Validation of the Oxford WebQ Online 24-hour Dietary Questionnaire Using Biomarkers

Darren C. Greenwood, Laura J. Hardie, Gary S. Frost, Nisreen A. Alwan, Kathryn E. Bradbury, Michelle Carter, Paul Elliott, Charlotte E.L. Evans, Heather E. Ford, Neil Hancock, Timothy J. Key, Bette Liu, Michelle A. Morris, Umme Z. Mulla, Katerina Petropoulou, Gregory D.M. Potter, Elio Riboli, Heather Young, Petra A. Wark, and Janet E. Cade.

Correspondence to: Dr Darren C Greenwood, School of Medicine, University of Leeds, Leeds LS2 9JT, UK. (e-mail: d.c.greenwood@leeds.ac.uk)

Author affiliations: School of Medicine, University of Leeds, Leeds LS2 9JT, UK (Darren C. Greenwood, Laura J. Hardie, and Gregory D.M. Potter); Leeds Institute for Data Analytics, University of Leeds, Leeds LS2 9JT, UK (Darren C. Greenwood and Michelle A. Morris); Nutrition and Dietetic Research Group, Division of Diabetes, Endocrinology and Metabolism, Faculty of Medicine, Imperial College London, London W12 ONN, UK (Gary S. Frost, Heather E. Ford, and Katerina Petropoulou); Academic Unit of Primary Care and Population Sciences, Faculty of Medicine, University of Southampton SO16 6YD, UK (Nisreen A. Alwan); National Institute for Health Research (NIHR) Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton SO16 6YD, UK (Nisreen A. Alwan); Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK (Kathryn E. Bradbury, Timothy J. Key, Bette Liu, and Heather Young); National Institute for Health Innovation, University of Auckland, Auckland 1072, New Zealand (Kathryn E. Bradbury); Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds, Leeds LS2 9JT, UK (Michelle Carter, Charlotte E.L. Evans, Neil Hancock, and Janet E. Cade); Medical Research Council – Public Health England (MRC-PHE) Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London W2 1PG, UK (Paul Elliott and Elio Riboli); NIHR Imperial Biomedical Research Centre, Imperial College London, London, UK (Paul Elliott); School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia (Bette Liu); Global eHealth Unit, Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, UK (Umme Z.

Mulla and Petra A. Wark); Centre for Innovative Research Across the Life Course, Faculty of Health and Life Sciences, Coventry University, Coventry, UK (Petra A. Wark).

Author notes: Darren C. Greenwood, Laura J. Hardie, and Gary S. Frost contributed equally as joint first authors; Petra A. Wark and Janet E. Cade contributed equally as joint senior authors.

Funding: This work was supported by the Medical Research Council (UK) grant MRC G1100235. Prof. Gary Frost acknowledges support from a NIHR Senior Investigator award. Prof. Paul Elliott acknowledges support from the Medical Research Council and Public Health England (MR/L01341X/1) for the MRC-PHE Centre for Environment and Health; and the National Institute for Health Research Health Protection Research Unit in Health Impact of Environmental Hazards (HPRU-2012-10141) and additional support from the NIHR Imperial College Biomedical Research Centre in collaboration with Imperial College NHS Healthcare Trust. Prof. Janet Cade acknowledges additional support from the Medical Research Council (MR/L02019X/1).

Conflict of Interest: Prof. Janet E. Cade is a director of a University of Leeds spin out private company, Dietary Assessment Ltd., supporting the development of myfood24. Dr. Michelle Carter, Neil Hancock and Dr. Michelle A. Morris are shareholders in Dietary Assessment Ltd. The company has been running for 1 year and has not made a profit

(or loss). Prof. Gary Frost has worked as a consultant for a number of food companies including Nestlé, Unilever, Quorn, New Food Innovations and Malaysian Palm Oil.

Running head: Validation of the Oxford WebQ Dietary Assessment

ABSTRACT

Oxford WebQ is an online dietary questionnaire covering 24 hours, appropriate for repeated administration in large-scale prospective studies including UK Biobank and the Million Women Study. We compared performance of the Oxford WebQ and a traditional interviewer-administered multi-pass 24-hour recall against biomarkers for protein, potassium and total sugar intake, and total energy expenditure estimated by accelerometry. 160 participants were recruited between 2014 and 2016 in London, UK, and measured at 3 non-consecutive time-points. The measurement error model simultaneously compared all 3 methods. Attenuation factors for protein, potassium, sugars and total energy intake estimated by the mean of 2 Oxford WebQs were 0.37, 0.42, 0.45, and 0.31 respectively, with performance improving incrementally for the mean of more measures. Correlation between the mean of 2 Oxford WebQs and estimated true intakes, reflecting attenuation when intake is categorised or ranked, was 0.47, 0.39, 0.40, and 0.38 respectively, also improving with repeated administration. These were similar to the more administratively burdensome interviewer-based recall. Using objective biomarkers as the standard, Oxford WebQ performs well across key nutrients in comparison with more administratively burdensome interviewer-based 24-hour recalls. Attenuation improves when the average is taken over repeated administration, reducing measurement error bias in assessment of diet-disease associations.

KEYWORDS: Dietary assessment; Million Women study; nutrition assessment; recall; recovery biomarkers; UK Biobank; validation

ABBREVIATIONS
BMI, body mass index;
CI, confidence interval;
FFQ, food frequency questionnaire;
MPR, multiple pass 24-hour dietary recall;
PABA, 4-aminobenzoic acid;
SD, standard deviation;
TEE, total energy expenditure.

Dietary intakes estimated from self-reported dietary assessments are prone to measurement error, introducing potentially substantial bias and loss of power(1,2). It is therefore important to calibrate self-reported intakes against objective biomarkers, where measurement errors can be assumed independent(3,4). Most cohort studies have used food frequency questionnaires (FFQs) designed to assess diet over the long-term, but short-term recalls may have less bias from measurement error(5-7), other than for episodically-consumed foods. Repeated application of short-term recalls may offer longer-term coverage, but is administratively burdensome. Online dietary assessment offers repeated administration with reduced administrative costs(8), but to facilitate this, must be convenient for the participant to use(9,10).

