Received: 21 March 2022 | Accepted: 28 November 2022

DOI: 10.1111/aphw.12430

APPLIED PSYCHOLOGY: Health and Well-Being IAAP

ORIGINAL ARTICLE

Evaluating self-reported retrospective average daily fruit, vegetable, and egg intake: Trustworthy—Sometimes!

Viktoria S. Egele | **Eric Klopp** | **Robin Stark**

Department of Educational Research, Saarland University, Saarbrücken, Germany

Correspondence
Viktoria S. Egele, Department of Educational Research, Saarland University, Saarbrücken 66123, Germany.
Email: viktoria.egele@uni-saarland.de

Abstract

Retrospective self-reports are commonly used to assess dietary intake. Yet, their use is criticized as it is unclear whether the underlying assumptions for valid self-reports are met: Individuals have to consider the behavior of all days in the retention interval and weigh the behavior of all days equally. This study examines whether these assumptions for retrospective self-reports are met and whether interindividual differences in self-report performance are relevant regarding these assumptions. Ninety-two participants aged 18–61 years participated in seven sequential 24-h recalls and one retrospective 7-day recall concerning their intake of fruit, vegetables, and eggs. A multiple linear regression approach was used to examine the relation between the daily reported dietary intake and the 7-day recall. In the overall sample, the requirements for retrospective self-reports were not tenable. Distinguishing good and poor self-reporters based on a rational criterion showed that the requirements can be taken as given for good self-reporters, whereas poor self-reporters base their retrospective self-reports mostly on recency effects. The underlying requirements for retrospective self-reports appear to be met in two thirds of the sample, supporting the use of retrospective self-reports to capture dietary behavior. Future research should investigate characteristics separating good from poor self-reporters.

KEYWORDS
assessment of dietary intake, assumptions of self-reports, fruit and vegetable intake, reliability of self-reports, retrospective self-reports

INTRODUCTION

The rapidly increasing prevalence of obesity and poor eating behavior has called for worldwide action to understand underlying causes and mechanisms (Jackson et al., 2020). The main research targets are the assessment of dietary intake, the communication of recommendations for a healthy diet, and the creation of interventions to foster healthy eating behavior (de Ridder et al., 2017). As the accurate assessment of dietary intake is essential (Gibson et al., 2017), numerous new technologies in dietary assessment are being explored, for example, multisensory devices and digital photography (McClung et al., 2018). Yet, due to the complexity of this behavior, there is still a lack of simple objective measures (Conner & Norman, 2017; de Ridder et al., 2017), and the majority of studies still rely on self-reports (Thompson & Subar, 2017). Long-term retrospective self-reports (LTRSRs) are the most frequently used measure to assess dietary intake (McClung et al., 2018; Rutishauser, 2005). For example, in nutritional epidemiological studies, Diet History Questionnaires in which respondents are asked to recall their usual food consumption retrospectively for prolonged periods are often used (Mensink et al., 2001; Merten et al., 2011; Naska et al., 2017; Straßburg et al., 2019). In these retrospective self-reports, participants report the frequency and amount of food eaten during a certain period.

As self-reports are based on complex cognitive processes, it is important to beware of the strengths and weaknesses of self-report measures (Thompson & Subar, 2017). Therefore, scholars recently engaged in a debate on the usefulness of LTRSR in dietary assessment. Some defend the usefulness of these self-reports (Foster & Bradley, 2018; Hébert et al., 2014; Subar et al., 2015), and some claim that LTRSR should not be used anymore (Archer et al., 2013; Archer & Blair, 2015; Dhurandhar et al., 2015; Ioannidis, 2013; Mitka, 2013), questioning how retrospective self-reports should be placed in perspective to alternative assessment methods.

Strengths and weaknesses of LTRSR

A major advantage is the effective and efficient application and evaluation of those LTRSRs (Foa et al., 1997; McClung et al., 2018) and the collection of a large amount of information within a short period (Wilfley et al., 1997). Moreover, people are considered the best experts on their behavior. Often, no one else has access to more detailed information, and even when information is available to oneself as well as to others, people tend to remember more of what is relevant to themselves (Rogers et al., 1977; Symons & Johnson, 1997). Therefore, the value of self-reports in dietary assessment is widely acknowledged (McClung et al., 2018; Subar et al., 2015).

Yet, skepticism arises not only because of inherent measurement flaws but also from large differences between self-reported dietary intake and more objectively measured intake (Kristal & Potter, 2006). For example, comparing self-reports of energy intake to more objective measures like biomarkers shows significant differences between the two measures (Freedman et al., 2014; McClung et al., 2018).

Limited validity of LTRSR due to biases

A possible explanation for these differences is that self-reports are prone to conscious and unconscious biases that might limit the validity of self-reported data (Thompson & Subar, 2017). It was shown previously that the reporting of health behavior is prone to faking and willful response distortion, leading to invalid self-reports. According to Levashina and Campion (2006), capacity, willingness, and opportunity to modify the information given influence the likelihood of dishonest reporting. The three factors are assumed to be linked multiplicatively. Previous research showed that participants are willing to modify information on their health behavior. For example, patients admit withholding information from their clinician and altering the reports of their behavior to create the desired impression (Levy et al., 2018). Conscious response distortion arises if respondents wish to benefit from the creation of the desired impression because the behavior in question seems undesirable or socially unacceptable (Mazar & Ariely, 2006). Also, the capacity to fake health behavior data was shown previously, indicating potential large effects of conscious response distortion in two directed faking studies, both in a within-subjects design and in a between-subjects design (Egele et al., 2021). As it is often very difficult to verify the validity of self-reported data, the opportunity to modify information without being caught can be taken as given. Thus, conscious response distortion may threaten the validity of self-reported data and explain differences between LTRSR and more objective measures of dietary intake. Nevertheless, Paulhus and Vazire (2007) assume conscious response distortion to be an exception, with respondents usually tending to engage intensively with the question and answering openly and insightfully in practice.

Assuming that respondents attempt to answer honestly and to the best of their knowledge and beliefs, unconscious biases may occur in LTRSR. In retrospective self-reports, information has to be reproduced after a certain retention interval. Respondents are asked to recall numerous situations in which the behavior in question was shown and then make complex inferences on the average frequency or intensity. Following the instruction to report an average daily behavior, people need to recall numerous situations in which the behavior was shown and then make complex inferences on the average of the behavior in question (Wirfält, 1998).

In a nutshell, memory can be regarded as structured in three phases: an encoding phase, in which information is recorded, a retention phase, in which information is stored, and retrieval, in which information is accessed (Baddeley et al., 2004). To report their dietary intake, participants must have encoded information about what they ate during a certain period and they must be able to assess, retrieve, and report this information (Smith et al., 1991). It was shown early on that individuals differ in memory performance (Cohen, 1994), recall, and forgetting (Jonker, 2016; Unsworth, 2019; Wierzba et al., 2018). Interindividual differences in the ability to encode, assess, retrieve, and report information correctly might contribute to explaining differences between LTRSR and more objective assessment methods as LTRSR by subjects with lower memory performance could be flawed and therefore different from more objective measurement methods.

Another source of unconscious bias is the susceptibility to errors in the memory process. For example, decreasing accessibility of information over time, shallow processing, and temporal inaccessibility of information stored hinder accurate memory (Schacter, 1999). Additionally, memory distortions like misattribution and suggestibility impair memory (Schacter, 1999). Imperfect memory may lead to a limited validity of self-reports. For example, Smith et al. (1991) assume that memories about dietary intake are linked to specific dietary information like episodes of eating and generic dietary information like knowledge about food and one's diet. Imperfect memory leads to compromises between reports of literal dietary intake and inferences retrieved from general knowledge. In this case, responses are given based on beliefs about "typical" behavior, which depends on self-perception and may not accurately reflect actual behavior (Stone et al., 1998). Memories of specific episodes of dietary intake thus contribute very little to self-reports with long retention intervals (Smith et al., 1991). In that case, participants would derivate from the item instruction to report the frequency and amount of food eaten, which might also explain differences between LTRSR and more objective measures.

