**QUEEN'S
UNIVERSITY
BELFAST**

Combining biomarker and food intake data: calibration equations for citrus intake

D'Angelo, S., Gormley, I., McNulty, B., Nugent, A., Walton, J., Flynn, A., & Brennan, L. (2019). Combining biomarker and food intake data: calibration equations for citrus intake. *American Journal of Clinical Nutrition*, 110(4), 977. <https://doi.org/10.1093/ajcn/nqz168>

Published in:

American Journal of Clinical Nutrition

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2019 Oxford University Press. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

1 **COMBINING BIOMARKER AND FOOD INTAKE DATA: CALIBRATION EQUATIONS FOR CITRUS INTAKE**

2

Silvia D'Angelo^{1,2}, Isobel Claire Gormley², Breige A McNulty¹, Anne P Nugent³, Janette Walton^{4, 5}, Albert Flynn⁴, and Lorraine Brennan¹.

¹ Institute of Food and Health, School of Agriculture and Food Science, University College Dublin, Dublin, Ireland

² School of Mathematics and Statistics, University College Dublin, Dublin, Ireland

³ School of Biological Sciences, QUB, Belfast

⁴ School of Food and Nutritional Sciences, University College Cork, Cork, Ireland

⁵ Cork Institute of Technology, Cork, Ireland

Address correspondence to L Brennan, Institute of Food and Health, Belfield, University College Dublin, Dublin 4, Ireland. E-mail: lorraine.brennan@ucd.ie

Supported by a research grant from The European Research Council ERC (647783).

Running heading: Calibration equations for citrus intake

3

4

5

6 Abstract

7 Background

8 Measurement error associated with self-reported dietary intake is a well-documented issue.
9 Combining biomarkers of food intake and dietary intake data is a high priority.

10 Objectives

11 The present work introduces a general methodology for the development of calibration equations for
12 food intake, illustrated with an application on citrus intake. Further, a simulation-based framework
13 was developed to determine the portion of biomarker data needed in large populations for stable
14 calibration equations estimation.

15 Methods

16 Calibration equations were developed using mean daily self-reported citrus intake (4 days semi-
17 weighted food diaries) and biomarker-derived intake (urinary proline betaine biomarker) data from
18 participants (n=565) as part of the NANS study. Different functional specifications and biomarker
19 transformations were tested to derive the optimal calibration equations specification. The simulation
20 study was developed using linear regression for the calibration equations. Stability in the calibration
21 equations estimation was investigated for varying portions of biomarker and intake data “qualities”.

22 Results

23 With citrus intake, linear regression on non-transformed biomarker data resulted in the optimal
24 calibration equations specification and produced good quality predicted intakes. Also, cross validation
25 was used to estimate distributions of calibrated intakes instead of point estimates, to better
26 incorporate the uncertainty inherent in statistical modelling of measurement error. The simulation
27 study suggested that in large population studies, biomarker data on 20-30% of the subjects are enough
28 to guarantee a stable calibration equations estimation. The paper is accompanied by a web application
29 (“Bio-Intake”), developed to perform measurement error correction in self-reported mean daily citrus
30 intake data.

31 Conclusion

32 Calibration equations proved to be a useful instrument to correct measurement error in self-reported
33 food intake data. The simulation study demonstrated that the use of food intake biomarkers may be
34 feasible and helpful also in the context of large population studies.

35

36

37 Keywords

38 Biomarkers, calibration equations, measurement error, citrus, proline betaine

39

40

41 INTRODUCTION

42

43 Problems surrounding our ability to measure accurately dietary intake has been well documented in
44 recent years in the literature (1-3). To overcome the measurement error associated with self-reported
45 dietary intake the concept of dietary biomarkers has emerged. These biomarkers can deliver objective
46 measures of intake and examples include urinary sodium and nitrogen for salt and protein intake (1).
47 Furthermore, there is evidence that use of biomarkers to develop calibrated equations to correct self-
48 reported intake for estimates for protein and energy intake can positively impact on one's ability to
49 examine diet-disease associations: for example inclusion of biomarker calibrated data in the Women's
50 Health Initiative cohorts allowed for disease associations to be revealed that otherwise would not
51 have been identified (4, 5). Using urinary nitrogen and doubly labelled water as biomarkers to
52 calibrate intakes of protein and energy in a group of postmenopausal women revealed associations
53 between protein and energy intake with diabetes risk. It is noteworthy that these associations were
54 not identified in the uncalibrated data.

55 The majority of the work to date in this field has focused on energy and protein intake using urinary
56 nitrogen and doubly labelled water as biomarkers. However, recent data has demonstrated that a
57 series of biomarkers including carotenoids, tocopherols, folate, vitamin B12, and phospholipid fatty
58 acids performed as well as established energy (doubly labelled water) and protein (urinary nitrogen)
59 biomarkers in representing nutrient intake (6). This highlights the potential of widening the nutrients
60 that could be corrected for measurement error using biomarkers. Furthermore, data emerging from
61 the application of metabolomics to biomarker discovery has revealed that food intake biomarkers may
62 be used to estimate food intake opening up the possibility of development of calibration equations
63 for intake of foods (7-9).

64 The objective of the present work was to develop calibration equations for citrus intake using urinary
65 proline betaine as the biomarker. Furthermore, we demonstrate through a stimulation study the
66 required portion of biomarker measures necessary to enable development of calibration equations in
67 a large population group thus laying the foundations for future work. Finally, a web application ("Bio-
68 Intake") was developed to perform correction of measurement error in self-reported mean daily citrus
69 intake data, based on the findings of this paper (<https://adiet.shinyapps.io/Bio-Intake>.)

