intake, all indicators were able to differentiate participants' intake of fruits, whole grains and dietary fiber to a certain degree. On the other hand, no indicator managed to differentiate participants' added sugar intake (which was estimated using products' category affiliation and sugar content, as added sugar is not declared by food producers in Switzerland). In general, the FSA-NPS DI again outperformed the other indicators in differentiating participants' absolute nutritional intake as 17 out of the 36 statistical tests were significant ( $P < .05$ ). It was the only indicator that was capable of differentiating participants' intake of sweets, sodium and saturated fat, which

Table 4: Correlations between normalized purchase indicators and *absolute* individual daily nutritional intake, N = 89.

Normalized indicator <sup>a</sup>	HEI-2015	HETI	GPQI	FSA-NPS DI <sup>b</sup>
Meat, portions/day	-.083	-.099	-.094	<b>-.246*</b>
Vegetables, portions/day	.190	.181	.194	<b>.235*</b>
Fruits, portions/day	<b>.254*</b>	<b>.274**</b>	<b>.258*</b>	<b>.239*</b>
Whole grains, portions/day	<b>.322**</b>	<b>.232*</b>	<b>.226*</b>	.161
Sweets, portions/day	<b>-.124</b>	-.026	-.094	-.111
Sodium, g/day	.023	.072	-.004	<b>-.121</b>
Dietary fibers, g/day	<b>.329**</b>	<b>.296**</b>	.190	<b>.312**</b>
Saturated fatty acids, g/day	.006	.034	-.058	<b>-.144</b>
Added sugar, g/day	-.193	-.137	<b>-.216*</b>	-.101
Points	3	1	1	4

<sup>a</sup> Highest absolute value marked bold to indicate best calibrated purchase indicator for each nutrient or food group.

<sup>b</sup> We inverted the original FSA-NPS DI scores for an easier to read comparison with the other indices, as the FSA-NPS DI scale ranges from -15 (healthiest) to +40 (unhealthiest).

\*  $P < .05$ .

\*\*  $P < .01$ .

\*\*\*  $P < .001$ .

Table 5: Correlations between normalized purchase indicators and *relative* individual daily nutritional intake, N = 89.

Normalized indicator <sup>a</sup>	HEI-2015	HETI	GPQI	FSA-NPS DI <sup>b</sup>
Meat, portions/1000 kcal	-.241*	-.240*	-.109	<b>-.359***</b>
Vegetables, portions/1000 kcal	.191	.108	.214*	<b>.321**</b>
Fruits, portions/1000 kcal	<b>.238*</b>	.197	<b>.262*</b>	<b>.354***</b>
Whole grains, portions/1000 kcal	<b>.267*</b>	.193	.173	<b>.231*</b>
Sweets, portions/1000 kcal	<b>-.197</b>	-.139	-.090	-.097
Sodium, g/1000 kcal	-.178	-.055	-.051	<b>-.244*</b>
Dietary fibers, g/1000 kcal	<b>.342**</b>	<b>.235*</b>	.206	<b>.500***</b>
Saturated fatty acids, g/1000 kcal	-.228*	-.177	-.207	<b>-.367***</b>
Added sugar, g/1000 kcal	<b>-.316**</b>	<b>-.306**</b>	<b>-.291**</b>	-.093
Points	3	0	0	6

<sup>a</sup> Highest absolute value marked bold to indicate best calibrated purchase indicator for each nutrient or food group.

<sup>b</sup> We inverted the original FSA-NPS DI scores for an easier to read comparison with the other indices, as the FSA-NPS DI scale ranges from -15 (healthiest) to +40 (unhealthiest).

\*  $P < .05$ .

\*\*  $P < .01$ .

\*\*\*  $P < .001$ .

Table 6: Discrimination potential of relevant purchase data indicators to differentiate health-relevant individual daily nutritional intake behaviour (*absolute*, i.e. weight-based), N = 89.

Normalized indicator	HEI-2015	HETI	GPQI	FSA-NPS DI
Meat, portions/day	○	○	○	●
Vegetables, portions/day	●	○	○	○
Fruits, portions/day	●	●	○	○
Whole grains, portions/day	●	●	●	○
Sweets, portions/day	○	○	○	●
Sodium, g/day	○	○	○	●
Dietary fiber, g/day	●	●	○	○
Saturated fatty acids, g/day	○	○	○	●
Added sugar, g/day	○	○	○	○
Points	11	10	7	17

○ No test was significant.

