A bivariate measurement error model for nitrogen and potassium intakes to evaluate the performance of regression calibration in the European Prospective Investigation into Cancer and Nutrition study

P Ferrari[1,25], A Roddam[2], MT Fahey[3], M Jenab[4], C Bamia[5], M Ocké[6], P Amiano[7], A Hjartåker[8], C Biessy[1], S Rinaldi[1], I Huybrechts[1], A Tjønneland[9], C Dethlefsen[10], M Niravong[11], F Clavel-Chapelon[11], J Linseisen[12,26], H Boeing[13], E Oikonomou[5], P Orfanos[5], D Palli[14], M Santucci de Magistris[15], HB Bueno-de-Mesquita[6], PHM Peeters[16], CL Parr[17], T Braaten[18], M Dorronsoro[7], T Berenguer[19], B Gullberg[20], I Johansson[21], AA Welch[22,27], E Riboli[23], S Bingham[22,24,*] and N Slimani[1]

[1]*Dietary Exposure Assessment Group, International Agency for Research on Cancer, Lyon, France;* [2]*Cancer Epidemiology Unit, University of Oxford, Oxford, UK;* [3]*Biostatistics Unit, Medical Research Council and University of Cambridge, Cambridge, UK;* [4]*Lifestyle and Cancer Group, International Agency for Research on Cancer, Lyon, France;* [5]*Department of Hygiene Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece;* [6]*National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands;* [7]*Public Health Division of Gipuzkoa, Basque Government, San Sebastian and CIBER Epidemiología y Salud Pública (CIBERESP), Spain;* [8]*Cancer Registry of Norway, Oslo, Norway;* [9]*Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark;* [10]*Department of Cardiology, Center for Cardiovascular Research, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark;* [11]*Inserm, ERI 20, Institut Gustave Roussy, Villejuif, France;* [12]*Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany;* [13]*Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbrücke, Germany;* [14]*Molecular and Nutritional Epidemiology Unit, ISPO, Florence, Italy;* [15]*Department of Clinical and Experimental Medicine, University of Naples, Federico II, Naples, Italy;* [16]*Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands;* [17]*Institute of Basic Medical Sciences, Department of Biostatistics, University of Oslo, Oslo, Norway;* [18]*Institute of Community Medicine, University of Tromsø, Tromsø, Norway;* [19]*Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of Oncology (ICO), Barcelona, Spain;* [20]*Department of Clinical Sciences, Lund University, Malmö, Sweden;* [21]*Department of Odontology, Faculty of Medicine, University of Umeå, Umeå, Sweden;* [22]*Department of Public Health and Primary Care, MRC Centre for Nutritional Epidemiology in Cancer Prevention and Survival, University of Cambridge, Cambridge, UK;* [23]*Department of Epidemiology, Public Health and Primary Care, Imperial College, London, UK and* [24]*Diet and Cancer Group, MRC Mitochondrial Biology Unit, Cambridge, UK*

Objectives: Within the European Prospective Investigation into Cancer and Nutrition (EPIC) study, the performance of 24-h dietary recall (24-HDR) measurements as reference measurements in a linear regression calibration model is evaluated critically at the individual (within-centre) and aggregate (between-centre) levels by using unbiased estimates of urinary measurements of nitrogen and potassium intakes.

Correspondence: Dr N Slimani, Dietary Exposure Assessment Group, International Agency for Research on Cancer (IARC), WHO, Lyon, France.
E-mail: slimani@iarc.fr
*The author is deceased.
[25]Current address: Data Collection and Exposure Unit (DATEX), European Food Safety Authority, Parma, Italy.
[26]Current address: Institute of Epidemiology, Helmholtz Centre Munich, Neuherberg, Germany.
[27]Current address: School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, UK.
Guarantor: P Ferrari.
Contributors: PF conducted the statistical analysis, prepared the tables and wrote the paper, taking into account comments from all co-authors. NS was the overall coordinator of this project and the EPIC Nutrient Database (ENDB) project. AR, MF, MJ, CB, MO, PA, AH and CB were members of the writing group and gave input on the statistical analyses, drafting of the paper and interpretation of results. The other co-authors were local EPIC collaborators who participated in the collection of dietary and other data and in the EPIC nutritional database (ENDB) project. ER is the overall coordinator of the EPIC study.

Methods: Between 1995 and 1999, 1072 study subjects (59% women) from 12 EPIC centres volunteered to collect 24-h urine samples. Log-transformed questionnaire, 24-HDR and urinary measurements of nitrogen and potassium intakes were analysed in a multivariate measurement error model to estimate the validity of coefficients and error correlations in self-reported dietary measurements. In parallel, correlations between means of 24-HDR and urinary measurements were computed. Linear regression calibration models were used to estimate the regression dilution (attenuation) factors.
Results: After adjustment for sex, centre, age, body mass index and height, the validity coefficients for 24-HDRs were 0.285 (95% confidence interval: 0.194, 0.367) and 0.371 (0.291, 0.446) for nitrogen and potassium intakes, respectively. The attenuation factors estimated in a linear regression calibration model were 0.368 (0.228, 0.508) for nitrogen and 0.500 (0.361, 0.639) for potassium intakes; only the former was different from the estimate obtained using urinary measurements in the measurement error model. The aggregate-level correlation coefficients between means of urinary and 24-HDR measurements were 0.838 (0.637, 0.932) and 0.756 (0.481, 0.895) for nitrogen and potassium intakes, respectively.
Conclusions: This study suggests that 24-HDRs can be used as reference measurements at the individual and aggregate levels for potassium intake, whereas, for nitrogen intake, good performance is observed for between-centre calibration, but some limitations are apparent at the individual level.

Introduction

The accuracy of dietary assessment instruments used in nutritional epidemiology studies, that is, questionnaires such as food frequency questionnaires or dietary histories, has been repeatedly questioned (Freedman *et al.*, 1990; Schatzkin *et al.*, 2003; Kristal *et al.*, 2005), as it is recognized that they are subject to random and systematic measurement errors (Kipnis *et al.*, 2003). In the past decade, epidemiological investigations aimed at quantifying the association between diet and risk of chronic disease(s) have often integrated data from dietary questionnaires, completed by all study subjects, with reference instruments, usually 24-h dietary recalls (24-HDRs), administered to a subsample of the study (Thompson *et al.*, 1997; Riboli and Kaaks, 2000; Stram *et al.*, 2000). The joint use of questionnaire and reference measurements in a calibration study allows relative risk estimates, expressing the association between exposure and risk of disease, to be corrected for measurement errors (Rosner *et al.*, 1989).

