Journal of Clinical Epidemiology 140 (2021) 79–92

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Meta-analysis for individual participant data with a continuous exposure: A case study

Darsy Darssan[a], Gita D. Mishra[a], Darren C. Greenwood[b], Sven Sandin[c,d,e], Eric J. Brunner[f], Sybil L. Crawford[g], Samar R. El Khoudary[h], Maria Mori Brooks[h], Ellen B. Gold[i], Mette Kildevæld Simonsen[j], Hsin-Fang Chung[a], Elisabete Weiderpass[k], Annette J. Dobson[a,*]

[a] *University of Queensland, School of Public Health, Faculty of Medicine, Queensland, Australia*
[b] *School of Medicine, University of Leeds, Leeds, UK*
[c] *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*
[d] *Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, USA*
[e] *Seaver Autism Center for Research and Treatment at Mount Sinai, New York, USA*
[f] *Department of Epidemiology and Public Health, University College London, London, UK*
[g] *Graduate School of Nursing, University of Massachusetts Medical School, Worcester, MA*
[h] *Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, PA, USA*
[i] *Department of Public Health Sciences, University of California, Davis, CA, USA*
[j] *The Parker Institute and Aarhus University, Frederiksberg Hospital, 2000 Frederiksberg, Denmark*
[k] *International Agency for Research on Cancer, World Health Organisation, Lyon, France*

Accepted 30 August 2021; Available online 4 September 2021

Abstract

Objective: Methods for meta-analysis of studies with individual participant data and continuous exposure variables are well described in the statistical literature but are not widely used in clinical and epidemiological research. The purpose of this case study is to make the methods more accessible.

Study Design and Setting: A two-stage process is demonstrated. Response curves are estimated separately for each study using fractional polynomials. The study-specific curves are then averaged pointwise over all studies at each value of the exposure. The averaging can be implemented using fixed effects or random effects methods.

Results: The methodology is illustrated using samples of real data with continuous outcome and exposure data and several covariates. The sample data set, segments of Stata and R code, and outputs are provided to enable replication of the results.

Conclusion: These methods and tools can be adapted to other situations, including for time-to-event or categorical outcomes, different ways of modelling exposure-outcome curves, and different strategies for covariate adjustment.

Key words: Meta-analysis; Individual participant data; Continuous variables; Fractional polynomials

1. Introduction

For categorical exposure variables meta-analysis methods for summary statistics, such as relative risks or hazard ratios, are well-known [1]. The meta-analysis involves calculating weighted averages of the estimates from each study, with weights inversely proportional to their precision (or standard errors). The methods can take into account within-study correlation, heterogeneity across studies, and non-linear exposure-outcome associations [2,3]. However, if individual participant data (IPD) are available there are other opportunities for meta-analysis [4]. In particular, if continuous exposure data are available, it is preferable to model the exposure-outcome association continuously rather than to categorise the exposure [5]. If the association is linear, or has some other simple form, a single stage analysis can be conducted by pooling the IPD for

Declarations of interest: none.

Funding: InterLACE project is funded by the Australian National Health and Medical Research Council project grant (APP1027196). GDM is supported by Australian National Health and Medical Research Council Principal Research Fellowship (APP1121844). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

* Corresponding author. +61417214501.
E-mail address: a.dobson@sph.uq.edu.au (A.J. Dobson).

https://doi.org/10.1016/j.jclinepi.2021.08.033

all studies and fitting a random effects model to take account of within-study correlation. If the exposure-outcome association is non-linear, relevant methods have been published in the statistical literature but are not widely used in epidemiological research.

In this tutorial paper we explain an approach proposed by Sauerbrei and Royston [6] and further examined by White et al. [7]. These authors used a two-stage method. Firstly, they modelled the exposure-outcome curve for each study separately and calculated the predicted outcome values and their standard errors each observed exposure value. Secondly, they calculated pointwise weighted averages across all the study-specific curves using weights inversely proportional to the standard errors of the predictions. This approach provides considerable flexibility as various methods, such as fractional polynomials, can be used to fit curves with a variety of shapes, and covariates (which may differ across studies) can be included in the models [8]. The authors illustrated the method using time-to-event outcome variables and continuous prognostic (exposure) variables. Their approach is, however, much more widely applicable for categorical or continuous outcomes and using other types of functions for the exposure and covariates.