○ The Kruskal-Wallis-Test showed significant differences between all three tertiles of the indicator score distribution ( $P < .05$ ).

○ The Mann-Whitney U test between the 2<sup>nd</sup> & the 3<sup>rd</sup> tertiles was significant ( $P < 0.05$ ).

○ The Mann-Whitney U test between the 1<sup>st</sup> & the 3<sup>rd</sup> tertiles was significant ( $P < .05$ ).

○ The Mann-Whitney U test between the 1<sup>st</sup> & the 2<sup>nd</sup> tertiles was significant ( $P < .05$ ).

●●●● Multiple above-mentioned statistical tests were significant ( $P < .05$ ).

are all important health-influencing intake factors. When it comes to differentiating the *relative* food intake, FSA-NPS DI also outperformed the other indicators. HEI-2015 and HETI performed equally well when the nutritional intake is assessed on a *relative* basis (i.e. on a per 1000 kcal scale). Noteworthy, when looking into specific food categories or nutrients, the best performance might not always be achieved by the same indicator. The detailed performance of FSA-NPS DI differentiating *absolute* and *relative* nutritional intake can be found in Tables 8 and 9.

Table 7: Discrimination potential of relevant purchase data indicators to differentiate health-relevant individual daily nutritional intake behaviour (*relative*, i.e. calorie-adjusted), N = 89.

Normalized indicator	HEI-2015	HETI	GPQI	FSA-NPS DI
Meat, portions/1000 kcal	●	●	○	●
Vegetables, portions/1000 kcal	●	○	○	●
Fruits, portions/1000 kcal	●	○	●	●
Whole grains, portions/1000 kcal	●	●	●	●
Sweets, portions/1000 kcal	○	○	○	○
Sodium, g/1000 kcal	●	○	○	●
Dietary fibers, g/1000 kcal	●	●	●	●
Saturated fatty acids, g/1000 kcal	○	●	●	●
Added sugar, g/1000 kcal	●	●	○	○
Points	12	12	7	19

○ No test was significant.

● The Kruskal-Wallis-Test showed significant differences among all three tertiles ( $P < .05$ ).

● The Mann-Whitney U test between the 2<sup>nd</sup> & the 3<sup>rd</sup> tertiles was significant ( $P < .05$ ).

● The Mann-Whitney U test between the 1<sup>st</sup> & the 3<sup>rd</sup> tertiles was significant ( $P < .05$ ).

● The Mann-Whitney U test between the 1<sup>st</sup> & the 2<sup>nd</sup> tertiles was significant ( $P < .05$ ).

●●● Multiple above-mentioned statistical tests were significant ( $P < .05$ ).

## 4. Discussion

### 4.1. Summary

This manuscript represents the first ever quantitative validation study on the calibration and discrimination ability of previously suggested purchase quality indicators. In total, digital receipts from two loyalty card systems and validated FFQs from N=89 individuals in Switzerland were collected to validate four indicators that were previously developed and validated in different regions in the world separately. The presented results demonstrate that the HETI, GPQI, HEI-2015 as well as the FSA-NPS DI feature moderate correlations between purchased product logs captured from digital receipts and individual daily nutritional intake. Compared to absolute daily dietary intake, the calibration capacity of the purchase indicators is especially strong on a relative energy density basis. Similarly, compared to absolute dietary intake, the indicators also tend to show higher discrimination capacities when using density-based nutritional intake. These results show that purchase quality indices are indeed transferable from other regions to a new region such as Switzerland and can in fact give a somewhat reliable indication on the nutritional quality of individual diets by assessing their digital receipts. The British FSA-NPS DI which also forms the basis for the French Nutri-Score framework consistently outperformed the other purchase indicators in calibrating and differentiating participants' nutritional intake, on both - absolute and relative density-based scales.