In the European Prospective Investigation into Cancer and Nutrition (EPIC), which is a cohort study covering ~500 000 individuals, each of whom completed a country-specific dietary food frequency questionnaire at recruitment, a calibration study was designed using a single replicate 24-HDR per study subject as the reference measurement. The main objectives of this study were to (1) correct for systematic bias in different dietary questionnaires at the aggregate (between-centre) level and (2) correct for measurement error bias in the exposure–disease relationship (within-centre level). An important statistical requirement of any reference measurement is that it must be unbiased, that is, any deviation from the unknown true intake must be entirely random, and errors in the reference measurement must be uncorrelated with those in the questionnaire measurement (Kaaks *et al.*, 1994).

It has been observed that error correlations are likely to exist in self-reported estimates of intake as a consequence of the study subject's tendency to consistently underestimate or overestimate the dietary intake (Plummer and Clayton, 1993; Day and Ferrari, 2002; Kipnis *et al.*, 2003). This motivates the comparison of questionnaires and 24-HDR measurements with 'objective' measurements. Recovery-based biomarkers of absolute dietary intakes suit this purpose (Kaaks *et al.*, 2002), as they provide measurements of intake whose errors are independent of the unknown true intakes and uncorrelated with errors in self-reported dietary intakes, either from questionnaire or from 24-HDR. A limited number of recovery biomarkers are currently available for use as reference measurements in validation studies (Bingham *et al.*, 1997; Day *et al.*, 2001; Kipnis *et al.*, 2003; Slimani *et al.*, 2003).

In this paper, 24-h urinary measurements, which provide unbiased estimates of nitrogen and potassium intakes (Bingham *et al.*, 1997), were compared with dietary questionnaires and 24-HDRs in a bivariate measurement error model, with the aim of evaluating the magnitude of measurement errors and error correlations between self-reported dietary assessment tools. Moreover, estimates of attenuation factors from a bivariate linear regression calibration model for nitrogen and potassium intakes, obtained using questionnaire and 24-HDR measurements as the reference (Ferrari *et al.*, 2008), were compared with the corresponding estimates obtained from the measurement error model, which includes the recovery biomarkers. In this manner, the performance of linear regression calibration was evaluated, in the absence of biomarker data, to provide accurate solutions for measurement error corrections to the parameters that quantify the exposure–disease relationship. In addition, a comparison of 24-HDR with urinary measurements was undertaken at the aggregate level by computing correlation coefficients for centre- and gender-specific means, to evaluate further the capacity of 24-HDRs for use as reference dietary measurements for between-centre calibration in the EPIC study.

Materials and methods

Dietary data

Information on habitual dietary intakes was assessed at baseline using different quantitative dietary questionnaires (from France, the Netherlands, Germany, Greece and Italy, except Naples) or a semi-quantitative food frequency questionnaire (from the United Kingdom and Naples), developed and validated in each participating country (Riboli *et al.*, 2002). In addition, a single 24-HDR interview was conducted in EPIC to obtain a reference measurement from a subsample (~8%) of each cohort (Slimani *et al.*, 2002). In contrast to the baseline questionnaires, the 24-HDR interviews were highly standardized across countries using the same structure and interview procedure and a common computer program (EPIC-SOFT). More details on the concept of standardization and the structure of EPIC-SOFT are described in detail elsewhere (Slimani *et al.*, 2000).

Food composition tables

Whereas country-specific nutrient databases were used to estimate nutrient intakes from baseline EPIC dietary questionnaires, standardized databases were developed to compute nutrient intakes from the 24-HDRs (Slimani *et al.*, 2007). These nutrient databases were compiled within the framework of the ENDB (EPIC Nutrient Database project) (Slimani *et al.*, 2007) with the main aim of improving the comparability of nutrient databases across the 10 countries participating in EPIC. The calculation of nutrient contents for each food was done by following a standardized procedure, developing on the nutrient information available in national food composition tables. Other common rules and calculations were adopted for computing the nutrient values of mixed recipes and foods that were missing from the national food composition tables (Slimani *et al.*, 2007).

24-h urine samples

A convenient subsample from 12 EPIC centres, Paris (France), Varese, Turin, Florence, Naples and Ragusa (Italy), Greece, Cambridge and Oxford (United Kingdom), Bilthoven (The Netherlands) and Heidelberg and Postdam (Germany), were re-contacted between 1995 and 1999 and asked to collect a 24-h urine specimen. Those who agreed to participate signed a consent form. More details are available elsewhere (Slimani *et al.*, 2003). To evaluate the completeness of the 24-h urine collections, the subjects were given three 80-mg tablets of para-amino benzoic acid (PABA) to take with meals on the day of urine collection, in order to evaluate the completeness of collection (Bingham *et al.*, 1997). Details on the collection, storage and treatment of urine samples are provided elsewhere (Slimani *et al.*, 2003). In urine, total nitrogen was determined using the Kjeldahl technique (Tecator 1002; Perstorp Analytical Ltd., Bristol, UK); potassium was determined using an IL 943 flame photometer (Instrumentation Laboratory, Warrington, UK), with calibrators and internal standards included in every run (Tasevska *et al.*, 2006); and PABA was measured using colorimetry (Bingham *et al.*, 1988).

Twenty-four hour urine collections containing between 85 and 110% of the PABA markers were considered complete. Urine specimens containing <70% PABA recovery (when <3 tablets may have been taken) and >110% PABA recovery (when subjects may have taken drugs that interfere with the colorimetric analysis) were excluded from the analysis. Additional urine samples (with 70–84% PABA recovery) were used after correction for urinary electrolytes up to the expected values of 93% PABA recovery (Johansson *et al.*, 1999).

Statistical analyses

Adjusted overall geometric means and inter-quintile range (10th–90th percentiles) of the questionnaire, 24-HDR and urinary measurements of nitrogen and potassium intakes were computed by averaging out the effects of gender and centre.

