Validation of a web-based self-administered non-consecutive-day dietary record tool against urinary biomarkers

Running title: Validity of web-based dietary records

Camille Lassale *, Katia Castetbon, François Laporte, Géraldine Maelle Camilleri, Valérie Deschamps, Michel Vernay, Patrice Faure, Serge Hercberg, Pilar Galan, Emmanuelle Kesse-Guyot

Authors affiliations:

Nutritional Epidemiology Research Team, Research Center in Epidemiology and Biostatistics Sorbonne Paris Cité, Inserm U1153, Bobigny, France (Camille Lassale, Géraldine M Camilleri, Serge Hercberg, Pilar Galan, Emmanuelle Kesse-Guyot);

Nutrition Surveillance and Epidemiology Unit, University Paris 13, Sorbonne Paris Cité, Bobigny, France (Katia Castetbon, Valérie Deschamps, Michel Vernay, Serge Hercberg);

Nutrition Surveillance and Epidemiology Unit, French Institute for Health Surveillance, Saint Maurice, France (Katia Castetbon, Valérie Deschamps, Michel Vernay, Serge Hercberg);

Department of Biochemistry, University Hospital of Grenoble, Grenoble, France (François Laporte, Patrice Faure);

University Grenoble Alpes, HP2 Laboratory, Grenoble, France (Patrice Faure);

Public Health Department, Hôpital Avicenne, Université Paris 13, Bobigny, France (Serge Hercberg)

***To whom correspondence should be addressed:**

Camille LASSALE

Nutritional Epidemiology Research Team, Research Center in Epidemiology and Biostatistics Sorbonne Paris Cité, U1153, SMBH, 74 rue Marcel Cachin, Bobigny, F-93017 France

Tel number: +44 207 594 3783 / Fax number: +33 1 48 38 89 31

E-mail: c.lassale@imperial.ac.uk; c.lassale@uren.smbh.univ-paris13.fr

Word count: **overall text**: 4612 **/ abstract**: 263

Number of Tables: 6 / **Figures**: 0 / **Online Supporting Material**:1

Keywords: Validation study; Dietary record; Internet; Urinary Biomarker;

Abstract

New technologies are promising for the use of short term instruments for dietary data collection, but innovative tools should be validated against objective biomarkers. The aim of this study was to investigate the validity of a web-based self-administered dietary record (DR) tool using protein, potassium and sodium intake against 24h urinary (24hU) biomarkers.

199 adult volunteers (104 men and 95 women, mean age 50.5 years (23 to 83y) of the NutriNet-Santé study were included in the protocol: they completed three non-consecutive-day DRs, and two 24h Us on the first and third DR days.

Relative differences between reported (DR) and measured (24hU) intake were calculated from the log ratio (DR/24hU) for protein, potassium and sodium intake, respectively: -14.4%, +2.6% and -2.1% for men and -13.9 %, -3.7% and -8.3% for women.

The correlations between reported and true intake were 0.61, 0.78 and 0.47 for men and 0.64, 0.42 and 0.37 for women, for protein, potassium and sodium respectively. Attenuation factors, that represent attenuation of the true diet-disease relationship due to measurement error (a value closer to 1 indicating lower attenuation), ranged from 0.23 (sodium, women) to 0.60 (potassium, men).

We showed that the web-based DR tool used in the NutriNet-Santé cohort study performs well in estimating protein and potassium intake and fairly well for sodium intake. Furthermore, three non-consecutive days of DR appear to be valid to estimate usual intake in protein and potassium, although caution is advised regarding the generalizability of these findings to other nutrients and the general population.

Introduction

Collection of high-quality dietary data in large populations is a challenging priority in nutritional epidemiology, in both etiological research and surveillance studies. Bias due to measurement error of dietary factors is now widely acknowledged because no instrument to assess dietary intake is perfectly accurate ([1; 2]). Beyond unreliable descriptions of usual intakes, estimates of relationships between diet and disease may be attenuated or biased towards the null, and measurement error causes a loss of power to detect significant associations ([3]).

The main dietary tools used in nutritional epidemiology are either contemporaneous dietary records, or retrospective instruments such as multiple 24 hour recalls or Food Frequency Questionnaires (FFQs). Until recently, repeated 24h recalls or records on non-consecutive days were not used as main instruments for assessing diet in many cohort studies, because of the substantial costs of repeated assessment to ensure reliable usual intake estimation. Instead, dietary exposure was mostly assessed through FFQs ([4]), despite evidence that repeated 24h recalls, taking into account the day to day variation, outperform FFQs in the accurate assessment of individual usual intake ([5-7]).

The development of new technologies has led to an increasing number of innovative assessment tools, including online options, which are promising for applying in large-scale epidemiological studies ([8; 9]). In this context, web-based self-administered tools for dietary records or 24h dietary recalls could allow for accessing accurate dietary data on large samples with substantial resource savings. However, it is first necessary to validate such tools against objective markers of dietary intake.

'Recovery biomarkers' such as urinary nitrogen, potassium and sodium, are likely to closely reflect true dietary intake of these nutrients, and errors in measuring intake and urinary biomarkers are likely to be independent of each other ([10]). This contrasts 'concentration biomarkers', such as plasma vitamins or fatty acids, which are subject to metabolic regulation and do not always correlate closely with intakes of their corresponding nutrients ([11]).

Recovery biomarkers have been used in various dietary instrument validation studies (FFQs and 24h recalls), including the OPEN study ([4; 12]), the EFCOVAL study ([13]), the Women's Health Initiative Nutritional Biomarker Study ([14]) and the AMPM Study ([15]), where estimates of the difference between reported and measured intakes could be estimated, as well as correlations between intakes and biomarker values.

NutriNet-Santé is the first web-based prospective cohort study that aims to investigate the relationship between nutrition and health ([16]). Diet is assessed by three non-consecutive days

of records at baseline, and again at each year of follow-up. The dietary recording is self-administered through a specific web-based tool, which has shown high agreement with an interview with a dietician as shown by median intra-class correlation and Pearson's correlation coefficients of 0.7 to 0.8 ([17]). However, this comparison study was not able to estimate the ability of the tool to assess true intake.

In this study, we aimed to investigate the validity of a web-based self-administered dietary record (DR) tool of protein, potassium and sodium intake, as assessed by three non-consecutive DR days, against two non-consecutive measures of 24h urinary biomarkers (24h U) of these nutrients.