The Oxford WebQ is an online dietary questionnaire covering the previous day's intake(11), developed to provide an easy-to-complete dietary assessment appropriate for repeated use in large-scale prospective studies. It is currently used in UK Biobank(12-14) and the Million Women Study(15,16).

The Oxford WebQ has previously been shown to provide similar results to an interviewer-administered self-report 24-hour dietary recall but is quicker to complete(17,18). However, the comparison tool was itself self-reported, providing an inadequate basis for validation, because self-report tools are prone to correlated person-specific biases(19-21). These biases may differ by personal characteristics such as age, sex or body mass index (BMI).

We therefore aimed to provide the first validation of the Oxford WebQ tool against established recovery and predictive nutritional biomarkers and a reference measure of energy expenditure free from these person-specific biases. In doing so, we present the degree to which diet–disease relationships assessed using Oxford WebQ are attenuated and the extent to which statistical power to detect these comparisons is reduced even in large-scale studies such as the UK Biobank and the Million Women Study.

METHODS

Recruitment

Participants were enrolled into a study designed to validate both the Oxford WebQ dietary assessment tool and the myfood24 dietary assessment tool(22) against nutritional biomarkers, comparing these with a standard interviewer-based multiple pass 24-hour dietary recall(MPR)(23). Eligibility criteria were aimed at recruiting participants broadly representative of the adult general population. Participants were eligible for the

study if they were aged between 18 and 65 years and were maintaining a stable weight, confirmed by no substantive weight loss or weight gain over the study (>5% weight change from first clinic appointment). Further criteria included regular access to high-speed internet, use of a telephone and ability to speak and read English so they could complete the online questionnaires and 24-hour recalls. Participants had to be willing to visit the Clinical Research Facility at Hammersmith Hospital, London (Imperial College Healthcare NHS Trust) to provide blood and urine samples. Participants were identified between 2014 and 2016 through a multidisciplinary network of primary care professionals and practices, the North West London Primary Care Research Network, and individuals known to the Clinical Research Facility who had previously expressed an interest in participating in research projects. Participants were also identified from a list of local addresses provided by the post office. Participants were not a sub-sample of UK Biobank, but an independent sample designed to be of similar age and sex distribution. On completion, participants were provided with modest financial reimbursement for their time. The recruitment target was 200 participants with complete information collected (see Web Appendix 1).

Overview of study design

Each participant provided 3 sets of urine samples for reference measures (recovery biomarkers, predictive biomarkers and total energy expenditure), completed 3 MPRs and 3 Oxford WebQ online dietary questionnaires, all spread over a 5 week period. This data collection was achieved in 3 separate cycles, each 2 weeks apart (Figure 1). At the start of each cycle the participants provided the set of reference measures, followed by

a dietary assessment 1 to 3 days later, and another dietary assessment, 2 to 4 days after that. The order of the dietary assessments within each cycle was allocated by simple randomisation to reduce order effects. Each of the assessments is described in detail below.

Biomarkers

Participants provided 24-hour urine samples, discarding the first morning void, then collected every subsequent urine specimen for the remaining 24 hours, ending with the last specimen the following morning. Urine specimens were then returned to the clinic on the day that collection ended. Urine volumes were recorded and then aliquoted into separate 50 ml aliquots before being stored at -20°C and transported to the Molecular Epidemiology Unit at the University of Leeds. The Kjeldahl method(24) was used to measure the total urinary nitrogen content of the samples. Participants took 3 80mg 4-Aminobenzoic acid (PABA) tablets with meals during the course of the 24-hour urine sample period to confirm completeness of the samples(25). PABA concentration in the urine was measured using high performance liquid chromatography. We considered 93% PABA indicated complete urine collection over the 24 hours, but 85-110% was permissible, consistent with previous research(25).

Protein intake was estimated based on the assumption that 81% of nitrogen is excreted within 24 hours(26). Potassium intakes were estimated from the amount excreted in the urine, as measured by the Clinical Biochemistry Department at the Leeds Teaching

Hospitals NHS Trust using an ADVIA 2400 Clinical Chemistry System (Siemens AG, Munich, Germany) with ion selective electrode detection. We assumed that 80% of potassium intake is excreted in the urine(27).

Urinary concentrations of fructose and sucrose were measured using a Sucrose/D-Glucose/D-Fructose assay (Boehringer Mannheim/R-Biopharm AG, Darmstadt, Germany) scaled down to a microplate format. Daily excretion of urinary sucrose and fructose was then estimated based on total urine volume collected over 24 hours. The predicted intake of total sugars for each individual, allowing for age and sex, was then estimated using a calibration equation derived from previous feeding studies comprising 30-days' intervention in a metabolic suite under controlled conditions (28,29).

Total Energy Expenditure

Resting Energy Expenditure was measured using open-loop indirect calorimetry (Gas Exchange Monitor, GEM Nutrition, Daresbury, UK), assessed at the research facility when participants came for their clinic visit. The calorimeter was calibrated and volunteers lay in a semi-recumbent position. Following stabilisation of measurements, oxygen consumption (VO2) and carbon dioxide production (VCO2) were recorded every minute for 15 minutes. The means of the last ten sets of measurements were used to estimate Resting Energy Expenditure (30). Activity Energy Expenditure was also estimated, using three-plane accelerometry, by a SenseWear armband mini accelerometer (BodyMedia Inc., Pittsburgh, USA). This was worn for 24 hours on the

left upper arm on one of the days before each clinic visit. The thermic effect of food was assumed to be 10% of Total Energy Expenditure (TEE)(31). TEE was estimated by summing Resting Energy Expenditure, Activity Energy Expenditure and the thermic effect of food, with estimated TEE indicating total energy intake, providing individuals remained in energy balance. This method has previously demonstrated close agreement with energy expenditure estimated using doubly-labelled water(32). TEE estimates for participants with more than 5% weight change over the study were excluded. Within-person variability was taken into account in statistical analysis for all repeated measures.