Measurement error model. We evaluated the validity of self-reported dietary measurements for nitrogen and potassium intakes by computing the validity coefficient, which estimates the correlation between the observed measurements of exposure (X), whether questionnaire or 24-HDR, and the unknown true level (T). A measurement error model was defined that assumes linear relationships between questionnaire (Q_k), 24-HDR (R_k) and urinary marker (M_k) measurements and the true level (T_k) of intake, where $k=1, 2$, for nitrogen and potassium, respectively. The model can be defined as

$$\begin{aligned} &\text{Questionnaire:} && Q_k = \alpha_{Q_k} + \beta_{Q_k} T_k + \varepsilon_{Q_k} \\ &\text{24-HDR:} && R_k = \alpha_{R_k} + \beta_{R_k} T_k + \varepsilon_{R_k} \\ &\text{Urinary biomarker:} && M_k = T_k + \varepsilon_{M_k} \end{aligned} \qquad (1)$$

where we assume that $E(\varepsilon_{Q_k}|T_k)=0$, $E(\varepsilon_{R_k}|\ T_k)=0$, $E(\varepsilon_{M_k}|T_k)=0$, and $\mathrm{Var}(\varepsilon_{Q_k})=\sigma^2_{\varepsilon Q_k}$, $\mathrm{Var}(\varepsilon_{R_k})=\sigma^2_{\varepsilon R_k}$ and $\mathrm{Var}(\varepsilon_{M_k})=\sigma^2_{\varepsilon M_k}$. For questionnaire and 24-HDR measurements, the coefficients α_X and β_X ($X=Q_k, R_k$) in model (1) express constant and proportional scaling biases (Kaaks *et al.*, 1994), while the ε_X terms model the random measurement errors in Q_k and R_k, and are assumed to be uncorrelated with the true level, T_k (Kaaks *et al.*, 2002; Ferrari *et al.*, 2007).

In this analysis, the urinary measurements were chosen to be the 'reference measurements', and the relationship in model (1) assumes that any errors are strictly random. Therefore, variation around an individual's unknown true values is entirely attributable to random within-person variation in true intake levels over time (that is, $\mathrm{Cov}(\varepsilon_{M_k}, T_k)=0$).

Model (1) also allowed for a possible non-zero correlation between the random errors in questionnaires and in 24-HDRs ($\mathrm{Cov}(\varepsilon_{Q_k}, \varepsilon_{R_\ell}) \neq 0$, for k, $\ell = 1, 2$), which in practice may occur as a result of inter-individual differences in tendencies either to under-report or to over-report the intake levels consistently. Furthermore, correlation between errors in questionnaire-based assessments may occur because these measurements are likely to share common sources of variability. Similar assumptions were made for error correlations of the same type of dietary assessment, that is, questionnaire and 24-HDR measurements ($\mathrm{cov}(\varepsilon_{Xi_1}, \varepsilon_{Xi_2}) \neq 0$ for $X = Q, R$).

The potential confounding role of several adjustment factors was taken into account in the measurement error model. Dietary and urinary measurements were regressed on the study subjects' age, body mass index, height, recruitment centre and gender, as well as the residuals used in model (1). Age, body mass index and height were chosen because previous studies (Heitmann and Lissner, 1995; Prentice, 1996; Ferrari *et al.*, 2002) have suggested that they are significant determinants of the accuracy of questionnaire measurements. In addition, to remove any ecological correlation between the observed measurements and true intakes, taking into account the multi-centre nature of the data, and to remove systematic differences between genders, adjustments for centre and gender were included.

Calibration model. For centre $j = 1, \ldots, J$, let $\boldsymbol{T}_j = (T_{1j}, T_{2j})$ denote the vector of true values of the dietary variables, $\mathbf{Q}_j = (Q_{1j}, Q_{2j})$ be the corresponding questionnaire measurements and $\boldsymbol{Z}_\mathbf{j} = (\mathbf{Z_{1j}}, \ldots, \mathbf{Z_{1K_1}})$ be the $1 \times K_1$ vector of variables assumed to be measured without errors. In an extension to Rosner *et al.* (1989, 1990), the linear regression calibration model can be defined for the first variable, for simplicity, as

$$\begin{aligned} T_{1j} &= \alpha_{1j} + \lambda_{11j} Q_{1j} + \lambda_{12} Q_{2j} + \boldsymbol{Z_j}\boldsymbol{\Gamma_1} + \varepsilon_{1j} \\ \alpha_{1j} &= \alpha_1 + u_{10j},\ \lambda_{11j} = \lambda_{11} + u_{11j} \\ \begin{pmatrix} u_{10j} \\ u_{11j} \end{pmatrix} &\sim \mathrm{MVN}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{B_1} & \sigma_{B_1\lambda_{11}} \\ \sigma_{B_1\lambda_{11}} & \sigma^2_{\lambda_{11}} \end{pmatrix} \right] \end{aligned} \qquad (2)$$

and

$$\varepsilon_{1j} \sim N(0, \sigma^2_{\varepsilon_1})$$

The term 'ε_{1j}' captures the residual random variation within each group and the coefficient vector $\boldsymbol{\Gamma_1}$ models the effect of confounder variables ($\mathbf{Z}_j$), which are also included in the measurement error model (1). Model (2) has fixed-effect parameters (α_1, λ_{11}, λ_{12}, $\boldsymbol{\Gamma_1}$), and three random effects (u_{10j}, u_{11j} and ε_{1j}), which give rise to four variance components, reflecting, respectively, variation between groups ($\sigma^2_{B_1}$), variability in the attenuation factor across groups ($\sigma^2_{\lambda_{11}}$), their covariance ($\sigma_{B_1\lambda_{11}}$) and within-centre random variation ($\sigma^2_{\varepsilon_1}$). As previously detailed in multivariate models (Rosner *et al.*, 1990), coefficients λ_{11} and λ_{12} can be referred to as the attenuation and contamination factor, respectively (Kipnis *et al.*, 1997). It should be noted that relationships similar to model (2) apply for the second dietary variable (T_{2j}), in which λ_{21} and λ_{22} indicate contamination and attenuation factors, respectively. For the sake of simplicity, in the EPIC calibration model only attenuation factors (λ_{11} and λ_{22} for Q_{1j} and Q_{2j}, respectively) were assumed to have a random component (u_{k1j}, $k = 1, 2$).

