American Journal of Epidemiology, 2024, **193**, 1482–1493
https://doi.org/10.1093/aje/kwae039
Advance access publication date May 13, 2024
Practice of Epidemiology

Regression calibration of self-reported mobile phone use to optimize quantitative risk estimation in the COSMOS study

Marije Reedijk[†,1,2], Lützen Portengen[†,1], Anssi Auvinen[3,4], Katja Kojo[4], Sirpa Heinävaara[5], Maria Feychting[6], Giorgio Tettamanti[6], Lena Hillert[6], Paul Elliott[7,8,9,10], Mireille B. Toledano[7,8,9,10], Rachel B. Smith[7,8,9,10], Joël Heller[7,8,9], Joachim Schüz[11], Isabelle Deltour[11], Aslak Harbo Poulsen[12], Christoffer Johansen[12,13], Robert Verheij[14], Petra Peeters[2], Matti Rookus[15], Eugenio Traini[1], Anke Huss[1], Hans Kromhout[1], Roel Vermeulen[*,1,2,7,8], The COSMOS Study Group

[1]Institute for Risk Assessment Sciences, Utrecht University, 3584CM Utrecht, the Netherlands
[2]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht (UMCU), 3584CG Utrecht, the Netherlands
[3]Tampere University, Faculty of Social Sciences, Tampere FI-33014, Finland
[4]STUK – Radiation and Nuclear Safety Authority, 01370 Vantaa, Finland
[5]Cancer Society of Finland/Finnish Cancer Registry, 00500 Helsinki, Finland
[6]Karolinska Institutet, Institute of Environmental Medicine, 171 77 Stockholm, Sweden
[7]MRC Centre for Environment and Health, School of Public Health, Imperial College London, London W12 0BZ, United Kingdom
[8]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London W2 1PG, United Kingdom
[9]National Institute for Health and Care Research Health Protection Research Unit in Chemical and Radiation Threats and Hazards, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London W2 1PG, United Kingdom
[10]Mohn Centre for Children's Health and Wellbeing, School of Public Health, Imperial College London, London SW7 2AZ, United Kingdom
[11]International Agency for Research on Cancer (IARC/WHO), Environment and Lifestyle Epidemiology Branch, CS 90627 69366 Lyon, France
[12]Danish Cancer Society Research Center, 2100 Copenhagen, Denmark
[13]CASTLE Cancer Late Effect Research Oncology Clinic, Center for Surgery and Cancer, Rigshospitalet, 2100 Copenhagen, Denmark
[14]Netherlands Institute for Health Services Research (NIVEL), 3513CR Utrecht, the Netherlands
[15]The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, 1066CX Amsterdam, The Netherlands
[*]Corresponding author: Roel Vermeulen, Utrecht University, Institute for Risk Assessment Sciences, Yalelaan 2, 3584CM Utrecht, the Netherlands (r.c.h.vermeulen@uu.nl)
[†]Equal contribution.

Abstract

The Cohort Study of Mobile Phone Use and Health (COSMOS) has repeatedly collected self-reported and operator-recorded data on mobile phone use. Assessing health effects using self-reported information is prone to measurement error, but operator data were available prospectively for only part of the study population and did not cover past mobile phone use. To optimize the available data and reduce bias, we evaluated different statistical approaches for constructing mobile phone exposure histories within COSMOS. We evaluated and compared the performance of 4 regression calibration (RC) methods (simple, direct, inverse, and generalized additive model for location, shape, and scale), complete-case analysis, and multiple imputation in a simulation study with a binary health outcome. We used self-reported and operator-recorded mobile phone call data collected at baseline (2007-2012) from participants in Denmark, Finland, the Netherlands, Sweden, and the United Kingdom. Parameter estimates obtained using simple, direct, and inverse RC methods were associated with less bias and lower mean squared error than those obtained with complete-case analysis or multiple imputation. We showed that RC methods resulted in more accurate estimation of the relationship between mobile phone use and health outcomes by combining self-reported data with objective operator-recorded data available for a subset of participants.

Key words: regression calibration; mobile phone use; measurement error; cohort analysis; health outcomes; exposure assessment.

Introduction

Exposure measurement error is a threat to the validity of environmental epidemiologic research, and may be especially important for studies that rely on self-reported exposure information.[1-3] Systematic and random measurement error may result in substantial bias and loss of precision in estimated exposure-outcome relationships.[1,3] Validation studies have shown substantial error in self-reported estimates of mobile phone use.[4-10] Measurement error correction methods, such as regression calibration (RC), simulation extrapolation, and likelihood-based approaches have been widely used in nutritional epidemiology,[2,11-13] and are increasingly used in environmental epidemiology.[3,14-18]

Received: February 24, 2023. **Accepted:** April 3, 2024

For mobile phone use, multiple bias modeling has been used in the Interphone Study,[19] bias correction in the CEFALO study,[6] and regression calibration by Tokola et al.[20] Redmayne et al developed a measurement error correction method for the number of weekly text messages.[16] The prospective Cohort Study of Mobile Phone Use and Health (COSMOS)[9,21] of more than 310 000 participants from 6 countries aimed to collect both self-reported and operator-recorded data on mobile phone use. Operator data could not be obtained for all participants, eg, because of prepaid service or data-sharing issues, or because subscriptions were held by employers or participants did not consent to the provision of their mobile traffic data to the researchers.[21]

Deciding which information to use for the primary analyses of exposure-outcome associations within COSMOS requires a careful consideration of the available data sources on mobile phone use.[10] Use of operator-recorded information alone may lead both to imprecision in estimates, because data are available for only a subset of participants, and to bias if data are not missing at random. Meanwhile, the use of more readily obtained (but error-prone) self-reported mobile phone data is likely to result in biased risk estimates and precision loss due to both systematic and random measurement errors.

Here we combined self-reported and operator-recorded mobile phone use data aiming to obtain acceptable precision of exposure-outcome estimates while minimizing the bias. We implemented regression calibration (RC),[3,22,23] complete-case (CC) analyses, and multiple imputation (MI)[24] in a simulation study, comparing both bias and precision within and across methods, to optimize quantitative risk estimation in the COSMOS study.

Methods

Study design

We designed our simulation study using real-world data from the COSMOS study, which has been described in detail elsewhere.[21,25,26] Briefly, information on mobile phone use was collected for participants aged 18 and over from 2007 onward in 6 countries: Denmark, Finland, France, the Netherlands, Sweden, and the United Kingdom.[21,25,26] In Denmark, Finland, and Sweden, potential participants were selected using stratified random sampling from subscribers of major mobile phone network operators. In the Netherlands, recruitment was from the general population and a nurses cohort.[26] In the United Kingdom, approximately 65% of participants were recruited by stratified random sampling from subscribers of major mobile phone network operators, while 35% of participants were recruited from the electoral register[25] (Table S1). The COSMOS study protocol was approved by ethical committees in each country. Written or electronic informed consent to link to operator-recorded mobile phone use data was requested from each study participant.

Questionnaire data

The COSMOS baseline questionnaire was administered between 2007 and 2012 either on paper or in most countries electronically. It was administered in 2019 in France, and the French data were therefore not considered in the current analysis. Key topics included mobile phone use, cordless phone use, use of other wireless devices, demographic and social characteristics and several self-reported health outcomes.[21] Characteristics considered as potential predictors of mobile phone use (or affecting self-report) were sex, age, marital status (living together, living apart, and not being in a relationship), educational level (elementary versus secondary school or higher), and employment status (active versus inactive).

Self-reported mobile phone use

Self-reported duration of mobile phone use (REPORT) at baseline[21] was based on answers to the question: "Over the last three months, on average, how much time per week did you spend talking on a mobile phone?" The following response options were provided: "< 5 minutes/week," "5-29 minutes/week," "30-59 minutes/week," "1-3 hours/week," "4-6 hours/week," and "> 6 hours/week." In the Netherlands and the United Kingdom, 2 further categories of call duration were included, "7-9 hours/week" and "10 or more hours/week," but these were combined into "> 6 hours/week" for the present analyses.

Operator-recorded mobile phone use

Operator-recorded duration of mobile phone use (RECORD) was collected for all participants who provided consent and had a subscription under their own name. Operator-recorded data were used only when available for all mobile phones that were reported (up to 2 phones in Denmark, the Netherlands, Sweden, Finland, and up to 3 in the United Kingdom) and for at least 3 full months at the time the baseline questionnaire was administered. Participants with mobile phones that were also used by others were excluded. Operator-recorded data were available for 21% of participants in Denmark, 76% in Finland, 40% in the Netherlands, 63% in Sweden, and 76% in the United Kingdom.