70 METHODS

71 STATISTICAL ANALYSIS

72

73 Development of calibration equations for estimation of citrus intake

74 The calibration equations method relies on a central assumption, that both self-reported and
75 biomarker-derived estimates of intake of study participant i are linearly related to the so-called "true"
76 intake X_i , with the relations being perturbed by Gaussian errors. However, as true intakes are
77 unknown, directly estimating such relationships is not possible. The calibration method overcomes
78 this issue by deriving a conditional prediction of X_i , $\hat{X}_i$, given the biomarker-derived estimates of
79 intake M_i (10). Such conditional predictions are derived from the calibrated intakes W_i , i.e. the self-
80 reported intakes adjusted for measurement error:

$$M_i = \beta_0 + \beta_1 W_i + \epsilon_i ; \quad \widehat{X}_i = \widehat{\beta}_0 + \widehat{\beta}_1 W_i$$

The data analyzed herein refer to a study conducted on $N = 1500$ subjects, indexed by $i = 1, \dots, N$ and collectively denoted A_N , where food consumption data was recorded using 4 day semi weighed food diaries. Mean daily citrus intake W_i (g/day) was computed for each subject. Also, biomarker-derived estimates of citrus intake M_i were computed for a subsample of A_N , here denoted A_n , with $n = 565$. Self-reported data and biomarker estimated intake data for these n subjects were used to develop calibration equations. The estimated calibration equations were then employed to correct self-reported mean daily citrus intake W_i for the remaining $(N - n) = n^* = 935$ subjects, collectively denoted A_{n^*} .

The supplementary material provides a more detailed description of the calibration method. Common problems that may be encountered with food intake data include zero inflation (11), due the presence of non-consumers, and non-Gaussianity of the errors ϵ_i . Potentially, such issues could result in poor correction of measurement error. In the literature, possible non-Gaussianity of the errors has been addressed using different transformations of the biomarker-derived estimates of intake, M_i , such as the log transformation (12) or Box-Cox transformations (13). In previous work (13), zero-inflated data have been commonly modeled with the assistance of covariates. However, when such subject-specific information is not available, such an approach cannot be pursued.

In the present work, transformations of the biomarker-derived intake data have been considered for the estimation of calibration equations, as well as original scale biomarker-derived estimates of intake, to address potential non-Gaussianity of subject-specific errors. Additionally, to address the issue of zero-inflated data in the absence of covariates, zero-inflated models have been explored and compared with a standard linear regression model, to select the optimal model specification for the calibration equations. The average mean squared error (MSE) between biomarker-derived estimates of intake and calibrated mean-daily self-reported intake has been used to compare different model specifications and different transformations of the biomarker-derived intake data. The supplement presents more detail on the different models explored and gives a thorough discussion of the results obtained.

108 Estimation of the required portion of biomarker data in large studies

Acquisition of biomarker derived estimates of intake for all subjects may not be feasible in large population studies, due to high costs and limited biological samples availability. In order to estimate the required quantity of biomarker data needed in large studies a simulation study has been conducted. A set of $P = 100$ datasets of $N = 100,000$ subjects was created. For each dataset p ($p = 1, \dots, P$) biomarker-derived estimates of intake and self-reported intake data have been simulated, according to the results obtained from the analysis of mean-daily citrus intake data. Specifically, true mean-daily citrus intake values (g/day) have been simulated from a Gaussian distribution, with mean and variance fixed as the empirical mean and variance of the calibrated mean-daily self-reported citrus intakes. Self-reported and biomarker-derived intakes (g/day) have been then constructed from the simulated true intakes, following a classical calibration equations framework (14). To represent different situations that could be encountered with real data, simulated self-reported and biomarker intakes may either have moderate or high variability around the “true intake”. Such variation is regulated by the parameters α_W and α_M for self-reported and biomarker-derived intakes respectively. Also, different strengths of the association between self-reported and true intakes have been accounted for, to represent different self-reported data qualities. This feature is regulated by the parameter β_W ; further detail on parameters α_W, α_M and β_W are given in the supplement. Biomarker derived intakes are assumed to have been measured only for n_g of the N subjects, with $n_g = gN$ and $g = (0.01, 0.02, \dots, 0.99, 1)$. For each value of n_g , simulated intakes were used to estimate calibration equations, with the corresponding parameters being estimated via cross-validation (100 repetitions,

excluding 20% of the observations at each round). Standard deviations of the estimated parameters, $sd(\hat{\beta}_0)$ and $sd(\hat{\beta}_1)$, were computed and compared for different values of n_g , i.e. for different sizes of the sub-sample of subjects for which biomarker data are available. A stabilization of the parameter estimates indicates that, after a given n_g , further increasing the number of subjects for which biomarker-derived intakes are computed would not yield a substantial improvement in measurement error correction. More details regarding the construction of the simulation study are given in the supplement, while the results are presented in the next section.

RESULTS

Estimating citrus intake using calibration equations

Summary statistics for self-reported and biomarker-derived mean daily citrus intakes (g/day) are reported in Table 1. All examined measures of citrus intake have similar interquartile ranges and similar means, $\overline{W_{A_n}} = 75.42$ g/day and $\overline{M_{A_n}} = 83.26$ g/day. Biomarker-derived estimates of intake range in a wider interval ($\text{Max}(M_{A_n}) = 1226.79$) compared to self-reported intakes ($\text{Max}(W_{A_n}) = 845.75$). Non-consumers represent around 40% of subjects in self-reported intake data, while zero value biomarker-derived intakes are only 17% of the total. Concordance between self-reported and biomarker-derived intakes was confirmed by the statistics summarizing the conditional distributions $M_{A_n|W_{A_n}=0}$ and $W_{A_n|M_{A_n}=0}$ in Table 1. These distributions represent, respectively, biomarker-derived intake values corresponding to null self-reported intake values and vice-versa. Such measures of intake display good agreement, as both take low values when the other is null.

Summaries of the predicted intakes obtained using different model specifications and biomarker transformations are compared in Table 2. Models using a log-transform of the biomarker-derived estimates of intake in the estimation of the calibration equations, or of a Box-Cox transformation with small λ parameter, result in high-variability in the calibrated intakes. Indeed, regardless of the functional specification of the calibration equations, in these cases the standard deviation of the calibrated intakes is always greater than 792, much larger than the standard deviations of the observed self-reported intakes ($sd(W_{A_n^*}) = 107$). Also, when using these biomarker transformations, poor agreement is found between the interquartile ranges of the observed self-reported intakes and the calibrated intakes. The calibrated intakes from the remaining model specifications and biomarker transformations range in similar intervals, closer to that of the self-reported mean daily citrus intakes. Among these, the calibrated intakes computed using a linear regression model for the calibration equations with original-scale biomarker-derived estimates of intake (M_i) are those whose mean is closer to the mean of $W_{A_n^*}$. Indeed, a linear regression model was optimal, in terms of MSE, to estimate calibration equations of citrus intake. Table 3 details the average MSE values associated with different model specifications and transformations of biomarker-derived estimates of intake. Mean squared errors have been computed between biomarker-derived estimates of intakes and calibrated mean-daily self-reported intakes, computed with different specifications of the calibration equations. The lowest MSE (14354) corresponds to a linear regression model, defined with biomarker-derived estimates of intakes on the original scale, M_i . The preference for a simple model is confirmed by the second lowest MSE, associated with a linear regression model, with unitary parameter Box-Cox transformation of M_i , coinciding with a mere translation of the data. Given the selected optimal linear model, mean-daily calibrated citrus intakes were computed using the estimated calibration equations:

$$\hat{X}_i = 33.60 + 0.63W_i$$

It is worth noting that the estimated intercept is close to the average biomarker-derived intake for non-consumers (see Table 1). Figure 1 offers a comparison between calibrated and self-reported

mean-daily citrus intakes for the n^* subjects in the validation study. Overall, there is good agreement between self-reported and calibrated mean-daily intakes, however, differences emerge at high intakes. Self-reported non-consumers were assigned to a low baseline citrus consumption level, corresponding to $\hat{\beta}_0 = 33.60$.

What portion of biomarker data is required in large population studies?

Sixteen distinct scenarios of simulated intake data were generated by considering different combinations of variability of self-reported data and biomarker data around the true intake, quantified through $\alpha_W, \alpha_M = (0.5, 2)$, and self-reported data quality levels, quantified through $\beta_W = (0.1, 0.5, .08, 1)$. Among these, the group of “best-case” scenarios corresponds to those with a $\alpha_W = \alpha_M = 0.5$, as both biomarker-derived intakes and self-reported intakes have very little variability around true intakes. Figure 3 reports the standard deviations of the estimated calibration equations slope parameter $\hat{\beta}_1$ under such scenarios, for the different self-reported data quality levels. Standard deviations stabilize quickly, for relatively small values of n_g (the number of subjects for which biomarker-derived estimates of intake are available), regardless of the β_W value. However, β_W does influence the level at which the standard deviation stabilizes, with poor-quality self-reported data ($\beta_W = 0.1$) leading to slightly higher variability in the estimate of β_1 . Similar results are found also for the set of “worst-case” scenarios (Figure 3), represented by the $\alpha_W = \alpha_M = 2$ configuration, that is, when large measurement error is present in both measures of intake. Standard deviations tend to stabilize quite rapidly around $n_g \approx 25000$, as happened in the “best-case” scenarios. However, this time the quality of self-reported data does not have a large impact. Indeed, there is little visible differences between the cases $\beta_W = 0.1$ and $\beta_W = 1$. The remainder of the sixteen scenarios represent less extreme, hybrid situations in terms of data quality and measurement error, with corresponding results in-line with those described for the two extreme scenarios. Similar conclusions were drawn for the estimated calibration equations intercept parameter β_0 ; standard deviations for the estimates of the intercept parameter stabilized jointly with those of β_1 , around the same n_g values. More detail regarding the simulation study, including results and MSEs associated with the different simulation scenarios, are provided in the supplement. In general, the standard deviations of the calibration equation parameters tend to stabilize when biomarker data is obtained from 25% of the study population, in this case for 25000 out of the 100000 subjects.

DISCUSSION

The present study presents an important advancement in the use of intake biomarkers in Nutrition research. Using a well-established biomarker of citrus intake calibration equations were developed to enable correction of self-reported intake for measurement error. Development and demonstration of such an approach for food intake clearly shows how one could use food intake biomarkers in nutrition research. To date the literature is lacking real examples of the utility of such food intake biomarkers with the majority of studies stopping at an association between food intake and biomarker levels. Within this context, the use of urinary proline betaine to correct self-reported citrus intake is a clear example of the potential of the field of food intake biomarkers. Furthermore, we developed a framework for assessing the amount of biomarker data needed in a large population study to correct for measurement error in food intake. Interestingly, in the present simulation study biomarker data was necessary for around 20-30% of the population highlighting an important aspect in the translation of this work into large epidemiological studies.

To date, calibration equations for correction of measurement error have focused on nutrients and to the best of our knowledge there is no example in relation to specific food intake. As dietary guidelines move towards food-based guidelines it is imperative that we also have good assessment of food intake. Use of food intake biomarkers has gained increased traction in recent years with the promise

of improvement in dietary assessment. The current development of calibration equations for citrus intake clearly paves the path forward in this regard. The ability to correct for measurement error for food intake has the potential to enhance our understating of the relation between food and disease. Previous literature demonstrated that correction of measurement error in nutrient intake using biomarkers led to the identification of relationships that were not observed without the calibration. Using biomarkers to correct for measurement error, energy intake was associated with increased risk of breast cancer, all cancer and type 2 diabetes and protein intake was associated with increased risk of type 2 diabetes (4, 15-17). Building upon the present work, the next step will be to examine relationships between calibrated food intake and certain health parameters.

An important aspect of our approach was the development of calibrated intake distributions which incorporate the uncertainty inherent in statistical modelling of measurement error. Such an approach results in calibrated intake distributions instead of single point-estimates. For researchers the potential distribution of intake delivers valuable information considering the uncertainty surrounding self-reported data. To aid the end user to implement the citrus intake calibration equations herein we developed a freely available web application, "Bio-Intake". Bio-intake is an open-source, easy to use web application designed to allow reproducibility of the methodology proposed in this paper.

Typically, for biomarker-based studies the biomarker data is acquired on a sub-population to develop the calibration equations. In this context we developed a framework for the assessment of the portion of food intake biomarker data necessary to develop accurate calibration equations. The framework advocates for examination of the quality of the self-reported data and the biomarker derived data. The framework was illustrated through a stimulation study where the portion of biomarker measures (n_g) necessary to enable accurate development of calibration equations in a large population group (N individuals) was estimated. For this purpose, standard deviations of the calibration equations parameter estimates are analyzed. A stabilization of such standard deviations occurring for low values of n_g , $n_g \ll N$, would indicate that the collection of biomarker data for a small portion of the population of interest and that for the whole population are practically equivalent in terms of stability of calibration equation estimation.