Oxford WebQ online dietary questionnaire

The development of the Oxford WebQ online dietary questionnaire has been described in full elsewhere(11,17,18). Briefly, the tool was designed as a web-based dietary questionnaire that was easy to use both by the participants and the researchers in large-scale observational studies, through extensive piloting and iterative improvement. The Oxford WebQ presents participants with 21 broad food groups, with options then expanding to offer over 200 commonly consumed foods and drinks. The participants are prompted to select the amount consumed over the previous 24 hours, mostly from pre-defined categories offered to them. To facilitate large-scale automatic coding of nutrient information, use of free-text boxes is minimised. On completion of the tool, the participants are presented with a summary page of all the food and drink items they reported consuming, together with the amounts reported, and the participants are asked to make any necessary amendments. Completed questionnaires are coded automatically by multiplying the amount consumed by the nutrient content specified in

standard UK food composition tables(33) to provide a profile of 21 separate nutrients without any additional intervention required by nutritionists.

Interviewer-administered 24-hour recall

To facilitate comparison of the Oxford WebQ with an equivalent interviewer-administered tool, participants also completed an MPR which was conducted using a prompt sheet based on the 5-step multipass method over the telephone by a trained researcher(34). Participants were asked to provide details on cooking methods, brand names and portion sizes. Nutrient intake was estimated using Dietplan6.7 software (Forestfield Software, Horsham, UK), based on the same food composition tables as Oxford WebQ(33). Trained researchers matched the food and drink items recorded to the food composition tables and applied the portion sizes using a standard operating protocol described fully elsewhere(35).

Statistical analysis

Urine samples with 2 or more voids missed during the 24-hour period were excluded. Apart from this, the main analyses included all participants(36). The robustness of urinary biomarker results to completeness of the urine samples was assessed by conducting a sensitivity analysis including only participants who had complete PABA recovery (85-110%) or whose PABA recovery was 50-85% with their urinary nitrogen and potassium rescaled to the 93% PABA recovery expected for complete recovery, consistent with previous research(37).

We present results for both nutrients and nutrient densities. The densities are defined as the ratio of nutrient intake (g) to energy intake (MJ) measured by the same dietary assessment tool to represent energy-adjusted quantities derived from the tool. All nutrient intake and nutrient density data were log-transformed prior to statistical analysis to better approximate normal distributions. All statistical analyses were performed in Stata version 14.2(38).

Measurement error models

A similar measurement error structure was assumed to that used by the OPEN study and EPIC-Norfolk(20,39), including linear associations between the longer-term true intake and both the biomarkers and self-reported intakes. We assumed person-specific systematic biases for both self-report tools, which were assumed to be correlated. We also assumed a systematic bias related to level of intake.

Our measurement error model follows that proposed by Kipnis *et al*.(19,39). For Oxford WebQ estimate Q_{ij}, interviewer-based MPR F_{ij} and biomarker M_{ij} on person i at occasion j:

$$Q_{ij} = \mu_{Qj} + \beta_{Q0} + \beta_{Q1}T_i + r_i + \varepsilon_{ij}$$

$$F_{ij} = \mu_{Fj} + \beta_{F0} + \beta_{F1}T_i + s_i + u_{ij}$$

$$M_{ij} = \mu_{Mj} + T_i + v_{ij}$$

where T_i is the true intake for individual i, μ_{Qj} and μ_{Fj} represent possible drift over time between measures; β_{Q0}, β_{Q1}, β_{F0} and β_{F1} are biases where β_{Q0} and β_{F0} are additive components associated with each tool, and β_{Q1} and β_{F1} are multiplicative components; r_i and s_i model the person-specific biases for each tool. We allow these person-specific biases to be correlated with $\rho(r,s)\neq 0$, because the same mechanisms may be influencing both r_i and s_i. We assume independent within-person errors ε_{ij} and u_{ij} that follow normal distributions with zero mean and variances σ_ε^2 and σ_u^2 respectively.

We assume that there is no person-specific bias associated with biomarker M_{ij} and within-person error v_{ij} follows a normal distribution with zero mean, variance σ_v^2, and is independent of the true intake and other error components. For analyses assessing estimated intake from Oxford WebQ based on the average of k serial measurements, variance σ_ε^2 is replaced by σ_ε^2/k.

We assume that correlation between biomarkers and dietary assessment measures does not vary by proximity in time because, for each cycle, the biomarker measure is completed before the dietary assessment day, and the gap between final dietary assessment of a previous cycle and the biomarker collection for the next is short. Subsequent exploration of the observed correlation structure was consistent with this assumption (data not shown).

We assume associations between urinary sucrose/fructose excretion and total sugars intake was similar to previously published feeding studies(28,29), allowing us to apply calibration equations derived from those studies:

$$M^*_{ij} = M_{ij} - 1.67 - 0.02S_i + 0.71A_i$$

where M^*_{ij} is the calibrated biomarker value, M_{ij} is the observed biomarker value, S_i is 0 for men and 1 for women, and A_i is a log-transformed age. M^*_{ij} was then used in place of M_{ij} in the measurement error model defined above. Participants in the feeding study from which this calibration equation was derived were healthy adults aged 23 to 66 years(28), similar to the OPEN study population of healthy adults (40 to 69 years)(29) and UK Biobank (40 to 69 years)(12).