Providers collected information on number and duration separately for outgoing and incoming calls. Across countries, outgoing calls made up 55%-72% of the total number. Where the discrepancy between incoming and outgoing calls was large, this could suggest under-recording of incoming calls, possibly because calls between subscribers from the same provider were not always charged and may have gone unrecorded in some billing systems. For this reason, to evaluate the best RC method, we based our simulations on outgoing call duration only. Data for the 3 months at baseline were used to calculate average duration of calls (minutes per week).

Simulation study

We compared results from different approaches that combine self-reported and operator-recorded mobile phone use: 4 variants of regression calibration (simple RC, generalized additive model for location, shape, and scale (GAMLSS)-based RC, direct RC, and inverse RC), a complete-case (CC) analysis, and multiple imputation (MI) (Figure 1, Appendix S1).

All simulations used data from the subset of participants for whom both self-reported and operator-recorded data were available. For each simulation, we assigned a binary health outcome with probability $P(Y_i = 1)$ according to the following equation:

$$P(Y_i = 1) = \exp(\alpha + \beta \times RECORD_i)/(1 + \exp(\alpha + \beta \times RECORD_i))$$

for participant i with recorded call duration $RECORD_i$ (in minutes/week).

For our main analyses, the outcome was simulated using a slope coefficient (β) of 0.001 (ie, assuming an odds ratio of $\exp(0.001 \times 30) = 1.03$ per additional 30 minutes call-time per week) and with a balanced ratio of cases to noncases (by tuning α). The value of 0.001 was chosen to provide a reasonably strong signal-to-noise ratio across the country-specific datasets. Additional results using a slope coefficient (β) of 0.005 are presented in Table S2.

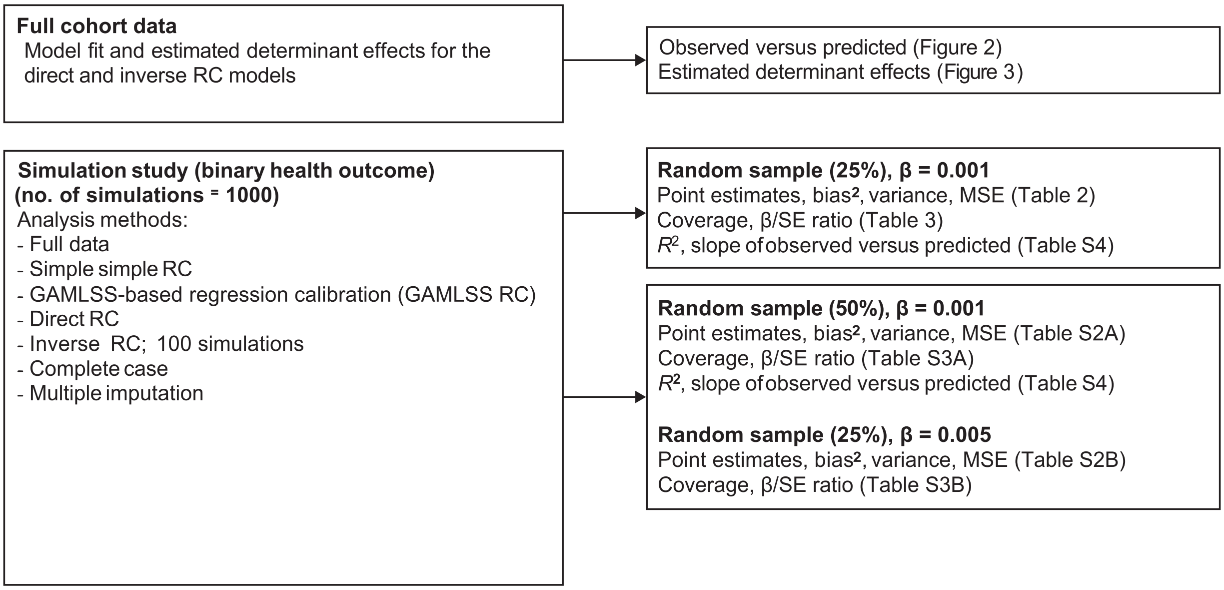

Figure 1. Roadmap to the main results from this paper, using data from the Cohort Study of Mobile Phone Use and Health (COSMOS), multiple countries, 2007-2012. This includes evaluation of model fit and estimated determinant effects for regression calibration (RC) models fitted to the available cohort data and results of the simulation study (including sensitivity analyses). GAMLSS, generalized additive model for location, shape, and scale; MSE, mean squared error; RC, regression calibration; SE, standard error.

We randomly assigned 25% of the participants to the training set used to develop the regression calibration models and for fitting the complete-case model, while the remainder was used as a test set for fitting the "calibrated" health outcome model (mimicking the situation that provider data were missing for 75% of the population). Results obtained using 50% of participants as the training set are presented in Table S3.

The performance of each approach was evaluated by comparing the average estimated slope with the "true" slope coefficient used in the simulations; specifically, we estimated squared bias, variance, and mean squared error (MSE). We calculated 95% CIs using bootstrapping to reflect Monte-Carlo error from using a limited number of simulations. We also estimated the coverage of 95% CIs calculated using the estimated standard error (SE) of the slope coefficient (Wald-type 95% CIs) to investigate whether these correctly accounted for the additional (sampling) variability (Tables S4 and S5). We calculated ratios of estimated slopes to estimated standard errors (β/SE) as a surrogate for statistical power and included results for the full data health model in the tables as an "ideal-case" reference method.

We used 1000 simulations to evaluate performance for all approaches, except for the inverse RC method, where 100 simulations were used to avoid excessive running times. For the same reason, we used model-based ("naive") SEs to obtain precision-weighted estimates, as we found these to be only marginally different from bootstrapped SEs in a subset of 100 simulations.

Full data and complete-case analysis

Logistic regression models were fitted to the simulated health outcome in the combined training and test datasets ("full data") or the training set only ("complete-case"). Note that the "full data" option would not be available for COSMOS.

Operator-recorded mobile phone use was treated as gold standard for "true" mobile phone use, and the goal of our analyses was to estimate the slope coefficient (β) in the following health outcome model:

$$g(E(Y)) = \alpha + \beta.RECORD$$

where g is the logistic link function in the case of a binary outcome Y, and RECORD is operator-recorded mobile phone use.

Regression calibration

Because RECORD is only available for a subset of the participants, we consider using self-reported mobile phone use (REPORT) as an (error-prone) proxy for operator-recorded use (RECORD) and consider different implementations of RC to correct for the measurement error.[27] All involve fitting the so-called "calibrated" health outcome model:

$$g(E(Y)) = \alpha \times +\beta \times .E(RECORD|REPORT, Z)$$

where $E(RECORD|REPORT,Z)$ is the expected value of "true" operator-recorded mobile phone use, conditional on REPORT and other covariates (Z).

We evaluated 4 different RC approaches, differing by the amount of information used and in assumptions regarding the distribution of RECORD or on the relationship between REPORT and RECORD. Our simple RC approach was the most basic, using the empirical average RECORD for each REPORT category. REPORT estimates appeared to follow an approximate log-normal distribution, and for the model-based approaches we indirectly estimated $E(RECORD|REPORT,Z)$ using the relationship between arithmetic mean (AM) and geometric mean (GM) plus geometric standard deviation (GSD) of a log-normally distributed variable (ie, $AM = GM \times \exp(\log(GSD)^2/2)$). Our second approach, GAMLSS RC, used maximum likelihood in a generalized additive modeling framework to estimate both GM and GSD as a function of REPORT categories and covariates, allowing for nonlinear effects of continuous covariates. Our third approach, direct RC, used regression splines to achieve the same flexibility in a Bayesian setting, but in addition relaxed distributional assumptions by modeling the residuals nonparametrically as a Dirichlet process mixture of (log)normal distributions, and with the AM estimated by averaging across draws from that distribution. Finally, our fourth approach, which we called inverse RC, was a Bayesian structural method and featured a probit regression model for REPORT categories using RECORD and all other covariates as predictor variables. By modeling the prior (conditional) distribution of RECORD as a Dirichlet process mixture of (log)normal distributions, we obtained estimates of RECORD conditional on REPORT and other covariates. The GAMLSS RC model was fitted using the R (R Foundation for Statistical Computing) package gamlss,[27] while the Bayesian RC models (direct, inverse) were fitted using the R package rjags.[28]

Because health data in the COSMOS study is available for most of the participants that contributed to the RECORD models, we followed the approach suggested by Spiegelman et al[29] for studies that have an internal validation study. We combined the slope estimate from the complete-case model (ie, the health model for the 25% of participants where RECORD was available) with that from the "calibrated" model fitted to data from the remainder of the population (the test set), using inverse-precision weighting. Although robust SEs for the slope estimate from the "calibrated" model could be obtained using bootstrapping, we did not do that here for computational reasons.