Indeed, simulation results showed rapidly decreasing standard deviations, independently of the scenario considered, that is of the "observed data-quality". Such promising results suggest that substantial improvements in the correction of self-reported intake data may be achieved even with little biomarker data. Indeed, collecting biomarker-derived estimates of intake for a larger portion of the population of interest may not lead to more precise calibration equations estimates. In the current example of citrus intake, measurement of the biomarker in 20-30% of the population was deemed sufficient. Obtaining biomarker data from just a small portion of the population suggests that calibration equations can prove to be powerful tools in large cohort studies, where cost and limited biological samples are barriers to their development. In the literature, previous studies have used biomarker data in 1-3 % of the total larger cohort to assess relationships between intake and disease risk (add 4 and 20).

The findings of the current simulation study were partly in line with previous large population studies for nutrients, in that it is not necessary to collect biomarker data for the whole population of interest to develop calibration equations. Nonetheless, simulation results showed a substantial decrease in the standard deviations of the calibration equations parameter estimates for portion of biomarker data in between $n_g = 0.01N$ and $n_g = 0.25N$, suggesting that a value of 1% may be suboptimal. However, the simulation study was developed to analyze calibration equations estimates for food biomarkers and its results may not have direct applicability in the context of nutrient biomarkers. Lastly, mean squared error values from the simulation study demonstrated that the quality of analyzed self-reported data is quite relevant, even with high quality biomarker data. Indeed, mean squared error values associated with "low data quality" scenarios ($\beta_W = 0.1$) were always much larger than those related to "higher data quality" ones ($\beta_W = (0.5, 0.8, 1)$). More importantly, mean squared

error distributions in “low data quality” scenarios did not vary particularly with different variability levels in biomarker data ($\alpha_W = (0.5, 2)$). In conclusion, the simulation study findings suggest that calibration equations may be stably estimated even with a relatively small portion of biomarker data, but that the quality of calibrated intake predictions is strongly dependent on that of self-reported intake data. Large biomarker data samples and excellent quality biomarkers will not be enough to get reliable estimates of intake (calibrated intake) if the observed self-reported data are not reliable.

This study has a number of noteworthy strengths. The development of calibration equations was based on reliable biomarker data and good quality food intake data (estimate from 4-day semi weighted food diaries). Furthermore, the calibration equations framework developed (and the corresponding “Bio-Intake” web app) allows estimation of calibrated intake distributions, instead of single point estimates. Hence, the proposed framework allows to better account for uncertainty inherent to calibrated intakes estimation: self-reported intake values are associated to a range of possible calibrated values, providing richer information on the quantity of interest.

A number of limitations need to be acknowledged. The proposed framework only accounts for self-reported intake data in the estimation of calibration equations. When available, the introduction of subject-specific covariates in the modelling framework would allow a more flexible estimation of the calibrated intakes. However, as a drawback, controlling for different factors with covariates would lead to the estimation of less generalizable calibration equations, which would have narrower applicability. Finally, the simulation study provided results concerning the stability of calibration equations parameter estimates, in terms of portion of biomarker data needed for the estimation. This portion does not guarantee convergence of such estimates, that is, reliable estimated calibrated intakes. As shown in Figure 2, reliability of the calibrated intakes depends on the quality of both the biomarker-derived and self-reported intake data, which does not depend on the modelling framework adopted for the development of calibration equations. Good data quality should always be assessed separately and before carrying out calibration equations estimation

In conclusion, the present study demonstrates the potential of food intake biomarkers in nutritional epidemiology. Future work will include development of calibration equations for other foods which will in turn improve the accuracy of dietary assessment. A further important aspect to be explored is whether the calibrated food intakes could be used in disease-risk studies in larger cohorts in a similar fashion to the work performed with calibrated nutrient intake.

298

299 SUPPLEMENTARY MATERIAL

300 THE CALIBRATION METHOD

301 How it works

302 The calibration method is a measurement error correction technique widely used in nutrition
303 literature (5, 15, 18). The main assumption behind it states that self-reported intake data W_i are
304 linearly related to true intakes X_i :

$$305 \quad W_i = \mu_w + \beta_w X_i + \epsilon_{w,i} \quad (1)$$

306 for subjects $i = 1, \dots, N$. The relation between the observed data W_i and the latent quantity of interest
307 X_i is mainly influenced by two factors: the slope coefficient β_w and the subject-specific errors $\epsilon_{w,i}$.
308 The β_w parameter may be seen as a measure of the “quality” or the intensity of the relationship
309 between self-reported and true intake data. Also, it expresses the direction of the relation between
310 the two variables, with negative values of β_w indicating disagreement. Subject-specific errors $\epsilon_{w,i}$
311 represent individual variability in self-reported intakes. Such errors are assumed to be normally
312 distributed with zero mean and homoscedastic, i.e. with equal variances σ_w^2 . Thus, the calibration
313 method assumes there is no interaction between the subjects when reporting the intakes, a
314 reasonable assumption in many nutrition studies, as reports are collected separately. Also, by
315 assuming equal error variances σ_w^2 , the calibration method implies that subjects are equally likely to
316 misreport their consumption, hence all subjects are equally reliable.

317 The linear relation between self-reported and true intake data cannot be estimated, as the latter is
318 not observed. To overcome such issue, the calibration method relies on a second assumption: another
319 linear relation exists between biomarker-derived (M_i) and true intake data:

$$320 \quad M_i = \mu_M + \beta_M X_i + \epsilon_{M,i} \quad (2)$$

321 Further, it is assumed that biomarker-derived estimates of intake provide a more reliable
322 representation of true intakes, setting $\beta_M = 1$ and $\mu_M = 0$, thus following the classical measurement
323 error model REF here Carroll book 2006. Subject-specific errors $\epsilon_{M,i}$ are assumed to be independently
324 distributed, according to a Gaussian distribution with zero mean and equal variance σ_M^2 .