In situations in which two variables are measured with error, the level of bias in the parameter that captures the association between the first exposure variable and the disease has a multiplicative component captured by the attenuation factor (λ_{11}), and an additive component captured by the contamination factor of the second exposure variable (λ_{21}) (Rosner *et al.*, 1990; Kipnis *et al.*, 1997).

In nutritional epidemiology, apart from a few expensive and logistically challenging biomarkers of dietary intake, the use of which in large-scale investigations is not practical, measurements of the true intake are not available. In the EPIC study, R measurements (that is, 24-HDR) are used instead of T in model (2) to correct for errors in dietary measurements when evaluating the relationship between food groups and/or nutrients and cancer incidence (Ferrari *et al.*, 2008). For this purpose, R measurements are assumed to be the reference measurements (that is, $\alpha_{R_k} = 0$ and $\beta_{R_k} = 1$ in (1)) and errors in $\boldsymbol{R}_k$ and $\boldsymbol{Q}_k$ measurements are assumed to be uncorrelated.

In this study, the performance of the EPIC linear regression calibration model for nitrogen and potassium intakes was evaluated by comparing the estimates of attenuation factors obtained using $\boldsymbol{R}_k$ and $\boldsymbol{Q}_k$ measurements in (2) with the estimates obtained from the linear regression calibration model using $\boldsymbol{T}_k$ and $\boldsymbol{Q}_k$ measurements and parameter estimates from the measurement error model (1). Computational details are provided in Appendix.

The means of questionnaire and 24-HDR measurements were compared at the aggregate level with urinary nitrogen and potassium measurements by computing Pearson's correlation coefficients of centre- and gender-specific means adjusted for age, body mass index and height. The correlation coefficients were adjusted for gender and weighed by the centre-specific sample size. The 95% confidence intervals (CIs) were computed by applying Fisher's z transformation (Fisher, 1970). The significance of the difference between the correlations of urinary measurements with Q and R measurements was tested using the Hotelling test for correlated correlations (Cohen, 1989).

In all analyses, self-reported and urinary measurements were log-transformed to improve normality. Gender-specific and overall models were fitted, but only the latter were presented, as results were similar in men and women. Parameter estimates for the measurement error model were obtained using the CALIS procedure in SAS, Cary, NC, USA (SAS Institute, 1999), which applies a maximum likelihood estimation algorithm. Approximately 95% CIs for estimates of the validity coefficients, error correlations, correlations between true intakes in the measurement error model and

estimates of the attenuation factors were obtained by computing the 2.5th and 97.5th percentiles of a distribution determined by bootstrap sampling (Efron and Tibshirani, 1998). A total of 1000 repetitions gave sufficiently stable intervals. For the estimates of the true variability of nitrogen and potassium intakes, 95% CIs were obtained using s.e. from the CALIS procedure.

Parameters in the regression calibration model were estimated through the MIXED procedure of the SAS software (SAS Institute, SAS Online Doc, Version 8), using the restricted maximum likelihood method. The statistical significance of the difference between the attenuation factors obtained using the parameters of model (1) and those obtained from a standard regression calibration model was determined using the bootstrap sampling distribution. For all analyses, P-values <0.05 were considered statistically significant.

Results

A total of 1386 urine samples were collected for this study (59% women). After the exclusion of subjects with missing questionnaires or 24-HDRs ($n=45$), those who were on a diet at the time of the 24-HDR interview ($n=52$), those with missing potassium measurements in urine samples ($n=31$) and those with incomplete PABA recovery ($n=189$), information obtained from 1072 subjects was retained for statistical analyses.

Overall and centre-specific geometric means and inter-quintile range (10th–90th percentiles) for questionnaire, 24-HDR and urinary measurements of nitrogen and potassium intakes are presented in Table 1. For nitrogen intake, geometric means were consistently higher for urinary measurements than for 24-HDRs and questionnaires, except for those from France, Naples and Cambridge. A less clear pattern was apparent for potassium intake, with good agreement between geometric means of self-reported assessments and urinary measurements, with the exception of Naples (low 24-HDR mean), and Ragusa and Athens (high questionnaire means). A remarkably good agreement was observed between the overall 24-HDR and questionnaire means and the urinary measurement means.

Estimates of the validity coefficients were higher for 24-HDR (0.285 and 0.371), than for questionnaire measurements (0.222 and 0.287) for nitrogen and potassium intakes, respectively (Table 2). High error correlations were observed for nitrogen and potassium intakes in questionnaire (0.751) and 24-HDR measurements (0.664). Relatively lower values were observed for correlations between errors in questionnaire and 24-HDR measurements for nitrogen (0.206) and potassium intakes (0.213). In the same table, the estimated true correlation between nitrogen and potassium intakes was 0.432 (95% CI: 0.308, 0.531).

For nitrogen intake, the attenuation factor obtained using parameter estimates in the measurement error model (1) was 0.238 (95% CI: 0.171, 0.313), whereas the same estimates from the linear regression calibration model (2) using the 24-HDR as the reference measurement was 0.368 (95% CI: 0.228, 0.508) ($P_{diff}=0.042$) (Table 3). After inclusion of energy intake in the calibration model, the attenuation factor was 0.317 (95% CI: 0.151, 0.482). For potassium intake, the equivalent estimates were 0.498 (95% CI: 0.393, 0.594) and 0.500 (95% CI: 0.361, 0.639), respectively ($P_{diff}=0.889$) (Table 3).

For nitrogen intake, estimates of the correlation coefficient between adjusted means of urinary and dietary measurements were 0.838 for 24-HDRs and 0.605 for questionnaires ($P_{diff}=0.034$). For potassium intake, the correlations were 0.756 for 24-HDR and 0.283 for questionnaires ($P_{diff}=0.018$) (Table 3).