### 4.2. Contribution

This study has multiple contributions to research and practice. First, this study represents the first validation of previously suggested purchase indicators from different regions in the world within a thorough comparative quantitative, comparative assessment. Second, the consistent out-performance of the FSA-NPS DI in nutritional intake calibration and discrimination might inspire researchers and practitioners when looking for a grocery quality indicator. Third, given that some health-related dietary characteristics such as estimating added sugar intake were not well captured by the assessed purchase indicators, there might be room for the development of different purpose-specific purchase quality indicators (e.g. estimating the risk for diabetes type two by estimating intake levels of added sugar and carbohydrate quality). Fourth, researchers, practitioners and designers for food choice monitoring systems and behavior change interventions should take away that

Table 8: Median absolute nutritional intake by FSA-NPS DI tertiles (T) (Median and interquartile range (IQR))

	FSA-NPS DI score tertile											
	Overall(N=89)			T1(N=30)			T2(N=29)			T3(N=30)		
	Median(IQR)	Median(IQR)	Median(IQR)	Median(IQR)	Median(IQR)	Median(IQR)	Median(IQR)	Median(IQR)	P	P <sub>T1-T2</sub>	P <sub>T1-T3</sub>	P <sub>T2-T3</sub>
<b>Meat, portions/day</b>	1.02(1.22)	1.17(1.24)	1.05(1.44)	0.77(1.01)	<.001***	.744	<.001***	.003**				
<b>Vegetables, portions/day</b>	2.30(1.99)	1.97(1.76)	2.57(1.56)	2.30(1.98)	.135	.128	.064	.722				
<b>Fruits, portions/day</b>	1.17(1.28)	0.71(1.21)	1.46(1.29)	1.22(1.40)	.063	.200	.018*	.336				
<b>Whole grains, portions/day</b>	0.25(0.45)	0.09(0.44)	0.25(0.43)	0.36(0.37)	.049*	.367	.012*	.185				
<b>Sweets, portions/day</b>	2.53(2.24)	2.90(2.41)	2.35(1.86)	2.33(1.75)	.038*	.471	.030*	.032*				
<b>Sodium, g/day</b>	1.87(1.00)	1.84(0.93)	1.97(1.72)	1.88(0.61)	.017*	.529	.022*	.011*				
<b>Dietary fibers, g/day</b>	22.20(17.30)	17.55(8.65)	28.60(16.00)	28.80(25.08)	.018*	.084	.006**	.262				
<b>Saturated fatty acids, g/day</b>	31.70(17.10)	35.30(16.13)	28.70(38.40)	31.25(14.90)	.022*	.970	.011*	.028*				
<b>Added sugar, g/day</b>	8.01(7.82)	10.86(9.33)	8.47(5.54)	6.58(5.71)	.444	.897	.304	.252				

P The Kruskal-Wallis-Test was done to compare all three tertiles.

P<sub>T1-T2</sub>, P<sub>T1-T3</sub>, P<sub>T2-T3</sub> The Mann-Whitney U tests were done to compare pairwise tertiles.

\* P<.05.

\*\* P<.01.

\*\*\* P<.001

Table 9: Median density-based relative nutritional intake per 1000 kcal and interquartile range(IQR) of FSA-NPS DI tertiles (T).

	FSA-NPS DI tertile										
	Overall(N=89)	T1(N=30)			T2(N=29)			T3(N=30)			
	Median(IQR)	Median(IQR)	Median(IQR)	Median(IQR)	Median(IQR)	Median(IQR)	Median(IQR)	P	P <sub>T1-T2</sub>	P <sub>T1-T3</sub>	P <sub>T2-T3</sub>
<b>Meat, portions/1000 kcal</b>	0.57(0.57)	0.70(0.56)	0.62(0.55)	0.48(0.59)	<.001***	.611	<.001***	.002**			
<b>Vegetables, portions/1000 kcal</b>	1.19(1.06)	1.07(1.23)	1.20(0.81)	1.24(0.91)	.040*	.120	.014*	.321			
<b>Fruits, portions/1000 kcal</b>	0.60(0.63)	0.49(0.73)	0.66(0.55)	0.71(0.53)	.008**	.190	.002**	.080			
<b>Whole grains, portions/1000 kcal</b>	0.13(0.22)	0.05(0.21)	0.13(0.20)	0.18(0.24)	.007**	.299	.002**	.043*			
<b>Sweets, portions/1000 kcal</b>	1.50(0.89)	1.64(0.99)	1.49(0.65)	1.28(0.97)	.165	.779	.115	.097			
<b>Sodium, g/1000 kcal</b>	1.07(0.36)	1.11(0.48)	1.11(0.39)	1.00(0.28)	.003**	.897	.003**	.004**			
<b>Dietary fibers, g/1000 kcal</b>	13.23(8.21)	10.48(4.98)	13.99(8.99)	14.00(9.44)	<.001***	.057	<.001***	.012*			
<b>Saturated fat, g/1000 kcal</b>	19.54(6.27)	20.97(5.22)	19.10(5.92)	18.65(5.02)	.002**	.190	<.001***	.020**			
<b>Added sugar, g/1000 kcal</b>	4.72(2.64)	5.61(3.23)	4.64(2.23)	4.13(2.92)	.851	.767	.631	.688			

P The Kruskal-Wallis-Test was done to compare all three tertiles.