Model fitting

The measurement error models were fitted as structural equation models using maximum likelihood estimation, assuming any missing data points were missing at random. Results were presented as attenuation factors indicating the extent to which estimated diet-disease associations are diluted using Oxford WebQ. Attenuation factors closer to 1 indicate less bias in diet-disease estimates. The correlation between Oxford WebQ and the latent variable in the structural equation model estimating true longer-term intake was also presented to indicate the amount of power lost in prospective studies using Oxford WebQ. This correlation also represents the attenuation of log relative risks between equal-sized categories of intake estimated by Oxford WebQ(5,6,20,40). Oxford WebQ is designed for repeated administration(17,18), and in UK Biobank participants were invited to complete it on up to 5 separate occasions over

a 16 month period, with the majority of responders completing it twice(41). We therefore present the predicted attenuation factors for the mean of several repeat administrations and derived from our estimated measurement model parameters. This takes the same approach used by Schatzkin et al, 2003(40). We focus on the mean of 2 administrations to reflect current use in UK Biobank.

The mean differences between Oxford WebQ and recovery biomarkers (for protein, sodium and potassium), the predictive biomarkers (sugars), and total energy intake (accelerometry) were presented. For each participant, this was based on the mean intake over the repeated cycles of the Oxford WebQ measures minus the mean over the repeated cycles of the biomarker and energy expenditure measures, back-transformed and expressed as a percentage. This is equivalent to the mean difference estimated by the Bland-Altman method of assessing agreement(42).

Subgroup analyses

Analyses were repeated stratified on sex, on age (<40 vs 40+ years) and on BMI (<25 vs 25+ kg/m^2), to quantify the robustness of results to different participant characteristics and explore possible impacts of differences in person-specific biases.

Ethics

The validation study was conducted according to the guidelines laid down in the Declaration of Helsinki. Full written informed consent was obtained from all participants

included. The procedures of the validation study and associated documentation were reviewed and approved by the West London NHS Research Ethics Committee (14/SC/1267).

RESULTS

In total, 225 participants were invited for screening for eligibility. Of these, 7 (3%) were ineligible, 30 (13%) did not consent and 27 (12%) subsequently withdrew consent. The remaining 161 (72%) participants completed Oxford WebQs, MPRs and provided samples for biomarkers on at least 1 occasion. After excluding missed voids, data from 160 participants were available for analysis, 152 (95%) of whom completed the Oxford WebQ after visit 1, 146 (91%) after visit 2, and 147 (92%) after visit 3, with 130 (81%) completing all 3 WebQs. Of these, 434 (98%) were completed on weekdays. The median time to complete Oxford WebQ was 10 minutes (inter-quartile range: 10 to 15 minutes).

Demographic characteristics of the participants at recruitment are shown in Table 1. Participants appeared metabolically stable over the course of the study, with weights only changing by more than 5% of weight at booking for 6 (4%) participants, whose energy expenditure readings were excluded from the analysis.

Estimated geometric mean intakes of protein, potassium and total sugars and their associated nutrient densities are shown in Table 2 for Oxford WebQ, MPR, biomarkers

and reference tools relating to the first clinic visit. Estimated intakes from Oxford WebQ were broadly similar to MPR for all nutrients. Compared to biomarker measures, Oxford WebQ over-estimated protein and potassium intakes and under-estimated total sugars, with estimated total energy intake less than the estimated TEE.

Attenuation factors and correlations between the self-report tools and estimated true longer-term intake for a single application of Oxford WebQ are shown in Table 3. For nutrient densities, attenuation factors were slightly higher and correlations were slightly lower than for unadjusted nutrient intakes. The full list of parameters estimated from the measurement models is shown in Web Table 1. Table 3 also shows the mean percentage difference between the self-report tools and the biomarker measures (Table 3). Mean percentages differences for Oxford WebQ were similar to those for the MPR.

Using the mean of a series of 2, 3, 4 or 5 repeat administrations of Oxford WebQ would substantially improve measurement properties (Table 4), with associated reduction in bias. With 2 repeats of the tool, the most likely use within UK Biobank as it currently stands, the attenuation factors and the correlation with true intake would improve markedly. With more repeats of the tool the attenuation and correlation with true intake would improve further.

When urinary biomarker concentrations were adjusted for completeness of urine samples and samples with PABA recovery less than 50% or more than 110% excluded, attenuation factors and correlations were essentially unchanged (see Web Appendix 2).

There was some variation between subgroups defined by age group, by sex and by BMI (Web Tables 2, 3 and 4). Attenuation factors for protein, potassium and sugar intake in men were higher than in women. Attenuation was worse in older people (≥40 years) than younger people (<40 years) for protein, but similar between age groups for total sugars and better in older people for potassium and total energy intake. Participants with BMI ≥25 kg/m^2 had broadly similar attenuation to those with BMI <25 kg/m^2, but with generally greater disparities for correlation with the truth.

DISCUSSION

Our findings show that the Oxford WebQ dietary assessment tool being used in UK Biobank and a sample of the Million Women study have good measurement error properties, improving further when taking the mean of several measures.

Oxford WebQ tends to over-estimate potassium intake and under-estimate total sugar intake, but these figures were similar for the interviewer-administered MPR. Additionally, Oxford WebQ is of broadly equivalent validity to the MPR in terms of attenuation of diet-disease associations. This held across 3 nutrients that could be measured by recovery biomarkers and other objective reference tools. For total sugars the Oxford WebQ

performed better than the interviewer-administered MPR. However, the Oxford WebQ is substantially quicker and cheaper to implement(17,18).

Oxford WebQ compares well to a recently validated online 24-hour recall in the UK(23). The validity of Oxford WebQ is also broadly similar to 24-hour recalls that have been validated in the US(5,6), though our finding that protein is over-reported and potassium under-reported in the UK contrasts with the under-reporting of protein and unbiased reporting of potassium in the US on average. This may reflect the shorter length of assessment in our study compared to most US studies or different cultural perceptions of foods with high concentrations of those nutrients.

FFQs generally estimate diet over a longer timescale than the 24-hour period covered by the Oxford WebQ. Similarly, in common with 24-hour recalls, Oxford WebQ cannot estimate diet in the past, while FFQs may be used for this purpose. However, repeated measures of the web-based Oxford WebQ tool throughout follow-up, covering different seasonal intakes and reflecting dietary changes as the cohort ages, can provide a estimate of long-term diet in a more prospective manner. The correlations we found between the mean of 2 to 5 Oxford WebQ estimates and the truth were also better than those previously reported for FFQs(5,6,29), though no better for nutrient densities. The improved measurement properties on repeat administration reflects how Oxford WebQ is currently being used in UK Biobank(18,41). It is possible that a mobile-optimised tool could be used in a more prospective manner, further improving performance.