Table 1 Overall and centre-specific sample size (n), gender-adjusted geometric means and inter-quintile range (10th–90th percentiles[a]) of questionnaire, 24-HDR (one replicate per study subject) and urinary measurements of nitrogen and potassium intakes (g/day)

	n	*Questionnaire*		*24-HDR*		*Urine*	
		Mean	*10th–90th*	*Mean*	*10th–90th*	*Mean*	*10th–90th*
Nitrogen							
France[b]	107	15.3	10.5–22.9	13.3	8.8–19.5	13.7	10.5–18.7
Florence	49	14.6	10.1–19.9	13.1	8.9–18.9	15.2	11.7–20.7
Varese	51	14.0	9.6–19.7	13.9	9.4–25.0	15.9	11.6–21.2
Ragusa	18	14.5	9.4–23.1	16.2	9.6–26.3	16.5	13.6–19.6
Turin	43	13.8	10.0–18.7	14.9	10.3–21.0	16.2	12.2–21.6
Naples[b]	20	16.0	11.7–20.5	11.6	8.6–17.5	14.9	11.5–20.9
Cambridge	282	13.8	10.2–18.8	12.6	8.5–18.0	13.8	10.2–17.8
Oxford	84	13.2	10.1–17.2	11.9	8.5–17.3	13.6	10.8–17.1
Bilthoven	177	12.8	9.1–17.5	13.7	8.8–22.4	15.2	10.2–20.3
Athens	61	11.3	8.2–15.3	9.7	5.5–16.7	13.6	9.1–19.0
Heidelberg	81	11.3	8.3–15.9	12.4	7.1–20.6	14.3	10.3–22.9
Potsdam	99	12.2	8.7–16.8	12.1	7.4–18.5	14.2	10.9–18.5
Overall[c]	1072	13.4	9.5–18.6	13.0	8.4–20.3	14.8	11.2–19.9
Potassium							
France[b]	107	3.7	2.6–5.2	3.3	2.3–4.9	3.4	2.3–5.1
Florence	49	3.6	2.8–4.7	3.3	2.3–5.0	3.1	2.2–5.1
Varese	51	3.1	2.3–4.5	3.4	2.2–5.2	3.1	2.1–4.3
Ragusa	18	4.1	2.7–7.6	3.7	2.6–6.5	3.4	2.5–4.5
Turin	43	3.2	2.5–4.0	3.7	2.5–4.8	3.2	2.4–4.4
Naples[b]	20	3.7	2.9–5.1	3.0	2.0–4.5	3.9	3.0–5.7
Cambridge	282	3.9	2.9–5.3	3.5	2.5–5.0	3.7	2.5–5.4
Oxford	84	3.9	3.0–5.1	3.5	2.5–4.7	3.7	2.6–5.3
Bilthoven	177	3.6	2.7–4.9	3.6	2.3–5.7	3.7	2.4–5.2
Athens	61	3.2	2.4–4.4	2.5	1.4–4.2	2.3	1.5–3.9
Heidelberg	81	2.8	2.1–3.6	3.6	2.2–5.3	3.9	2.7–5.3
Potsdam	99	3.0	2.1–4.0	3.3	2.2–5.0	3.5	2.1–5.1
Overall[c]	1072	3.5	2.6–4.8	3.4	2.3–5.1	3.4	2.3–4.9

Abbreviation: 24-HDR, 24-h dietary recall.
[a]Computed as $e^{(10th\ percentile)}$ and $e^{(90th\ percentile)}$ respectively, based on centre-specific log-transformed values.
[b]Women only.
[c]Gender- and centre-adjusted geometric means and inter-quintiles range.

Discussion

In the EPIC study, much effort has been devoted to obtaining reference measurements of intakes using 24-HDRs, by

Table 2 Estimates and approximate 95% CI of validity coefficients ($\hat{\rho}_{X_T}$, $X = Q,R,M$), error correlation between questionnaire and 24-HDR measurements ($\hat{\boldsymbol{\rho}}_{\varepsilon_{Q_k},\varepsilon_{R_k}}$, $k = 1, 2$), error correlation between the same type of measurements ($\hat{\boldsymbol{\rho}}_{X_k}$, ε_{X_l}, $X = Q, R$), true variability ($\hat{\sigma}^2_{T_k}$) and true correlation (ρ_{T_1,T_2}) between nitrogen and potassium intakes[a] according to model (1)

	Validity coefficient	*Error correlations*		*True (co)variability*	
	$\hat{\rho}_{X_T}$	$\hat{\boldsymbol{\rho}}_{\varepsilon_{Q_k},\varepsilon_{R_k}}$[b]	$\hat{\boldsymbol{\rho}}_{\varepsilon_{X_k},\varepsilon_{X_l}}$[c]	$\hat{\sigma}^2_{T_k}$	$\hat{\rho}_{T_1,T_2}$
Nitrogen					
Q[d]	0.222 (0.157, 0.289)	0.206 (0.138, 0.266)	0.751 (0.719, 0.780)	0.049 (0.028, 0.069)	0.432 (0.308, 0.531)
R[e]	0.285 (0.194, 0.367)	—	0.664 (0.610, 0.712)	—	—
Potassium					
Q	0.287 (0.214, 0.354)	0.213 (0.136, 0.287)	—	0.070 (0.048, 0.091)	—
R	0.371 (0.291, 0.446)	—	—	—	—

Abbreviations: BMI, body mass index; CI, confidence interval; 24-HDR, 24-h dietary recall.
[a]Log-transformed values of Q, R and M were regressed on age, BMI, height, centre and gender and residuals computed.
[b]Estimates of error correlations between R and Q measurements of nitrogen and potassium, respectively.
[c]Estimates of error correlations between nitrogen and potassium intakes in R and Q measurements, respectively.
[d]Dietary questionnaire on habitual long-term intakes.
[e]Single 24-HDR, referring to the diet of the day before the interview.