Requirements for valid retrospective self-reports

To evaluate whether the method to collect dietary data is assessing the intended measure, it is a common validation study procedure to assess the match of different self-report measures. Most often, retrospective self-reports with long retention intervals are compared with 24-h recalls that are seen as less biased (Thompson & Subar, 2017). A major benefit of 24-h recalls is the immediacy of the recall period. It is assumed that retention intervals of a few hours lead to reports of specific memories instead of descriptions of typical dietary intake (Kristal et al., 2006; Smith et al., 1991). Therefore, participants can generally recall most of their food intake (Thompson & Subar, 2017). Previous studies also found reports of 24-h recalls to be similar to observed intakes of nutrients (Gersovitz et al., 1978; Madden et al., 1976). Conway et al. (2004) showed that participants can report their intakes of energy, protein, carbohydrate, and fat accurately by comparing a direct observation and a 24-h recall. Also, energy underreporting was lower in 24-h recalls than in food frequency questionnaires (Freedman et al., 2014). Thus, it seems like self-reports with shorter retention intervals are a good alternative to conventional LTRSR.

Comparing self-reports with longer retention intervals to self-reports with shorter retention intervals, significant differences arise. For example, there are significant differences between diet history interviews and 24-h recalls in repeated-measures studies like the National Nutrition Survey II, a study conducted by the German Federal Ministry of Food, Agriculture, and Consumer Protection (Heuer et al., 2015). Participants reported consuming significantly more fruit and vegetables in the dietary history interviews than in the 24-h recall. Similar results were found by Sjöberg and Hulthen (2004) and van Liere (1997). Heuristics and recall biases are discussed as reasons for this discrepancy. It seems unclear whether participants report their average daily behavior in retrospective self-reports. Assuming that the 24-h recall is the more valid measure, the finding that these two measures yield different results raises questions about whether retrospective self-reports may not do justice to the implicitly assumed processes. This question can be considered in more detail from an empirical perspective on 24-h recalls and LTRSR.

Usually, the item instruction in LTRSR is like "How many portions of fruit did you eat during the last seven days?", a typical retrospective 7-day recall (e.g. Chapman & Armitage, 2010). Given that individuals remember their behavior and correctly form a mean value, a retrospective self-report should correspond to the mean of the memory fragments of the behavior in question in the given period (Wirfält, 1998). That is, if individuals fill out 24-h recalls for the entire retention interval and then complete a 7-day recall, the mean of the 24-h recalls should correspond to the 7-day recall. Valid retrospective self-reports thus rely on three requirements:

Individuals have to recall the behavior of all days in the retention interval, consider the behavior of all days in the retention interval in their responses, and weigh the behavior of all days in the retention interval equally. Imperfect memory might lead to constraints in these assumptions and potentially explain significant differences between 24-h recalls and LTRSR that have been shown previously.

Given these rather strong assumptions underlying the long-time self-report, the question arises whether the underlying assumptions are tenable. From an empirical point of view, these assumptions should find their expressions in a certain pattern in a regression model. (1) In case individuals recall all days and (2) consider them for the long-time self-report, then all days should be relevant predictors for the criterion 7-day recall. (3) If additionally, the assumption holds that all days are equally weighted, the regression coefficients should approximately show the same size for each predictor. In contrast, a pattern of regression coefficients in which not all days are relevant predictors, or a pattern in which some predictors differ considerably from others, that is, they do not all affect the criterion equally, is not following such assumptions. For example, effects of imperfect memory like effects of salience may bias memory retrieval, as the most frequent behavior or the first or last memory of behavior might be reported instead of reporting the average behavior (Kahneman et al., 1982). A recency effect would either yield the last days being relevant predictors or would manifest itself in increasing magnitudes of the regression coefficients toward the end of the measurement period.

Thus, the first research question of the present study is to empirically scrutinize the assumptions of the long-term self-report, that is, to examine whether the abovementioned assumptions for retrospective self-reports are met. Assuming perfect memory, research question 1a (RQ1a) examines whether participants consider all days of the retention interval for their 7-day mean. Research question 1b (RQ1b) examines whether the days that are used for the 7-day mean are weighted equally. Because interindividual differences in memory performance were shown previously, these differences in self-report performance are taken into consideration, and we examine in a second research question (RQ2) whether they are relevant regarding the assumptions of retrospective self-reports. Therefore, we apply a rational criterion to distinguish between good and bad self-reporters and scrutinize the regression pattern in each of these groups.

The current research is an exploratory study that aims to give a first glimpse of the processes involved in answering the 7-day recall for the assessment of food intake. In the current research, we focus on the assessment of fruit, vegetable, and egg intake. Food categories were chosen because of their relation to health and current research. It is widely acknowledged that fruit and vegetable intake is a fundamental element of a healthy diet (Joffe & Robertson, 2001). There is not only evidence that sufficient fruit and vegetable intake can reduce the risk for multiple diseases like cancer and cardiovascular diseases (van Veer et al., 2000), but it is also favorable for bone health (Gundgaard et al., 2003) and mental health (Ocean et al., 2019). Therefore, it is a World Health Organization (WHO) goal to promote fruit and vegetable intake (Agudo & Joint FAO/WHO Workshop on Fruit and Vegetables for Health (2004: Kobe, 2005). Similarly, eggs are an important element of most western diets (Münger et al., 2018), but contrarily to fruit and vegetable consumption, for which research agrees that a higher consumption is desirable, there is debate regarding egg consumption as to whether it has a positive or negative impact on health. Until a few years ago, the assumption was that high egg consumption was more likely to be detrimental to health (Li et al., 2012); more recently, there is evidence that egg consumption may not be as harmful as assumed (Clayton et al., 2017; Mah et al., 2020). Thus, the assessment of egg consumption is the focus of current research and for this reason is considered in more detail in this study.

METHOD

Sample

Because the criterion (the retrospective self-report) was assessed on the eighth day, only participants who had completed all 7 days of the retention interval, as well as the retrospective self-report, were included in the analyses. Out of 148 subjects who started the survey, 92 subjects completed all questionnaires completely and were included in the analyses. The final sample included 59 women and 33 men between the ages of 18 and 61 ($M = 28.47$, $SD = 11.14$). Participants were recruited via notices on campus and various online platforms. Participation in the study was not monetarily rewarded.

Design

To assess daily dietary intake, participants filled out seven 24-h recalls and one retrospective self-report concerning their average daily dietary intake on Day 8. Thus, the study consisted of eight sequential measurements. All questionnaires were implemented online via SoSci Survey (Leiner, n.d.). The conduct of the study complied with the ethical standards of the responsible committee (The Ethics Committee of the Faculty of Empirical Human and Economic Sciences of Saarland University). Written informed consent was obtained from all subjects before the study.

Instruments

In the 24-h recalls, participants were asked to report the amount of fruit, vegetables, and eggs they ate during the last 24 h in open questions (e.g.: “How many portions of fruit did you eat during the last 24 hours?”). This item formulation complies with established practices (Jovanovic et al., 2021; Kristjansdottir et al., 2006). Subjects were instructed regarding the definition of portion sizes as well as the foods that fall into each category (e.g. potatoes do not count as vegetables). For this purpose, the specifications of the German Nutrition Society were followed (Ernährungskreis, n.d.), which are very similar to the recommendations of the WHO (Agudo & Joint FAO/WHO Workshop on Fruit and Vegetables for Health (2004: Kobe, 2005). In the retrospective self-report, a 7-day recall, participants reported the number of days out of the last seven on which they consumed a certain category of fruit, vegetables, and eggs, as well as the amount of food eaten. This recording method is in line with the approach of previous studies (Baker & Wardle, 2003; Heuer et al., 2015). The number of days was assessed on a scale ranging from 0 to 7, and the amount of food was assessed in open questions. Definitions of portion sizes and foods of each category were identical to the instructions of the 24-h recalls. In addition, sociodemographic data including gender, age, education, and household structure were collected. Before data analysis, exclusion criteria to prevent leverage effects in the following regression analyses was applied (e.g. participants reporting more than seven portions of fruit or vegetables per day and more than five eggs per day either in the 24-h recalls or in the 7-day recall).