We have focussed on the use of Oxford WebQ to estimate true longer-term diet, and the extent to which measurement error in this estimated exposure could lead to attenuated estimates of the association with disease outcomes. This application of the tool for estimating longer-term diet is the most relevant to large-scale cohort studies with long follow-up. Dietary exposures are often categorised to simplify presentation, but also to recognise that absolute intake is harder to estimate than ranking intakes from low to high. We therefore present the correlation between the Oxford WebQ and estimated true longer-term intake, which reflects the attenuation in diet-disease estimates based on ranked exposures. Oxford WebQ generally performed slightly better by this criterion. Oxford WebQ performed well compared with other tools assessed using the same statistical methodology(21,23).

In assessing validity of Oxford WebQ, we have used objective biomarkers free from person-specific biases shared by self-report tools. Our validation is therefore more robust than using another self-report tool that may agree well partly because it shares this same bias. However, 45% of urine samples contained less than 85% PABA recovery, which could have led to under-estimation of the agreement between self-reported diet and urinary biomarkers. Biomarkers were collected at clinic visits prior to completing the dietary recalls, but were not informed of results, minimising potential recall bias. Had biomarker collection coincided with dietary recall measures, there may have been greater agreement between them.

We did not use doubly labelled water to estimate TEE, which is a potential weakness in our study. In addition to estimated energy intake this could also affect nutrient density estimates and could partly explain why our results differed from previous studies which generally found using densities improved measurements. However, use of activity monitor equipment provided an equally objective measure, which we used instead. It is a potential weakness that activity monitors were only worn for 1 day in each cycle, but within-person variation was still estimable because of the repeated cycles.

Unfortunately, not all nutrients have adequate reference tools such as recovery biomarkers(43). This is another potential weakness of our study shared by other validation studies of dietary assessment tools. It is therefore possible that Oxford WebQ performs better or worse for other nutrients than those we were able to validate it against, particularly those deriving from episodically-consumed foods, for which 24-hour recalls are not well-suited. Where this is particularly important, combination with other dietary assessment tools is recommended(9,44-47).

Our measurement error models also only consider one error-prone variable at a time. In the presence of additional error-prone covariates, the error structure becomes more complex and the direction of bias may change. This commonly occurs with a nutrient and total energy intake in the same model. We therefore present estimates for nutrient densities as well.

Internet applications such as Oxford WebQ are potentially more accessible to some groups such as the younger or better educated. To address this concern, our validation study included both men and women with a spread of ages and a range of educational backgrounds. Additionally, we repeated our analyses by age, sex and BMI. Results were broadly comparable between men and women. Results suggested the online format was not a deterrent to the quality of reporting in older participants. Participants with higher BMI had similar attenuation factors, but correlation with the truth was worse for total sugars and total energy intake, suggesting greater person-specific bias in reporting certain food types in this group. This provides some support for taking BMI into account in measurement error models, as others have proposed(48,49). Whilst Oxford WebQ was specifically developed for UK Biobank and the Million Women study, the wide age ranged used in our validation, and exploration within demographic subgroups, provide a basis for its use in other large-scale prospective studies.

Our results indicate that repeat applications of Oxford WebQ in large-scale projects such as UK Biobank and the Million Women study should provide high quality dietary information, at least for total energy, protein, sugars and potassium. Oxford WebQ provides broadly similar results to using the more researcher-intensive and expensive to administer 24-hour recall delivered and coded by a trained researcher. This should facilitate additional dietary assessments repeated over time to measure long-term diet with greater precision, providing a platform for better estimates of the relationships between diet and disease.

ACKNOWLEDGEMENTS

This work was supported by the Medical Research Council (UK) grant MRC G1100235. Prof. Gary Frost acknowledges support from a National Institute for Health Research (NIHR) Senior Investigator award. Prof. Paul Elliott acknowledges support from the Medical Research Council and Public Health England (MR/L01341X/1) for the Medical Research Council - Public Health England (MRC-PHE) Centre for Environment and Health; and the National Institute for Health Research Health Protection Research Unit in Health Impact of Environmental Hazards (HPRU-2012-10141) and additional support from the NIHR Imperial College Biomedical Research Centre in collaboration with Imperial College NHS Healthcare Trust. Prof. Janet Cade acknowledges additional support from the Medical Research Council (MR/L02019X/1).

We thank Helen Brown for project administration, Prof. Laurence S Freedman for advice on statistical methods, Claire McLoughlin for project administration, Dr. Amy F. Subar for her role as external scientific advisor, advising on study design and statistical methods, Kay L. White for help with laboratory methods development and PABA analysis, Cybele P. Wong and Kamal Wahab for data entry.

Conflict of Interest: Prof. Janet E. Cade is a director of a University of Leeds spin out private company, Dietary Assessment Ltd., supporting the development of myfood24. Dr. Michelle Carter, Neil Hancock and Dr. Michelle A. Morris are shareholders in Dietary Assessment Ltd. The company has been running for 1 year and has not made a profit (or loss). Prof. Gary Frost has worked as a consultant for a

number of food companies including Nestlé, Unilever, Quorn, New Food Innovations and Malaysian Palm Oil.

REFERENCES

1. Thomas D, Stram DO, Dwyer JH. Exposure measurement error: influence on exposure-disease relationships and methods of correction. *Annu Rev Public Health*. 1993;1469-93.

2. Clayton D. Measurement Error - Effects and *Remedies in Nutritional Epidemiology. Proc Nutr Soc*. 1994;53(1):37-42.

3. Plummer M, Clayton D. Measurement error in dietary assessment - an investigation using covariance structure models - 2. *Stat Med*. 1993;12(10):937-948.