To make these methods more accessible we demonstrate their use with a simple worked example. We start with a sample data set of IPD from several studies. Then we describe how the two-stage meta-analysis can be performed. The mathematical details are in Appendix 1 and segments of code and output for both Stata 16.0 (StataCorp, USA) and R are provided in Appendix 2. Finally, we discuss how the method can be extended to more complicated situations and mention other available software.

2. Methods and results

2.1. Sample data set

To illustrate the methodology, we use data on the association between two continuous variables, age at natural menopause (the outcome) and body mass index (BMI) before menopause (the exposure of interest). The data were assembled for the International Collaboration for a Life Course Approach to Reproductive Health and Chronic Disease Events (InterLACE) [9]. Zhu et al. examined the association using harmonized data from 11 longitudinal cohort studies with data from more than 24,000 women who were premenopausal at the baseline survey and experienced menopause during the follow-up period [10]. Covariates included age at the baseline survey, smoking status, level of education and number of children. For the original analysis both the outcome and exposure variables were categorized and multinomial logistic regression models were fitted with adjustment for clustering within studies.

For this paper, to respect data sovereignty we used random samples from four of the larger studies. From each study a simple random sample of data from 1,500 participants was selected. The sample data set (InterLACE4sample.csv) is available as supplementary material.

2.2. Exploratory analysis

Exploratory analyses of the association between age at natural menopause and baseline BMI are shown in the scatter plots and lowess (local weighted scatterplot smoothing) curves in Fig. 1. Notably age at natural menopause has a ceiling at 55 years for Study 4 corresponding to the last available follow-up for that study. Overall, the patterns are generally similar for the four studies although the ranges differ for both variables and the extent of curvature differs. The descriptive statistics in Table 1 show the broad similarities between the studies. The associations between age at natural menopause and baseline age, and between BMI and baseline age, are approximately linear (results not shown here).

2.3. Modelling

The strategy is to model the association between the outcome and exposure of interest for each study separately (taking the covariates into account) and use each study-specific model to calculate estimates of the outcome. The individual study-specific estimates are then pooled pointwise using standard meta-analysis methods. Full details are provided in Appendix 1.

For the sample data we fitted multiple linear regression models for each study. The dependent variable was age at natural menopause. The independent variables were a curved function for BMI, a linear term for age at baseline, and indicator variables for the categories of smoking, level of education and number of children. The curved functions we used were fractional polynomials which are sums of polynomial and logarithmic terms [8] – see Appendix 1. For Study 1 the results obtained using the Stata command *fp* are: predicted age at natural menopause = 49.38 - 1134.30 × $(1/\text{BMI})^2$ - 2.93 × *ln*(BMI) + 0.30 × baseline age + (0 if the participant was a never smoker, or 0.19 for a former smoker, or -0.09 for a current smoker) + and so on. Details of the study-specific models are shown in Table 2. The models differ in: the functional forms of terms for BMI, coefficients for the covariates and adequacy of fit as measured by adjusted R-squared values. Notably, the model for Study 1 has the poorest fit and the model for Study 4 has the best fit (as expected from the ceiling effect for age at natural menopause due to last available follow-up data for that study).