Analysis strategy

The first research question (RQ1) consists of two joint research questions, RQ1a and RQ1b, asking if the two assumptions of the LTRSR hold. For RQ1a, a regression model with all 24-h recalls as predictors for the 7-day recall is estimated. Afterward, a statistical procedure is applied to select the predictors which are relevant predictors for the 7-day recall. The assumptions of this regression model are checked. To answer RQ1b, a statistical test is applied that checks if the regression coefficients of the selected predictors are equal. This test is not applied if the model selection process results in a model with only one predictor.

Concerning RQ2, a rational criterion is applied to separate good self-reporters (GSRs) from bad self-reporters (BSRs) and divide the sample into two subsamples accordingly. Then, employing a statistical test, it is examined if there are distinct regression models in each of the two subsamples. If this test is statistically significant, then the procedure described above for RQ1 is applied separately in both the GSR subsample and the BSR subsample. If this test fails statistical significance, then there is no empirical evidence for separate regression models in each subsample, and the analysis stops at this point.

To sum up, this analysis strategy provides insight if the assumptions of the LTRSR are tenable for the whole sample and in the two subsamples consisting of GSR and BSR. To answer the research questions, the above-described set of analyses is conducted for each type of food, that is, vegetables, fruits, and eggs, separately. In the following section, the statistical methods applied are described separately for RQ1 and RQ2 in detail.

Statistical methods for RQ1

To examine the general relationship between the daily 24-h recalls and the 7-day recall, a multiple linear regression approach is used, in which the 7-day recall is regressed on the seven 24-h recalls. To tackle RQ1, a two-step procedure is used to examine RQ1a in the first place and afterward RQ1b. The first step starts with estimating a regression model for the whole sample, which is called hereafter the initial model. This initial model is then used to conduct an exhaustive search for the best subset regression (e.g. James et al., 2017, p. 205). The best subset regression aims to select the set of predictors that best describes the relationship between the seven 24-h recalls and the 7-day recall. The best subset model indicates whether, as addressed in RQ1, all days are taken into account for the 7-day recall. The maximum number of predictors in the model was set to 7. The best subset is selected based on the Bayesian Information Criterion (BIC; Schwarz, 1978). The BIC tends to select the true model if the true model is among the list of candidate models (Vrieze, 2012). This procedure avoids the selection of predictors through its statistical significance which can lead to biased results (e.g. Heinze & Dunkler, 2017). To ensure that the subset of predictors does not worsen the model in terms of explained variance, the initial model is compared with the best subset model through an F-test that should be statistically insignificant. For the best subset model, the regression assumptions are checked using the Jarque–Berra Test (JB; normality assumption; e.g. Heij et al., 2004), the Durbin–Watson Test (DW; autocorrelation of residuals; e.g. Fox, 2015, p. 494), and the Breusch–Pagan Test (BP, heteroscedasticity; e.g. Heij et al., 2004, p. 345). As serial dependencies between the seven 24-h recalls are possible, the variance inflation factors (VIFs; James et al., 2017, p. 101) are checked to detect multicollinearity. Multicollinearity provides information on possible serial dependencies between the predictors of a regression model. A VIF higher than 5 can be considered problematic, as it indicates a high correlation between the predictors, and a VIF below 5 can be considered unproblematic. If multicollinearity is present, then the statistical test of the regression coefficients may be biased. Yet, a VIF larger than 2.5 indicates a considerable degree of multicollinearity (cf., Johnston et al., 2018). VIFs exceeding 2.5 are interpreted as indicative of serial correlations among the predictors and interpret VIFs smaller than 5 as an absence of problems arising from multicollinearity. Also, the power of the tests of the regression coefficients is analyzed by using a Monte Carlo simulation (e.g. Beaujean, 2014). The power is the percentage of the repetitions for which the hypothesis test is significant at the significance level α. Here, the nominal significance level is set at $\alpha = .05$, and 10,000 repetitions are conducted. Power should at least be .50, at best exceeding a value of .80 and above (Kyriazos, 2018).

In the second step, RQ1b is examined, which aims to scrutinize if all 7-day recalls are equally weighted, the hypothesis of equal regression coefficients in the best subset model is tested with a General Linear Hypothesis (GLH; e.g. Fox, 2015, p. 219).

Statistical methods for RQ2

To examine RQ2 which takes into account differences in self-report performance, a rational criterion is used to separate GSR from BSR. This criterion is based on the absolute difference between the reported 7-day mean and the mean of the seven 24-h recalls. If the absolute difference is smaller or equal to half a portion (for vegetables and fruit and one portion for eggs), the participant is classified as a GSR. If the absolute difference is larger than half a portion, the participant is classified as a BSR.[1] This criterion is based on the result of a discussion of experts, namely, that an individual can reliably judge half a portion of food like fruit and vegetables and one egg. The criteria serve as an operational definition in the present study. Depending on the contents and contexts of other studies, the criteria should be adapted accordingly. To check if such a distinction is empirically tenable, a Chow Test (e.g. Heij et al., 2004, p. 305) and a series of *t*-tests are conducted. The Chow Test indicates the presence of a structural break. In the case of a statistically significant Chow Test, the two-step procedure described for RQ1 is conducted for the GSR and BSR separately. To examine how well these criteria empirically separate GSR and BSR, the reported 7-day mean and the mean of the seven 24-h recalls are compared separately for GSR and BSR, using a *t*-test. Cohen's *d* is used to judge the size of the mean difference, with the usual classification.

RESULTS

The research questions were examined for vegetables, fruit, and eggs. These food categories serve as structural criteria in the analysis below, both research questions are evaluated per food. Data are available on request due to privacy/ethical restrictions.

Vegetable intake

From the initial data set, 13 participants drop out because of the exclusion criteria, and the final data set consisted of 79 participants. The results for the reported vegetable intake are shown in Tables 1–4. Further descriptive statistics are displayed in Tables S1 to S4. For RQ1a, the initial model is significant, $F(7, 71) = 21.98$, $p < .001$, and explains 68% of the variance. The BIC shows

[1]Let M denote the reported 7-day mean and let R denote the mean of the seven 24-h recalls with R, then the criterion formally reads as $|M\text{-}R| \leq 0.5$.

TABLE 1 Model selection (Bayesian Information Criterion) for all models and all foods

	Vegetables			Fruits	Eggs		
	All	GSR	BSR	All	All	GSR	BSR
Model 1	186.43	101.75	76.16	204.78	141.53	41.38	57.79
Model 2	174.89	74.09	78.23	192.21	133.04	26.18	58.95
Model 3	171.37	61.94	80.26	184.41	130.59	9.38	58.54
Model 4	173.33	54.83	83.59	183.49	128.78	0.53	60.85
Model 5	176.52	43.37	86.95	185.82	130.77	1.43	63.33
Model 6	180.23	43.11	90.31	189.43	133.26	3.92	66.19
Model 7	184.45	45.72	93.67	193.78	137.14	7.62	69.1

Abbreviations: BSR, bad self-reporter; GSR, good self-reporter.

TABLE 2 Regression table for all participants, vegetables

	B	*SE*	β	*t*	*p*	Power	VIF
(Intercept)	0.35	0.15	0	2.38	.02	.65	
Day 1	0.2	0.07	0.23	2.81	.006	.79	1.54
Day 3	0.23	0.07	0.29	3.52	<.001	.93	1.51
Day 6	0.34	0.07	0.45	5.2	<.001	1.00	1.71
	$F(3, 75) = 49.90$, $p < .001$, $R^2 = .67$						

Abbreviation: VIF, variance inflation factor.

TABLE 3 Regression table for good self-reporters, vegetables

	B	*SE*	β	*t*	*p*	Power	VIF
(Intercept)	0.06	0.09	0	0.71	.483	.11	
Day 1	0.19	0.04	0.23	4.71	<.001	.99	1.77
Day 3	0.24	0.05	0.28	4.75	<.001	.99	2.66
Day 4	0.09	0.04	0.11	1.93	.06	.47	2.51
Day 5	0.17	0.05	0.2	3.72	<.001	.94	2.19
Day 6	0.16	0.04	0.21	3.76	<.001	.95	2.39
Day 7	0.15	0.04	0.18	3.56	<.001	.92	1.96
	$F(6, 43) = 117.60$, $p < .001$, $R^2 = .94$						

Abbreviation: VIF, variance inflation factor.