4. Prentice RL. Measurement error and results from analytic epidemiology: Dietary fat and breast cancer. *J Nat Cancer Inst*. 1996;88(23):1738-1747.

5. Freedman LS, Commins JM, Moler JE, *et al*. Pooled Results From 5 Validation Studies of Dietary Self-Report Instruments Using Recovery Biomarkers for Energy and Protein Intake. *Am J Epidemiol.* 2014;180(2):172-188.

6. Freedman LS, Commins JM, Moler JE, *et al*. Pooled Results From 5 Validation Studies of Dietary Self-Report Instruments Using Recovery Biomarkers for Potassium and Sodium Intake. *Am J Epidemiol.* 2015;181(7):473-487.

7. Freedman LS, Midthune D, Arab L, Prentice RL, Subar AF, Willett W, et al. Combining a Food Frequency Questionnaire With 24-Hour Recalls to Increase the Precision of Estimating Usual Dietary Intakes - Evidence From the Validation Studies Pooling Project. *Am J Epidemiol.* 2018;187:2227-32.

8. Thompson FE, Subar AF, Loria CM, et al. Need for Technological Innovation in Dietary Assessment. *J Am Diet Assoc*. 2010;110(1):48-51.

9. Schatzkin A, Subar AF, Moore S, et al. Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiol Biomarkers Prev*. 2009;18(4):1026-1032.

10. Illner AK, Freisling H, Boeing H, et al. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. *Int J Epidemiol*. 2012;41(4):1187-1203.

11. UK Biobank. Questions on Diet. https://www.ukbiobank.ac.uk/wp-content/uploads/2011/07/diet_questionnaire.pdf. Published April 6, 2009. Accessed February 21, 2019.

12. Collins R. What makes UK Biobank special? *Lancet.* 2012;379(9822):1173-1174.

13. Manolio TA, Weis BK, Cowie CC, *et al*. New Models for Large Prospective Studies: Is There a Better Way? *Am J Epidemiol.* 2012;175(9):859-866.

14. Sudlow C, Gallacher J, Allen N, *et al*. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* 2015;12(3):e1001779.

15. The Million Women Study Collaborative Group. The Million Women Study: design and characteristics of the study population. *Breast Cancer Res.* 1999;1(1):73-80.

16. Green J, Reeves GK, Floud S, et al. Cohort Profile: the Million Women Study. *Int J Epidemiol*. 2019;48(1):28-29e.

17. Liu B, Young H, Crowe FL, *et al*. Development and evaluation of the Oxford WebQ, a low-cost, web-based method for assessment of previous 24 h dietary intakes in large-scale prospective studies. *Public Health Nutr.* 2011;14(11):1998-2005.

18. Galante J, Adamska L, Young A, *et al*. The acceptability of repeat Internet-based hybrid diet assessment of previous 24-h dietary intake: administration of the Oxford WebQ in UK Biobank. *Public Health Nutr.* 2016;115(4):681-686.

19. Kipnis V, Midthune D, Freedman LS, *et al*. Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol.* 2001;153(4):394-403.

20. Kipnis V, Midthune D, Freedman L, *et al*. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr.* 2002;5(6A):915-923.

21. Kipnis V, Subar AF, Midthune D, *et al*. Structure of dietary measurement error: Results of the OPEN Biomarker Study. *Am J Epidemiol.* 2003;158(1):14-21.

22. Carter MC, Albar SA, Morris MA, *et al*. Development of a UK Online 24-h Dietary Assessment Tool: myfood24. *Nutrients.* 2015;7(6):4016-4032.

23. Wark PA, Hardie LJ, Frost GS, *et al*. Validity of an online 24-hour recall tool (myfood24) for dietary assessment in population studies: comparison with biomarkers and standard interviews. *BMC Med.* 2018;16136.

24. Bingham SA. Urine Nitrogen as a Biomarker for the Validation of Dietary Protein Intake. *J Nutr.* 2003;133(3):921S.

25. Bingham S, Cummings JH. The use of 4-aminobenzoic acid as a marker to validate the completeness of 24h urine collections in man. *Clin Sci.* 1983;64(6):629-635.

26. Bingham SA, Cummings JH. Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. *Am J Clin Nutr.* 1985;42(6):1276-1289.

27. Freedman LS, Midthune D, Carroll RJ, *et al*. Adjustments to Improve the Estimation of Usual Dietary Intake Distributions in the Population. *J Nutr.* 2004;134(7):1836-1843.

28. Tasevska N, Runswick SA, McTaggart A, *et al*. Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer Epidemiol Biomarkers Prev.* 2005;14(5):1287-1294.

29. Tasevska NA, Midthune D, Potischman N, *et al*. Use of the predictive sugars biomarker to evaluate self-reported total sugars intake in the Observing Protein and Energy Nutrition (OPEN) study. *Cancer Epidemiol Biomarkers Prev.* 2011;20(3):490-500.

30. Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol.* 1949;109(1-2):1-9.

31. Tataranni PA, Larson DE, Snitker S, *et al*. Thermic effect of food in humans: methods and results from use of a respiratory chamber. *Am J Clin Nutr.* 1995;61(5):1013-1019.

32. Johannsen DL, Calabro MA, Stewart J, *et al*. Accuracy of Armband Monitors for Measuring Daily Energy Expenditure in Healthy Adults. *Med Sci Sports Exerc.* 2010;42(11):2134-2140.

33. Royal Society of Chemistry and Ministry of Agriculture FaF. *McCance and Widdowson's the Composition of Foods.* Cambridge, UK: Royal Society of Chemistry, 2002.

34. Raper N, Perloff B, Ingwersen L, *et al*. An overview of USDA's Dietary Intake Data System. *Journal of Food Composition and Analysis.* 2004;17(3-4):545-555.

35. Gibson R, Eriksen R, Lamb K, *et al*. Dietary assessment of British police force employees: a description of diet record coding procedures and cross-sectional evaluation of dietary energy intake reporting (The Airwave Health Monitoring Study). *BMJ Open.* 2017;7(4):e012927.