Table 3 Estimates and approximate 95% CI of attenuation factors ($\hat{\lambda}$) from a bivariate linear calibration model (1), and from regressing 24-HDR on questionnaire measurements, as in model (2)

	Model (1)	*Linear calibration (2)*					
	$\hat{\lambda}$[a]	$\hat{\lambda}$[b]	$\hat{\sigma}_{\lambda}$[c]	P_{diff}[d]	$\hat{\rho}_{\bar{M},\bar{R}}$[e]	$\hat{\rho}_{\bar{M},\bar{Q}}$[e]	P_{diff}[f]
Nitrogen	0.238 (0.171, 0.313)	0.368 (0.228, 0.508)	0.115	0.042	0.838 (0.637, 0.932)	0.605 (0.234, 0.822)	0.034
Potassium	0.498 (0.393, 0.594)	0.500 (0.361, 0.639)	0.129	0.889	0.756 (0.481, 0.895)	0.283 (−0.170, 0.637)	0.018

Abbreviations: BMI, body mass index; CI, confidence interval; 24-HDR, 24-h dietary recall.
[a]Adjustment for confounding factors was obtained using residuals computed after regressing dietary and urinary nitrogen and potassium measurements on study subjects' gender, centre, age, BMI and height. The 95% CI was obtained using the 2.5th and 97.5th percentiles of a bootstrap distribution.
[b]Adjustment for confounding factors was obtained by inclusion of study subjects' gender, centre, age, BMI and height in the linear regression calibration model.
[c]Square root of variance component that estimates the between-centre variability of attenuation factors in model (2).
[d]P-value for difference between the attenuation factor estimated from (2) and the correct estimate from (1) was determined by dividing the mean of the difference by the s.d. from the bootstrap distribution, and approximating a standardized normal distribution.
[e]Weighed correlation coefficients (using centre-specific sample sizes) estimated using gender- and centre-specific means ($n = 22$) adjusted by age, BMI and height. The 95% CI is determined using Fisher's transformation.
[f]P-value for difference between $\hat{\rho}_{\bar{M},\bar{R}}$ and $\hat{\rho}_{\bar{M},\bar{Q}}$ from a Student's t-distribution for a two-tailed test statistic for correlated correlations.
Correlation coefficients between urinary (M) and self-reported measurements of nitrogen and potassium intakes at the aggregate level. All estimates were obtained after adjustment for age, BMI, height, centre and gender.

adopting a list of standardized procedures to ensure that dietary information with a high degree of geographical specificity is comparable across countries (Slimani *et al.*, 2003). These measurements are used to complement country-specific dietary questionnaires through linear regression calibration models, which allow correction of the estimates of the association between dietary exposures and risk of chronic diseases for random and systematic errors (Kaaks *et al.*, 1994; Ferrari *et al.*, 2008). Given the multi-centre nature of the EPIC study, the calibration model has both an aggregate-level component, to make dietary exposures readily comparable across centres (between-centre calibration), and an individual-level component, to correct the relationships between dietary exposures and risk of chronic diseases for measurement errors (within-centre calibration) (Ferrari *et al.*, 2004, 2008).

A previous validation study has shown high agreement between 24-HDR and urinary measurements for nitrogen intake (Slimani *et al.*, 2000), thus indicating the good capability of 24-HDRs to serve as reference measurements for between-centre calibration of nitrogen intake. We extended this work to include a critical evaluation of the performance of 24-HDRs for measuring potassium intake, and to consider both the individual and the aggregate components. For the former, a sophisticated bivariate measurement error model, which uses latent factors, was used to capture the complex measurement error structure of nitrogen and potassium intake measurements (Day *et al.*, 2001) using unbiased 'objective' measurements of nitrogen and potassium intakes in urine.

The ability of 24-HDRs to act as reference measurements for dietary intakes in linear regression calibration models

relies on their providing unbiased measurements, the errors of which can be assumed to be independent of the errors in questionnaires (Plummer and Clayton, 1993; Day *et al.*, 2001). Such error correlation can produce bias in both the attenuation and contamination factors. Conversely, error correlations between the questionnaire measurements of different nutrients, which have the potential to bias relative risk estimates severely in situations with more than one error-prone variable (Day *et al.*, 2004), can be fully controlled for if multivariate linear regression calibration models are used.

The results of this study indicate a strong correlation between errors in nitrogen and potassium intakes estimated from questionnaires, and a moderate correlation between errors in 24-HDR and questionnaire measurements. The attenuation factor for potassium intake using 24-HDR as a reference measurement was not different from the estimate using urinary measurements as reference, neither were the contamination factors of both potassium and nitrogen (data not shown). On the contrary, the attenuation factor for nitrogen intake using 24-HDRs as the reference measurement was higher than the true estimate using urinary measurements. This suggests that even moderate amounts of error correlation between questionnaire and 24-HDR measurements can result in a substantial bias in the attenuation factors. On including energy into the calibration model for nitrogen, the attenuation factor was not different from the estimate obtained using urinary measurements as the reference, which is consistent with previous evidence of the benefits of energy adjustment in reducing the effect of error correlations (Willett, 2001; Schatzkin *et al.*, 2003; Day *et al.*, 2004).

An important assumption for calibration of the linear regression is that the reference measurements, in this case 24-HDR, are unbiased. This requires the variation around the unknown true intake to be attributable to either random day-to-day variability or random measurement error in reporting individual diet (that is, $R_k = T_k + \varepsilon_{R_k}$ and $\mathrm{Cov}(T_k, \varepsilon_{R_k}) = 0$). In the measurement error model used in this study, the proportional scaling factors for 24-HDR measurements were 0.461 and 0.457 for nitrogen and potassium intakes, respectively, which are very similar to the estimates for protein and energy intakes from the OPEN (Observing Protein and Energy Nutrition) study (Kipnis *et al.*, 2003). It has been suggested that, for dietary measurements to provide valid reference measurements for between-centre calibration, it is necessary that the level of bias be constant across population subgroups (Kaaks *et al.*, 1994; Riboli and Kaaks, 2000). Results from country-specific centre-adjusted measurement error models showed that scaling factors were not statistically heterogeneous across countries for both nitrogen and potassium intakes (results not shown). However, the limitations of using 24-HDRs as reference measurements at the individual level need to be acknowledged when measurement error correction procedures are applied to the diet–disease relationship. These issues have been addressed by Kipnis *et al.* (2003) and Ferrari *et al.* (2008).