Materials and methods

Study population and ethics statement

Participants were volunteers participating in the NutriNet-Santé study, an on-going web-based cohort study launched in France in May 2009, whose aims and methods have been described elsewhere ([16]). Briefly, using a dedicated website, adult volunteers (aged >18 years) are followed for at least 10 years (recruitment still on-going). Informed consent is obtained electronically from all participants. All procedures were approved by the International Research Board of the French Institute for Health and Medical Research (IRB Inserm n° 0000388FWA00005831) and the French National Information and Citizen Freedom Committee "CNIL" (n° 908450 and n° 909216). At inception, participants complete a set of questionnaires assessing demographic, socioeconomic and lifestyle factors, dietary intake measurements (three non-consecutive DR days), physical activity (PA), anthropometry and health status. Dietary intake is evaluated again annually and questionnaires on health status are sent on a regular basis.

A randomly selected sample of 1400 NutriNet-Santé study participants living in Paris and greater area (for logistical reasons), stratified by gender, age (<45y, >45y) and educational level (primary and secondary up to some college, university graduate), were invited by e-mail to take part in the Dietary Validation Study. The objective was to recruit 200 participants. Since recovery biomarkers have been shown to be robust markers of dietary intake in individuals who are weight-stable and not experiencing illness ([18]), exclusion criteria were: self-reported metabolic disease (diabetes, heart failure, kidney failure, or intestinal malabsorption e.g. Crohn's disease); adherence to a weight-loss diet with observed weight loss >1.5kg/week over the past 4 weeks; and currently pregnant or breastfeeding.

To ensure the validity of biomarkers derived from 24-hour urine collections using Para-Amino Benzoic Acid (PABA), allergy to PABA was also an exclusion criterion. Participants were already enrolled in the NutriNet-Santé study, and thus all had at least basic computer knowledge and no difficulty in understanding or reading French. The protocol was approved by the Consultation committee for the Protection of Participants in Biomedical Research of Paris Saint-Louis (n°2011/22) and the "CNIL" (DR-2012-467). Participants who completed the study received 100 euros as compensation for the burdensome protocol.

Study design

Recruitment was carried out between October 2012 and April 2013. Interested subjects responded by e-mail, and were subsequently contacted by telephone to check eligibility and to schedule their clinic visits and dates of DRs and 24h Us. The study consisted of two visits at the clinical centre (Hôtel Dieu hospital, Paris), both in a fasting state (6 hours minimum). At the first visit, clinical measurements were taken (blood pressure and heart pulse, height and weight). Participants were given instructions for the 24h U collection, and a physical activity questionnaire (PAQ) on occupational, transport and leisure time PA during the last 4 weeks to fill in at home (paper, self-administered) before the second visit. To complete the three DR days, a specific login and password was given to the participants. The second visit was scheduled approximately 3 weeks later. Between the two visits, three DRs on non-consecutive days were self-administered through the specific web-based tool. Two 24h urine samples were collected per participant, covering the same 24h periods as the first and the third DR days, with a time-lag of approximately 2 weeks between first and third DR. This scheme corresponds to the design participants follow in the NutriNet-Santé study: three DR days randomly allocated over 2 weeks.

Dietary data collection

The web-based tool is designed for self-administration and based on a secured user-friendly interface, designed by Medical Expert Systems © (Paris, France). Participants report all foods and beverages (type and quantity) consumed during all eating occasion during 24 hours from midnight to midnight. Participants first enter a list of every food item consumed at all eating occasions that they can recall via one of two ways: a food browser (foods are grouped by category) or a search engine that accepts spelling errors. Participants then estimate portion sizes with the help of photographs, derived from a previously validated picture booklet that represent more than 250 generic foods ([19]), corresponding to more than 2000 specific food

items, presented in three different portions sizes. Along with the two intermediate and two extreme quantities, there are seven choices of amounts. Participants could also directly enter the quantity of foods consumed in grams or a measure of volume, use purchased units or describe intake in standard household units (e.g. teaspoons, tablespoons). Finally, after all food items and quantities have been entered, a summation is provided and participants have to review and describe if additional salt was consumed, and if so, in what quantity (household units or grams). For each participant, daily nutrient intakes were calculated using the ad-hoc NutriNet-Santé composition table ([20]). An intake below 500kcal per day for women, or 800kcal for men was considered implausible and excluded ([21]), and the final analyses included only participants with at least two valid DRs. Two DRs were collected on weekdays and one on a weekend day.

24h urine collections and recovery biomarkers

At the first clinic visit, participants received instructions, materials (containers, 4 PABA pills) and a questionnaire for each 24h urine collection. They were instructed to discard the first urine of the day of collection, then to collect all urine passed during the next 24 hours, up to the first urine passed on the next morning which was also collected. During the day of collection, the container was kept at room temperature with the instruction to keep it in a dark place. To verify the collection samples, participants were asked to take two 100 mg PABA tablets on the day of collection and were informed that this process was to check the completeness of the collection as it may aid the collection of accurate samples ([22]). On the questionnaire, participants had to provide the times when collection started and finished (the following day), the time at which PABA pills were taken, any missing void, and medications taken on that day. Urine samples were processed straight after collection the following morning: they were weighed, carefully mixed and aliquoted into 1 mL samples and stored at -80°C. In May 2013, all samples were transported to appropriate laboratories.

Urinary nitrogen was measured by pyrochemoluminescence on an Antek 9000 analyzer, which produces results very well correlated to the reference method (the Kjeldahl technique) ([23]), at Cochin Hospital, Université Paris Descartes. Potassium (K) and Sodium (Na) were measured by ion-selective electrodes (Siemens Dimension Vista, Saint-Denis, France) at the laboratory of Nutrition Hormonology in the CHU of Grenoble. Creatinine, used as a marker to check for validity of urine collection, was measured by alkaline picrate kinetic (Siemens Dimension Vista, Saint-Denis, France) also in Grenoble. The CV of these analyses (intra-assay precision) was <3%.

Covariate assessment

Height was measured for shoeless participants to the nearest 0.5 cm by a trained technician, using a wall-mounted stadiometer ([24]). Weight (to the nearest 0.1 kg) of participants (wearing underwear solely) was measured with a calibrated impedance body composition analyzer (BC-418MA, TANITA ©, Tokyo, Japan). Body mass index (BMI) was calculated as the weight (kg) divided by the squared height (m^2). Dietary supplement use, frequency and type were determined by questionnaire.