36. Subar AF, Midthune D, Tasevska N, *et al*. 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.* 2013;67(8):863-867.

37. Johansson G, Bingham S, Vahter M. A method to compensate for incomplete 24-hour urine collections in nutritional epidemiology studies. *Public Health Nutr.* 1999;2(4):587-591.

38. StataCorp. *Stata statistical software: Release 14.2*. College Station, TX: Stata Corporation, 2015.

39. Kipnis V, Carroll RJ, Freedman LS, et al. Implications of a new dietary measurement error model for estimation of relative risk: Application to four calibration studies. *Am J Epidemiol*. 1999;150(6):642-651.

40. Schatzkin A, Kipnis V, Carroll RJ, *et al*. 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.* 2003;32(6):1054-1062.

41. UK Biobank. 24-hour Dietary Recall Questionnaire. Version 1.1. http://biobank.ctsu.ox.ac.uk/crystal/docs/DietWebQ.pdf. Published October 16, 2012. Accessed February 21, 2019.

42. Altman DG, Bland JM. Measurement in Medicine: the Analysis of Method Comparison Studies. *Statistician.* 1983;32307-32317.

43. Bingham SA. Biomarkers in nutritional epidemiology. *Public Health Nutr.* 2002;5(6A):821-827.

44. Kipnis V, Midthune D, Buckman DW, *et al*. Modeling Data with Excess Zeros and Measurement Error: Application to Evaluating Relationships between Episodically Consumed Foods and Health Outcomes. *Biometrics.* 2009;65(4):1003-1010.

45. Keogh RH, Park JY, White IR, *et al*. Estimating the alcohol-breast cancer association: a comparison of diet diaries, FFQs and combined measurements. *Eur J Epidemiol.* 2012;27:547-559.

46. Carroll RJ, Midthune D, Subar AF, *et al*. Taking Advantage of the Strengths of 2 Different Dietary Assessment Instruments to Improve Intake Estimates for Nutritional Epidemiology. *American Journal of Epidemiology.* 2012;175(4):340-347.

47. Subar AF, Freedman LS, Tooze JA, *et al*. Addressing Current Criticism Regarding the Value of Self-Report Dietary Data. *J Nutr.* 2015;145(12):2639-2645.

48. Prentice RL. Measurement error and results from analytic epidemiology: Dietary fat and breast cancer. *JNCI*. 1996;88(23):1738-1747.

49. Tooze JA, Subar AF, Thompson FE, et al. Psychosocial predictors of energy underreporting in a large doubly labeled water study. *Am J Clin Nutr*. 2004;79(5):795-804.

Table 1. Demographic and Lifestyle Characteristics of Participants by Sex, between 2014 and 2016 in London, UK.

Participant characteristic	Male (n=68) [a]		Female (n=92) [a]	
	No.	%	No.	%
Age (years) [b]	43 (16)		43(16)	
Ethnicity				
White	50	74	65	71
Black	1	1	7	8
Asian	4	6	5	5
Mixed and other	12	18	12	13
Age of leaving the educational system (years)				
16 or under	8	12	8	9
17 to 18	18	26	25	27
19+	42	62	57	62
Smoking status				
Non-smoker	52	77	72	78
Smoker	10	15	8	9
Weight (kg)	81 (13)		66 (12)	
Body mass intake (kg/m^2) [c]				
< 25	30	44	53	58
25 to 29	26	38	27	29
30+	12	18	12	13

[a] Note, where numbers in each category do not sum to the totals for the column, this is because of incomplete data for that characteristic.

[b]. Values are expressed as mean (standard deviation)

[c] Weight (kg) / height (m)^2

Table 2. Geometric Means and 95% Confidence Intervals for Protein, Potassium and Total Sugar Intake per day and Density as Assessed by Oxford WebQ, Interviewer-Based 24-Hour Recall, and Biomarkers Relating to the First Clinic Visit.

Nutrient	Oxford WebQ			Interviewer-based 24-hour recall			Biomarker / reference tool		
	n	Geometric mean	95% CI	n	Geometric mean	95% CI	n	Geometric mean	95% CI
Nutrient intake:									
Protein (g)	152	85.0	79.3, 91.1	154	82.0	77.0, 87.4	152	70.2	65.7, 75.1
Potassium (g)	152	3.3	3.1, 3.5	154	3.1	3.0, 3.3	152	2.1	2.0, 2.3
Total sugars (g)	152	100.8	92.9, 109.4	154	88.9	82.0, 96.3	151	133.5	116.3, 153.2
Energy expenditure:									
Total energy expenditure (MJ)	152	8.7	8.1, 9.2	154	8.5	8.1, 9.0	144	11.0	10.4, 11.5
Nutrient density [a]:									
Protein (g / MJ)	152	9.8	9.4, 10.2	154	9.6	9.2, 10.1	142	6.4	6.0, 6.9
Potassium (g / MJ)	152	0.38	0.36, 0.40	154	0.37	0.35, 0.39	142	0.19	0.18, 0.21
Total sugars (g / MJ)	152	11.6	10.9, 12.4	154	10.4	9.7, 11.2	141	12.1	10.4, 14.0

Abbreviations: CI, confidence interval.