TABLE 4 Regression table for bad self-reporters, vegetables

	B	*SE*	β	*t*	*p*	Power
(Intercept)	0.79	0.27	0	2.95	.007	.81
Day 6	0.43	0.11	0.61	3.95	<.001	.95
	$F(1.27) = 15.63$, $p = .001$, $R^2 = .37$					

that the model with three predictors is the most parsimonious. This best subset model is significant and explains 67% of the variance. The predictors are the first, third, and sixth days. A model comparison does not indicate a significant effect of omitting the predictors, $F(1, 75) = 1.017$, $p = .405$. The best subset model does not show nonnormality, autocorrelated residuals, or heteroscedasticity issues either, JB = 0.37, $p = .830$; DW = 1.93, $p = .728$; BP = 4.04, $df = 3$, $p = .257$. The VIFs are all below the thresholds of 2.5 and 5 and indicate no serial correlation and no problems due to multicollinearity. The power was above the minimum. For RQ1b, the GLH shows the assumption that all days contribute equally to the reported 7-day mean holds, $F(2, 75) = 0.88$, $p = .421$. Thus, in the overall sample, a model including the first, third, and sixth days fits the data best, indicating that participants did not fulfill the assumed assumptions underlying LTRSR of their vegetable consumption.

Concerning RQ2, the criterion to separate GSR from BSR is applied to the reported vegetable intake and leads to subsamples of size $nGSR = 50$ and $nBSR = 29$. For the GSR, there is no significant difference between the reported 7-day mean and the actual mean reported vegetable intake, $t(49) = -0.14$, $p = .887$, $d = 0.02$, whereas there is a significant difference for the BSR, $t(28) = -2.90$, $p = .007$, $d = 0.53$. The Chow Test is significant, $F(8, 63) = 4.35$, $p < .001$, which supports the appropriateness of distinguishing GSR from BSR in LTRSR of vegetable consumption.

For the GSR, the initial model including all 7 days is significant, $F(7, 42) = 101.20$, $p < .001$, and explains 94% of the variance. According to the BIC, the best subset model with six predictors is the most parsimonious. In the best subset model for GSR, the second day is omitted from the predictors. The best subset model is significant, $F(6, 43) = 117.60$, $p < .001$, and explains 94% of the variance. The model comparison does not indicate a significant effect of omitting a predictor, $F(1, 43) = 1.11$, $p = .298$. As for the initial model, there are no nonnormality, autocorrelated residuals, or heteroscedasticity issues, JB = 3.12, $p = .210$; DW = 2.10, $p = .776$; BP = 2.22, $df = 6$, $p = .898$. A noticeable finding is that all but the fourth day are significant predictors. The VIFs indicate serial dependencies but are not problematic. The power is higher than .90 for all but the fourth day. The GLH shows that in the restricted model, all days contribute equally to the reported 7-day mean, $F(5, 43) = 1.19$, $p = .332$, indicating that participants did not consider all days for their LTRSR but weighted the days they considered equally.

For BSR, the initial model is not significant, $F(7, 21) = 21$, $p < .077$, and explains 43% variance. The BIC indicates that the best subset model has only one predictor, the sixth day. The best subset model is significant, $F(1, 27) = 15.63$, $p < .001$, and explains 37% of the variance. The model comparison does not indicate a significant effect of omitting the other predictors, $F(6, 27) = 0.34$, $p = .907$, and there is no nonnormality, autocorrelation, and heteroscedasticity, JB = 0.18, $p = .911$; DW = 1.77, $p = .532$; BP = 0.445, $df = 1$, $p = .505$. The power for the sixth day is very high, .95. Thus, BSR consider only the sixth day of the retention interval for their LTRSR of vegetable consumption.

Fruit intake

From the initial data set, seven participants drop out because of the exclusion criteria, and the final data set consisted of 85 participants. The results for the reported fruit intake are shown in Tables 1 and 5. Again, further descriptive statistics are displayed in Tables S5 and S6. Concerning RQ1a, the initial model for the whole sample is significant, $F(7, 77) = 2.95$, $p < .001$, and explains 67% of the variance. The best subset model contains the first, fourth, sixth, and seventh days according to the BIC. The best subset model is significant, $F(4, 80) = 37.81$, $p < .001$, and explains

TABLE 5 Regression table for all participants, fruits

	B	SE	β	t	p	Power	VIF
(Intercept)	0.37	0.12	0	2.99	.004	.84	
Day 1	0.2	0.06	0.26	3.24	.002	.89	1.46
Day 4	0.24	0.07	0.33	3.7	<.001	.95	1.88
Day 6	0.14	0.06	0.18	2.28	.025	.61	1.4
Day 7	0.2	0.07	0.25	2.67	.009	.75	2.05
	$F(4, 80) = 37.81$, $p < .001$, $R^2 = .65$						

Abbreviation: VIF, variance inflation factor.

65% of the variance, and omitting predictors does not deteriorate the amount of explained variance, $F(3, 80) = 0.94$, $p = .427$. For the best subset model, there are nonnormal regression residuals, but there are no autocorrelated residuals or heteroscedasticity issues either, JB = 27.10, $p < .001$; DW = 1.97, $p = .876$; BP = 5.37, $df = 4$, $p = .251$. In the best subset model, all predictors are significant. The VIFs are below the 2.5 threshold, and the power is sufficient. For RQ1b, the GLH shows all predictors contribute equally to the criterion, $F(3, 80) = 0.37$, $p = .756$. Thus, for LTRSR of fruit consumption, a model including the first, fourth, sixth, and seventh days fits the data best, indicating that participants again did not consider all days of the retention interval but weighted the days they considered equally.

Concerning RQ2, applying the criterion to separate GSR from BSR concerning their reported fruit intake leads to subsamples of size $nGSR = 62$ and $nBSR = 23$. However, there are no significant differences between the self-reported 7-day mean and actual mean, neither for the GSR, $t(61) = -0.04$, $p = .964$, $d = 0.01$, nor the BSR, $t(22) = 0.29$, $p = .771$, $d = 0.06$. The Chow Test is not significant either, $F(8, 69) = 1.46$, $p = .189$. Thus, no further regression models are estimated. Thus, for LTRSR of fruit consumption, there is no empirically justifiable distinction between GSR and BSR. In line with our procedure, the analysis for RQ2 was not pursued further.

Egg intake

From the initial data set, eight participants dropped out because of the exclusion criteria, and the final data set consisted of 84 participants. The results for the reported egg intake are shown in Table 1, as well as in Tables 6–8. Once more, further descriptive statistics are displayed in Tables S7 to S10. For RQ1a, the initial model for the whole sample is significant, $F(7, 76) = 14.71$, $p < .001$, and explains 58% of the variance. The BIC indicates selecting a model with four predictors, these are the same as the significant predictors. The best subset model is significant, $F(4, 79) = 24.11$, $p < .001$, and explains 55% of the variance. All predictors are significant, and the omission of the predictors does not significantly deteriorate the amount of explained variance, $F(3, 79) = 1.53$, $p = .213$. Again, there are nonnormal regression residuals, JB = 31.58, $p < .001$, but there are no autocorrelated residuals or heteroscedasticity issues, DW = 2.11, $p = .546$; BP = 3.24, $df = 4$, $p = .519$. The VIFs did not show any degrees of considerable serial dependencies. The power is adequate, too. Concerning RQ1b, the GLH indicates that all predictors contribute equally to the criterion, $F(3, 79) = 0.24$, $p = .867$. Thus, for LTRSR of egg consumption, a model including the second, third, fourth, and seventh days fits the data best, indicating that participants again did not consider all days of the retention interval but weighted the days they considered equally.