Statistical analysis

Description of study participants' characteristics (mean ± SD or n, %) were compared between men and women through Kruskal-Wallis (when normality was not met) or t-test for continuous and χ^2 tests for categorical variables.

Assuming that approximately 81% of nitrogen is excreted via urine in 24h, and that proteins contain 16% of N ([25]), that 77% of potassium ([26]) and 86% of sodium ([15]) are excreted in 24h, we could calculate biomarker-based intakes:

Protein (g/d) = Urinary N (mol/L) × Volume 24h U (L) ×14 (g/mol) ×6.25/0.81

Potassium (mg/d) = Urinary Potassium (mol/L) × Volume 24h U (L) × 39 (g/mol) ×1000/0.77

Sodium (mg/d) = Urinary Sodium (mol/L) × Volume 24h U (L) × 23 (g/mol) ×1000/0.86

24h urine collections were determined as valid using the following criteria: collection time between 22 and 26 hours, urine volume ≥ 500mL ([15]) , reported missing urine (estimated volume missed void>5% total volume) and creatinine >10 or >15 mg/kg for women and men respectively ([27]). If one or more of the listed criteria was not met, then the 24h U collection was considered invalid. The following sensitivity analyses were conducted: 1) exclusion of urine samples with > 1 reported missing void because people admitting 1 missing void might be actually more diligent or have missed only a small volume compared to those reporting more than 1 missing void ([22]); and 2) exclusion of participants with one invalid urine measure.

All intake and excretion values were log transformed to improve normality. Intra-cluster correlation coefficients between U1 and U2 (using the mean of 2 measurements), and between the three DRs (using the mean of 3 measurements), were calculated with the SAS macro %ICC9 ([28]). Mean protein, potassium and sodium intake based on up to 3 days of DR (R_{ij} for an individual i on a day j), and excretion on up to 2 days of 24h U (M_{ij}), were calculated on the log-transformed values and exponentiated to obtain geometric means and 95% CI. For an individual i, the log-ratio $\log(\overline{R_i}/\overline{M_i})$ was calculated, where $\overline{R_i}$ is the individual mean of up to three DRs and $\overline{M_i}$ is the mean of up to two 24h Us. After exponentiation of the sample mean log-ratio, with a ratio of 1 representing no difference between intake and excretion, we

expressed the distance to the reference in percent, e.g. a ratio of 0.90 (90%) is equivalent to a relative difference of -10%. Misreporting refers to presence of a significant difference.
A ratio below 70% indicated the presence of severe underreporting, between 70% and 80% moderate underreporting, between 80% and 120% correct reporting, and above 120% overreporting bias ([29]).
We calculated the ratio across age categories (≤45 years old, > 45), and across BMI categories (<25, 25-29.9, ≥30 kg/m^2) and compared them using ANOVA after assumptions were checked.
To assess validity of the dietary record tool, we calculated Pearson correlation coefficients and their confidence interval using the Fisher's Z transformation; both unadjusted and adjusted for age, BMI, physical activity and energy intake (by the residual method ([21])).
To examine the structure of the measurement errors, a complex measurement error was assumed ([10]). It is described in Online Supporting Material. This allowed for the calculation of the correlation between reported and *true* intake *on the same given day* (assesses if the instrument measures what it is supposed to measure), and correlation coefficients between *usual* reported intake and true intake, as well as attenuation factors (λ) ([10]). Attenuation factors represent the attenuation of the strength of the relationship between nutrient intake and a disease of interest; a value closer to 1 meaning that there is less attenuation (with 1 representing no attenuation at all). Although no exact cut-off exists to interpret correlation and attenuation coefficients, a value of at least 0.40 would avoid needing hugely inflated sample sizes to observe significant diet-disease relationship ([30]) hence values ≥0.40 were deemed acceptable/fair, ≥0.60 as high and <0.40 as low.

All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC, USA), the significance level was two-sided and set at 0·05.

Results

Subjects' characteristics

Of the 1400 individuals contacted by e-mail, 237 (16.9%) responded. Of these, 7 (3%) were ineligible and 31 (13%) were not able to attend the planned clinic visits; hence 199 participants were included.
A total of 398 24h U specimens were available. Both 24h U measurements were invalid for four female participants and one male participant; hence these five participants were excluded from the analysis. One man had 1 invalid 24h U and 2 implausible DRs, and was thus

excluded. This meant 193 subjects were included in the analysis sample. 25 subjects had data for only one 24h U because 14 (7.3%) first 24h U and 11 (5.7%) second 24h U were considered invalid.

Participants' characteristics are presented in **Table 1**. The sample was 47.7% female, who did not differ from males in terms of age (mean ±SD: 50.5±16.4 years) or BMI (24.0±3.5 kg/m^2). Obesity (BMI≥30 kg/m^2) was more common in women than men (12% vs 3%), but overweight (25≤BMI<30 kg/m^2) was more common in men (36% vs 18%). Women had a higher frequency of dietary supplement use (36% vs 24%). Men had higher energy intake (10 000 kJ in men vs 7172 kJ in women). Energy from protein was lower for men than women but energy from fat and carbohydrates were not appreciably different.

Intakes of protein, potassium and sodium and misreporting

Intakes of protein, potassium and sodium based on three DR days and two 24h U excretion are presented in **Table 2**. Intra-cluster correlation coefficients between U1 and U2 were 0.60 for proteins, 0.45 for potassium and 0.36 for sodium, and between three diet records 0.52, 0.54, and 0.47 for protein, potassium and sodium respectively.

Men and women underreported their protein intake (-14.4% and -13.9%, respectively, NS between-gender difference p=0.88). Men showed non-significant overreporting for potassium and sodium intake, while women underreported these two nutrients.

Misreporting was greater in women aged >45y than those aged ≤45y for intake of protein (-17% vs -8%, p=0.047) and of sodium (-15% vs +3%, p=0.04), but no significant difference across age categories was observed for potassium, and no misreporting differences were observed for males. By BMI categories, misreporting of sodium intake was greater for obese women than those overweight or normal weight although the difference did not reach statistical significance (-25% in obese, -2% in overweight and -7% in normal weight, p=0.13).

Frequency of misreporting is presented in **Table 3**. The difference between men and women was non-significant, but a trend was observed for potassium with more men overreporting (24.5%) than women (20.9%), and for sodium with more women severely underreporting (29.7%) than men (16.7%).