The correlation coefficients between urinary measurements and 24-HDR measurements showed good agreement at the aggregate level. Correlations were significantly higher than for questionnaires, indicating a relative gain when rescaling questionnaire values on 24-HDR means, as is customarily done in EPIC analyses using a linear regression calibration model (Ferrari *et al.*, 2008). Estimates of the error correlations from this study were similar to previous findings on nitrogen and potassium intakes (Day *et al.*, 2001), with substantial correlation between errors in measurements using the same dietary assessment instrument, and lower values for error correlation between 24-HDR and questionnaire measurements. Similar trends were observed in the OPEN study for nitrogen (protein) and energy intake (Kipnis *et al.*, 2003).

When multiple exposures are measured with errors, it is necessary to use a multivariate linear regression calibration model (Rosner *et al.*, 1990) to ensure that the error correlations in Q measurements are properly taken into account and corrected for in the disease model, thus obtaining calibrated (corrected) parameter estimates of the exposure–disease association. In principle, aetiological considerations regarding the possible link between exposures and the condition of interest should drive the decisions on whether to use univariate or bivariate linear regression calibration models, although it should be recognized that associations between exposures and disease could be masked by measurement errors. However, this work focused *a priori* on the impact of exposure measurement errors in the context of a bivariate model, independently of any aetiological evaluations, because this allows all the relevant variances and covariances of nitrogen and potassium intake to be estimated, which would not be possible by analysing two independent univariate models.

In general, in the case of two exposures measured with errors, the level and direction of bias in parameter estimates that capture the exposure–disease association depend on both the attenuation (multiplicative) and the contamination (additive) factors, as well as on the true unknown parameters between each exposure and the risk of disease (Rosner *et al.*, 1990; Kipnis *et al.*, 1997). In multivariate scenarios, therefore, bias due to measurement errors does not necessarily lead to attenuation of relative risk estimates towards the null. This is particularly true in cases in which one of the exposures is not associated with the risk of disease.

This work has critically evaluated the performance of 24-HDR measurements for use as reference measurements in linear regression calibration models. The results for both nitrogen and potassium intakes highlight the strengths and limitations of this approach for measurement error correction. The utilization and development of biomarkers of absolute dietary intake, of which there are few, will allow the techniques described herein to be applied in order to have a better understanding of diet–disease associations.

Conflict of interest

A Roddam has received grant support and owns stock/equity ownership in Amgen Ltd. M Jenab has received grant support from the World Cancer Research Fund. CL Parr has received grant support from the Norwegian Foundation for Health and Rehabilitation. S Bingham has received grant support from MRC Centre. The remaining authors have declared no financial interests.

Acknowledgements

The work described in this paper was carried out with the financial support of the European Commission: Public Health and Consumer Protection Directorate 1993–2004; Research Directorate-General 2005; Ligue contre le Cancer (France); Société 3M (France); Mutuelle Générale de l'Education Nationale; Institut National de la Santé et de la Recherche Médicale (INSERM); Institut Gustave Roussy; German Cancer Aid; German Cancer Research Center; German Federal Ministry of Education and Research; Danish Cancer Society; Health Research Fund (FIS) of the Spanish Ministry of Health; Spanish Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra and the Catalan Institute of Oncology; and ISCIII RETIC (RD06/0020), Spain; Cancer Research UK; Medical Research Council, UK; The Stroke Association, UK; British Heart Foundation; Department of Health, UK; Food Standards Agency, UK; The Wellcome Trust, UK; Greek Ministry of Health; Hellenic Health Foundation; Italian Association for Research on Cancer; Italian National Research Council, Regione Sicilia (Sicilian government); Associazione Iblea per la Ricerca Epidemiologica—ONLUS (Hyblean Association for Epidemiological Research, NPO); Dutch Ministry of Health, Welfare and Sport; Dutch Prevention Funds; LK Research Funds; Dutch ZON (Zorg Onderzoek Nederland); World Cancer Research Fund (WCRF); Swedish Cancer Society; Swedish Research Council; Regional Government of Skane and the County Council of Vasterbotten, Sweden; Norwegian Cancer Society; the Norwegian Research Council and the Norwegian Foundation for Health and Rehabilitation. We thank Sarah Somerville, Nicole Suty and Karima Abdedayem for their assistance with editing, and Kimberley Bouckaert and Heinz Freisling for their technical assistance.

References

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

Bingham SA, Williams R, Cole TJ, Price CP, Cummings JH (1988). Reference values for analytes of 24-h urine collections known to be complete. *Ann Clin Biochem* **25**, 610–619.

Cohen A (1989). Comparison of correlated correlations. *Stat Med* **8**, 1485–1495.

Day N, McKeown N, Wong M, Welch A, Bingham S (2001). Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol* **30**, 309–317.

Day NE, Ferrari P (2002). Some methodological issues in nutritional epidemiology. In: Riboli E, Lambert R (eds). *Nutrition and Lifestyle: Opportunities for Cancer Prevention*. IARC Sci Publ No 156. International Agency for Research on Cancer: Lyon. pp 5–10.

Day NE, Wong MY, Bingham S, Khaw KT, Luben R, Michels KB *et al.* (2004). Correlated measurement error—implications for nutritional epidemiology. *Int J Epidemiol* **33**, 1373–1381.

Efron B, Tibshirani RJ (1998). *An Introduction to the Bootstrap*. Chapman & Hall: London.

Ferrari P, Day NE, Boshuizen HC, Roddam A, Hoffmann K, Thiébaut A *et al.* (2008). The evaluation of the diet/disease relation in the EPIC study: considerations for the calibration and the disease models. *Int J Epidemiol* **37**, 379–381.

Ferrari P, Friedenreich C, Matthews C (2007). The role of measurement error in estimating levels of physical activity. *Am J Epidemiol* **166**, 832–844.

Ferrari P, Kaaks R, Fahey MT, Slimani N, Day NE, Pera G *et al.* (2004). Within- and between-cohort variation in measured macronutrient intakes, taking account of measurement errors, in the European Prospective Investigation into Cancer and Nutrition study. *Am J Epidemiol* **160**, 814–822.

Ferrari P, Slimani N, Ciampi A, Trichopoulou A, Naska A, Lauria C *et al.* (2002). Evaluation of under- and overreporting of energy intake in the 24-hour diet recalls in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Public Health Nutr* **5**(Suppl), S1329–S1345.

Fisher RA (1970). *Statistical Methods for Research Workers*, 14th edn. Hafner Publishing Company: Davien, CT.