325 As the true intakes X_i cannot be observed, the calibration method combines equations (1) and (2) to
326 derive estimates of such quantities. Specifically, given the assumptions outlined hold, X_i may be
327 conditionally predicted given biomarker-derived estimates of intakes and self-reported intakes in two
328 steps:

$$329 \quad M_i = \beta_0 + \beta_1 W_i + \epsilon_i, \quad i = 1, \dots, n$$

$$330 \quad \hat{X}_i = \hat{\beta}_0 + \hat{\beta}_1 W_i, \quad i = 1, \dots, n$$

331 The errors ϵ_i are within-subject random effects, distributed independently according to a standard
332 Gaussian distribution. The above equations are the so-called “calibration equations” and the predicted
333 true intakes $\hat{X}_i$ are the calibrated self-reported intakes. Indeed, the estimates of the fixed effect
334 population level parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ may be used to correct self-reported data and get more reliable
335 estimates of true intake; note that such correction incorporates the information coming from
336 biomarker-derived data, but it does not directly depend on it. (10) provides detail on the construction,
337 implementation and theoretical justification of the calibration method.

338 Alternative model formulations

339 Intake data is often characterized by a large presence of zero values. Depending on whether the
 340 measure of intake is self-reported or biomarker-derived, zero values either indicate non-consumers
 341 (subjects that reported null consumption of a given food) or null biomarker-derived estimates. There
 342 may be a partial mismatch between subjects self-reporting zero intake values and those having zero
 343 biomarker-derived estimates of intake values. When subject-specific covariates have been collected,
 344 several approaches, eg. (13), have been proposed to incorporate them in the modelling framework
 345 and to try modelling “true” and “false” (those misreporting) non-consumers. In the absence of
 346 covariate data such zero self-reported intakes are more challenging to address, but modelling
 347 frameworks designed for zero-inflated data can be exploited eg. (19). In particular, the calibration
 348 equations may be re-defined according to either a Tobit model (20) or a two-part model (21); such
 349 models are designed specifically for a regression framework where the response variable, M_i , is semi-
 350 continuous. Then, the calibration equations may be rewritten as:

$$351 \quad M_i = f_l(W_i) + \epsilon_i, \quad i = 1, \dots, n$$

$$352 \quad \hat{X}_i = \hat{f}_l(W_i)$$

353 The function $f_l(\cdot)$ is a link function, corresponding to the linear regression, the Tobit regression or the
 354 two-part regression, respectively for $l = 1, 2, 3$.

355 Biomarker transformation

356 The assumption of Gaussian errors ϵ_i may not be completely satisfied in empirical studies, and some
 357 skewness may be found in observed intake data. As can be seen from the summary statistics in Table
 358 1, this is the case for the observed mean daily self-reported intakes and biomarker-derived estimates
 359 of citrus intake. In particular, when deriving calibration equations for citrus intake, the skewness in
 360 the distribution of biomarker-derived estimates of intake may imply non-Normality of the errors ϵ_i .
 361 To try overcome this potential issue, in the context of episodically consumed foods (13) propose to
 362 transform the response variable using the Box-Cox family of transformations (22),

$$363 \quad M_i^* = BC(M_i, \lambda) = \frac{M_i^\lambda - 1}{\lambda}, \quad i = 1, \dots, n$$

364 so that the distribution of the transformed marker data M_i^* would more closely resemble that of a
 365 Gaussian. Using a Box-Cox transformation of biomarker-derived intakes, the calibration equations may
 366 be defined as:

$$367 \quad M_i^* = f_l(W_i) + \epsilon_i, \quad i = 1, \dots, n$$

$$368 \quad \hat{X}_i^* = \hat{f}_l(W_i)$$

369 where $\hat{X}_i^*$ is the transformed predicted true intake and $f_l(\cdot)$ is any given model specification for the
 370 calibration equations. Predicted intakes on the original scale may be found by computing the inverse
 371 of the Box-Cox transformation, $\hat{X}_i = BC^{-1}(\hat{X}_i^*, \lambda)$, using the Delta method (23):

$$372 \quad \hat{X}_i = (\lambda \hat{X}_i^* + 1)^{\frac{1}{\lambda}} \left[1 + \frac{(1 - \lambda)}{2(\lambda \hat{X}_i^* + 1)^2} \right]$$

373 Note that when biomarker-derived intake data are transformed using a Box-Cox transformation, the
 374 two-part and Tobit model can no longer be estimated, as structural zeros “disappear” in the
 375 transformed variable.

376 A different transformation that attempts to normalize the data while leaving the structural zeros
 377 unchanged is the log-transformation $M_i^* = \log(M_i + 1)$. This type of log-transformation is widely used
 378 in dose-response study data, where there is often the need to rescale the data without losing
 379 information on the structural zero values. Indeed, for $M_i = 0$, $M_i^* = \log(1) = 0$. As for Box-Cox
 380 transformed intake data, when using log-transformed biomarker-derived estimates of intake to
 381 estimate calibration equations, the computed calibrated intakes are defined on a new, transformed,
 382 scale. Calibrated intake on the original scale may be obtained using the back transformation $\hat{X}_i =$
 383 $\frac{3}{2} \exp(\hat{X}_i^*) - 1$.

384 Estimating calibration equations: cross validation and mean squared error

385 In order to select the optimal model specification, $\hat{f}_l(\cdot)$, $l = 1, 2, 3$, and data transformation (original
 386 scale, Box-Cox or log-transformation), different calibration equation models have been estimated,
 387 resulting from all possible combinations of model specifications and data transformations. For the
 388 Box-Cox transformation, parameter values $\lambda = 0.05, 0.1, 0.15, \dots, 1.2$ have been considered, as these
 389 yield the most Gaussian-like distributions of transformed biomarker derived estimates of intake. As a
 390 result, 30 different models have been estimated. To obtain stable estimates of the calibration
 391 equations, each model $\mathcal{M}_k$, $k = 1, \dots, K$ ($K = 30$), has been estimated using “leave-p-out cross-
 392 validation”, with “p” corresponding to 20% of the data. Such operation has been repeated $T = 5000$
 393 times, $t = 1, \dots, T$, each time randomly splitting the observations in A_n into a training set (containing
 394 80% of the data) and a test set. For each $\mathcal{M}_k$ model, calibration equations have been estimated using
 395 training data only; the corresponding estimated coefficients are then used to compute calibrated
 396 intakes from the self-reported intakes in the test set:

$$397 \quad M_{i,k,t}^{train} = f_l(W_{i,k,t}^{train}) + \epsilon_{i,k,t}^{train} \quad (3)$$

$$398 \quad \hat{X}_{i,k,t}^{test} = \hat{f}(W_{i,k,t}^{test}) \quad (4)$$

399 When using transformed biomarker-derived estimates of intake, in equations (3) and (4), $M_{i,k,t}^{train}$ may
 400 be substituted with the transformed-intake $M_{i,k,t}^{train,*}$, and $\hat{X}_{i,k,t}^{test}$ would then be the transformed-
 401 calibrated intake $\hat{X}_{i,k,t}^{test,*}$.