Concerning RQ2, applying the distinction criterion to differentiate GSR from BSR concerning their reported egg intake provides subsamples of size $nGSR = 62$ and $nBSR = 22$. For the

TABLE 6 Regression table for all participants, eggs

	B	*SE*	β	*t*	*p*	Power	VIF
(Intercept)	0.09	0.06	0	1.42	.159	.30	
Day 2	0.29	0.08	0.34	3.64	<.001	.94	1.51
Day 3	0.19	0.08	0.22	2.47	.016	.69	1.4
Day 4	0.19	0.08	0.21	2.5	.014	.70	1.2
Day 7	0.22	0.08	0.24	2.75	.007	.77	1.31
	$F(4, 79) = 24.11$, $p < .001$, $R^2 = .55$						

Abbreviation: VIF, variance inflation factor.

TABLE 7 Regression table for good self-reporters, eggs

	B	*SE*	β	*t*	*p*	Power	VIF
(Intercept)	−0.01	0.03	0	−0.4	.694	.07	
Day 1	0.29	0.04	0.43	6.74	<.001	1.00	1.31
Day 3	0.28	0.04	0.42	7.01	<.001	1.00	1.16
Day 4	0.2	0.06	0.24	3.64	<.001	.94	1.41
Day 6	0.18	0.05	0.23	3.77	<.001	.95	1.23
	$F(4, 57) = 67.55$, $p < .001$, $R^2 = .83$						

Abbreviation: VIF, variance inflation factor.

TABLE 8 Regression table for bad self-reporters, eggs

	B	*SE*	β	*t*	*p*	Power
(Intercept)	0.73	0.2	0	3.72	.001	.94
Day 7	0.41	0.18	0.45	2.25	.036	.58
	$F(1, 20) = 5.06$, $p = .036$, $R^2 = .20$					

GSR, there is no significant difference between the reported 7-day mean and the actual mean, $t(61) = -0.88$, $p = .380$, $d = 0.11$, and the same result emerges for the BSR, $t(21) = 1.49$, $p = .149$, $d = 0.31$. Although the difference is not significant for the BSR, the effect size is medium, so the insignificance might be a result of the small sample size. The Chow Test is significant, $F(8, 63) = 4.35$, $p < .001$, and hints at the appropriateness of distinguishing GSR from BSR in LTRSR of egg consumption. Therefore, separate regression analyses for GSR and BSR are performed.

The initial model for GSR is significant, $F(7, 54) = 40.52$, $p < .001$, and explains 84% of the variance. The BIC indicates that the best subset model has four predictors, which are the first, third, fourth, and sixth days. The best subset model is significant, $F(5, 57) = 67.55$, $p < .001$, and explains 83% of the variance. The omission of the predictors does not reduce the amount of explained variance, $F(3, 57) = 1.60$, $p = .199$. As for the initial model, the model checks indicate nonnormal residuals JB = 6.52, $p = .038$), but there are no autocorrelated residuals or heteroscedasticity, DW = 1.90, $p = .754$; BP = 4.92, $df = 4$, $p = .296$. The VIFs are all below 2.5, and the power is almost perfect. The GLH indicates that all predictors contribute equally to the criterion, $F(3, 57) = 1.36$, $p = .265$. Thus, GSRs consider the first, third, fourth, and sixth days for their LTRSR and weigh these equally.

For the BSR, the initial model is not significant, $F(7, 14) = 1.48$, $p = .251$, and explains 43% of the variance. The BIC indicates that the model with only the seventh day is the best subset model. This model is significant, $F(1, 20) = 5.06$, $p < .001$, and explains 20% of the variance. Although there is a large drop in explained variance, the omission of the predictors did not significantly reduce the amount of explained variance, $F(6, 20) = 0.90$, $p = .516$. The model checks do not indicate nonnormality, autocorrelated residuals, or heteroscedasticity issues, JB = 1.11, $p = .576$; DW = 2.65, $p = .120$; BP < 0.01, $df = 1$, $p = .998$. Thus, for BSR, a model including only the seventh day fits the data best.

DISCUSSION

The research questions examined first hints of whether the requirements for retrospective self-report are met in general and with regard to differences in self-report performance. For vegetables, fruit, and eggs, it seems like in the overall sample, the requirements for valid retrospective self-reports are not tenable. The present study did not yield evidence for participants taking into consideration all days of the retention interval but rather indicated that participants relied on shortcuts. It seems like they do consider a day at the beginning of the retention interval, one in the middle of the retention interval and one at the end of the retention interval for all three food categories instead of considering each day meticulously. This seems to be a case of cognitive parsimony or cognitive effort avoidance (Kool et al., 2010; Strobel et al., 2020; Westbrook et al., 2013).

Taking into account interindividual differences in self-report performance

Considering differences in self-report performance, for LTRSR of vegetables and eggs, it was possible to distinguish participants that gave more accurate self-reports from participants whose self-reports were less accurate. Interestingly, GSR continuously considered more days of the retention interval for their LTRSR. Thereby, GSR differed from BSR, which seem to report one of the last days of the retention interval. Thus, for BSR, there is evidence for recency effects in their reports, it seems like they report the last memory of behavior instead of reporting the average behavior (Kahneman et al., 1982). Also, it seems like BSRs were especially prone to reporting their general dietary information instead of an average behavior, as none of the seven 24-h recalls was a significant predictor for the 7-day recall. This aligns with the assumption of Schacter (1999) and Smith et al. (1991) that responses are sometimes based on beliefs about "typical" behavior, and memories of specific episodes of dietary intake contribute very little to self-reports with long retention intervals. Nevertheless, GSR also did not consider all days of the retention interval either. This might be an effect of the rather small sample or the rather long retention interval. As shown by Tulving et al. (1982), 7 days are a rather long retention interval; in their learning studies, participants were hardly able to remember the words they had learned a week earlier. Perhaps, replication with a larger sample could provide further insights on that matter.

A differentiation of GSRs and BSRs seems reasonable and shows that implicitly underlying requirements of retrospective self-reports seem to be untenable at least for some participants. Overall, in two thirds of the sample, retrospective self-reports may be a valid measure, but, for some individuals, retrospective self-reports do not capture what they aim to. This might explain the differences that arise when comparing 24-h recalls to retrospective self-reports in previous studies (Heuer et al., 2015; Sjöberg & Hulthén, 2004; van Liere, 1997). At the same time, this raises the question of differences between the subjects who report well on their own as opposed to those who are more prone to cognitive shortcuts or parsimony. Ideally, future research could identify which characteristics distinguish GSRs from BSRs. Possibly, memory capacity, need for cognition, or conscientiousness may impact self-report performance. Initial exploratory analyses in this regard were conducted as part of this study and can be found in Supporting Information

S1. However, no systematic effects were found that predicted either the categorization as GSR or BSR or the absolute deviation of the reported 7-day mean from the mean of the seven 24-h recalls. In the long run, it might be possible to identify a criterion to determine whether retrospective self-reports are a useful measure for an individual subject.

Taking into account differences in food categories

Interestingly, all participants were able to report their fruit intake accurately, contrary to reporting their vegetable and egg intake, where GSRs and BSRs were identifiable. This indicates a need to consider differences in reported food categories. The reporting of fruit and vegetables in retrospective self-reports may differ in complexity, as it requires the grouping of several foods. As similar foods are included in a single question (e.g. strawberries, apples, or bananas), the grouping process might render recall more complex (Thompson & Subar, 2017). This might be less problematic for fruit intake but more of a problem for vegetable intake. Also, definitions of food items may complicate self-reports of dietary intake (e.g. are legumes vegetables or should they be excluded from the reported amount of vegetable intake?). However, this does not explain why then egg consumption was not also very well reported.

Another explanation may be that food consumed in clearly defined portions (e.g. an apple and a banana) is easier to report than amorphous food (e.g. salad/spinach, vegetable stews, and eggs in pastries) (De Keyzer et al., 2011; Subar et al., 2010; Thompson & Subar, 2017). Because vegetables have larger variations in portion sizes, their portion size may be more difficult to judge (Wirfält, 1998). As eggs are usually consumed in very different forms and egg products are highly diverse (Münger et al., 2018), this might complicate reporting egg consumption retrospectively.