Multiple imputation

Measurement error can also be regarded as a missing data problem,[30] and we therefore considered multiple imputation (MI) in addition to the RC and complete-case analyses for the simulations. MI of operator-recorded phone use was performed by chained equations with fully conditional specification, using predictive mean matching as implemented in the R package mice.[24] Participant characteristics considered as predictors in the imputation model were the same as for the RC approaches, but also included the (simulated) health outcome (Y). We imputed a total of 10 different datasets and combined exposure slope coefficients from the health outcome model using Rubin's rule.[31]

Model fit

We evaluated model fit for RC models using data from all participants with complete RECORD, REPORT, and the covariate data, for each country separately, by comparing predicted with observed average weekly call-time. We calculated the proportion of variance in observed call durations that could be explained by predicted call durations and evaluated the relationship between predicted and observed call durations in detail by fitting a GAMLSS model that allowed for nonlinear effects and nonhomogeneous residual variance.

Results

Participant characteristics

Table 1 lists study attrition and personal characteristics for included participants. Complete information on covariates, self-reported data (REPORT), and operator-recorded data (RECORD) were available for < 50% of participants for most countries, except for Finland and the United Kingdom. In the Netherlands, where sampling was from the general population and a nurses cohort, REPORT was not filled out by 33% of participants, likely due to the high percentage of nonusers. Slightly more women than men participated, except in the Netherlands, where most participants were women. There were notable differences in self-reported and operator-recorded data between countries. Operator-recorded minutes were lowest in the Netherlands (GM (GSD) = 23.4 (2.6) minutes/week) and highest in Finland (81.5 (3.0) minutes/week).

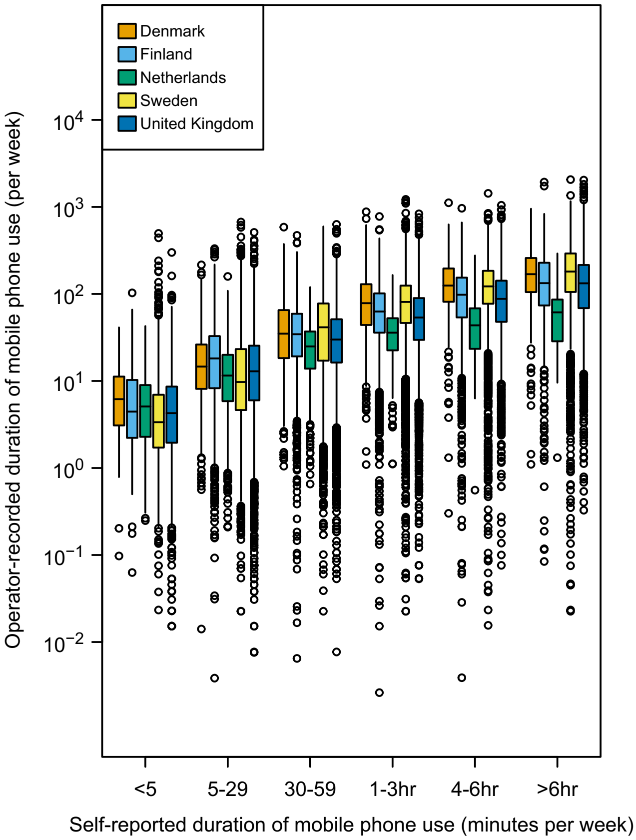

Figure 2. Outgoing operator-recorded mobile phone use according to country and categories of self-reported use, using data from the Cohort Study of Mobile Phone Use and Health (COSMOS), multiple countries, 2007-2012.

The probability of having missing self-reported data depended on the value of operator-recorded data in all countries, but absolute differences were small and the direction across countries was inconsistent. A detailed comparison of participants with or without information on self-reported and operator-recorded mobile phone use and covariates per country is provided in Table S6.

Figure 2 shows the distribution of self-reported data for outgoing calls only, according to country and categories of self-reported use. Self-reported durations tended to be lower for participants from the Netherlands and the United Kingdom, even within categories of self-reported use, and this was most pronounced for higher operator-reported categories.

Model fit and parameter estimates from RC models on the full data

Estimated covariate effects for the direct and inverse RC models fitted to the full cohort data, by country, are presented in the Supplementary Material (Figures S1 and S2). Being in a relationship but not living together was a strong predictor for self-reported durations in all countries and was also associated with higher operator-reported categories (conditional on self-reported) in the probit regression model for the inverse RC method (Figure S1). Older age was associated with lower operator-recorded and self-reported durations in most countries (Figure S2). The underlying variable for self-reported durations in the probit regression

Table 1. Overview of available data and personal characteristics for participants where both self-reported and operator-recorded data were available, using data from the COSMOS Study, multiple countries, 2007-2012.

	Denmark	Finland	Netherlands	Sweden	United Kingdom
Total no. of participants[a]	25 912	13 062	88 466	50 678	98 685
Covariate data missing, no. (%)	706 (3%)	260 (2%)	3428 (4%)	2215 (4%)	12 008 (12%)
Self-reported use missing, no. (%)	1871 (7%)	550 (4%)	29 160 (33%)	4959 (10%)	6336 (6%)
Operator data missing or incomplete, no. (%)	20 342 (79%)	3090 (24%)	52 839 (60%)	18 623 (37%)	23 479 (24%)
Complete data, no. (%)	2993 (11%)	9162 (70%)	3039 (3%)	24 881 (49%)	56 862 (58%)
Age in years, mean (SD)	46.9 (12.9)	49.3 (14.0)	45.0 (12.3)	43.3 (13.4)	45.1 (14.7)
Sex, no. (%)					
Male	1350 (45%)	3808 (42%)	280 (9%)	11 352 (46%)	26 044 (45.8%)
Female	1643 (55%)	5354 (58%)	2759 (91%)	13 529 (54%)	30 818 (54.2%)
Marital status, no. (%)					
Living together	2004 (67%)	6386 (70%)	2282 (75%)	16 784 (67%)	37 304 (65.6%)
Not living together	321 (11%)	860 (9%)	131 (4%)	2768 (11%)	7445 (13.1%)
Not in a relationship	668 (22%)	1916 (21%)	626 (21%)	5329 (21%)	12 113 (21.3%)
Employment status, no. (%)					
Active	2212 (74%)	4721 (52%)	2562 (84%)	18 033 (72%)	39 854 (70.1%)
Unemployed/inactive[b]	781 (26%)	4441 (48%)	477 (16%)	6848 (28%)	17 008 (29.9%)
Educational level, no. (%)					
Elementary school[c]	423 (14%)	4014 (44%)	165 (5%)	3331 (13%)	6166 (10.8%)
At least secondary school	2570 (86%)	5148 (56%)	2874 (95%)	21 550 (87%)	50 696 (89.2%)
Self-reported call duration, no. (%)					
<5 minutes/week	132 (4%)	101 (1%)	378 (12%)	1021 (4%)	2471 (4.3%)
5-29 min/week	1086 (36%)	1634 (18%)	1555 (51%)	7023 (28%)	15 651 (27.5%)
30-59 minutes/week	661 (22%)	2175 (24%)	642 (21%)	4894 (20%)	12 754 (22.4%)
1-3 hours/week	676 (23%)	3573 (39%)	359 (12%)	6647 (27%)	15 610 (27.5%)
4-6 hours/week	268 (9%)	1181 (13%)	75 (2%)	3075 (12%)	6030 (10.6%)
>6 hours/week	170 (6%)	498 (5%)	30 (1%)	2221 (9%)	4346 (7.6%)
Recorded call duration in minutes per week, GM (GSD)	60.6 (3.3)	81.5 (3.0)	23.4 (2.6)	78.3 (3.9)	46.5 (3.9)

Abbreviations: COSMOS, Cohort Study of Mobile Phone Use and Health; GM, geometric mean; GSD, geometric standard deviation.
[a]In Sweden 55 471 participants filled in the baseline questionnaire, but only 50 678 participants provided consent for health register data and were therefore included in this study.
[b]Includes retired workers, students, disabled persons, and homemakers.
[c]In Finland the elementary school category also includes participants with lower-level vocational education and nonvocational high school.

model for the inverse RC method tended to decrease with age, but increased with increasing operator-recorded durations (Figure S2).