Correlation and attenuation

Correlation coefficients between intake (DR) and excretion (24h U) are given in **Table 4**. Higher correlations were observed for men than for women for all three nutrients. For men, crude correlations ranged from 0.45 (sodium) to 0.63 (potassium), and for women from 0.27 (sodium) to 0.54 (protein). Adjusted correlations for age, BMI, level of education and energy intake were higher than the crude coefficients for women, but lower for men.

Sensitivity analyses taking into account only the first and third dietary records, which correspond to the days of 24h U collection, showed overall similar results for relative differences and correlations; the only notable exception was a lower correlation between sodium intake and excretion in men (r=0.17).

Taking into account the complex measurement error model, we calculated the correlations between reported intake by one DR and *true* intake *on the same day* (**Table 5**). These coefficients were higher than crude correlations for women, and similar to those for men. Finally, correlations between intake of the *average* of three DRs and *true usual* intake (**Table 6**) were high for protein for both men and women (>0.60), very high for potassium in men, while only fair for women, and fair (men) to poor (women) for sodium. Attenuation factors ranged from 0.23 (sodium, women) to 0.60 (potassium, men).

Discussion

This validation study is the first to examine the structure of the measurement error with repeated web-based self-administered non-consecutive-day dietary records, allowing for the estimation of the correlation with true intake of protein, potassium and sodium. Only a few studies have assessed the validity of repeated short-term instruments, like 24h recalls, against biomarkers ([4; 13; 29; 31-35]) and none have validated web-based self-administered non-consecutive dietary records.

Misreporting of protein, potassium and sodium intake

We found that on average, men underreported protein but slightly overreported their potassium and sodium intake, whereas women underreported protein, potassium and sodium intake. Correlation coefficients indicated that three non-consecutive 24h diet records self-administered via the web-based tool perform well for the estimation of intakes of protein and potassium, and fairly well for estimating sodium intake.

The EFCOVAL and the OPEN studies aimed to validate two 24h recalls, administered by a dietitian, against urinary biomarkers. Results in the French EFCOVAL center showed underreporting of -12.1% for protein and -17.1% for potassium in men and -12.8% and -13.0% respectively in women ([13]). For protein, the results are similar to our findings, but for potassium, underreporting was much more prominent in the EFCOVAL study than in this study. In the American OPEN study, underreporting of protein was also similar (-11% to -12%) ([4]). Regarding sodium, the USDA AMPM Validation study ([15]), with two 24h urine collections covering the same time period as two 24h recalls, showed greater underreporting (-7% for men and -10% for women) than in our study. Protein, potassium and sodium find

their main source in very different food groups, and represent different aspects of diet quality so it is not surprising that dietary misreporting differs across nutrients, as suggested elsewhere ([36]).

Crude correlation coefficients in EFCOVAL were 0.65 (protein) and 0.62 (potassium) in men and 0.46 and 0.61 respectively in women, which is slightly higher than in our study. However, correlation coefficients for protein found in the present study are somewhat higher than usually reported in other validation studies including short term instruments (24h recalls), like in OPEN (r=0.41 for men and r=0.26 for women) ([4]), the DEARR study (r=0.29) ([34]), or the UK arm of EPIC (0.10 for one 24h recall)([31]), and are more similar to the one observed with a 7-day diary (r=0.65) ([31]).

Greater misreporting and lower correlation coefficients for all three nutrients (protein, potassium, sodium) were observed in women than in men in the present study, which is fairly consistent with most of the validation studies of short term instruments for protein ([4; 13]), potassium ([13]) or sodium ([15]). Although the present study does not allow exploring this aspect in depth, differences in social desirability is a potential explanation, because of the societal pressure placed on women to be slim. Women, more than men, may underreport to prevent being seen as indulging in an undesirable behavior, like eating unhealthy food or overeating ([37; 38]).

We found no significant difference in misreporting of protein, potassium or sodium according to BMI categories. However, for protein the trend was towards more underreporting of intake among the overweight or obese than normal weight individuals. Given the very low number of obese men (n=3) in the study, we carried out the analyses between normal weight (BMI<25) and overweight/obese (BMI≥25) and showed the same non-significant trend (-18% in overweight vs -12% in normal weight, p=0.16). This follows the trend observed in the OPEN Study: lower correlation coefficients between reported protein intake (average of two 24h recalls) and biomarkers in obese than in non-obese men (r=0.217 vs 0.483, p=0.05) ([12]).

For potassium, BMI classification did not seem to influence misreporting. For sodium, the AMPM Validation study found that overweight and obese men and women underreported more than their normal weight counterparts. This finding is similar to the trend observed in the present study for women. Across age categories, in the AMPM study, females under 50y tended to underreport sodium intake more than their elder counterparts (-15% vs -5%) whereas we found the opposite. This can be explained by a lower computer knowledge among the older participants ([39]), and these results are consistent with the comparison study of our

tool with a 24h recall assessment by a dietitian, where the proportion of "novice or inexperienced with computer" was higher among women than men ([17]).

Besides, it is known that dietary misreporting (particularly energy underreporting) is more frequent among the elderly ([40; 41]). Our study includes six participants aged ≥75years old (3 men and 3 women). When we excluded them from the main analysis, the results remained unchanged. However, among these 6 participants we observed greater underreporting of potassium (-13.4% in men and -14.6% in women), protein for men (-19.3%) and sodium for women (-36.8%), although the Kruskal-Wallis test showed no significant difference (all $p>0.05$), which is likely to be due to a lack of power. These results may imply that extra attention should be paid to the quality of dietary data when studying diet-disease associations among the elderly.

Correlation with true intake and structure of the measurement error

Correlations between reported intake and true intake were not estimated in the EFCOVAL or AMPM studies, but they were in the OPEN study ([7]). It was estimated that four 24h recalls could lead to a correlation coefficient of 0.508 (men) and 0.440 (women) with true intake of protein. The correlation between the average of three non-consecutive-day records and true intake observed in the present study (0.61 in men and 0.64 in women) are higher and actually outperform the prediction by Schatzkin *et al.* with a theoretically infinite number of 24h recalls (0.597 for men and 0.584 for women) ([7]).

Attenuation factors found in the present study are similar to estimates from four 24h recalls in the OPEN study for protein in men (0.37), and higher in women (0.43 in our study vs 0.32 in OPEN) ([7]); a higher value indicating less bias in estimating diet-health relationships. For potassium, we found a higher attenuation factor, i.e. less bias, than in OPEN for men (0.60 vs 0.32) but a slightly lower factor for women (0.29 vs 0.33) ([4]). No comparison can be made for sodium since, to our knowledge, no other study has estimated attenuation factors for this nutrient.