Limitations

A major limitation of our approach is that we could only examine which days of the retention interval were taken into account in forming an LTRSR and whether they were weighted equally, but it remains unclear whether the 24-h recalls were correct, or not. Optimally, a more objective criterion would have allowed verifying the correctness of the 24-h recalls. It is assumed that inquiries about current behaviors, that is, self-reports with a shorter retention interval, are less biased than self-reports with a longer retention interval (Arab et al., 2010; Conway et al., 2004). Because self-reports with shorter retention intervals depend on the ability to assess relevant information and the willingness to report it unadulterated just like self-reports with longer retention intervals but they go without extensive memory retrieval, they are considered more accurate than retrospective self-reports concerning a prolonged period (Petermann & Eid, 2006). It is assumed that retention intervals of a few hours lead to reports of specific memories instead of descriptions of typical dietary intake (Kristal et al., 2006; Smith et al., 1991). Therefore, participants can generally recall most of their food intake in 24-h recalls (Thompson & Subar, 2017). Previous studies also found reports of 24-h recalls to be similar to observed intakes of nutrients (Gersovitz et al., 1978; Madden et al., 1976). Conway et al. (2004) showed that participants are able to report their intakes of energy, protein, carbohydrate, and fat accurately by comparing a direct observation and a 24-h recall. Also, energy underreporting was lower in 24-h recalls than in food frequency questionnaires (Freedman et al., 2014). Thus, it seems like 24-h recalls can be seen as accurate, despite this not being examined in the current study.

The difficulty with this way of examining retrospective self-reports using 24-h recalls, however, is that the 24-h recalls may have functioned as a form of memory recall training. Likewise, they may have led to increased attention to one's eating behavior. Both possibilities would potentially lead to an overestimation of the validity of the retrospective self-reports. For these reasons as well, it would be beneficial to examine the utility of retrospective self-reports using objective measures.

Another limitation of the current study is the relatively small sample size. Especially in the comparison of GSRs and BSRs, the BSR subsamples are rather small. Small sample sizes in combination with a violation of the normality assumption, as indicated by the significant JB test, may potentially increase the standard error of the regression coefficient. Consequently, potentially increased type II error rates may in turn affect the power of the estimates. This seems not to be an issue in the regression model for the BSR-vegetable intake. However, in the regression model for BSR-egg intake, the power is just below the minimum required threshold. Thus, this model should be interpreted carefully because its validity is questionable. However, the normality assumption is the least important of all assumptions (Gelman & Hill, 2006). Additionally, Knief and Forstmeier (2021) showed that the *p*-values remain reliable even in the presence of nonnormal errors. Thus, the interpretation of the regression coefficient's significance is not affected by the potential nonnormality of the errors. Concerning the other models for which there is evidence of a violation of the residuals' normality assumption by a significant JB test, the resulting power values are much larger and in an acceptable range, and the validity of these models is not at stake. In addition, the small sample size means that the best subset model selection might select a different model as the best in a larger sample. Therefore, it needs to be emphasized again that we do not intend to interpret the individual days as predictors but rather to evaluate in general whether all days are used as predictors or not.

Although the hypotheses for retrospective self-assessment are derived based on the aforementioned framing, it is unclear whether these processes are elementary for the correct reporting of a health behavior or whether these processes are different in reality. For this purpose, it would be useful to conduct a qualitative study based on the Thinking Aloud Method (Eccles & Arsal, 2017) in which participants verbalize their thoughts during the completion of a retrospective self-report to understand which cognitive processes occur during the process. This method might also shed light on another issue: Although the current research indicates that participants differ in the days of the retention interval they take into account in forming an LTRSR, it remains unclear whether this is an issue of retrieval or consideration, that is, whether participants do not recall a certain day or whether they do recall the day but do not consider it for their retrospective self-report.

The current study advocates for a use of retrospective self-reports for dietary assessment, as it seems to be an appropriate tool for the majority of participants. Nevertheless, participants in the study were relatively young and not representative of the general population, so replication of the study with a larger and more representative sampling could serve to strengthen the findings of the present study. Also, the requirements for valid self-reports were only examined for a 7-day recall period. Often, dietary assessment concerns more than the last 7 days, most often the last year (Naska et al., 2017). It is unclear whether the requirements for valid self-reports remain the same across all retention intervals, and thus, the generalizability of the current study might be limited. Further research therefore should test the generalizability of our findings for different retention intervals to replicate these results and extend the findings.

Conclusion

We contribute to the body of literature by responding to George Beaton, who claimed: "There will always be error in dietary assessments. The challenge is to understand, estimate, and make use of the error structure during analysis" (Beaton et al., 1997, 1100S). Nearly 20 years later, this claim still is supported and research concerning the strengths and limitations of dietary assessment methods is strongly encouraged (Straßburg et al., 2019). The findings of the present study cautiously support the use of retrospective self-reports for dietary assessment. It further highlights the specific need for research regarding interindividual differences between GSRs and BSRs affecting the appropriateness of the use of retrospective self-reports.

ACKNOWLEDGEMENT

Open Access funding enabled and organized by Projekt DEAL.

ETHICS STATEMENT

Informed consent was obtained from all participants included in the study. The conduct of the study complied with the ethical standards of the Ethics Committee of the Faculty of Empirical Human and Economic Sciences of Saarland University.

CONFLICT OF INTEREST

We have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Data are available on request due to privacy/ethical restrictions.

ORCID

Viktoria S. Egele https://orcid.org/0000-0003-4354-6751
Eric Klopp https://orcid.org/0000-0002-6638-8912
Robin Stark https://orcid.org/0000-0002-3807-2355

REFERENCES

Agudo, A., & Joint FAO/WHO Workshop on Fruit and Vegetables for Health (2004: Kobe, J. (2005). *Measuring intake of fruit and vegetables [electronic resource]*. World Health Organization. https://apps.who.int/iris/handle/10665/43144

Arab, L., Wesseling-Perry, K., Jardack, P., Henry, J., & Winter, A. (2010). Eight self-administered 24-hour dietary recalls using the internet are feasible in African Americans and Caucasians: The energetics study. *Journal of the American Dietetic Association*, *110*(6), 857–864. https://doi.org/10.1016/j.jada.2010.03.024

Archer, E., & Blair, S. N. (2015). Reply to LS Freedman et al. *Advances in Nutrition*, *6*(4), 489–490. https://doi.org/10.3945/an.115.009183

Archer, E., Hand, G. A., & Blair, S. N. (2013). Validity of U.S. nutritional surveillance: National Health and Nutrition Examination Survey caloric energy intake data, 1971–2010. *PLoS ONE*, *8*(10), e76632.

Baddeley, A. D., Kopelman, M. D., & Wilson, B. A. (Eds.). (2004). *The essential handbook of memory disorders for clinicians*. John Wiley & Sons.

Baker, A. H., & Wardle, J. (2003). Sex differences in fruit and vegetable intake in older adults. *Appetite*, *40*(3), 269–275. https://doi.org/10.1016/S0195-6663(03)00014-X

Beaton, G. H., Burema, J., & Ritenbaugh, C. (1997). Errors in the interpretation of dietary assessments. *The American Journal of Clinical Nutrition*, *65*(4), 1100S–1107S. https://doi.org/10.1093/ajcn/65.4.1100S

Beaujean, A. A. (2014). Sample size determination for regression models using Monte Carlo methods in R. *Practical Assessment, Research and Evaluation*, *19*(1), 12.