P<sub>T1-T2</sub>, P<sub>T1-T3</sub>, P<sub>T2-T3</sub> The Mann-Whitney U tests were done to compare pairwise tertiles.

\* P<.05.

\*\* P<.01.

\*\*\* P<.001

these indicators capture the calorie-adjusted nutritional intake better than absolute dietary intake. Finally, the study demonstrates that regulatory mandates such as the European General Data Privacy Regulation (GDPR) [28] and the mandate for nutritional declaration for food products sold online (EU)1169 [29] can pave the way for novel tools for scalable, non-invasive monitoring of food choices and tailored digital behavior change interventions, using digital receipts and food composition databases.

#### 4.3. Limitations

This study does have certain limitations. Foremost, the number of participants restricts the generalizability of the results. Compared with the actual composition of the Swiss population [51,52], the female participation ratio, the average age, and the mean BMI of the present sample were lower. Multiple factors could have led to the sample not being representative. For once, connecting the digital receipts to the study required having an account with the BAM service and an above-average technical expertise and a certain interest in data itself, which might appeal to male users more than female users. Second, participants were mainly health-conscious individuals in or beyond tertiary education, considering where our advertisements were placed (e.g. university shuttle buses and billboards). In addition, the study is biased towards loyalty card holders of the leading two retailers in Switzerland, since they are the only group that is able to participate to retrieve their digital receipts which were the focal aspects of this study. Compared to non-card-holders, they might have slightly different eating or purchase habits. Next, although the FFQ has been validated previously in Switzerland, the tool has certain limitations, such as potentially missing food items or the fact that vitamin and mineral supplements, which can be very important for vegetarians, are not covered in the FFQ. In addition, inaccurate or missing information in the food composition database forms another limitation of the study. The manual matching process between article identifiers in digital receipts and their corresponding GTINs is challenging and can sometimes lead to ambiguous results. To mitigate this limitation, we invested a significant effort in the food composition database and manually mapped the most frequent 5'950 products' article identifiers, which is far beyond the top 3'000 products mapped and identified in similar studies.

#### 4.4. Future work

First and foremost, it is pivotal for the future of this and related studies to recruit a larger and more representative sample. In addition, data quality of digital receipt and food composition databases are another important factor that contains gaps in research and practice. Machine learning based tools could for example support the (semi-)automatic correction of errors in food composition databases and support the correct identification of products present in digital receipts. The manuscript's authors also suggest regulators to mandate the use of industry standards, such as GTINs in digital receipts to support the development of novel and scalable tools for food choice monitoring and interventions aimed at improving food choices. Data portability as mandated by the GDPR [28] itself is not sufficient, but if products' GTINs were present in digital receipts, novel post-purchase services could help consumers around the world make healthier and potentially more sustainable food choices. To overcome the limitations of the FFQ, bio-sampling (e.g. sodium excretion or blood sampling) could be added to offset the potential of under-reporting or reporting biases. Given that the FSA-NPS DI has been primarily designed to discriminate single food items rather than aggregated purchase baskets, future research could assess the potential of even superior purchase indicators, or models for the reliable and accurate estimation of specific nutrients, e.g. via data science and machine learning models to estimate and potentially even predict absolute nutritional intake and health states of consumers.