Finally, this is the first study to assess correlation between web-based self-reported and true intake on a given day, which is a method for evaluating how well the instrument measures its target, without penalizing the correlation for the fact that dietary intake may exhibit considerable daily variability. The correlation coefficients were high for protein in both gender, high for potassium in men and fair in women, and fair for sodium in both men and women. Coefficients were lower for women than men, indicating a lower intrinsic validity of the instrument for women than for men.

Methodological considerations

The main strength of this study is the use of objective biomarkers, namely 24h urinary protein, potassium and sodium, collected on the same day of diet record, in a repeated fashion, which allowed for the estimation of the extent of misreporting, as well as same-day correlations and for usual intake with a complex measurement error model. Accuracy – i.e. completeness – of the 24h urine collections was assessed comprehensively by different criteria: creatinine (5 invalid), total volume (1 invalid) and self-report of missing voids (23 invalid). Also, although PABA was not assayed, participants were asked to take the PABA pills during the collection which potentially has a "placebo effect" to engage in more compliant behavior ([22]). Results of both sensitivity analyses using different criteria for exclusion were identical for women, but there were slightly lower correlation coefficients and attenuation factors were observed for men. This seems to imply that our strategy of exclusion of invalid urine was an adequate balance between accuracy and statistical power.
Finally, as our strategy of excluding DR days with implausibly low energy intake may introduce bias, we repeated the analyses including the three implausible DRs, which did not change the results.
The main limitation of this study is the absence of use of a recovery biomarker for energy intake, namely doubly labeled water, which requires a much more costly and burdensome protocol. Hence, although protein intake, given its caloric content, can be used as a proxy of energy intake, we cannot extrapolate the results on protein intake to other macronutrients or total energy intake, as suggested by the OPEN results ([4; 10]). An important issue in validating dietary assessment tool is the current paucity of valid recovery biomarkers, but emerging food metabolomics studies may be a promising way to assess nutritional intake through biomarkers ([42]).
Caution is advised when extrapolating from the results of the present validation study to the general population because it was carried out on a relatively small sample of subjects. These were volunteers and likely differed in terms of socioeconomic, demographic, and lifestyle characteristics from the general population. However, we carried out our sampling strategy in order to have a wide spectrum of age, education level and equal numbers of men and women so that validity could be assessed irrespective of these parameters.

We showed that the web-based repeated non-consecutive-day DR tool used in the NutriNet-Santé cohort study performs well in estimating protein and potassium intake and fairly well for sodium intake. Furthermore, three repeated DRs appear to be valid to estimate usual intake

in protein and potassium, although caution is advised regarding the generalizability of these findings to other nutrients and to the general population.

Acknowledgements

We thank Dr Amy Subar from the US National Cancer Institute, Bethesda, MD, for her help in designing this study protocol and Drs Victor Kipnis, Douglas Midthune and Pr Laurence Freedman for kindly adapting their measurement error model to our study design and providing SAS codes. We thank all the staff involved in this study, especially Karine Prevost, technician, the team of dieticians, as well as Mehdi Menai and Rachida Mehroug.

Details of measurement error models are available as Online Supporting Material with the online posting of this paper.

Funding: This work was funded by Institut de Veille Sanitaire (InVS) and supported by grants from the Région Ile de France (CORDDIM).

The NutriNet-Santé study is being supported by the following institutions: Ministère de la Santé (DGS), InVS, Institut National de la Prévention et de l'Education pour la Santé (INPES), Fondation pour la Recherche Médicale (FRM), Institut National de la Santé et de la Recherche Médicale (INSERM), Institut National de la Recherche Agronomique (INRA), Conservatoire National des Arts et Métiers (CNAM) and Université Paris 13.

Conflict of interest: The authors declare no conflicts of interest.

Author Contribution Statement: CL, EKG, KC, VD, MV, PG and SH were responsible for developing the design and protocol of the study. CL conducted research, carried out data checking and analyses and was responsible for drafting the manuscript. KC, VD, MV, GC, PG, SH, EKG, FL and PF were involved in interpreting results and editing the manuscript. FL and PF carried out the biomarker analyses. All authors read and approved the final manuscript.

References

1. Kipnis V, Midthune D, Freedman L, *et al.* (2002) Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr* **5**, 915-923.

2. Schatzkin A, Subar AF, Moore S, *et al.* (2009) Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiol Biomarkers Prev* **18**, 1026-1032.

3. Kipnis V, Midthune D, Freedman LS, *et al.* (2001) Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol* **153**, 394-403.

4. Subar AF, Kipnis V, Troiano RP, *et al.* (2003) Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol* **158**, 1-13.

5. Carroll RJ, Midthune D, Subar AF, *et al.* (2012) Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *Am J Epidemiol* **175**, 340-347.

6. Prentice RL, Mossavar-Rahmani Y, Huang Y, *et al.* (2011) Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *Am J Epidemiol* **174**, 591-603.

7. Schatzkin A, Kipnis V, Carroll RJ, *et al.* (2003) A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. *Int J Epidemiol* **32**, 1054-1062.

8. Hercberg S (2012) Web-based studies: The future in nutritional epidemiology (and overarching epidemiology) for the benefit of public health? *Prev Med* **55**, 544-545.

9. Illner AK, Freisling H, Boeing H, *et al.* (2012) Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. *Int J Epidemiol* **41**, 1187-1203.

10. Kipnis V, Subar AF, Midthune D, *et al.* (2003) Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* **158**, 14-21.

11. Freedman LS, Kipnis V, Schatzkin A, *et al.* (2010) Can we use biomarkers in combination with self-reports to strengthen the analysis of nutritional epidemiologic studies? *Epidemiol Perspect Innov* **7**, 2.

12. Lissner L, Troiano RP, Midthune D, *et al.* (2007) OPEN about obesity: recovery biomarkers, dietary reporting errors and BMI. *Int J Obes (Lond)* **31**, 956-961.

13. Crispim SP, de Vries JH, Geelen A, *et al.* (2011) Two non-consecutive 24 h recalls using EPIC-Soft software are sufficiently valid for comparing protein and

potassium intake between five European centres--results from the European Food Consumption Validation (EFCOVAL) study. *Br J Nutr* **105**, 447-458.