Chapman, J., & Armitage, C. J. (2010). Evidence that boosters augment the long-term impact of implementation intentions on fruit and vegetable intake. *Psychology & Health*, *25*(3), 365–381. https://doi.org/10.1080/08870440802642148

Clayton, Z. S., Fusco, E., & Kern, M. (2017). Egg consumption and heart health: A review. *Nutrition*, *37*, 79–85. https://doi.org/10.1016/j.nut.2016.12.014

Cohen, R. L. (1994). Some thoughts on individual differences and theory construction. *Intelligence*, *18*(1), 3–13. https://doi.org/10.1016/0160-2896(94)90018-3

Conner, M., & Norman, P. (2017). Health behaviour: Current issues and challenges. *Psychology & Health*, *32*(8), 895–906. https://doi.org/10.1080/08870446.2017.1336240

Conway, J. M., Ingwersen, L. A., & Moshfegh, A. J. (2004). Accuracy of dietary recall using the USDA five-step multiple-pass method in men: An observational validation study. *Journal of the American Dietetic Association*, *104*(4), 595–603. https://doi.org/10.1016/j.jada.2004.01.007

De Keyzer, W., Huybrechts, I., De Maeyer, M., Ocké, M., Slimani, N., van't Veer, P., & De Henauw, S. (2011). Food photographs in nutritional surveillance: Errors in portion size estimation using drawings of bread and photographs of margarine and beverages consumption. *British Journal of Nutrition*, *105*(7), 1073–1083. https://doi.org/10.1017/S0007114510004551

de Ridder, D., Kroese, F., Evers, C., Adriaanse, M., & Gillebaart, M. (2017). Healthy diet: Health impact, prevalence, correlates, and interventions. *Psychology & Health*, *32*(8), 907–941. https://doi.org/10.1080/08870446.2017.1316849

Dhurandhar, N. V., Schoeller, D., Brown, A. W., Heymsfield, S. B., Thomas, D., Sørensen, T. I. A., Speakman, J. R., Jeansonne, M., & Allison, D. B. (2015). Energy balance measurement: When something is not better than nothing. *International Journal of Obesity*, *39*(7), 1109–1113. https://doi.org/10.1038/ijo.2014.199

Eccles, D. W., & Arsal, G. (2017). The think aloud method: What is it and how do I use it? *Qualitative Research in Sport, Exercise and Health*, *9*(4), 514–531. https://doi.org/10.1080/2159676X.2017.1331501

Egele, V. S., Kiefer, L. H., & Stark, R. (2021). Faking self-reports of health behavior: A comparison between a within- and a between-subjects design. *Health Psychology and Behavioral Medicine*, *9*(1), 895–916. https://doi.org/10.1080/21642850.2021.1991803

Ernährungskreis. (n.d.). Retrieved February 8, 2021, from https://www.dge.de/ernaehrungspraxis/vollwertige-ernaehrung/ernaehrungskreis/?L=0

Foa, E. B., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The Posttraumatic Diagnostic Scale. *Psychological Assessment*, *9*(4), 445–451. https://doi.org/10.1037/1040-3590.9.4.445

Foster, E., & Bradley, J. (2018). Methodological considerations and future insights for 24-hour dietary recall assessment in children. *Nutrition Research*, *51*, 1–11. https://doi.org/10.1016/j.nutres.2017.11.001

Fox, J. (2015). *Applied regression analysis and generalized linear models*. SAGE Publications.

Freedman, L. S., Commins, J. M., Moler, J. E., Arab, L., Baer, D. J., Kipnis, V., Midthune, D., Moshfegh, A. J., Neuhouser, M. L., Prentice, R. L., Schatzkin, A., Spiegelman, D., Subar, A. F., Tinker, L. F., & Willett, W. (2014). Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American Journal of Epidemiology*, *180*(2), 172–188. https://doi.org/10.1093/aje/kwu116

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Gersovitz, M., Madden, J. P., & Smiciklas-Wright, H. (1978). Validity of the 24-hr. dietary recall and seven-day record for group comparisons. *Journal of the American Dietetic Association*, *73*(1), 48–55. https://doi.org/10.1016/S0002-8223(21)05669-8

Gibson, R. S., Charrondiere, U. R., & Bell, W. (2017). Measurement errors in dietary assessment using self-reported 24-hour recalls in low-income countries and strategies for their prevention. *Advances in Nutrition*, *8*(6), 980–991. https://doi.org/10.3945/an.117.016980

Gundgaard, J., Nielsen, J. N., Olsen, J., & Sørensen, J. (2003). Increased intake of fruit and vegetables: Estimation of impact in terms of life expectancy and healthcare costs†. *Public Health Nutrition*, *6*(1), 25–30. https://doi.org/10.1079/PHN2002355

Hébert, J. R., Hurley, T. G., Steck, S. E., Miller, D. R., Tabung, F. K., Peterson, K. E., Kushi, L. H., & Frongillo, E. A. (2014). Considering the value of dietary assessment data in informing nutrition-related health policy. *Advances in Nutrition* (Bethesda, Md., *5*(4), 447–455. https://doi.org/10.3945/an.114.006189

Heij, C., de Boer, P., Franses, P. H., Franses, R. F. P. H., Kloek, T., & van Dijk, H. K. (2004). *Econometric methods with applications in business and economics*. Oxford University Press.

Heinze, G., & Dunkler, D. (2017). Five myths about variable selection. *Transplant International*, *30*(1), 6–10. https://doi.org/10.1111/tri.12895

Heuer, T., Krems, C., Moon, K., Brombach, C., & Hoffmann, I. (2015). Food consumption of adults in Germany: Results of the German National Nutrition Survey II based on diet history interviews. *British Journal of Nutrition*, *113*(10), 1603–1614. https://doi.org/10.1017/S0007114515000744

Ioannidis, J. P. A. (2013). Implausible results in human nutrition research. *BMJ*, *347*(nov14 3), f6698–f6698. https://doi.org/10.1136/bmj.f6698

Jackson, S. E., Llewellyn, C. H., & Smith, L. (2020). The obesity epidemic—Nature via nurture: A narrative review of high-income countries. *SAGE Open Medicine*, *8*, 205031212091826. https://doi.org/10.1177/2050312120918265

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning with applications in R*. New York: Springer.

Joffe, M., & Robertson, A. (2001). The potential contribution of increased vegetable and fruit consumption to health gain in the European Union. *Public Health Nutrition*, *4*(4), 893–901. https://doi.org/10.1079/PHN2001126

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity*, *52*(4), 1957–1976. https://doi.org/10.1007/s11135-017-0584-6

Jonker, T. R. (2016). Individual differences in incorrect responding and the ability to discriminate the source of the products of retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(7), 1078–1089. https://doi.org/10.1037/xlm0000223

Jovanovic, C. E., Whitefield, J., Hoelscher, D. M., Chen, B., Ranjit, N., & van den Berg, A. E. (2021). Validation of the FRESH Austin food frequency questionnaire using multiple 24-h dietary recalls. *Public Health Nutrition*, 1–9.

Kahneman, D., Slovic, S. P., Slovic, P., Tversky, A., & Press, C. U. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809477

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, *53*(6), 2576–2590. https://doi.org/10.3758/s13428-021-01587-5

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology. General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Kristal, A., Peters, U., & Potter, J. (2006). Is it time to abandon the food frequency questionnaire? *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *14*, 2826–2828.

Kristal, A. R., & Potter, J. D. (2006). Not the time to abandon the food frequency questionnaire: Counterpoint. *Cancer Epidemiology, Biomarkers & Prevention*, *15*(10), 1759–1760. https://doi.org/10.1158/1055-9965.EPI-06-0727

Kristjansdottir, A. G., Andersen, L. F., Haraldsdottir, J., de Almeida, M. D. V., & Thorsdottir, I. (2006). Validity of a questionnaire to assess fruit and vegetable intake in adults. *European Journal of Clinical Nutrition*, *60*(3), 408–415. https://doi.org/10.1038/sj.ejcn.1602332

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, *9*(08), 2207–2230. https://doi.org/10.4236/psych.2018.98126

Leiner, D. J. (n.d.). *SoSci survey (version 3.1.06)*. [Computer Software] https://www.soscisurvey.de/de

Levashina, J., & Campion, M. A. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment*, *14*(4), 299–316. https://doi.org/10.1111/j.1468-2389.2006.00353.x

Levy, A. G., Scherer, A. M., Zikmund-Fisher, B. J., Larkin, K., Barnes, G. D., & Fagerlin, A. (2018). Prevalence of and factors associated with patient nondisclosure of medically relevant information to clinicians. *JAMA Network Open*, *1*(7), e185293–e185293. https://doi.org/10.1001/jamanetworkopen.2018.5293

Li, K.-K., Concepcion, R. Y., Lee, H., Cardinal, B. J., Ebbeck, V., Woekel, E., & Readdy, R. T. (2012). An examination of sex differences in relation to the eating habits and nutrient intakes of university students. *Journal of Nutrition Education and Behavior*, *44*(3), 246–250. https://doi.org/10.1016/j.jneb.2010.10.002

Madden, J. P., Goodman, S. J., & Guthrie, H. A. (1976). Validity of the 24-hr. recall. Analysis of data obtained from elderly subjects. *Journal of the American Dietetic Association*, *68*(2), 143–147.