## 5. Conclusions

In this study, we present the first comparison of the calibration as well as the discrimination capacities of the widely-used purchase indicators HEI-2015, HETI, GPQI and FSA-NPS DI on a data set that was obtained from 89 participants and included responses to an FFQ as well as real grocery purchases data using digital receipts collected from two loyalty card systems in Switzerland. Our results show that, overall, the surveyed indicators correlate only moderately with absolute and slightly higher with relative individual daily nutritional intake. All indicators are moderately suitable to differentiate health-relevant daily nutritional intake behavior using digital receipts. However, among these indicators, the FSA-NPS DI in general outperforms the other three indicators on our dataset, having the best calibration and discrimination ability, thus contributing an empirically validated guideline regarding the selection of purchase indicators. This is counter-intuitive, as the FSA-NPS DI, and in particular its integration into the Nutri-Score framework were primarily designed to discriminate the nutritional quality of individual products' nutritional composition as well as their dietary fiber, fruit and vegetable content make the FSA-NPS DI a superior framework for deriving individual dietary intake behavior from digital receipts. The relatively inferior performance of the HETI and the GPQI might be because they were designed and calibrated within Australia by evaluating shares of wallet (in Australian Dollars), rather than assessing nutrients or food groups by their weight contribution. This might lead to an inferior performance of these indicators when translating them to other regions, as Australian prices and retail market dynamics might be different, and both indicators might require a re-calibration within other regions. In the future, we plan to enlarge the sample size, improve the data quality, and rerun the analysis using machine learning frameworks to estimate individual dietary intake deficits from digital receipts accurately. In addition, we plan a further investigation into the reasons behind the differences in the calibration and discrimination capacities of the surveyed indicators in the future.

**Author Contributions:** Conceptualization, Klaus Fuchs, Jing Wu and Jie Lian; Methodology, Jing Wu, Klaus Fuchs, Jie Lian; Formal analysis, Jing Wu; Software, Jie Lian, Mirella Haldimann; Resources, Klaus Fuchs, Elgar Fleisch, Christine Brombach, Roland Gassmann; Data curation, Klaus Fuchs; Writing—original draft preparation, Jing Wu; Writing—review and editing, all authors; Supervision, Simon Mayer, Tanja Schneider, Elgar Fleisch; Project administration, Klaus Fuchs; Funding acquisition, Klaus Fuchs, Jaewook Byun, Simon Mayer, Tanja, Schneider. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Swiss National Science Foundation under project grant number 188402 (see <http://p3.snf.ch/project-188402>).

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the Ethics Commission office, Stampfenbachstrasse 52/56, STE H 26, 8092 Zürich, of ETH Zurich (protocol code 2019-N-134, approved on October 15th, 2019).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The dataset used in this study, including the anonymized FFQ and anonymized digital receipt datasets can be found here: <https://gitlab.ethz.ch/jingwu/shopping-index-comparison>

**Acknowledgments:** First, we would like to the BitsaboutMe AG, Bollwerk 4, 3011 Bern, Switzerland, for providing the digital receipt infrastructure for this study. Second, we also would like to thank the former Bachelor and Master students for supporting the data collection of product data used in this study: Lucia Caiata, Sebastian von der Mark, Sven Brunner, Tristan Breyer. Third, we would like to thank Zurich University of Applied Science (ZHAW) for providing the web-mediated food frequency questionnaire for the study. Fourth, we would like to thank the food composition database operators GS1 trustbox, Codecheck and Openfoodfacts for offering their product datasets for research

purposes. Fifth, we would like to thank the European parliament for the introduction of the GDPR which significantly reduces the effort in collecting digital receipts for researchers and practitioners, and allows for the development of novel data-driven tools for scalable monitoring of food choices via loyalty cards. Last but not least, we would like to thank all study participants for the donation of their FFQ and digital receipt data, without which this study would not have been possible.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Sample Availability:** <https://gitlab.ethz.ch/jingwu/shopping-index-comparison>

### Abbreviations

The following abbreviations are used in this manuscript:

API	Application Programming Interface
BAM	BitsaboutMe
B2C	Business2Consumer
CHF	Swiss franc
FFQ	Food Frequency Questionnaire
FSA-NPS DI	Food Standards Agency Nutrient Profiling System Dietary Index
GDPR	General Data Protection Regulation
GPQI	Grocery Purchase Quality Index-2016
GTIN	Global Trade Item Number
HEI-2010	Healthy Eating Index 2010
HEI-2015	Healthy Eating Index 2015
HETI	Healthy Trolley Index
HPI	Healthy Purchase Index
NCD	Non-communicable Diseases