If the study-specific models all have the same terms (e.g., quadratic functions) an option for the meta-analysis is to calculate weighted averages of the parameter estimates from each study. If the study-specific models have different forms, an appropriate method is to calculate the predicted values of the outcome for each value of the exposure vari-

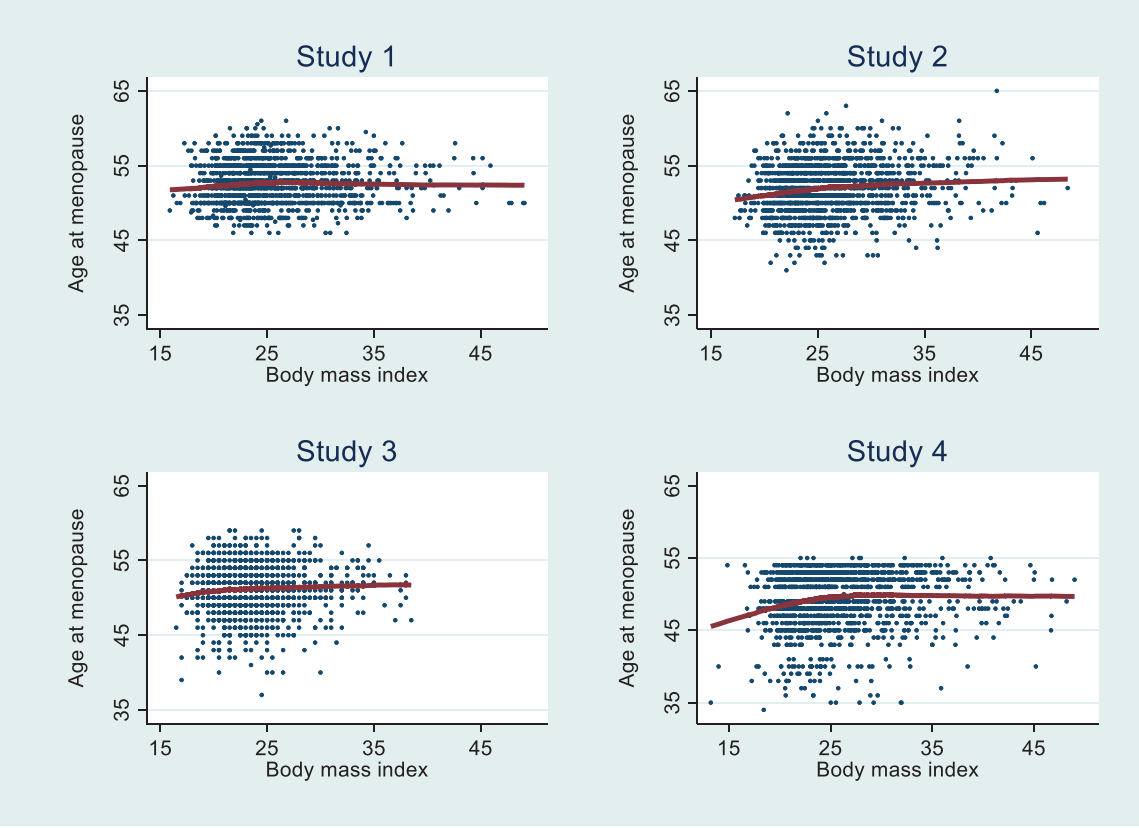

Fig. 1. Scatter plots and lowess fits for age at natural menopause and baseline body mass index for the sample data set.

Table 1. Summary statistics[a] for the sample data set

	Study 1	Study 2	Study 3	Study 4
Size of random sample, *n*	1500	1500	1500	1500
Age at natural menopause (y), mean (standard deviation)	52.55 (2.78)	51.93 (3.29)	51.18 (3.12)	49.38 (3.90)
Body Mass Index at baseline, mean (standard deviation)	25.44 (4.82)	26.16 (4.66)	23.58 (3.38)	25.82 (5.12)
Age at baseline (y), mean (standard deviation)	47.52 (1.43)	48.19 (4.05)	44.84 (3.44)	45.19 (5.19)
Smoking status at baseline, column %				
Never	57.73	70.47	41.53	47.40
Former	28.13	20.00	39.60	28.47
Current	14.13	9.53	18.87	24.13
Education (y), column %				
<=10	44.27	56.60	36.00	60.73
11-12	17.40	8.47	22.93	11.33
>12	38.33	34.93	41.07	27.93
Number of children, column %				
0	7.87	14.93	8.73	16.60
1	8.87	7.67	15.87	15.00
2	40.67	37.40	44.60	45.53
>=3	42.60	40.00	30.80	22.87

[a] There are small but statistically significant differences among the four studies ($P < 0.0001$ for all variables, based on one-way analysis of variance for the continuous variables and chi-squared tests for the categorical variables).