402 To compare calibrated intakes obtained with different calibration equations models $\mathcal{M}_k$, $k = 1, \dots, K$,
 403 mean squared errors are computed, between the known biomarker-derived intake values $M_{i,k,t}^{test}$ and
 404 the calibrated intake values $\hat{X}_{i,k,t}^{test}$.

$$405 \quad MSE_{k,t} = \frac{\sum_{i=1}^{0.20n} (M_{i,k,t}^{test} - \hat{X}_{i,k,t}^{test})^2}{0.20n}$$

406 Different models may then be compared using average mean square errors $\overline{MSE}_k = \sum_{t=1}^T MSE_{k,t} / T$,
 407 with lower values corresponding to better models. Thus, the optimal model $\mathcal{M}_{k^*}$ may be found as
 408 $k^* = \arg \min_k \overline{MSE}_k$.

409 SIMULATION STUDY

410 The calibration equations framework is built upon the assumption that both self-reported intakes and
 411 biomarker-derived estimates of intake are linearly related to the unknown true intakes. Normally,
 412 biomarker-derived intakes are collected only for a sub-sample of the population of interest, due to
 413 elevated extraction costs. Indeed, in the mean daily citrus intake data the population of interest
 414 counted $N = 1500$ subjects, but biomarker-derived intakes had been collected only for $n = 565$ of

415 them, approximately one third of the total. With large population studies, it may be infeasible to
 416 collect biomarker-derived intakes for a large subset of the subjects. Biomarker-derived data would be
 417 then collected only for a small number of n subjects, $n \ll N$. However, deriving calibration equations
 418 using $n \ll N$ observations may result in unstable and potentially unreliable intake predictions. Hence,
 419 how large, or small, n is required to be in order to obtain stable predictions is of key interest. To
 420 investigate this issue, a simulation study has been constructed. Assuming the population of interest
 421 has $N = 100,000$ subjects, $P = 100$ different sets of true, self-reported and biomarker-derived mean
 422 daily citrus intakes have been simulated according to the calibration equations framework:

$$423 \quad W_{i,p} = \mu_W + \beta_W X_{i,p} + \epsilon_{i,p,W}$$

$$424 \quad M_{i,p} = X_{i,p} + \epsilon_{i,p,M}$$

425 True intakes $X_{i,p}$ have been simulated from a Gaussian distribution $X_{i,p} \sim \mathcal{N}(\mu_X, \sigma_X^2)$, where $\mu_X = 81$
 426 and $\sigma_X^2 = (66)^2$ have been fixed to reflect the estimated mean and variance of the calibrated mean
 427 daily citrus intakes. The random errors, $\epsilon_{i,p,W}$ and $\epsilon_{i,p,M}$ are assumed to be normally distributed with
 428 zero mean and variances $\sigma_W^2 = \alpha_W \sigma_X^2$ and $\sigma_M^2 = \alpha_M \sigma_X^2$. The α_W and α_M scale parameters regulate
 429 the variation of the self-reported and biomarker-derived intake data around the true intake. These
 430 scale parameters have been set to $\alpha_W, \alpha_M = (0.5, 2)$, to represent both clean ($\alpha_W, \alpha_M = 0.5$) and
 431 noisy ($\alpha_W, \alpha_M = 2$) intake data. The β_W parameter measures the strength of the relation between
 432 self-reported and true intake data. To represent different levels of quality for self-reported data, this
 433 parameter has been set to $\beta_W = (0.1, 0.5, 0.8, 1)$, with $\beta_W = 0.1$ and $\beta_W = 1$ denoting poor and good
 434 data quality, respectively. Finally, the offset parameter μ_W measures the average difference between
 435 self-reported and true intake data. Here this parameter has been set to $\mu_W = 0$, as similar values were
 436 observed for the average self-reported and calibrated mean-daily intakes.

437 Different combinations of β_W , α_W and α_M parameter values give rise to sixteen different simulation
 438 scenarios. For each scenario, it has been further assumed that biomarker-derived estimates of intake
 439 have been collected for a subsample of n_g subjects, with $n_g = gN$ and $g = (0.01, 0.02, \dots, 0.99, 1)$. For
 440 each simulation scenario and for each value of n_g the calibration equations are estimated, using the
 441 optimal model from the citrus intake data analysis i.e. the linear model with the original-scale
 442 biomarker-derived intakes. As in the citrus intake analysis, calibration equations have been estimated
 443 using the “leave-p-out cross-validation” method. The standard deviations of the cross-validation
 444 estimates of the calibration equation parameters, $sd(\hat{\beta}_0)_{\alpha_W, \alpha_M, \beta_W, n_g, p}$ and $sd(\hat{\beta}_1)_{\alpha_W, \alpha_M, \beta_W, n_g, p}$ are
 445 then computed. Finally, these standard deviations are averaged across the P different simulated
 446 datasets giving $sd(\hat{\beta}_0)_{\alpha_W, \alpha_M, \beta_W, n_g}$ and $sd(\hat{\beta}_1)_{\alpha_W, \alpha_M, \beta_W, n_g}$. For each scenario, these averaged standard
 447 deviations are explored for increasing number of biomarker-derived intake estimates n_g . Small
 448 decrements of the standard deviations with increasing n_g would indicate no improvement in
 449 parameter estimate stability as a result of, expensively, collecting greater portions of biomarker data.