### References

1. Saeedi, P.; Petersohn, I.; Salpea, P.; Malanda, B.; Karuranga, S.; Unwin, N.; Colagiuri, S.; Guariguata, L.; Motala, A.A.; Ogurtsova, K.; others. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. *Diabetes research and clinical practice* **2019**, *157*, 107843.
2. Hales, C.M.; Fryar, C.D.; Carroll, M.D.; Freedman, D.S.; Ogden, C.L. Trends in obesity and severe obesity prevalence in US youth and adults by sex and age, 2007-2008 to 2015-2016. *Jama* **2018**, *319*, 1723–1725.
3. Roy, A.; Praveen, P.A.; Amarchand, R.; Ramakrishnan, L.; Gupta, R.; Kondal, D.; Singh, K.; Sharma, M.; Shukla, D.K.; Tandon, N.; others. Changes in hypertension prevalence, awareness, treatment and control rates over 20 years in National Capital Region of India: results from a repeat cross-sectional study. *BMJ open* **2017**, *7*, e015639.
4. Popkin, B.M.; Corvalan, C.; Grummer-Strawn, L.M. Dynamics of the double burden of malnutrition and the changing nutrition reality. *The Lancet* **2020**, *395*, 65–74.
5. Hawkes, C.; Harris, J.; Gillespie, S.; others. Changing diets: Urbanization and the nutrition transition. *IFPRI book chapters* **2017**, pp. 34–41.
6. Zhen, S.; Ma, Y.; Zhao, Z.; Yang, X.; Wen, D. Dietary pattern is associated with obesity in Chinese children and adolescents: data from China Health and Nutrition Survey (CHNS). *Nutrition journal* **2018**, *17*, 1–9.
7. Shi, Z.; Taylor, A.W.; Riley, M.; Byles, J.; Liu, J.; Noakes, M. Association between dietary patterns, cadmium intake and chronic kidney disease among adults. *Clinical nutrition* **2018**, *37*, 276–284.
8. Riddle, M.C.; Herman, W.H. The cost of diabetes care—an elephant in the room. *Diabetes Care* **2018**, *41*, 929–932.
9. Fallaize, R.; Forster, H.; Macready, A.L.; Walsh, M.C.; Mathers, J.C.; Brennan, L.; Gibney, E.R.; Gibney, M.J.; Lovegrove, J.A. Online dietary intake estimation: reproducibility and validity of the Food4Me food frequency questionnaire against a 4-day weighed food record. *Journal of medical Internet research* **2014**, *16*, e190.
10. Geelen, A.; Souverein, O.W.; Busstra, M.C.; de Vries, J.H.; van't Veer, P. Comparison of approaches to correct intake–health associations for FFQ measurement error using a duplicate recovery biomarker and a duplicate 24 h dietary recall as reference method. *Public health nutrition* **2015**, *18*, 226–233.
11. Denova-Gutiérrez, E.; Ramírez-Silva, I.; Rodríguez-Ramírez, S.; Jiménez-Aguilar, A.; Shamah-Levy, T.; Rivera-Dommarco, J.A. Validity of a food frequency questionnaire to assess food intake in Mexican adolescent and adult population. *salud pública de méxico* **2016**, *58*, 617–628.
12. König, L.M.; Sproesser, G.; Schupp, H.T.; Renner, B. Describing the process of adopting nutrition and fitness apps: Behavior stage model approach. *Journal of Medical Internet Research* **2018**, *20*. doi:10.2196/mhealth.8261.