Freedman LS, Schatzkin A, Wax Y (1990). The impact of dietary measurement error on planning sample size required in a cohort study. *Am J Epidemiol* **132**, 1185–1195.

Heitmann BL, Lissner L (1995). Dietary underreporting by obese individuals—is it specific or non- specific? *Br Med J* **311**, 986–989.

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

Kaaks R, Ferrari P, Ciampi A, Plummer M, Riboli E (2002). Uses and limitations of statistical accounting for random error correlations in the validation of dietary questionnaire assessments. *Public Health Nutr* **5**, 969–976.

Kaaks R, Plummer M, Riboli E, Esteve J, van Staveren W (1994). Adjustment for bias due to errors in exposure assessments in multicenter cohort studies on diet and cancer: a calibration approach. *Am J Clin Nutr* **59**(Suppl 1), S245–S250.

Kipnis V, Freedman LS, Brown CC, Hartman AM, Schatzkin A, Wacholder S (1997). Effect of measurement error on energy-adjustment models in nutritional epidemiology. *Am J Epidemiol* **146**, 842–855.

Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP *et al.* (2003). Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* **158**, 14–21.

Kristal AR, Peters U, Potter JD (2005). Is it time to abandon the food frequency questionnaire? *Cancer Epidemiol Biomarkers Prev* **14**, 2826–2828.

Plummer M, Clayton D (1993). Measurement error in dietary assessment: an investigation using covariance structure models, Part I. *Stat Med* **12**, 925–935.

Prentice R (1996). Measurement error and results from analytic epidemiology: dietary fat and breast cancer. *J Natl Cancer Inst* **88**, 1738–1747.

Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M *et al.* (2002). European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* **5**(Suppl), S1113–S1124.

Riboli E, Kaaks R (2000). Invited commentary: the challenge of multi-center cohort studies in the search for diet and cancer links. *Am J Epidemiol* **151**, 371–376.

Rosner B, Spiegelman D, Willett WC (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* **132**, 734–745.

Rosner B, Willett WC, Spiegelman D (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med* **8**, 1051–1069.

SAS Institute (1999). *SAS Online Doc*, Version 8. SAS Institute Inc.: Cary, NC.

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

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

Slimani N, Deharveng G, Unwin I, Southgate DA, Vignat J, Skeie G *et al.* (2007). The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *Eur J Clin Nutr* **61**, 1037–1056.

Slimani N, Ferrari P, Ocké M, Welch A, Boeing H, van Liere M *et al.* (2000). Standardisation of the 24-hour diet recall calibration method used in the European Prospective Investigation into Cancer and Nutrition (EPIC): general concepts and preliminary results. *Eur J Clin Nutr* **54**, 900–917.

Slimani N, Kaaks R, Ferrari P, Casagrande C, Clavel-Chapelon F, Lotze G *et al.* (2002). European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics. *Public Health Nutr* **5**(Suppl), S1125–S1145.

Stram DO, Hankin JH, Wilkens LR, Pike MC, Monroe KR, Park S *et al.* (2000). Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. *Am J Epidemiol* **151**, 358–370.

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

Thompson FE, Moler JE, Freedman LS, Clifford CK, Stables GJ, Willett WC (1997). Register of dietary assessment calibration-validation studies: a status report. *Am J Clin Nutr* **65**(Suppl 4), S1142–S1147.

Willett W (2001). Commentary: dietary diaries versus food frequency questionnaires—a case of undigestible data. *Int J Epidemiol* **30**, 317–319.

Appendix

Consider a bivariate linear regression model of the relationship between a vector of true exposure, $\boldsymbol{T} = (T_1, T_2)$, and the vector of observed exposure, $\boldsymbol{Q} = (Q_1, Q_2)$. For simplicity, if we consider the first exposure and omit any confounding variables, then the relationship has the form $T_1 = \lambda_{10} + \lambda_{11}Q_1 + \lambda_{12}Q_2 + \varepsilon_1$. The elements of the vector of parameters $\boldsymbol{\lambda}_1 = (\lambda_{11}, \lambda_{12})$ consist of the attenuation factor for the first exposure and the contamination factor for the second exposure, respectively. Let Σ_{AB} be the covariance matrix between random variables $\boldsymbol{A}$ and $\boldsymbol{B}$; then $\boldsymbol{\lambda}_1$ can be estimated as

$$\lambda_1 = (\Sigma_{QQ})^{-1}\Sigma_{QT_1} = \begin{pmatrix} \sigma^2_{Q_1} & \sigma_{Q_1Q_2} \\ \sigma_{Q_1Q_2} & \sigma^2_{Q_2} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{Q_1T_1} \\ \sigma_{Q_2T_1} \end{pmatrix} = \begin{pmatrix} \dfrac{\sigma_{Q_1T_1}\sigma^2_{Q_2} - \sigma_{Q_2T_1}\sigma_{Q_1Q_2}}{1-\rho^2_{Q_1Q_2}} \\ \dfrac{-\sigma_{Q_1T_1}\sigma_{Q_1Q_2} + \sigma_{Q_2T_1}\sigma^2_{Q_1}}{1-\rho^2_{Q_1Q_2}} \end{pmatrix} \quad (3)$$

Elements in matrices Σ_{QQ} and Σ_{QT_1} can be expressed as a function of parameters estimated from the measurement error model (1). For example, the variance between T_1 and Q_1 can be calculated by $\sigma_{Q_1T_1} = \hat{\beta}_{Q_1}\hat{\sigma}^2_{T_1} + \hat{\sigma}^2_{\varepsilon_{Q1}}$. The term '$\rho_{Q_1Q_2}$' represents the correlation between Q_1 and Q_2. Unlike the linear regression calibration model in (2), which uses $\boldsymbol{R}$ (instead of $\boldsymbol{T}$), $\boldsymbol{Q}$ and $\boldsymbol{Z}_{k1}$ variables, estimates of Σ_{QT_1} obtained from model (1) are not affected by error correlations in questionnaire and 24-HDR measurements. An identical approach is applied for the vector $\boldsymbol{\lambda}_2$ for the second error-prone variable, noting that $T_2 = \lambda_{20} + \lambda_{21}Q_1 + \lambda_{22}Q_2 + \varepsilon_2$.