Table 2. Models for age at natural menopause fitted to each data set separately: powers for fractional polynomials for BMI, coefficients and standard errors obtained using the Stata command ***fp***

	Study 1	Study 2	Study 3	Study 4
Powers for fractional polynomial for BMI	-2, 0	3, 3	-2, -2	-2, -2
First term for BMI	-1134.30 (555.94)	0. 00012 (0. 00015)	-4405.27 (3651.46)	-2736.25 (1579.61)
Second term for BMI	-2.93 (1.71)	-0. 00003 (0. 00004)	1689.89 (1380.56)	964.88 (588.62)
Age at baseline	0.30 (0.05)	0.55 (0.02)	0.58 (0.02)	0.65 (0.01)
Smoking status at baseline, column %				
Never (reference)	0	0	0	0
Former	0.19 (0.16)	-0.01 (0.16)	-0.29 (0.14)	-0.23 (0.11)
Current	-0.09 (0.21)	-1.00 (0.21)	-0.80 (0.17)	-0.31 (0.12)
Education (y), column %				
<=10 (reference)	0	0	0	0
11-12	0.10 (0.20)	0.10 (0.23)	0.27 (0.17)	0.22 (0.15)
>12	0.35 (0.16)	0.25 (0.14)	0.27 (0.14)	0.36 (0.11)
Number of children, column %				
0 (reference)	0	0	0	0
1	-0.24 (0.35)	0.08 (0.27)	0.49 (0.26)	-0.15 (0.17)
2	0.13 (0.28)	0.17 (0.19)	0.46 (0.23)	0.09 (0.14)
>=3	-0.19 (0.28)	0.12 (0.19)	0.42 (0.24)	0.12 (0.15)
Constant	49.38 (6.88)	24.59 (0.82)	23.29 (1.55)	19.24 (0.66)
Adjusted R-squared	0.03	0.49	0.41	0.78

able, and then calculate the weighted average across studies at each point, that is, pointwise averaging. To ensure that each study contributes to the predicted values at every exposure value and covariate pattern, the study-specific model is used to calculated predicted outcome values and their standard errors for every participant in every study, not only the participants in the study used for the study-specific model (this approach is supported by the empirical studies by White et al. [7]).

Fig. 2 shows lowess plots of the predicted values for age at natural menopause and their 95% confidence intervals (predicted value $\pm$ 1.96 $\times$ standard error) against BMI. Each plot depicts the whole dataset but using predictions derived from each of the four study-specific models (i.e., all 4 $\times$ 1,500 sets of exposure and covariate values). Notably, consistent with the larger adjusted R-squared value, Study 4 shows less variability (i.e., narrower confidence intervals across the range of BMI values).

Standard meta-analysis methods are now used for the pointwise averaging. The standard errors of the predicted values are used to calculate the inverse variance weights with different formulas for fixed effects or random effects models (see Appendix 1). For a fixed effects model the exposure-outcome pattern is assumed to be the same for all the study populations and the variation in estimates is only due to sampling variation. For the random effects model, it is assumed that there are differences between the study populations and the goal is to estimate the average effect, therefore there is variation between the studies as well as sampling variation and so the confidence intervals are wider. Fig. 3 shows lowess plots of the fixed effects and random effects weights for each study. Despite the identical sample size, there is considerable difference in the fixed effects weights over the range of BMI values and across the studies, with the largest weights usually for Study 4 which showed the most homogeneity (i.e., least variance) in Fig. 2. In contrast, the random effects weights are very similar across the BMI range and for all studies. The meta-analysis is conducted pointwise (i.e., at each value of BMI observed within the whole dataset) with weighted averaging of the predicted values from each study. Note that the predicted values depend on the observed values of the exposure and the covariates for each participant. The meta-analysis results for fixed or random effects are shown in Fig. 4. The pooled curves are similar for both methods of meta-analysis: low BMI was associated with early age at natural menopause, after adjusting for baseline age and other potential confounders. Age at natural menopause was highest for women with BMI around 30 and there was slight evidence of a decrease for more obese women. The confidence intervals are much wider for the random effects analysis, consistent with the underlying assumption of differences between the study populations.