13. Sluik, D.; Geelen, A.; de Vries, J.H.; Eussen, S.J.; Brants, H.A.; Meijboom, S.; van Dongen, M.C.; Bueno-de Mesquita, H.B.; Wijckmans-Duysens, N.E.; van't Veer, P.; others. A national FFQ for the Netherlands (the FFQ-NL 1.0): validation of a comprehensive FFQ for adults. *British Journal of Nutrition* **2016**, *116*, 913–923.
14. El Kinany, K.; Garcia-Larsen, V.; Khalis, M.; Deoula, M.M.S.; Benslimane, A.; Ibrahim, A.; Benjelloun, M.C.; El Rhazi, K. Adaptation and validation of a food frequency questionnaire (FFQ) to assess dietary intake in Moroccan adults. *Nutrition journal* **2018**, *17*, 1–12.
15. Noor Hafizah, Y.; Ang, L.C.; Yap, F.; Nurul Najwa, W.; Cheah, W.L.; Ruzita, A.T.; Jumuddin, F.A.; Koh, D.; Lee, J.A.C.; Essau, C.A.; others. Validity and reliability of a food frequency questionnaire (FFQ) to assess dietary intake of preschool children. *International journal of environmental research and public health* **2019**, *16*, 4722.
16. Lovell, A.; Bulloch, R.; Wall, C.R.; Grant, C.C. Quality of food-frequency questionnaire validation studies in the dietary assessment of children aged 12 to 36 months: a systematic literature review. *Journal of Nutritional Science* **2017**, *6*, e16. doi:10.1017/jns.2017.12.
17. Varghese, S. The UK's 11 billion yearly receipts are an environmental nightmare. Technical report, Wired UK, 2018.
18. Caicedo, O. eReceipts 101 : The What, Why, Where and How of B2C Tax Compliance., 2018.
19. Ernst and Young. Tax Administration Is Going Digital. Technical report, EY, 2016.
20. Johnson, M. The Big Picture On Digital Receipts. Technical report, Techcrunch, 2014.
21. Brewster, P.J.; Guenther, P.M.; Jordan, K.C.; Hurdle, J.F. The Grocery Purchase Quality Index-2016: An innovative approach to assessing grocery food purchases. *Journal of Food Composition and Analysis* **2017**, *64*, 119–126. doi:10.1016/j.jfca.2017.07.012.
22. Tran, L.; Brewster, P.; Chidambaram, V.; Hurdle, J. An innovative method for monitoring food quality and the healthfulness of consumers' grocery purchases. *Nutrients* **2017**, *9*. doi:10.3390/nu9050457.
23. Tang, W.; Aggarwal, A.; Liu, Z.; Acheson, M.; Rehm, C.D.; Moudon, A.V.; Drewnowski, A. Validating self-reported food expenditures against food store and eating-out receipts. *European Journal of Clinical Nutrition* **2016**, *70*, 352–357, [15334406]. doi:10.1038/ejcn.2015.166.
24. Ransley, J.; Donnelly, J.; Botham, H.; Khara, T.; Greenwood, D.; Cade, J. Use of supermarket receipts to estimate energy and fat content of food purchased by lean and overweight families. *Appetite* **2003**, *41*, 141–148. doi:10.1016/S0195-6663(03)00051-5.
25. Nelson, M. Family food purchases and home food consumption: comparison of nutrient contents. *British Journal of Nutrition* **1985**, *54*, 373–387.
26. Martin, S.L.; Howell, T.; Duan, Y.; Walters, M. The feasibility and utility of grocery receipt analyses for dietary assessment. *Nutrition Journal* **2006**, *5*, 6–12. doi:10.1186/1475-2891-5-10.
27. Chappuis, A.; Bochud, M.; Glatz, N.; Vuistiner, P.; Paccaud, F.; Burnier, M. Swiss Survey on Salt Intake: Main Results. *Service de Nephrologie et Institut Universitaire de Medecine Sociale et Preventive Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Suisse*. **2011**, pp. 1–32.
28. The European Parliament and the Council of European Union. Regulation (EU) 2016/679. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. (accessed on 17th June 2021).
29. EU-1169/2011. Regulation (EU) 1169/2011: Regulation on Food Information to Consumers, 2014.
30. Taylor, A.; Wilson, F.; Hendrie, G.A.; Allman-Farinelli, M.; Noakes, M. Feasibility of a Healthy Trolley Index to assess dietary quality of the household food supply. *British Journal of Nutrition* **2015**, *114*, 2129–2137.
31. Brewster, P.J.; Guenther, P.M.; Jordan, K.C.; Hurdle, J.F. The Grocery Purchase Quality Index-2016: An innovative approach to assessing grocery food purchases. *Journal of Food Composition and Analysis* **2017**, *64*, 119–126.
32. Tharrey, M.; Dubois, C.; Maillot, M.; Vieux, F.; Méjean, C.; Perignon, M.; Darmon, N. Development of the Healthy Purchase Index (HPI): a scoring system to assess the nutritional quality of household food purchases. *Public health nutrition* **2019**, *22*, 765–775.
33. Appelhans, B.M.; French, S.A.; Tangney, C.C.; Powell, L.M.; Wang, Y. To what extent do food purchases reflect shoppers' diet quality and nutrient intake? *International Journal of Behavioral Nutrition and Physical Activity* **2017**, *14*, 46.
34. Krebs-Smith, S.M.; Pannucci, T.E.; Subar, A.F.; Kirkpatrick, S.I.; Lerman, J.L.; Toozé, J.A.; Wilson, M.M.; Reedy, J. Update of the healthy eating index: HEI-2015. *Journal of the Academy of Nutrition and Dietetics* **2018**, *118*, 1591–1602.
35. Julia, C.; Méjean, C.; Touvier, M.; Péneau, S.; Lassale, C.; Ducrot, P.; Hercberg, S.; Kesse-Guyot, E. Validation of the FSA nutrient profiling system dietary index in French adults—findings from SUVIMAX study. *European journal of nutrition* **2016**, *55*, 1901–1910.
36. Julia, C.; Hercberg, S. Nutri-Score: Evidence of the effectiveness of the French front-of-pack nutrition label. *Ernährungs Umschau* **2017**, *64*, 181–187.
37. Accarda. Business Case Kundenkarte, 2005.
38. Handelszeitung.ch. Kundenkarten: Zweifel am Nutzen. Cumulus-, Supercard und Co. bringen den Detailhändlern kaum höhere Kundenfrequenz, 2004.
39. Brinkerhoff, K.M.; Brewster, P.J.; Clark, E.B.; Jordan, K.C.; Cummins, M.R.; Hurdle, J.F. Linking supermarket sales data to nutritional information: an informatics feasibility study. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium* **2011**, *2011*, 598–606.
40. Fuchs, K.; Zeltner, M.; Haldimann, M.; Ilic, A. Icon-based Digital Food Allergen Labels for Complementation of Text-based Declaration. 28th European Conference on Information Systems (ECIS 2020), 2020.
41. Fuchs, K.; Barattin, T.; Haldimann, M.; Ilic, A. Towards Tailoring Digital Food Labels: Insights of a Smart-RCT on User-specific Interpretation of Food Composition Data (in submission). CHIItaly19 (The 13th Biannual Conference of the Italian SIGCHI Chapter); , 2019. doi:10.1145/3347448.3357171.