[a] Nutrient density for protein, potassium and total sugars was expressed in grams per MJ of total energy intake

Table 3. Attenuation Factors, Correlation Between Dietary Assessment Tool and True Intake and Mean Difference Between Self-Report Tool and Reference Intake for Protein, Potassium and Total Sugar Intake and Density as Assessed by a Single Application of Oxford WebQ and the Interviewer-Based 24-Hour Recall.[a]

Nutrient	Attenuation factor	95% CI	Correlation with true intake	95% CI	Mean % difference vs reference tool	95% CI
Nutrient intake:						
Protein (g)						
Oxford WebQ	0.27	0.17, 0.36	0.40	0.27, 0.52	+12%	+6%, +19%
MPR	0.33	0.24, 0.43	0.46	0.36, 0.57	+8%	+3%, +14%
Potassium (g)						
Oxford WebQ	0.31	0.18, 0.44	0.34	0.20, 0.47	+53%	+42%, +64%
MPR	0.35	0.22, 0.48	0.37	0.25, 0.49	+47%	+37%, +57%
Total sugars (g)						
Oxford WebQ	0.31	0.18, 0.44	0.33	0.20, 0.46	-25%	-18%, -32%
MPR	0.16	0.01, 0.30	0.15	0.01, 0.30	-32%	-25%, -39%
Energy expenditure:						
Total energy expenditure (MJ)						
Oxford WebQ	0.22	0.12, 0.33	0.32	0.18, 0.46	-22%	-17%, -27%
MPR	0.30	0.17, 0.42	0.36	0.22, 0.49	-22%	-18%, -27%

Nutrient density [b]:						
Protein (g / MJ)						
Oxford WebQ	0.34	0.17, 0.51	0.29	0.16, 0.42	+46%	+37%, +55%
MPR	0.26	0.10, 0.42	0.23	0.09, 0.36	+41%	+32%, +50%
Potassium (g / MJ)						
Oxford WebQ	0.33	0.12, 0.54	0.23	0.09, 0.37	+99%	+83%, +115%
MPR	0.41	0.23, 0.59	0.33	0.19, 0.46	+91%	+77%, +106%
Total sugars (g / MJ)						
Oxford WebQ	0.32	0.15, 0.50	0.27	0.13, 0.41	-3%	+8%, -12%
MPR	0.16	-0.03, 0.35	0.13	-0.02, 0.28	-12%	-1%, -21%

Abbreviations: CI, confidence interval; MPR, interviewer-based Multiple-Pass 24-hour dietary Recall

[a] All dietary measures and estimates were log-transformed.

[b] Nutrient density for protein, potassium and total sugars was expressed in grams per MJ of total energy intake

Table 4. Attenuation Factors and Correlation Between Dietary Assessment Tool and True Intake for Protein, Potassium and Total Sugar Intake and Densities [a] Estimated from the Measurement Model for the Oxford WebQ Tool for Repeat Administrations.[b]

Number of repeat administrations for each nutrient	Attenuation factor	95% CI	Correlation with true intake	95% CI
Protein (g)				
1	0.27	0.17, 0.36	0.40	0.27, 0.52
2	0.37	0.24, 0.49	0.47	0.33, 0.61
3	0.42	0.28, 0.56	0.50	0.35, 0.65
4	0.45	0.30, 0.60	0.52	0.37, 0.67
5	0.48	0.32, 0.64	0.53	0.38, 0.69
Potassium (g)				
1	0.31	0.18, 0.44	0.34	0.20, 0.47
2	0.42	0.25, 0.60	0.39	0.24, 0.54
3	0.48	0.28, 0.68	0.42	0.26, 0.58
4	0.52	0.30, 0.73	0.44	0.27, 0.60
5	0.54	0.32, 0.77	0.45	0.28, 0.62
Total sugars (g)				
1	0.31	0.18, 0.44	0.33	0.20, 0.46
2	0.45	0.26, 0.64	0.40	0.24, 0.55
3	0.53	0.31, 0.75	0.43	0.27, 0.60
4	0.59	0.34, 0.83	0.45	0.28, 0.62
5	0.62	0.36, 0.88	0.47	0.29, 0.64
Total energy expenditure (MJ)				
1	0.22	0.12, 0.33	0.32	0.18, 0.46
2	0.31	0.16, 0.45	0.38	0.21, 0.54
3	0.35	0.19, 0.52	0.40	0.23, 0.58

4	0.38	0.20, 0.56	0.42	0.24, 0.60
5	0.40	0.21, 0.59	0.43	0.24, 0.62
Protein density (g / MJ) [c]				
1	0.34	0.17, 0.51	0.29	0.16, 0.42
2	0.51	0.27, 0.76	0.36	0.20, 0.51
3	0.62	0.32, 0.91	0.39	0.22, 0.56
4	0.69	0.36, 1.01	0.41	0.23, 0.59
5	0.73	0.38, 1.09	0.42	0.24, 0.61
Potassium density (g / MJ) [c]				
1	0.33	0.12, 0.54	0.23	0.09, 0.37
2	0.48	0.17, 0.78	0.28	0.11, 0.44
3	0.57	0.21, 0.93	0.30	0.12, 0.48
4	0.62	0.23, 1.02	0.31	0.12, 0.50
5	0.66	0.24, 1.09	0.32	0.13, 0.52
Total sugars density (g / MJ) [c]				
1	0.32	0.15, 0.50	0.27	0.13, 0.41
2	0.49	0.23, 0.75	0.33	0.16, 0.50
3	0.59	0.28, 0.91	0.36	0.18, 0.55
4	0.66	0.31, 1.01	0.38	0.19, 0.58
5	0.71	0.33, 1.09	0.40	0.20, 0.60

Abbreviations: CI, confidence interval; MPR, interviewer-based Multiple-Pass 24-hour dietary Recall

[a] All dietary measures and estimates were log-transformed.

[b] Estimates of measurement properties for mean of repeated administrations of the tool are based on the parameters provided in Web Table 1, using the approach described by Schatzkin et al, 2003 (27).

[c] Nutrient density for protein, potassium and total sugars was expressed in grams per MJ of total energy intake

Figure 1. Oxford WebQ Validation Study Design Overview.Each 24-hour dietary assessment (the Oxford WebQ online tool and interviewer-based multipass 24-hour recall in random order) and selected reference measures (recovery biomarkers, predictive biomarkers and total energy expenditure) were completed on 3 separate occasions separated by approximately 2 weeks. At each occasion, the reference measure was followed 1 to 3 days later by the first dietary assessment, which was followed approximately 2 to 4 days later by the second dietary assessment.