3. Discussion

The goal of this paper is to make meta-analysis methods for exposure-outcome associations with IPD and continuous exposure data more accessible. While our approach follows that of Sauerbrei, Royston, White et al. [6,7], a

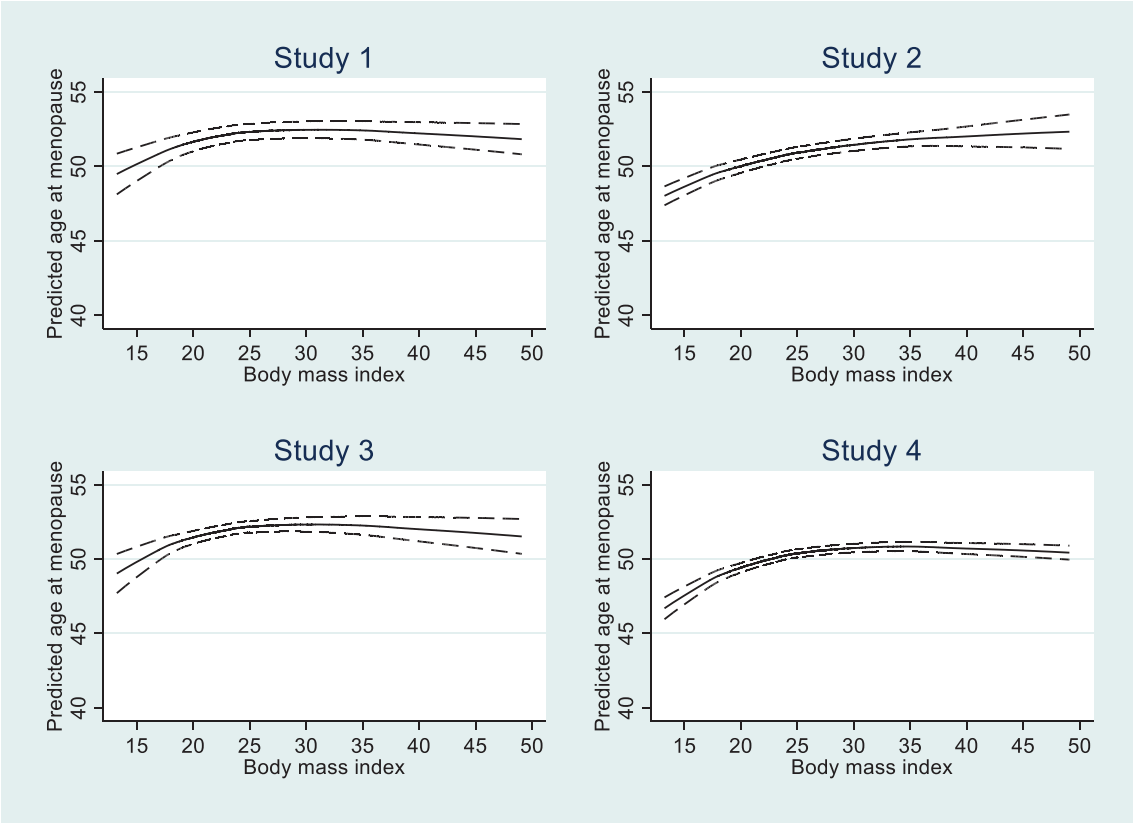

Fig. 2. Lowess plots of predicted values (and 95% confidence intervals) for age at natural menopause by body mass index for each study separately.