The relationship between observed and predicted operator-recorded data was close to linear for the direct RC model in most countries, but predictions in the high operator-recorded range tended to either underestimate (Finland, Sweden) or overestimate (Denmark, the Netherland, the United Kingdom) average observed values for the inverse RC model (Figure 3). Predictions from the GAMLSS-based RC model tended to be higher than observed in the high operator-recorded range for all countries (Figure S3). Differences in (in-sample) R^2 between models were smaller than differences between countries, but were consistently larger for the direct than the inverse RC method. R^2 was largest for RC models fitted to the Danish data (range 36.1%-38.7%) and lowest for models fitted to the Finnish data (range 23.6%-25.7%). Using any of the more complex RC methods resulted in slightly higher R^2s (increase < 2.6%) than the simple RC method (Table S7). Country-specific regression-calibrated estimates for the simple RC model based on outgoing call duration, and incoming and outgoing call duration combined, by category of self-report, are presented in Table S8.

Performance of RC models on the simulated data

Simulation results are presented in Table 2. This table shows the (range of) estimated coefficients across simulations, as well as the squared bias and variance that contribute to total MSE. Using any of the RC approaches resulted in MSEs that were approximately 50% lower than those from CC analyses. MI was inferior to all RC approaches on average. Both simple RC and inverse RC performed relatively well on data from Finland, Sweden, and the United Kingdom. MSEs were lower when a larger training set was used (50% of the sample; Table S3), especially for the inverse RC approach that performed well in almost all countries in both sets of simulations. Bias was a minor contributing factor to total MSE when the regression coefficient for operator-recorded data used in the simulations was relatively small (0.005) (Table S2). Point estimates obtained using GAMLSS-based RC tended to suffer from relatively large downward bias.

Table 3 shows coverage of Wald-type 95% CIs based on estimated SEs after inverse-precision weighting. Coverage was per expectation (95%) for the simple RC and direct RC methods, tended to be slightly lower for the inverse RC method (~93%), and was much lower for the GAMLSS-based RC in some countries (eg, 66% for Sweden). The average ratio of estimated slope (β) over SE was highest for the direct RC or inverse RC methods for all countries, confirming higher statistical efficiency of these methods over alternative approaches.

Discussion

We evaluated the performance of different approaches that combine operator-recorded mobile phone use and its more widely available, but error-prone proxy, self-reported use, in a simulation study using "real world" COSMOS mobile phone use and covariate data with a simulated binary health outcome. We show that RC models, especially the simple and direct RC methods, provide a

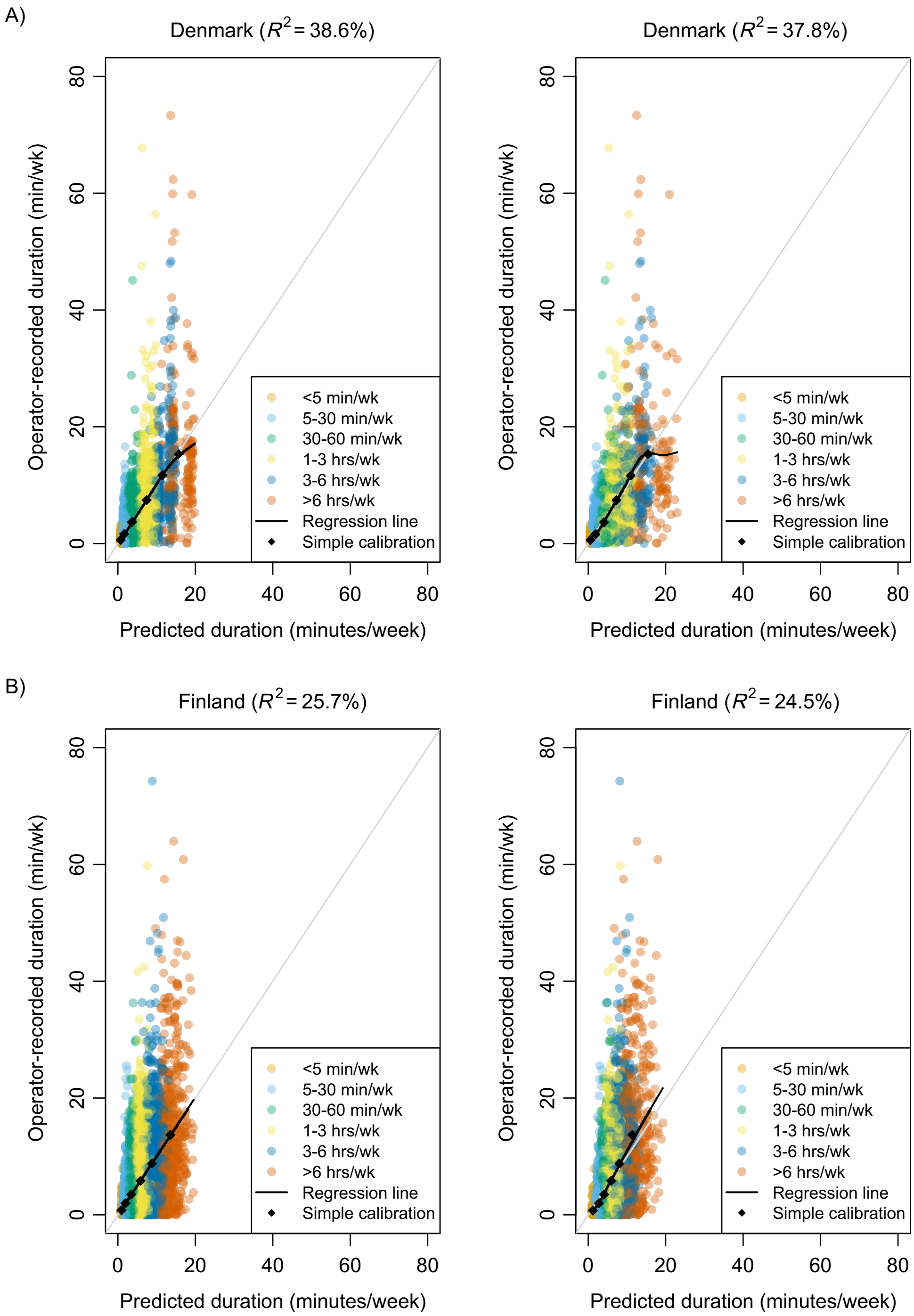

Figure 3. Continues.

good balance between minimizing bias and maximizing precision, and we therefore plan to use these methods in further analyses of the COSMOS data.

The results indicate that bias was a minor contributing factor to MSE of estimated slopes, when health outcomes were simulated under the assumption that the odds of disease increased by 2% for each additional 20 minutes of mobile phone call-time per week. Bias became more prominent when outcomes were simulated assuming a stronger exposure-outcome relationship. The direction of bias (upward or downward) varied by country and RC approach. RC is expected to produce unbiased point estimates only for linear models,[32] but our simulations indicate that bias was still limited for logistic regression models when the effect of (error-prone) reported use was not very strong or measurement error was relatively low by comparison,[22,33,34] even with heteroscedastic error.

Total MSE of estimated slope coefficients was lowest for the full-data analyses and highest for the CC and MI methods, with relatively minor differences between different RC approaches. Among RC approaches, bias tended to be largest for the GAMLSS-based method. The simple RC approach performed only slightly worse than approaches relying on more complex models. This may be explained by the fact that covariates did not explain much of the variation in operator-recorded data once self-reported data