14. Neuhouser ML, Tinker L, Shaw PA, *et al.* (2008) Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am J Epidemiol* **167**, 1247-1259.

15. Rhodes DG, Murayi T, Clemens JC, *et al.* (2013) The USDA Automated Multiple-Pass Method accurately assesses population sodium intakes. *Am J Clin Nutr* **97**, 958-964.

16. Hercberg S, Castetbon K, Czernichow S, *et al.* (2010) The Nutrinet-Sante Study: a web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status. *BMC Public Health* **10**, 242.

17. Touvier M, Kesse-Guyot E, Mejean C, *et al.* (2010) Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. *Br J Nutr*, 1-10.

18. Bingham SA (2002) Biomarkers in nutritional epidemiology. *Public Health Nutrition* **5**, 821-827.

19. Le Moullec N, Deheeger M, Preziosi P, *et al.* (1996) Validation of the photo manual used for the collection of dietary data in the SU.VI.MAX. study. *Cahier de Nutrition et Diététique*.

20. NutriNet-Santé coordination (2013) *Table de composition des aliments - Etude NutriNet-Santé*. Paris: Economica.

21. Willett WC (1998) *Nutritional Epidemiology*, 2nd ed. ed. New York, NY: Oxford University Press.

22. Subar AF, Midthune D, Tasevska N, *et al.* (2013) Checking for completeness of 24-h urine collection using para-amino benzoic acid not necessary in the Observing Protein and Energy Nutrition study. *Eur J Clin Nutr*.

23. Neveux N, David P & Cynober L (2004) Measurement of amino acid concentrations in biological fluids and tissues using ion exchange chromatography. In *Metabolic and Therapeutic Aspects of Amino Acids in Clinical Nutrition*, pp. 17-28 [L Cynober, editor]. Boca Raton: CRC Press.

24. Lohman T, Roche A & Martorell R (1988) *Anthropometric Standardization Reference Manual*, Human Kinetics Books ed. Champaign, Illinois.

25. Bingham SA (2003) Urine nitrogen as a biomarker for the validation of dietary protein intake. *J Nutr* **133 Suppl 3**, 921S-924S.

26. Tasevska N, Runswick SA & Bingham SA (2006) Urinary potassium is as reliable as urinary nitrogen for use as a recovery biomarker in dietary studies of free living individuals. *J Nutr* **136**, 1334-1340.

27. Stein J (1998) *Internal medicine*. St Louis, Maryland Heights, MO, USA: Elsevier Health Sciences.

28. Spiegelman D (2014) %icc9 SAS Program. Intraclass correlation coefficients and their 95 percent confidence intervals.

29. Arab L, Tseng CH, Ang A, *et al.* (2011) Validity of a multipass, web-based, 24-hour self-administered recall for assessment of total energy intake in blacks and whites. *Am J Epidemiol* **174**, 1256-1265.

30. Freedman LS, Commins JM, Moler JE, *et al.* (2014) Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol* **180**, 172-188.

31. Bingham SA, Gill C, Welch A, *et al.* (1997) Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* **26 Suppl 1**, S137-S151.

32. Kahn HA, Whelton PK, Appel LJ, *et al.* (1995) Validity of 24-hour dietary recall interviews conducted among volunteers in an adult working community. *Ann Epidemiol* **5**, 484-489.

33. Moshfegh AJ, Rhodes DG, Baer DJ, *et al.* (2008) The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. *Am J Clin Nutr* **88**, 324-332.

34. Shai I, Rosner BA, Shahar DR, *et al.* (2005) Dietary evaluation and attenuation of relative risk: multiple comparisons between blood and urinary biomarkers, food frequency, and 24-hour recall questionnaires: the DEARR study. *J Nutr* **135**, 573-579.

35. Slimani N, Bingham S, Runswick S, *et al.* (2003) Group level validation of protein intakes estimated by 24-hour diet recall and dietary questionnaires against 24-hour urinary nitrogen in the European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study. *Cancer Epidemiol Biomarkers Prev* **12**, 784-795.

36. Freedman LS, Commins JM, Moler JE, *et al.* (2014) Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol* **180**, 172-188.

37. Novotny JA, Rumpler WV, Riddick H, *et al.* (2003) Personality characteristics as predictors of underreporting of energy intake on 24-hour dietary recall interviews. *J Am Diet Assoc* **103**, 1146-1151.

38. Johnson RK, Goran MI & Poehlman ET (1994) Correlates of over- and underreporting of energy intake in healthy older men and women. *Am J Clin Nutr* **59**, 1286-1290.

39. Klovning A, Sandvik H & Hunskaar S (2009) Web-based survey attracted age-biased sample with more severe illness than paper-based survey. *J Clin Epidemiol* **62**, 1068-1074.

40. Bazelmans C, Matthys C, De HS, *et al.* (2007) Predictors of misreporting in an elderly population: the 'Quality of life after 65' study. *Public Health Nutr* **10**, 185-191.

41. Yannakoulia M, Tyrovolas S, Pounis G, *et al.* (2011) Correlates of low dietary energy reporting in free-living elderly: the MEDIS study. *Maturitas* **69**, 63-68.

42. Beckmann M, Lloyd AJ, Haldar S, *et al.* (2013) Dietary exposure biomarker-lead discovery based on metabolomics analysis of urine samples. *Proc Nutr Soc* **72**, 352-361.

Table 1. Characteristics of the Participants in the NutriNet-Santé Dietary Validation Study, France, 2013