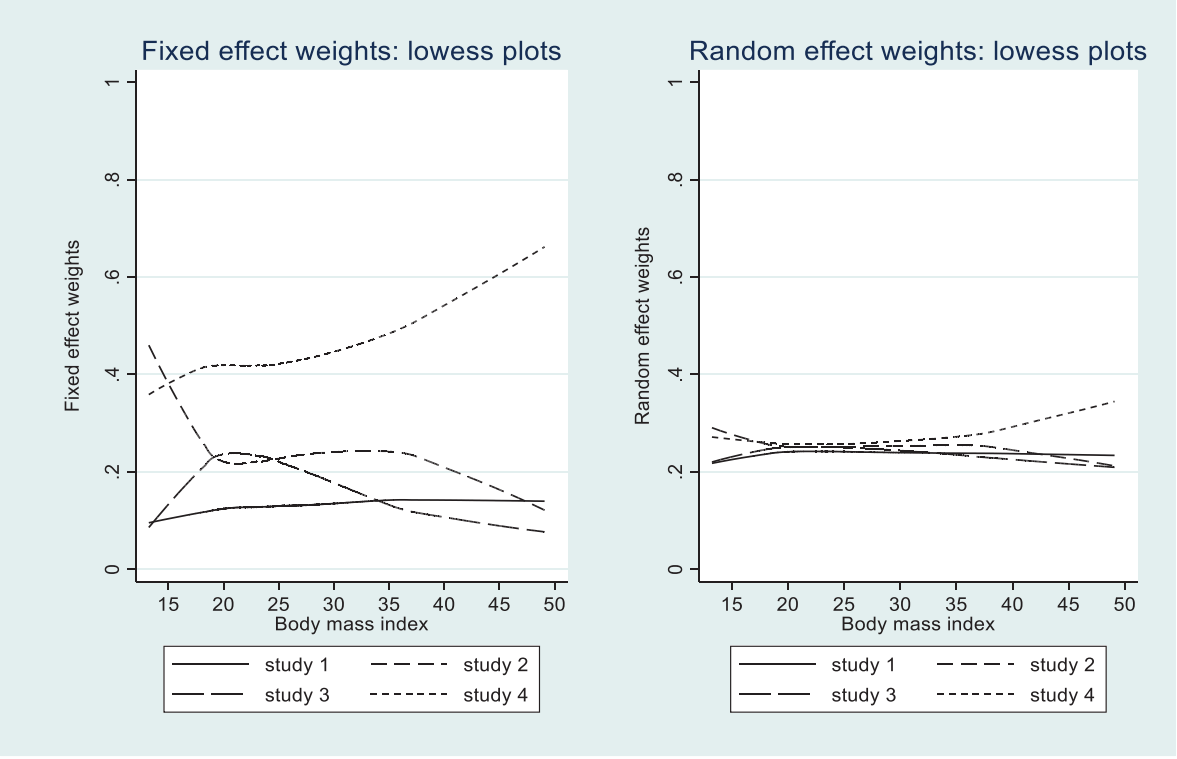

Fig. 3. Lowess plots of the weights for each study that are used in the fixed effects and random effects meta-analyses.