450

451 References

- 452 1. Bingham SA. Biomarkers in nutritional epidemiology. *Public Health Nutrition* 2002;5(6a):821-
453 7. doi: 10.1079/PHN2002368.
- 454 2. Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, Ferrari P, Carroll RJ. Bias in
455 dietary-report instruments and its implications for nutritional epidemiology. *Public Health*
456 *Nutrition* 2002;5(6a):915-23. doi: 10.1079/PHN2002383.
- 457 3. Dhurandhar NV, Schoeller D, Brown AW, Heymsfield SB, Thomas D, Sørensen TIA, Speakman
458 JR, Jeansonne M, Allison DB, Energy Balance Measurement W, et al. Energy balance
459 measurement: When something is not better than nothing. *International Journal of Obesity*
460 2015;39(7):1109-13. doi: 10.1038/ijo.2014.199.
- 461 4. Tinker LF, Sarto GE, Howard BV, Huang Y, Neuhouser ML, Mossavar-Rahmani Y, Beasley JM,
462 Margolis KL, Eaton CB, Phillips LS, et al. Biomarker-calibrated dietary energy and protein
463 intake associations with diabetes risk among postmenopausal women from the Women's
464 Health Initiative. *American Journal of Clinical Nutrition* 2011;94(6):1600-6. doi:
465 10.3945/ajcn.111.018648.
- 466 5. Prentice RL, Huang Y, Kuller LH, Tinker LF, Horn LV, Stefanick ML, Sarto G, Ockene J, Johnson
467 KC. Biomarker-calibrated Energy and Protein Consumption and Cardiovascular Disease Risk
468 Among Postmenopausal Women. *Epidemiology* 2011;22(2):170-9. doi:
469 10.1097/EDE.0b013e31820839bc.
- 470 6. Lampe JW, Huang Y, Neuhouser ML, Tinker LF, Song X, Schoeller DA, Kim S, Raftery D, Di C,
471 Zheng C, et al. Dietary biomarker evaluation in a controlled feeding study in women from the
472 Women's Health Initiative cohort. *American Journal of Clinical Nutrition* 2017;105(2):466-75.
473 doi: 10.3945/ajcn.116.144840.
- 474 7. Gibbons H, Michielsen CJR, Rundle M, McNulty BA, Nugent AP, Gibney MJ, Brennan L, al e.
475 Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as
476 an example. Wiley, 2017.
- 477 8. Garcia-Perez I, Posma JM, Chambers ES, Nicholson JK, C. Mathers J, Beckmann M, Draper J,
478 Holmes E, Frost G. An Analytical Pipeline for Quantitative Characterization of Dietary Intake:
479 Application to Assess Grape Intake. *Journal of Agricultural and Food Chemistry*
480 2016;64(11):2423-31. doi: 10.1021/acs.jafc.5b05878.
- 481 9. Garcia-Perez I, Posma JM, Gibson R, Chambers ES, Hansen TH, Vestergaard H, Hansen T,
482 Beckmann M, Pedersen O, Elliott P, et al. Objective assessment of dietary patterns by use of
483 metabolic phenotyping: a randomised, controlled, crossover trial. *The Lancet Diabetes &*
484 *Endocrinology* 2017;5(3):184-95. doi: 10.1016/S2213-8587(16)30419-3.
- 485 10. Gormley IC, Bai Y, Brennan L. *Combining biomarker and self-report dietary intake data: a*
486 *review of the state of the art and an overview of concepts*. *Statistical Methods in Medical*
487 *Research*. To appear, 2019.
- 488 11. Zhang S, Midthune D, Guenther PM, Krebs-Smith SM, Kipnis V, Dodd KW, Buckman DW,
489 Tooze JA, Freedman L, Carroll RJ. A new multivariate measurement error model with zero-
490 inflated dietary data, and its application to dietary assessment. *The Annals of Applied*
491 *Statistics* 2011;5(2B):1456-87. doi: 10.1214/10-AOAS446.
- 492 12. Prentice RL, Mossavar-Rahmani Y, Huang Y, Van Horn L, Beresford SAA, Caan B, Tinker L,
493 Schoeller D, Bingham S, Eaton CB, et al. Evaluation and comparison of food records, recalls,
494 and frequencies for energy and protein assessment by using recovery biomarkers. *American*
495 *Journal of Epidemiology* 2011;174(5):591-603. doi: 10.1093/aje/kwr140.
- 496 13. Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, Subar AF,
497 Tooze JA, Carroll RJ, Freedman LS. Modeling Data with Excess Zeros and Measurement Error:
498 Application to Evaluating Relationships between Episodically Consumed Foods and Health
499 Outcomes. *Biometrics* 2009;65(4):1003-10. doi: 10.1111/j.1541-0420.2009.01223.x.
- 500 14. Keogh RH, White IR. A toolkit for measurement error correction, with a focus on nutritional
501 epidemiology. *Statistics in Medicine* 2014;33(12):2137-55. doi: 10.1002/sim.6095.

15. Prentice RL, Pettinger M, Tinker LF, Huang Y, Thomson CA, Johnson KC, Beasley J, Anderson G, Shikany JM, Chlebowski RT, et al. Regression Calibration in Nutritional Epidemiology: Example of Fat Density and Total Energy in Relationship to Postmenopausal Breast Cancer. *American Journal of Epidemiology* 2013;178(11):1663-72. doi: 10.1093/aje/kwt198.
16. Prentice RL, Shaw PA, Bingham SA, Beresford SAA, Caan B, Neuhouser ML, Patterson RE, Stefanick ML, Satterfield S, Thomson CA, et al. Biomarker-calibrated Energy and Protein Consumption and Increased Cancer Risk Among Postmenopausal Women. *American Journal of Epidemiology* 2009;169(8):977-89. doi: 10.1093/aje/kwp008.
17. Zheng C, Beresford SA, Van Horn L, Tinker LF, Thomson CA, Neuhouser ML, Di C, Manson JE, Mossavar-Rahmani Y, Seguin R, et al. Simultaneous association of total energy consumption and activity-related energy expenditure with risks of cardiovascular disease, cancer, and diabetes among postmenopausal women. *American Journal of Epidemiology* 2014;180(5):526-35. doi: 10.1093/aje/kwu152.
18. Huang Y, Van Horn L, Tinker LF, Neuhouser ML, Carbone L, Mossavar-Rahmani Y, Thomas F, Prentice RL. Measurement Error Corrected Sodium and Potassium Intake Estimation Using 24-Hour Urinary Excretion. *Hypertension* 2014;63(2):238-44. doi: 10.1161/HYPERTENSIONAHA.113.02218.
19. Min Y, Agresti A. Modeling Nonnegative Data with Clumping at Zero: A Survey. *Journal of the Iranian Statistical Society* 2002;1(1):7-33.
20. Tobin J. ESTIMATION OF RELATIONSHIPS FOR LIMITED DEPENDENT VARIABLES: 1. INTRODUCTION. *Econometrica* (pre-1986) 1958;26(1):24.
21. Duan N, Manning WG, Morris CN, Newhouse JP. A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Business & Economic Statistics* 1983;1(2):115-26. doi: 10.1080/07350015.1983.10509330.
22. Box GEP, Cox DR. An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)* 1964;26(2):211-43. doi: doi:10.1111/j.2517-6161.1964.tb00553.x.
23. van der Vaart A.W. WJA. The Delta-Method. In *Weak Convergence and Empirical Processes*. New York: Springer, 1996.