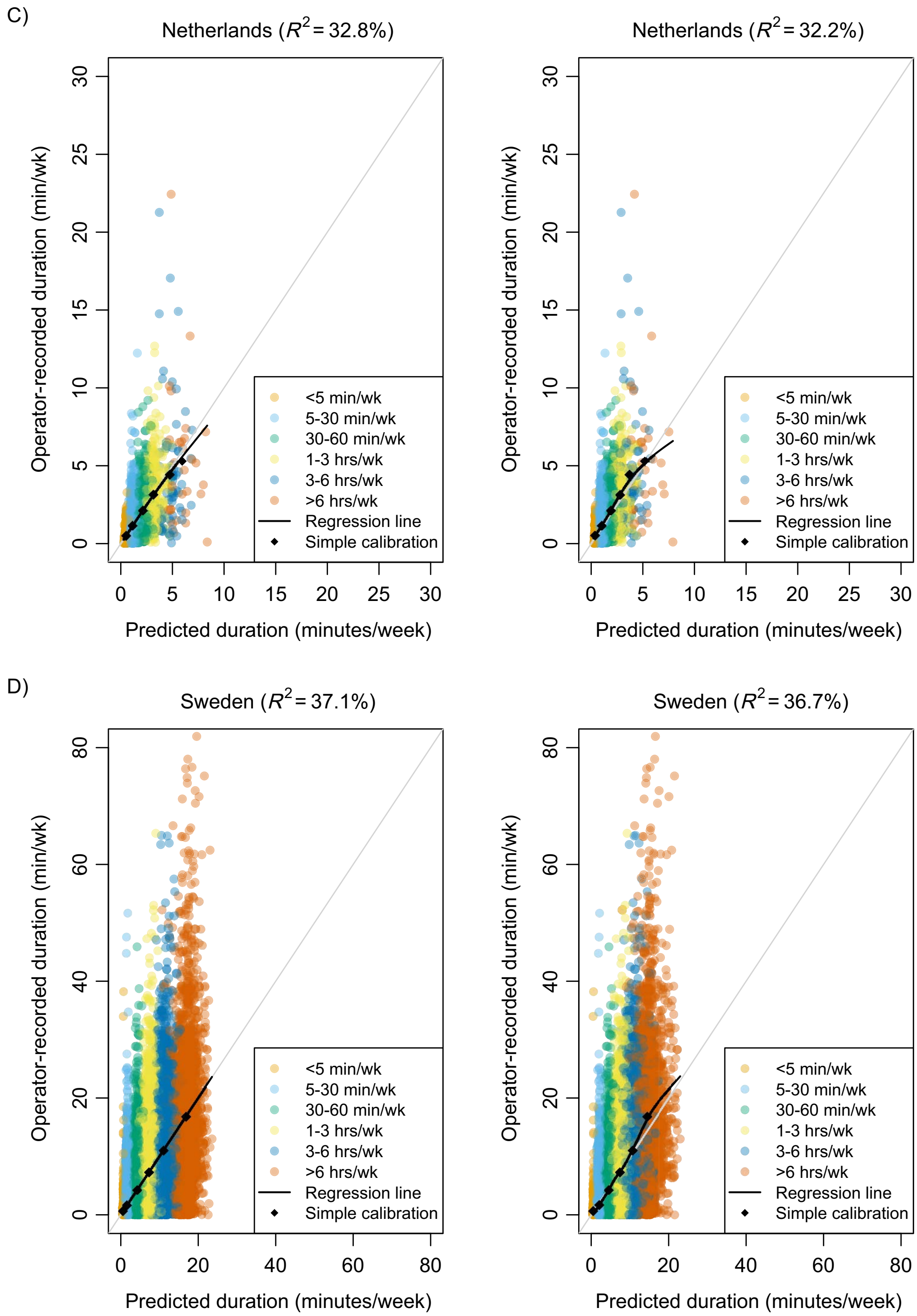

Figure 3. Continues.

was considered, as suggested by the modest increase in average R^2 for more complex models when compared with simple RC. An advantage of using simple RC is that no additional covariates need to be available, while any gain from using more covariates would be (partly) offset by the need to consider these covariates as potential confounders in the health-outcome model.[35,36] Also, requiring data on covariates to include in the model reduces the available sample size because of missing covariate data. We did not investigate this trade-off in this study, where health outcomes were simulated using only operator-recorded data, but this could be of practical relevance when some of the model predictors are also known risk factors for the health outcome. Our results were based on simulations that assumed a linear relationship between operator-recorded data and the log-odds of our simulated health outcome. This may be a reasonable assumption in many cases, but note that more complicated calibration models are required if the goal is to use the calibrated exposure estimate in a nonparametric or polynomial regression model.[37]

RC models were fitted using 2-stage regression, and SEs from the second-stage model do not fully account for the uncertainty in exposure estimates from the first-stage models. SEs could be estimated through (asymptotic) approximation or bootstrapping,[22,32] but for practical reasons we used model-based SEs for the simulations in our analyses. Coverage of estimated 95% CIs

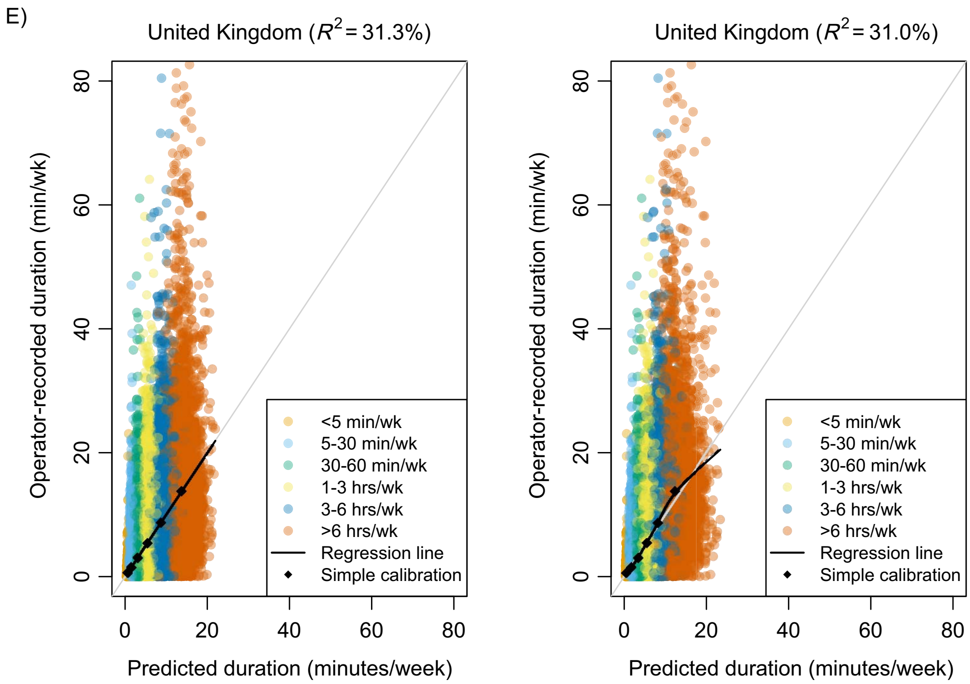

Figure 3. Observed versus predicted duration of outgoing calls according to country for the direct (left column) and inverse (right column) regression calibration (RC) approaches, using data from the Cohort Study of Mobile Phone Use and Health (COSMOS), multiple countries, 2007-2012. A) Denmark; B) Finland; C) The Netherlands; D) Sweden; E) United Kingdom. The regression line was estimated allowing for a nonlinear relationship between observed and predicted values and allowing the (residual) variance in observed durations to depend on predicted duration using penalized splines (P-splines) as implemented in the generalized additive model for location, shape, and scale software.[29] Note the different horizontal and vertical scales for the Netherlands.

was only slightly lower than expectation for most approaches, except for the GAMLSS-based RC, where bias in point estimates was substantial.

Simulated health outcomes were included in the MI models, but we did not consider fitting a full measurement error model that incorporates the health outcome regression.[38] Feedback between the health outcome model and the exposure and measurement error models could result in a more efficient analysis of the exposure-outcome relationship, but it could also result in bias amplification from model misspecification.[39] A full measurement error model would also require re-fitting the model separately for each health outcome or including all health outcomes in a single model, which is unattractive for a multicenter cohort study.

Other authors have suggested methods either to account for measurement error in self-reported mobile phone data or to assess the potential impact on study findings, but these too have their limitations. Redmayne et al[16] used a Bayesian forecasting method to correct for measurement error in the self-reported number of text messages. Their method shows similarities to the inverse RC approach applied here, but their model did not include any covariates and self-reported use was on a continuous scale. Vergnaud et al[40] imputed missing information on Terrestrial Truncated Radio (TETRA) use among police forces using machine learning techniques. They did not include health outcomes in their imputation model and primarily reported results based on personal exposure estimates averaged across 10 000 imputations, which made it similar to our GAMLSS-based RC method. They found, as we did, that this model tended to under-estimate use for higher exposures.

Complete information on covariates, self-reported phone use, and operator-recorded data was available for a third of participants (n = 96 937; 35.0%), and was collected prior to disease occurrence. This addresses a major limitation of earlier case–control studies, which relied on retrospective information on mobile phone use, although missing information could be associated with some of the health outcomes of interest. Operator data can also be problematic as they are collected by providers for billing or security reasons rather than scientific purposes. In some countries, calls made between a participant and another caller both using the same provider were registered as outgoing calls only, so some incoming calls may be missing. We therefore based our simulation study on outgoing call duration, even though we may choose to use calibrated call-duration estimates based on combined incoming and outgoing call duration data in COSMOS.

Collection of operator-recorded data within COSMOS will likely become more difficult over time due to attrition and participants changing network operators. Self-reported information on mobile phone use will therefore remain important,[10] but it may require the use of structural measurement error models to account for time-trends during follow-up, ever-changing patterns in mobile phone use, and newer network technologies. Notably, there are limitations inherent in capturing evolving mobile phone technology changes over time that may have an impact on the estimation of exposure-response relationships, which cannot be addressed with our approach.