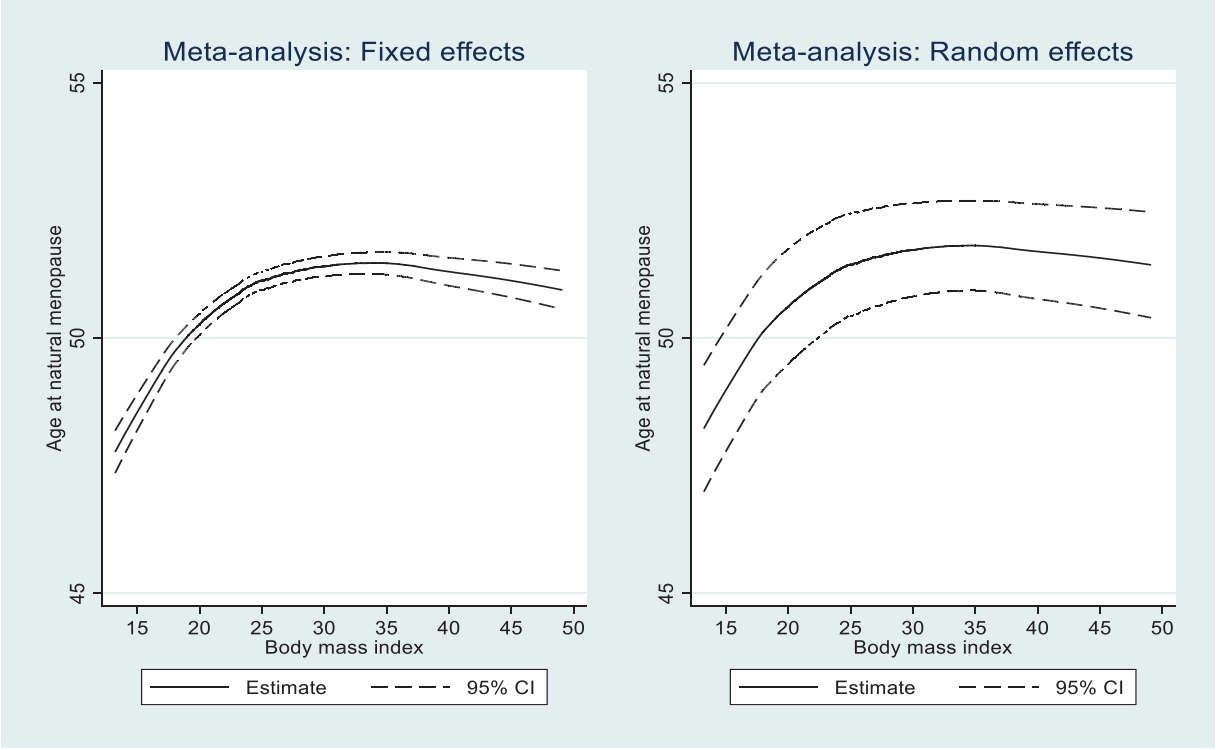

Fig. 4. Results of meta-analysis of the association between age at natural menopause and body mass index: lowess plots for estimates and 95% confidence intervals.

simplified version is used with the sample data set. Each step from the exploratory analysis to interpretation of the pooled results is explained.

In the example, the outcome is continuous and the curves of exposure-outcome association are estimated using multiple linear regression, including fractional polynomial terms. Other examples have involved time-to-event data and survival analysis [6,7]. However, the method is just as applicable for counts or categorical outcomes (e.g., proportions) and a variety of generalized linear models (e.g., logistic regression). The strength of the method is that continuous curves are estimated for each study; that is, the exposure variable is not categorized.

To allow the curves to vary in shape, fractional polynomials were used in the example. But there are other functional forms that can be used such as ordinary polynomials, splines, generalized additive models, or even discontinuous forms. In the example, the number of terms and orders of the polynomials for the fractional polynomial were chosen using the default for the Stata command *fp* The Stata command *fp* uses forward selection of the numbers and powers of terms which are chosen to minimise the deviance. The R procedure *mfp* uses different criteria. It uses backward elimination and family-wise *P*-values; this procedure is designed to protect against overfitting. In the example the R command *mfp* produced simpler (linear) functions but very similar values for the predicted outcomes (see Appendix 2). More generally, the choice of form for the exposure-outcome curve may be made using subject-matter knowledge, visual inspection of the curves, and comparisons of alternative forms (e.g., using criteria for model fit such as AIC or BIC). For any curve fitting there is a tension between selecting forms that are too simple (e.g., linear only) and overfitting with more complex ones.

In some cases, selecting the same form of curve for all studies may be appropriate. In this situation meta-analysis could be used to average the parameter estimates rather than pointwise pooling of the curves [11]. For meta-analyses of large numbers of studies with many participants this approach would be less computationally intensive, and White *et al.* have shown that the results are likely to be similar [7]. This strategy is also likely to have more power to model complicated curves [12]. Software for pooling parameter estimates is available in Stata and R programs both called mvmeta [7,12].