CITRUS INTAKE	W_{A_N}	$W_{A_n^*}$	W_{A_n}	M_{A_n}	$M_{A_n} _{W_{A_n}=0}$	$W_{A_n} _{M_{A_n}=0}$
Mean	75.42	72.41	80.39	83.26	30.52	24.18
Standard deviation	106.11	107.19	104.21	135.45	56.56	54.01
Min	0.00	0.00	0.00	0.00	0.00	0.00
First quartile	0.00	0.00	0.00	7.83	0.00	0.00
Median	31.88	22.00	43.75	34.48	12.81	0.00
Third quartile	117.31	113.00	122.75	98.86	35.07	7.12
Max	845.75	845.75	712.50	1226.79	441.20	302.50
Zero intakes	646 (43%)	424 (45%)	215 (38%)	94 (17%)	68(32%)	68(72%)

Table 1 Summary statistics for the citrus mean daily intake data (g/day). Self-reported mean daily intakes are reported for the entire sample (W_{A_N}), subjects for which no biomarker data is available ($W_{A_n^*}$) and subjects for which biomarker-derived estimates of intake are computed (W_{A_n}). The fourth column reports summary statistics for the biomarker-derived estimates of intake (M_{A_n}).

MODEL	Linear regression model								Tobit model		Two-part model	
	BIOMARKER TRANSFORMATION	M_t	$\log(M_t)$	$BC(M_t, 0.1)$	$BC(M_t, 0.5)$	$BC(M_t, 1)$	$BC(M_t, 1.1)$	$BC(M_t, 1.5)$	M_t	$\log(M_t)$	M_t	$\log(M_t)$
CALIBRATED MEAN DAILY CITRUS INTAKE (g/day)	Mean	81	103	116	58	81	86	101	81	66	64	128
	Standard deviation	66	792	1299	69	66	69	82	67	180	77	1332
	Min	34	19	3	22	34	35	33	33	17	8	15
	First quartile	34	19	3	22	34	35	33	33	17	8	15
	Median	54	25	4	31	53	58	74	52	25	32	20
	Third quartile	107	52	19	66	107	115	150	108	62	94	45
	Max	563	24124	38554	891	563	549	527	553	4313	625	42066
		590	588	585	583	581	578	576	574	572	569	567
		591	587	586	584	580	579	577	575	571	568	567
		592	586	585	583	581	578	576	574	572	569	567
		593	587	586	584	581	578	576	574	572	569	567
		594	588	585	583	581	578	576	574	572	569	567
		595	588	585	583	581	578	576	574	572	569	567

596

597

Table 2 Summary statistics for calibrated intake values using different calibration equations specifications. The calibrated intakes are those computed for the group of n^* subjects for which only self-reported mean daily citrus intake data (g/day) were available

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

Average Mean Square Error			
	BIOMARKER TRANSFORMATION		
MODEL	M_i	log(M_i + 1)	BC(M_i, 1)
Linear regression model	14354	49078	14480
Tobit model	15034	89645	-
Two-part model	21504	21010	-

Table 3 Average mean squared errors computed between calibrated intakes and biomarker-derived estimates of intakes, for the n subjects in the study and different specifications of the calibration equations. The average MSE values for different specification of the Box-Cox λ parameter have been omitted, being all greater than the one obtained with $\lambda = 1$.

Figure 1 Self-reported mean daily citrus intakes (g/day) and calibrated mean daily citrus intakes (g/day). For the calibrated intakes, 95% confidence intervals are reported (light grey). Intakes are ordered from lowest to highest values.

666

667 Figure 2 Simulations. This figure presents the distributions of the MSE computed in the different
 668 simulation scenarios, between simulated and calibrated mean daily intakes. As expected, the error is
 669 lower when both self-reported and biomarker-derived intakes have little variability around true
 670 intakes, and it decreases with increasing values of β_W . However, when self-reported intake data is of
 671 “poor” quality ($\beta_W = 0.1$), the MSE is not much affected by the noise level (α_W, α_M).

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691 Figure 3 Standard deviations of the estimated slope coefficient β_1 are reported, for different sample
 692 sizes n_g and different values of the β_W parameter. These simulated scenarios represent the “best
 693 possible” ones in terms of data quality, as both self-reported and biomarker-derived estimates of
 694 intakes have low variability around the true intake values. Notice that the standard deviations tend to
 695 stabilize quickly, even in the “worst-case” scenario, that is when $\beta_W = 0.1$.

696

697 Figure 4 Standard deviations of the estimated slope coefficient β_1 are reported, for different sample
 698 sizes n_g and different values of the β_W parameter. These simulated scenarios represent the “worst
 699 possible” outcomes in terms of data quality, as both self-reported and biomarker-derived estimates
 700 of intakes have high variability around the true intake values. Notice that the standard deviations tend
 701 to stabilize quickly, even in the “worst-case” scenario, that is when $\beta_W = 0.1$.

702

703 Figure 5 Standard deviations of the estimated intercept β_0 in the different scenarios, for different
 704 sample sizes n_g and different values of the β_W parameter. As for the estimated slope coefficient β_1 ,
 705 standard deviations tend to stabilize quickly, independently of the scenario considered.

706

707

708

709

710

711

712

713

714

715 Figure 6 Standard deviations of the estimated slope coefficient β_1 in the two intermediate scenarios,
 716 for different sample sizes n_g and different values of the β_W parameter. As for the other scenarios,
 717 standard deviations tend to stabilize quickly, even with “poor” self-reported data quality, that is for
 718 $\beta_W = 0.1$.

719

720

721

722