A major issue is how well mobile phone use predicts the exposure of interest, namely radiofrequency electromagnetic fields. While past validation studies have been carried out for the second mobile phone generation, showing fair agreement between amount of use and cumulated emission from the handset,[4] fewer data are available on the predictive power of mobile phones of the third, fourth, and fifth generations that have been and are being used by COSMOS participants.

Table 2. Point estimates, squared bias, variance and mean squared error for approaches based on the results from 1000 simulations[a] (100 simulations for the inverse regression calibration approach), using data from the COSMOS Study, multiple countries, 2007-2012.

Country	Model	β (×1000) (95% CI)[b]	Percentiles[c]: 2.5, 97.5	Bias^2 (95% CI)	Variance (95% CI)	MSE[d] (95% CI)
Denmark	Full data	1.02 (0.99-1.05)	0.34, 1.72	0.00 (0.00-0.00)	0.18 (0.17-0.20)	0.18 (0.17-0.20)
	Simple RC	0.98 (0.94-1.01)	0.05, 1.93	0.00 (0.00-0.00)	0.34 (0.31-0.37)	0.34 (0.31-0.37)
	GAMLSS RC	0.85 (0.82-0.88)	0.05, 1.70	0.02 (0.01-0.03)	0.26 (0.24-0.29)	0.28 (0.26-0.31)
	Direct RC	0.96 (0.92-0.99)	0.09, 1.91	0.00 (0.00-0.01)	0.31 (0.29-0.34)	0.31 (0.29-0.34)
	Inverse RC	0.90 (0.77-1.01)	−0.16, 1.77	0.01 (0.00-0.05)	0.35 (0.28-0.46)	0.36 (0.29-0.48)
	CC	1.02 (0.97-1.08)	−0.35, 2.46	0.00 (0.00-0.00)	0.75 (0.69-0.82)	0.75 (0.69-0.82)
	MI	0.95 (0.91-1.00)	−0.14, 2.10	0.00 (0.00-0.01)	0.48 (0.44-0.52)	0.48 (0.44-0.53)
Finland	Full data	1.00 (0.99-1.02)	0.55, 1.45	0.00 (0.00-0.00)	0.08 (0.07-0.08)	0.08 (0.07-0.08)
	Simple RC	1.01 (0.98-1.04)	0.33, 1.70	0.00 (0.00-0.00)	0.18 (0.16-0.19)	0.18 (0.16-0.19)
	GAMLSS RC	0.79 (0.76-0.81)	0.24, 1.37	0.05 (0.04-0.06)	0.11 (0.11-0.12)	0.16 (0.15-0.17)
	Direct RC	1.01 (0.98-1.03)	0.34, 1.70	0.00 (0.00-0.00)	0.17 (0.16-0.19)	0.17 (0.16-0.19)
	Inverse RC	1.13 (1.02-1.22)	0.40, 1.94	0.02 (0.00-0.05)	0.25 (0.19-0.32)	0.26 (0.20-0.35)
	CC	1.02 (0.98-1.05)	0.13, 1.95	0.00 (0.00-0.00)	0.31 (0.28-0.34)	0.31 (0.28-0.34)
	MI	1.05 (1.01-1.08)	0.16, 1.94	0.00 (0.00-0.01)	0.29 (0.27-0.32)	0.30 (0.27-0.33)
Netherlands	Full data	1.06 (0.96-1.16)	−1.48, 3.83	0.00 (0.00-0.03)	2.63 (2.44-2.87)	2.63 (2.43-2.88)
	Simple RC	1.06 (0.91-1.21)	−2.70, 5.05	0.00 (0.00-0.03)	5.62 (5.13-6.20)	5.62 (5.14-6.21)
	GAMLSS RC	0.97 (0.85-1.12)	−2.54, 4.64	0.00 (0.00-0.01)	4.83 (4.43-5.27)	4.83 (4.43-5.27)
	Direct RC	0.81 (0.70-0.94)	−2.13, 3.93	0.03 (0.00-0.09)	3.58 (3.26-3.91)	3.62 (3.28-3.95)
	Inverse RC	1.19 (0.77-1.62)	−3.24, 4.87	0.03 (0.00-0.37)	5.16 (3.75-7.89)	5.19 (3.72-7.59)
	CC	1.19 (0.99-1.39)	−4.02, 6.49	0.04 (0.00-0.15)	10.58 (9.59-11.56)	10.61 (9.62-11.62)
	MI	1.17 (0.99-1.36)	−3.96, 5.89	0.03 (0.00-0.13)	8.69 (8.02-9.55)	8.73 (8.06-9.59)
Sweden	Full data	1.00 (1.00-1.01)	0.78, 1.21	0.00 (0.00-0.00)	0.02 (0.02-0.02)	0.02 (0.02-0.02)
	Simple RC	1.00 (0.98-1.01)	0.72, 1.28	0.00 (0.00-0.00)	0.03 (0.03-0.03)	0.03 (0.03-0.03)
	GAMLSS RC	0.78 (0.77-0.79)	0.56, 1.02	0.05 (0.04-0.05)	0.02 (0.02-0.02)	0.07 (0.06-0.07)
	Direct RC	0.99 (0.98-1.00)	0.72, 1.28	0.00 (0.00-0.00)	0.03 (0.03-0.03)	0.03 (0.03-0.03)
	Inverse RC	1.04 (1.01-1.08)	0.80, 1.28	0.00 (0.00-0.01)	0.03 (0.02-0.04)	0.03 (0.02-0.04)
	CC	1.00 (0.99-1.02)	0.57, 1.46	0.00 (0.00-0.00)	0.07 (0.06-0.08)	0.07 (0.06-0.08)
	MI	1.05 (1.03-1.06)	0.62, 1.47	0.00 (0.00-0.00)	0.07 (0.06-0.07)	0.07 (0.06-0.08)
United Kingdom	Full data	1.00 (0.99-1.00)	0.81, 1.18	0.00 (0.00-0.00)	0.01 (0.01-0.01)	0.01 (0.01-0.01)
	Simple RC	0.99 (0.98-1.00)	0.74, 1.24	0.00 (0.00-0.00)	0.02 (0.02-0.03)	0.02 (0.02-0.03)
	GAMLSS RC	0.92 (0.91-0.93)	0.68, 1.15	0.01 (0.01-0.01)	0.02 (0.02-0.02)	0.03 (0.03-0.03)
	Direct RC	0.98 (0.97-0.99)	0.72, 1.23	0.00 (0.00-0.00)	0.02 (0.02-0.03)	0.02 (0.02-0.03)
	Inverse RC	1.05 (1.00-1.09)	0.68, 1.42	0.00 (0.00-0.01)	0.05 (0.04-0.06)	0.05 (0.04-0.07)
	CC	0.99 (0.98-1.01)	0.61, 1.35	0.00 (0.00-0.00)	0.05 (0.05-0.06)	0.05 (0.05-0.06)
	MI	1.00 (0.97-1.03)	0.74, 1.26	0.00 (0.00-0.00)	0.03 (0.02-0.03)	0.03 (0.02-0.03)

Abbreviations: CC, complete case; COSMOS, Cohort Study of Mobile Phone Use and Health; GAMLSS, generalized additive model for location, shape, and scale; MI, multiple imputation; MSE, mean squared error; RC, regression calibration.
[a]Simulations used data from 25% of the participants with complete information on both self-reported and operator-recorded data as the training set and the remainder as test set. The outcome was simulated using a slope coefficient (β) of 0.001 (ie, assuming an odds ratio of exp(0.001×30) = 1.03 for each additional 30 minutes call-time per week) and with a balanced ratio of cases to noncases. Bootstrapping was used to estimate 95% CIs for each statistic. The training set was used to fit the first-stage (exposure) models for the RC approaches, the health model for the CC analysis, and the multiple imputation model. All second-stage (health) models for the RC approaches were fitted to the test data only and results were precision-weighted with those from the CC approach before further analyses. The (non–full data) model that achieved lowest MSE and all models for which the estimated MSE fell within the 95% CI for that lowest MSE are highlighted in gray.
[b]95% CIs were estimated using bootstrapping and reflect Monte-Carlo error
[c](Empirical) quantiles of the distribution of β estimates across simulation
[d]MSE = (bias^2 + variance).