Mah, E., Chen, C.-Y. O., & Liska, D. J. (2020). The effect of egg consumption on cardiometabolic health outcomes: An umbrella review. *Public Health Nutrition*, *23*(5), 935–955. https://doi.org/10.1017/S1368980019002441

Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing*, *25*(1), 117–126. https://doi.org/10.1509/jppm.25.1.117

McClung, H. L., Ptomey, L. T., Shook, R. P., Aggarwal, A., Gorczyca, A. M., Sazonov, E. S., Becofsky, K., Weiss, R., & Das, S. K. (2018). Dietary intake and physical activity assessment: Current tools, techniques, and technologies for use in adult populations. *American Journal of Preventive Medicine*, *55*(4), e93–e104. https://doi.org/10.1016/j.amepre.2018.06.011

Mensink, G., Haftenberger, M., & Thamm, M. (2001). Validity of DISHES 98, a computerised dietary history interview: Energy and macronutrient intake. *European Journal of Clinical Nutrition*, *55*(6), 409–417. https://doi.org/10.1038/sj.ejcn.1601174

Merten, C., Ferrari, P., Bakker, M., Boss, A., Hearty, Á., Leclercq, C., Lindtner, O., Tlustos, C., Verger, P., Volatier, J. L., & Arcella, D. (2011). Methodological characteristics of the national dietary surveys carried out in the European Union as included in the European Food Safety Authority (EFSA) Comprehensive European Food Consumption Database. *Food Additives & Contaminants: Part A*, *28*(8), 975–995. https://doi.org/10.1080/19440049.2011.576440

Mitka, M. (2013). Do flawed data on caloric intake from NHANES present problems for researchers and policy makers? *Jama*, *310*(20), 2137–2138. https://doi.org/10.1001/jama.2013.281865

Münger, L. H., Garcia-Aloy, M., Vázquez-Fresno, R., Gille, D., Rosana, A. R. R., Passerini, A., Soria-Florido, M.-T., Pimentel, G., Sajed, T., Wishart, D. S., Andres Lacueva, C., Vergères, G., & Praticò, G. (2018). Biomarker of food intake for assessing the consumption of dairy and egg products. *Genes & Nutrition*, *13*(1), 26. https://doi.org/10.1186/s12263-018-0615-5

Naska, A., Lagiou, A., & Lagiou, P. (2017). Dietary assessment methods in epidemiological research: Current state of the art and future prospects. *F1000Research*, *6*, 926. https://doi.org/10.12688/f1000research.10703.1

Ocean, N., Howley, P., & Ensor, J. (2019). Lettuce be happy: A longitudinal UK study on the relationship between fruit and vegetable consumption and well-being. *Social Science & Medicine*, *222*, 335–345. https://doi.org/10.1016/j.socscimed.2018.12.017

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In *Handbook of research methods in personality psychology* (pp. 224–239). The Guilford Press.

Petermann, F., & Eid, M. (2006). *Handbuch der Psychologischen Diagnostik*. Hogrefe Verlag.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*(9), 677–688. https://doi.org/10.1037/0022-3514.35.9.677

Rutishauser, I. H. (2005). Dietary intake measurements. *Public Health Nutrition*, *8*(7a), 1100–1107. https://doi.org/10.1079/PHN2005798

Schacter, D. L. (1999). Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*(3), 182–203. https://doi.org/10.1037/0003-066X.54.3.182

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Sjöberg, A., & Hulthén, L. (2004). Assessment of habitual meal pattern and intake of foods, energy and nutrients in Swedish adolescent girls: Comparison of diet history with 7-day record. *European Journal of Clinical Nutrition*, *58*(8), 1181–1189. https://doi.org/10.1038/sj.ejcn.1601947

Smith, A. F., Jobe, J. B., & Mingay, D. J. (1991). Retrieval from memory of dietary information. *Applied Cognitive Psychology*, *5*(3), 269–296. https://doi.org/10.1002/acp.2350050308

Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., Paty, J., Porter, L. S., & Cruise, L. J. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology*, *74*(6), 1670–1680. https://doi.org/10.1037/0022-3514.74.6.1670

Straßburg, A., Eisinger-Watzl, M., Krems, C., Roth, A., & Hoffmann, I. (2019). Comparison of food consumption and nutrient intake assessed with three dietary assessment methods: Results of the German National Nutrition Survey II. *European Journal of Nutrition*, *58*(1), 193–210. https://doi.org/10.1007/s00394-017-1583-z

Strobel, A., Wieder, G., Paulus, P. C., Ott, F., Pannasch, S., Kiebel, S. J., & Kührt, C. (2020). Dispositional cognitive effort investment and behavioral demand avoidance: Are they related? *PLoS ONE*, *15*(10), e0239817. https://doi.org/10.1371/journal.pone.0239817

Subar, A. F., Crafts, J., Zimmerman, T. P., Wilson, M., Mittl, B., Islam, N. G., McNutt, S., Potischman, N., Buday, R., Hull, S. G., Baranowski, T., Guenther, P. M., Willis, G., Tapia, R., & Thompson, F. E. (2010). Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. *Journal of the American Dietetic Association*, *110*(1), 55–64. https://doi.org/10.1016/j.jada.2009.10.007

Subar, A. F., Freedman, L. S., Tooze, J. A., Kirkpatrick, S. I., Boushey, C., Neuhouser, M. L., Thompson, F. E., Potischman, N., Guenther, P. M., Tarasuk, V., Reedy, J., & Krebs-Smith, S. M. (2015). Addressing current criticism regarding the value of self-report dietary data. *The Journal of Nutrition*, *145*(12), 2639–2645. https://doi.org/10.3945/jn.115.219634

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*(3), 371–394. https://doi.org/10.1037/0033-2909.121.3.371

Thompson, F. E., & Subar, A. F. (2017). Dietary assessment methodology. *Nutrition in the Prevention and Treatment of Disease*, *4*, 5–48. https://doi.org/10.1016/B978-0-12-802928-2.00001-1

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 336–342.

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin*, *145*(1), 79–139. https://doi.org/10.1037/bul0000176

van Liere, M. (1997). Relative validity and reproducibility of a French dietary history questionnaire. *International Journal of Epidemiology*, *26*(90001), 128S–136S. https://doi.org/10.1093/ije/26.suppl_1.S128

van Veer, P., Jansen, M. C., Klerk, M., & Kok, F. J. (2000). Fruits and vegetables in the prevention of cancer and cardiovascular disease. *Public Health Nutrition*, *3*(1), 103–107. https://doi.org/10.1017/S1368980000000136

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLoS ONE*, *8*(7), e68210. https://doi.org/10.1371/journal.pone.0068210

Wierzba, M., Riegel, M., Wypych, M., Jednoróg, K., Grabowska, A., & Marchewka, A. (2018). Cognitive control over memory—Individual differences in memory performance for emotional and neutral material. *Scientific Reports*, *8*(1), 1, 3808–18. https://doi.org/10.1038/s41598-018-21857-1

Wilfley, D. E., Schwartz, M. B., Spurrell, E. B., & Fairburn, C. G. (1997). Assessing the specific psychopathology of binge eating disorder patients: Interview or self-report? *Behaviour Research and Therapy*, *35*(12), 1151–1159. https://doi.org/10.1016/S0005-7967(97)80010-1

Wirfält, E. (1998). Cognitive aspects of dietary assessment. *Näringsforskning*, *42*(1), 56–59. https://doi.org/10.3402/fnr.v42i0.1762

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Egele, V. S., Klopp, E., & Stark, R. (2023). Evaluating self-reported retrospective average daily fruit, vegetable, and egg intake: Trustworthy—Sometimes! *Applied Psychology: Health and Well-Being*, *15*(3), 1130–1149. https://doi.org/10.1111/aphw.12430