42. Steinemann, N.; Grize, L.; Ziesemer, K.; Kauf, P.; Probst-Hensch, N.; Brombach, C. Relative validation of a food frequency questionnaire to estimate food intake in an adult population. *Food & nutrition research* **2017**, *61*, 1305193.
43. Bochud, M.; Chatelan, A.; Blanco, J.M.; Beer-Borst, S. menuCH - National Survey on Nutrition.
44. Bochud, M.; Chatelan, A.; Blanco, J.M.; Beer-Borst, S. Anthropometric characteristics and indicators of eating and physical activity behaviors in the Swiss adult population Results from menuCH 2014-2015 **2015**.
45. Louie, J.C.Y.; Moshtaghian, H.; Boylan, S.; Flood, V.M.; Rangan, A.; Barclay, A.; Brand-Miller, J.; Gill, T. A systematic methodology to estimate added sugar content of foods. *European journal of clinical nutrition* **2015**, *69*, 154–161.
46. USDA. How much fiber is in whole grains versus refined grains/non whole grains? <https://ask.usda.gov/s/article/How-much-fiber-is-in-whole-grains-versus-refined-grains-non-whole-grains>. (accessed on 20th June 2021).
47. FDA. Health Claim Notification for Whole Grain Foods. <https://www.fda.gov/food/food-labeling-nutrition/health-claim-notification-whole-grain-foods>. (accessed on 20th June 2021).
48. USDA. Nutrients: Total Dietary Fiber (g). [https://www.nal.usda.gov/sites/www.nal.usda.gov/files/total\\_dietary\\_fiber.pdf](https://www.nal.usda.gov/sites/www.nal.usda.gov/files/total_dietary_fiber.pdf). (accessed on 20th June 2021).
49. Collins, G.S.; de Groot, J.A.; Dutton, S.; Omar, O.; Shanyinde, M.; Tajar, A.; Voysey, M.; Wharton, R.; Yu, L.M.; Moons, K.G.; others. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC medical research methodology* **2014**, *14*, 1–11.
50. Steyerberg, E.W.; Vickers, A.J.; Cook, N.R.; Gerds, T.; Gonen, M.; Obuchowski, N.; Pencina, M.J.; Kattan, M.W. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)* **2010**, *21*, 128.
51. Federal Statistical Office. Population. <https://www.bfs.admin.ch/bfs/en/home/statistics/population/effectif-change/population.html>. (accessed on 17th June 2021).
52. Federal Statistical Office. Overweight - Share of the population aged 15 and over who are overweight (BMI of 25 or more) - In percent (2019). <https://www.bfs.admin.ch/bfs/en/home/statistics/sustainable-development/monet-2030/indikatoren/overweight.assetdetail.10187695.html>. (accessed on 17th June 2021).