Conclusion

This study addressed an important concern in mobile phone research and more generally in environmental epidemiology: how to leverage self-reported exposure estimates that are often available but error-prone, with more objective measurements that may be obtained in only a subset of participants. Our simulation study indicated that RC approaches may improve estimation of exposure-outcome relationships between mobile phone use and health outcomes within COSMOS. The prospective design and improved exposure assessment within COSMOS compared with that in previous case–control studies are expected to lead to more robust conclusions about possible health effects from use of mobile phones.

Acknowledgments

We thank all the participants who have joined the COSMOS cohort study (http://www.thecosmosproject.org/). We thank mobile phone network operators in Denmark, Finland, France, the Netherlands, Sweden, and the United Kingdom for allowing invitation of their subscribers and/or provision of operator traffic data. COSMOS Study Group members not in the author list: Terhi Lampio, Susanna Lankinen, Turkka Näppilä, and Taru Vehmasto (Tampere University, Faculty of Social Sciences, Tampere, Finland); Inka Pieterson (Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, the Netherlands); David Muller, Margaret Douglass, and James Brook (Department of Epidemiology and Biostatistics, School of Public Health, Imperial

Table 3. Coverage of 95% CI and ratio of slope estimate (β) over its standard error as a proxy for efficiency modeling approaches based on the results from 1000 simulations[a] (100 simulations for the inverse regression calibration approach), using data from the COSMOS Study, multiple countries, 2007-2012.

Country	Model	Coverage, % (95% CI)[b]	Β/SE (95% CI)	Percentiles[c]: 2.5, 97.5
Denmark	Full data	95 (94-96)	2.37 (2.31-2.43)	0.82, 3.91
	Simple RC	96 (94-97)	1.67 (1.61-1.72)	0.08, 3.23
	GAMLSS RC	94 (92-95)	1.66 (1.60-1.71)	0.10, 3.23
	Direct RC	96 (94-97)	1.69 (1.63-1.75)	0.15, 3.24
	Inverse RC	93 (84-96)	1.62 (1.40-1.81)	−0.31, 3.07
	CC	95 (94-97)	1.16 (1.10-1.22)	−0.41, 2.68
	MI	97 (95-97)	1.28 (1.23-1.34)	−0.19, 2.77
Finland	Full data	95 (94-96)	3.58 (3.52-3.64)	2.02, 5.06
	Simple RC	96 (94-97)	2.38 (2.32-2.44)	0.81, 3.92
	GAMLSS RC	89 (87-90)	2.33 (2.27-2.39)	0.75, 3.92
	Direct RC	96 (95-97)	2.42 (2.36-2.48)	0.80, 3.95
	Inverse RC	92 (83-95)	2.51 (2.30-2.71)	0.88, 4.23
	CC	96 (94-97)	1.78 (1.72-1.84)	0.25, 3.25
	MI	94 (92-95)	1.90 (1.84-1.96)	0.31, 3.58
Netherlands	Full data	96 (95-97)	0.64 (0.58-0.70)	−0.90, 2.29
	Simple RC	95 (94-96)	0.44 (0.38-0.51)	−1.15, 2.11
	GAMLSS RC	95 (93-96)	0.44 (0.38-0.51)	−1.19, 2.05
	Direct RC	94 (92-95)	0.44 (0.38-0.51)	−1.20, 2.12
	Inverse RC	94 (85-97)	0.53 (0.35-0.71)	−1.43, 2.06
	CC	97 (95-98)	0.35 (0.29-0.40)	−1.22, 1.86
	MI	97 (96-98)	0.37 (0.31-0.42)	−1.20, 1.85
Sweden	Full data	95 (94-96)	7.70 (7.64-7.76)	6.04, 9.22
	Simple RC	96 (95-97)	5.57 (5.50-5.62)	4.02, 7.14
	GAMLSS RC	66 (63-69)	5.50 (5.45-5.57)	3.97, 7.04
	Direct RC	96 (95-97)	5.62 (5.56-5.68)	4.08, 7.20
	Inverse RC	96 (89-98)	5.53 (5.38-5.69)	4.30, 6.85
	CC	95 (93-96)	3.83 (3.78-3.89)	2.22, 5.46
	MI	94 (92-95)	4.04 (3.96-4.12)	2.26, 6.25
United Kingdom	Full data	95 (93-96)	9.10 (9.05-9.16)	7.52, 10.64
	Simple RC	96 (94-97)	6.35 (6.29-6.41)	4.76, 7.89
	GAMLSS RC	90 (88-91)	6.40 (6.34-6.46)	4.76, 7.98
	Direct RC	95 (93-96)	6.40 (6.34-6.46)	4.78, 7.98
	Inverse RC	94 (85-97)	5.14 (4.94-5.36)	3.32, 6.89
	CC	95 (93-96)	4.52 (4.45-4.57)	2.90, 6.03
	MI	96 (87-98)	6.28 (6.08-6.46)	4.74, 7.80

Abbreviations: CC, complete case; COSMOS, Cohort Study of Mobile Phone Use and Health; GAMLSS, generalized additive model for location, shape, and scale; MI, multiple imputation; RC, regression calibration; SE, standard error.
[a]Simulations used data from 25% of the participants with complete information on both self-reported and operator-recorded data as the training set and the remainder as test set. The outcome was simulated using a slope coefficient (β) of 0.001 (ie, assuming an odds ratio of $\exp(0.001 \times 30) = 1.03$ for each additional 30 minutes call-time per week) and with a balanced ratio of cases to noncases. Bootstrapping was used to estimate 95% CIs for each statistic.
[b]95% CIs were estimated using bootstrapping and reflect Monte-Carlo error.
[c](Empirical) quantiles of the distribution of β/SE ratios across simulations.

College London, United Kingdom); and Mats Talbäck (Unit of Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden). We thank Anders Ahlbom (Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden) for his essential role in commencing the COSMOS cohort collaboration.

Supplementary material

Supplementary material is available at the *American Journal of Epidemiology* online.

Funding

P.E. is Director of the MRC Centre for Environment and Health, supported by the Medical Research Council (MR/S019669/1). P.E. acknowledges funding from the NIHR Imperial Biomedical Research Centre, the NIHR Health Protection Research Unit in Chemical and Radiation Threats and Hazards (NIHR-200922). and the UK Dementia Research Institute supported by UK DRI Ltd, which is funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK (UKDRI-5001). P.E. is associate director of Health Data Research UK–London, which receives funding from a consortium led by the UK Medical Research Council. M.B.T.'s Chair and R.B.S.'s fellowship are supported by a donation from Marit Mohn to Imperial College London to support Population Child Health through the Mohn Centre for Children's Health and Wellbeing. M.R., L.P., R.V., and H.K. were supported by the Netherlands Organization for Health Research (ZonMW) within the programme Electromagnetic Fields and Health Research, under grant numbers 85200001, 85500003, and 85800001.

Conflict of interest

M.F. was vice chairman (2012-2020) of the International Commission on Non-Ionizing Radiation Protection, an independent body setting guidelines for non-ionizing radiation protection. She has served as advisor to a number of national and international public advisory and research steering groups concerning the potential health effects of exposure to non-ionizing radiation, currently for the World Health Organization (WHO). H.K. was the chair of the Committee on Electromagnetic Fields of the Health Council of The Netherlands till 2022. He currently is a member of the WHO Task Group for the Environmental Health Criteria Monograph on RF-EMF. A.H. is a member of the International Commission on Non-Ionizing Radiation Protection since 2020, and of the Committee on Electromagnetic Fields of the Health Council of The Netherlands, and chairs the Swedish Radiation Safety Authority's (SSM) Scientific Council on Electromagnetic Fields since 2020. A.A. currently is a member of the WHO Task Group for the Environmental Health Criteria Monograph on RF-EMF. M.B.T. is currently a member of the WHO groups tasked with systematic review of evidence on non-ionizing radiation and health, feeding into the Environmental Health Criteria Monograph on RF-EMF. All other authors declare they have no competing financial interests.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Data availability