	Men n=104		Women n=95		
	Mean	SD	Mean	SD	*p*-value[a]
Age (y)	50.3	16.1	50.7	16.8	0.9
Median, Q1-Q3	51	35-65	54	35-65	
BMI (kg/m²)	24.1	2.9	23.9	4.2	0.6
Weight (kg)	74.8	10.9	62.7	10.7	<.0001
Height (cm)	176.0	7.1	162.3	6.0	<.0001
Physical activity (MET-h/week)	85.1	48.7	81.6	50.1	0.6
LTPA (MET-h/week)	35.7	29.9	21.4	21.9	0.0002
	n	%	n	%	*p*-value[a]
Use of dietary supplement	25	24.0	34	35.8	0.07
BMI category					0.001
Underweight (<18.5)	1	1.0	7	7.4	
Normal (18.5-24.9)	63	60.6	60	63.2	
Overweight (25-29.9)	37	35.6	17	17.9	
Obese (≥30)	3	2.9	11	11.6	
Tobacco smoking					0.35
Smoker - regularly	9	8.7	10	10.5	
Smoker - occasionally	3	2.9	6	6.3	
Former smoker	39	37.5	26	27.4	
Never smoker	53	51.0	53	55.8	
Living with a partner	69	66.3	53	55.8	0.13
Occupation					0.04
Never employed	3	2.9	6	6.3	
Self-employed. farmers	2	1.9	1	1.1	
Managerial/professional position	45	43.3	30	31.6	
Manual workers	1	1.0	0	0.0	
Blue collar	15	14.4	27	28.4	
Retired	38	36.5	31	32.6	
Education					0.10
Up to high school	21	20.2	18	18.9	
Some college	34	32.7	25	26.3	
University graduate	49	47.1	52	54.7	
Dietary intake [b]	Mean	SD	Mean	SD	*p*-value[a]
Energy (kJ)	9999.8	2536.8	7172.2	1735.9	<.0001
Carbohydrate density [c]	42.2	6.7	41.2	6.9	0.31
Protein density [c]	16.7	3.5	17.8	3.8	0.03
Lipid density [c]	40.9	6.6	40.7	6.9	0.86
Alcohol (g)	13.9	16.4	7.3	8.6	0.001
Dietary fiber (g)	24.5	9.7	20.0	6.0	0.0001

Abbreviations: BMI, body mass index; LTPA, leisure time physical activity; SD, standard deviation
[a] P-value for the difference between men and women, t-test or χ^2 tests as appropriate
[b] Mean intake calculated from three non-consecutive DR days
[c] % of energy intake (excluding alcohol)

Table 2. Intake of Protein, Potassium and Sodium From Non-Consecutive-Day Dietary Records and 24h urine Excretions, NutriNet-Santé Dietary Validation Study, France 2013

	Men n=102				Women n=91				
	n	Mean [a]	95% CI		n	Mean [a]	95% CI		*p*-value[b]
Protein (g/day)									
24h U 1	96	104.8	61.4	179.0	86	82.9	49.0	140.3	<.0001
24h U 2	97	102.4	62.3	168.2	82	76.6	41.1	142.8	<.0001
Mean 24h U	102	101.7	62.3	166.2	91	77.4	45.8	130.5	<.0001
24h DR 1	102	90.3	84.5	96.5	91	70.3	66.1	74.7	<.0001
24h DR 2	102	88.9	82.7	95.6	90	68.4	63.1	74.0	<.0001
24h DR 3	101	86.9	81.2	93.1	90	67.6	63.2	72.4	<.0001
Mean 24h DR	102	88.6	83.9	93.7	91	68.8	65.1	72.8	<.0001
Difference % [c]	102	-14.4	-18.2	-10.3	91	-13.9	-18.3	-9.3	0.88
Potassium (mg/day)									
24h U 1	96	3407	3210	3616	86	3012	2814	3224	0.008
24h U 2	97	3353	3165	3552	82	2672	2486	2872	<.0001
Mean 24h U	102	3357	3189	3535	91	2843	2685	3010	<.0001
24h DR 1	102	3468	3266	3683	91	2800	2624	2988	<.0001
24h DR 2	102	3490	3279	3714	90	2684	2530	2847	<.0001
24h DR 3	101	3379	3191	3577	90	2717	2545	2900	<.0001
Mean 24h DR	102	3444	3279	3618	91	2739	2607	2879	<.0001
Difference % [c]	102	2.6	-1.7	7.1	91	-3.6	-8.9	1.9	0.08
Sodium (mg/day)									
24h U 1	96	3667	3355	4007	86	3105	2855	3377	0.009
24h U 2	97	3576	3295	3881	82	2836	2581	3118	0.0003
Mean 24h U	102	3578	3320	3856	91	2996	2790	3217	0.001
24h DR 1	102	3600	3308	3918	91	2812	2580	3065	<.0001
24h DR 2	102	3503	3195	3841	90	2703	2467	2962	0.0001
24h DR 3	101	3411	3139	3706	90	2706	2485	2948	0.0002
Mean 24h DR	102	3503	3271	3752	91	2747	2567	2941	<.0001
Difference % [c]	102	-2.1	-9.2	5.6	91	-8.3	-15.7	-0.2	0.26

Abbreviations: 95% CI; 95% Confidence Interval; DR, dietary record; 24h U, 24-hour urine collection

[a] Values are geometric means.

[b] P-value of t-test for the difference between men and women.

[c] Mean difference in % calculated from the log ratio of mean reported intake (non-consecutive DRs) over mean biomarker intake (24h Us) following the formula $100\left[exp\left(\frac{\sum_{i=1}^{n}\log\left(\frac{\overline{R_i}}{\overline{M_i}}\right)}{n}\right)-1\right]$ where $\overline{R_i}$ is the geometric mean of DRs for an individual *i* across the three measurements, $\overline{M_i}$ is the geometric mean of 24h U for an individual *i* across the two measurements, and n the number of individuals in the sample. A mean log ratio of zero would represent no difference in reporting compared with the biomarker measure. The exponentiation allows to express it as a ratio which reference value is 1 and we further expressed it as a percent difference, eg a ratio of 0.90 is a percent difference of -10%

Table 3. Frequency of misreporting[a] in Protein, Potassium and Sodium intake, NutriNet-Santé Dietary Validation Study, France 2013

	Men n=102		Women n=91		
	n	%	n	%	*p*-value [b]
Protein					0.88
Overreporter	7	6.9	9	9.9	
Correct reporter	58	56.9	51	56.0	
Moderate underreporter	17	16.7	13	14.3	
Severe underrepoter	20	19.6	18	19.8	
Potassium					0.22
Overreporter	25	24.5	19	20.9	
Correct reporter	63	61.8	51	56.0	
Moderate underreporter	11	10.8	12	13.2	
Severe underrepoter	3	2.9	9	9.9	
Sodium					0.19
Overreporter	30	29.4	24	26.4	
Correct reporter	38	37.3	26	28.6	
Moderate underreporter	17	16.7	14	15.4	
Severe underrepoter	17	16.7	27	29.7	

[a] Based on the log ratio of mean reported intake (non-consecutive DRs) over mean biomarker intake (24h Us) Ratio< 70% : severe underreporter ; 70% < ratio < 80% : moderate underreporter ; 80% < ratio < 120% : normo-reporter; ratio > 120% : overreporter