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

References

1. Carroll RJ, Ruppert D, Stefanski LA, et al. Measurement error in nonlinear models: a modern perspective. *Chapman & Hall/CRC*. 2006;39(2):231-232. https://doi.org/10.1080/00401706.1997.10485096
2. Freedman LS, Commins JM, Willett W, et al. Practice of epidemiology evaluation of the 24-hour recall as a reference instrument for calibrating other self-report instruments in nutritional cohort studies: evidence from the Validation Studies Pooling Project. *Am J Epidemiol*. 2018;186(1):73-82. https://doi.org/10.1093/aje/kwx039
3. Spiegelman D. Approaches to uncertainty in exposure assessment in environmental epidemiology. *Annu Rev Public Health*. 2010;31(1):149-163. https://doi.org/10.1146/annurev.publhealth.012809.103720
4. Vrijheid M, Cardis E, Armstrong BK, et al. Validation of short term recall of mobile phone use for the Interphone study. *Occup Environ Med*. 2006;63(4):237-243. https://doi.org/10.1136/oem.2004.019281
5. Vrijheid M, Armstrong BK, Bédard D, et al. Recall bias in the assessment of exposure to mobile phones. *J Expo Sci Environ Epidemiol*. 2009;19(4):369-381. https://doi.org/10.1038/jes.2008.27
6. Aydin D, Feychting M, Schüz J, et al. Impact of random and systematic recall errors and selection bias in case-control studies on mobile phone use and brain tumors in adolescents (CEFALO study). *Bioelectromagnetics*. 2011;32(5):396-407. https://doi.org/10.1002/bem.20651
7. Samkange-Zeeb F, Berg G, Blettner M. Validation of self-reported cellular phone use. *J Expo Anal Environ Epidemiol*. 2004;14(3):245-248. https://doi.org/10.1038/sj.jea.7500321
8. Berg G, Schüz J, Samkange-Zeeb F, et al. Assessment of radiofrequency exposure from cellular telephone daily use in an epidemiological study: German validation study of the international case-control study of cancers of the brain—INTERPHONE-study. *J Expo Anal Environ Epidemiol*. 2005;15:217-224. https://doi.org/10.1038/sj.jea.7500390
9. Heinävaara S, Tokola K, Kurttio P, et al. Validation of exposure assessment and assessment of recruitment methods for a prospective cohort study of mobile phone users (COSMOS) in Finland: a pilot study. *Environ Health*. 2011;10(1):14. https://doi.org/10.1186/1476-069X-10-14
10. Toledano MB, Auvinen A, Tettamanti G, et al. An international prospective Cohort Study of Mobile Phone Use and Health (COSMOS): factors affecting validity of self-reported mobile phone use. *Int J Hyg Environ Health Published Online First*. 2017;221(1):1-8. https://doi.org/doi:10.1016/j.ijheh.2017.09.008
11. Agogo GO, Van Der Voet H, van't Veer P, et al. Use of two-part regression calibration model to correct for measurement error in episodically consumed foods in a single-replicate study design: EPIC case study. *PloS One*. 2014;9(11):9. https://doi.org/10.1371/journal.pone.0113160
12. Freedman LS, Schatzkin A, Midthune D, et al. Dealing with dietary measurement error in nutritional cohort studies. *JNCI: J Natl Cancer Inst*. 2011;103(14):1086-1092. https://doi.org/10.1093/jnci/djr189
13. Agogo GO, van der Voet H, van't Veer P, et al. Evaluation of a two-part regression calibration to adjust for dietary exposure measurement error in the cox proportional hazards model: a simulation study. *Biom J*. 2016;58(4):766-782. https://doi.org/10.1002/bimj.201500009
14. Van RS, Li R, Hoek G, et al. Traffic-related outdoor air pollution and respiratory symptoms in children: the impact of adjustment for exposure measurement error. *Epidemiology*. 2008;19(3):409-416. https://doi.org/10.1097/EDE.0b013e3181673bab
15. Fraser GE, Stram DO. Regression calibration in studies with correlated variables measured with error. *Am J Epidemiol*. 2001;154(9):836-844. https://doi.org/10.1093/aje/154.9.836
16. Redmayne M, Smith E, Abramson MJ. A forecasting method to reduce estimation bias in self-reported cell phone data. *J Expo Sci Environ Epidemiol*. 2013;23(5):539-544. https://doi.org/10.1038/jes.2012.70
17. Petit C, Blangiardo M, Richardson S, et al. Association of environmental insecticide exposure and fetal growth with a Bayesian model including multiple exposure sources: the PELAGIE mother-child cohort. *Am J Epidemiol*. 2012;175(11):1182-1190. https://doi.org/10.1093/aje/kwr422
18. Bateson TF, Wright JM. Regression calibration for classical exposure measurement error in environmental epidemiology studies using multiple local surrogate exposures. *Am J Epidemiol*. 2010;172(3):344-352. https://doi.org/10.1093/aje/kwq123
19. Momoli F, Siemiatycki J, McBride ML, et al. Probabilistic multiple-bias modeling applied to the Canadian data from the Interphone study of mobile phone use and risk of glioma, meningioma, acoustic neuroma, and parotid gland tumors. *Am J Epidemiol*. 2017;186(7):885-893. https://doi.org/10.1093/aje/kwx157
20. Tokola K, Kurttio P, Salminen T, et al. Reducing overestimation in reported mobile phone use associated with epidemiological studies. *Bioelectromagnetics*. 2008;29(7):559-563. https://doi.org/10.1002/bem.20424

21. Schüz J, Elliott P, Auvinen A, et al. An international prospective Cohort Study of Mobile Phone Use and Health (COSMOS): design considerations and enrolment. *Cancer Epidemiol*. 2011;35(1):37-43. https://doi.org/10.1016/j.canep.2010.08.001
22. Spiegelman D, Logan R, Grove D. Regression calibration with heteroscedastic error variance. *Int J Biostat*. 2011;7(1):4. https://doi.org/10.2202/1557-4679.1259
23. van Buuren S, Groothuis-Oudshoorn K, Vink G, et al. mice: multivariate imputation by chained equations. 2023. Accessed Accepted April 3, 2024. https://cran.r-project.org/package=mice
24. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1-67. http://www.jstatsoft.org/v45/i03
25. Toledano MB, Smith RB, Chang I, et al. Cohort profile: UK COSMOS—a UK cohort for study of environment and health. *Int J Epidemiol*. 2017;46(3):775-787. https://doi.org/10.1093/ije/dyv203
26. Reedijk M, Lenters V, Slottje P, et al. Cohort profile: LIFEWORK, a prospective cohort study on occupational and environmental risk factors and health in the Netherlands. *Published Online First*. 2018;8(2):e018504. https://doi.org/10.1136/bmjopen-2017-018504
27. Rigby RA, Stasinopoulos DM, Lane PW. Generalized additive models for location, scale and shape. *J R Stat Soc Ser C Appl Stat*. 2005;54(3):507-554. https://doi.org/10.1111/j.1467-9876.2005.00510.x
28. Plummer M. A program for analysis of Bayesian graphical models using Gibbs sampling. DSC 2003 Working Papers. 2003. Accessed June 26, 2018. https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf
29. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr*. 1997;65(4):1179S-1186S. https://doi.org/10.1093/ajcn/65.4.1179S
30. Blackwell M, Honaker J, King G. A unified approach to measurement error and missing data: overview and applications. *Sociol Methods Res*. 2017;46(3):303-341. https://doi.org/10.1177/0049124115585360
31. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley & Sons, Inc; 1987. https://doi.org/10.1002/9780470316696
32. Spiegelman D, Carroll RJ, Kipnis V. Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. *Stat Med*. 2001;20(1):139-160. https://doi.org/10.1002/1097-0258(20010115)20:1<139::aid-sim644>3.0.co;2-k
33. Wang CY, Hsu L, Feng ZD, et al. Regression calibration in failure time regression. *Biometrics*. 1997;53(1):131-145.
34. Fraser GE, Stram DO. Regression calibration when foods (measured with error) are the variables of interest: markedly non-gaussian data with many zeroes. *Am J Epidemiol*. 2012;175(4):325-331. https://doi.org/10.1093/aje/kwr316
35. Liao X, Spiegelman D, Carroll RJ. Regression calibration is valid when properly applied. *Epidemiology*. 2013;24(3):466-467. https://doi.org/10.1097/EDE.0b013e31828b284b
36. Szpiro AA, Paciorek CJ. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environ*. 2013;24(8):501-517. https://doi.org/10.1002/env.2233
37. Carroll RJ, Maca JD, Ruppert D. Nonparametric regression in the presence of measurement error. *Biometrika*. 1999;86(3):541-554. https://doi.org/10.1093/biomet/86.3.541
38. Bartlett JW, Keogh RH. Bayesian correction for covariate measurement error: a frequentist evaluation and comparison with regression calibration. *Stat Methods Med Res*. 2018;27(6):1695-1708. https://doi.org/10.1177/0962280216667764
39. Szpiro AA, Sheppard L, Lumley T. Efficient measurement error correction with spatially misaligned data. *Biostatistics*. 2011;12(4):610-623. https://doi.org/10.1093/biostatistics/kxq083
40. Vergnaud AC, Aresu M, Kongsgård HW, et al. Estimation of TETRA radio use in the airwave health monitoring study of the British police forces. *Environ Res*. 2018;167:169-174. https://doi.org/10.1016/j.envres.2018.07.015