[b] P-value for Fisher's exact test

Table 4. Pearson Correlation Coefficients (r) Between Reported Intake by Three Non-Consecutive-Day DRs and Excretion in two 24h Us for Protein Potassium and Sodium Intake, NutriNet-Santé Dietary Validation Study, France 2013

	Men n=102			Women n=91		
	r	95% CI		r	95% CI	
Protein						
Unadjusted	0.61	0.47	0.72	0.54	0.37	0.67
Adjusted [a]	0.56	0.41	0.68	0.55	0.39	0.68
Potassium						
Unadjusted	0.63	0.50	0.74	0.45	0.27	0.60
Adjusted [a]	0.62	0.48	0.73	0.51	0.33	0.65
Sodium						
Unadjusted	0.45	0.28	0.59	0.27	0.06	0.45
Adjusted [a]	0.31	0.12	0.48	0.34	0.14	0.52

[a] Pearson correlation adjusted for energy intake by the residual method, age, BMI, and level of education.

Table 5. Estimated Correlation (r) Between one DR and True Intake on the Same Day for Protein, Potassium and Sodium Intake, NutriNet-Santé Dietary Validation Study, France 2013

	Men n=102			Women n=91		
	r [a]	95% CI		r [a]	95% CI	
Protein	0.60	0.46	0.75	0.59	0.41	0.76
Potassium	0.68	0.53	0.83	0.51	0.25	0.77
Sodium	0.45	0.26	0.63	0.39	0.11	0.66

[a] Correlation coefficient between DR and *true intake on the same given day* as estimated by the model accounting for the reference biomarkers (24hU) as reference measurement. For more detail on calculation see Online Supporting Material.

Table 6. Estimated Correlation Between the Average of Three non-consecutive-day DRs and True Usual Intake and Attenuation Factor for Protein, Potassium and Sodium Intake, NutriNet-Santé Dietary Validation Study, France 2013

	Men n=102			Women n=91		
	r [a]	95% CI		r [a]	95% CI	
Protein	0.61	0.43	0.78	0.64	0.43	0.85
Potassium	0.78	0.61	0.94	0.42	0.13	0.71
Sodium	0.47	0.23	0.71	0.37	0.03	0.70
	Atten[b]	95% CI		Atten[b]	95% CI	
Protein	0.37	0.24	0.50	0.43	0.26	0.59
Potassium	0.60	0.44	0.76	0.29	0.06	0.52
Sodium	0.37	0.17	0.56	0.23	0.01	0.45

[a] Correlation coefficient between the average of three non-consecutive-day DRs and *true usual intake* as estimated by the model accounting for the reference biomarkers (average of three 24hU) as reference measurement.

[b] Attenuation factor. Interpretation: a value closer to 1 indicates lower attenuation of the true relationship between intake and disease. For more detail on calculation see Online Supporting Material.

Online supporting material. Measurement error model for repeated non-consecutive-day dietary records

The following measurement error model was assumed:

$$\begin{aligned} R_{ij} &= \beta_0 + \beta_1 T_{ij} + \beta_2^T X_i + s_i + e_{ij}, \\ M_{ij} &= T_{ij} + \nu_{ij}, \\ T_{ij} &= T_i + d_{ij}, \\ T_i &= \gamma_0 + \gamma_1^T X_i + u_i, \end{aligned} \qquad (1)$$

Where, for subject i, T_i is the true usual dietary intake (average true intake over some specified time period), T_{ij} is the true dietary intake on day j, R_{ij} the self-reported dietary intake on day j (DR), M_{ij} the biomarker-measured dietary intake on day j (24h U), X_i = a (q×1) a vector of covariates measured without error. e_{ij} and ν_{ij} are random within-person errors with means zero and variances σ_e^2 and σ_ν^2, respectively, s_i is a person-specific bias (random effect) with mean zero and variance σ_s^2, d_{ij} is day-to-day variation in true intake with mean zero and variance σ_δ^2, and u_i is the residual error in the regression of T_i on X_i, with mean zero and variance σ_u^2. We assume that e_{ij}, ν_{ij}, d_{ij}, s_i, and u_i are independent of each other and independent of X_i. Under model (1), we have the following conditional variances and covariances of T_i and T_{ij} given X_i:

$$\begin{aligned} \mathrm{Var}(T_i \mid X_i) &= \sigma_u^2, \\ \mathrm{Var}(T_{ij} \mid X_i) &= \sigma_u^2 + \sigma_\delta^2, \\ \mathrm{Cov}(R_{ij}, T_i \mid X_i) &= \beta_1 \sigma_u^2, \\ \mathrm{Cov}(R_{ij}, T_{ik} \mid X_i) &= \beta_1 \sigma_u^2, \qquad j \neq k \\ \mathrm{Cov}(R_{ij}, T_{ij} \mid X_i) &= \beta_1 (\sigma_u^2 + \sigma_\delta^2), \end{aligned} \qquad (2)$$

The conditional **correlation of R_{ij} and T_{ij} (reported and true intakes on the same day)** given X_i is

$$\mathrm{corr}(R_{ij}, T_{ij}) = \frac{\beta_1 \sqrt{\sigma_u^2 + \sigma_\delta^2}}{\sqrt{\beta_1^2 (\sigma_u^2 + \sigma_\delta^2) + \sigma_s^2 + \sigma_e^2}}. \qquad (3)$$

The conditional (or partial) **correlation of R_{ij} and true usual intake T_i** given X_i is

$$\text{corr}(R_{ij}, T_i \mid X_i) = \frac{\text{cov}(R_{ij}, T_i \mid X_i)}{\sqrt{\text{var}(R_{ij} \mid X_i)\,\text{var}(T_i \mid X_i)}} = \frac{\beta_1 \sigma_u}{\sqrt{\beta_1^2(\sigma_u^2 + \sigma_\delta^2) + \sigma_s^2 + \sigma_e^2}}, \quad (4)$$

and the **attenuation factor λ** is

$$\text{atten}(R_{ij} \mid X_i) = \frac{\text{cov}(R_{ij}, T_i \mid X_i)}{\text{var}(R_{ij} \mid X_i)} = \frac{\beta_1 \sigma_u^2}{\beta_1^2(\sigma_u^2 + \sigma_\delta^2) + \sigma_s^2 + \sigma_e^2}. \quad (5)$$

The maximum likelihood estimates of the parameters were obtained fitting a linear mixed model using the MIXED procedure in SAS.