A notable difference between the method used above and the approach described by Sauerbrei, Royston, White and others [6,7] and implemented in the Stata program metacurve [13], is their use of an intermediate stage of fitting "confounder models." Instead of fitting a study-specific model with the outcome as the dependent variable and fractional polynomial terms of the exposure and covariates as the independent variables, they first fit a "confounder model" which has the exposure as the dependent variable and the covariates as the independent variables.

Next the linear predictor of this model is calculated for all individuals, this is called the "confounder index." Finally, they fit a study-specific model with the outcome as the dependent variable and the fractional polynomial of the exposure, adjusted for the confounder index. An advantage of using a confounder model is that it can accommodate more complex terms for the covariates. For instance, in the example above the covariate, baseline age, was treated as a linear term, but in a confounder model a fractional polynomial, or other functional form, for this variable could have been included. A nonstatistical researcher may initially find the concept of a confounder model confusing because the main exposure has the role of the dependent variable. This is why confounder models were not used in the example, but when they were used the final results were the same. A confounder model is analogous to a propensity score [14] with a model fitted for the exposure variable rather than the outcome but the coefficients may be less easily interpreted.

In the example, for simplicity, centring and scaling were not used for any of the continuous variables. Nevertheless, it is usually better statistical practice to standardize the exposure and other covariates, at least by centring them, as this can help interpretation of the estimates and reduce collinearity. In some situations, it is important that the results can be readily transformed back to the original scales (as in the example of age at natural menopause and BMI). In other situations, effect sizes relative to some fixed value are more interpretable, for example, risk of an outcome relative to a reference level of the exposure [7,12].

Using IPD to fit a continuous curve for the exposure-outcome association is preferable to categorizing the exposure. Categorizing continuous variables reduces the precision and power of an analysis [5]. If only published aggregate results, not IPD, are available for a continuous exposure variable, the effect estimates usually refer to categories of exposure, and these may be used to estimate the underlying continuous association [15]. Meta-analysis of these data is complicated by the correlation of estimates from the same study across the exposure range. Specialised software includes the SAS macro metadose [16] and the R program dosresmeta [17].

As with any meta-analysis it is important to consider whether the studies and their results are sufficiently similar to justify averaging them. Factors to be considered include differences in: study design, covariates measured, measurement scales, and ranges of exposure and outcome measures [18,19]. Recommendations for exploring heterogeneity for IPD include comparisons across studies of the distributions and associations between variables [19,20]. For the InterLACE consortium from which the example data were drawn [9], some studies collected age at menopause retrospectively and other prospectively, for some BMI was calculated from self-reported measures while others provided measured weight and height. Nevertheless, visual inspection of the plots in Fig. 1 and summary statistics in Table 1 suggest sufficient similarity in the sample data to justify meta-analysis.

The goal of making the sample data set publicly available, providing segments of Stata and R code, and output, is to facilitate replication of the results, comparison or alternative methods and software, and extension to other situations.

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.

CRediT authorship contribution statement

Darsy Darssan: Investigation, Formal analysis, Software, Writing – original draft. **Gita D. Mishra:** Conceptualization, Methodology, Writing – review & editing. **Darren C. Greenwood:** Methodology, Writing – review & editing. **Sven Sandin:** Writing – review & editing. **Eric J. Brunner:** Writing – review & editing. **Sybil L. Crawford:** Writing – review & editing. **Samar R. El Khoudary:** Writing – review & editing. **Maria Mori Brooks:** Writing – review & editing. **Ellen B. Gold:** Writing – review & editing. **Mette Kildevæld Simonsen:** Writing – review & editing. **Hsin-Fang Chung:** Writing – review & editing, Data curation. **Elisabete Weiderpass:** Writing – review & editing. **Annette J. Dobson:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision.

Acknowledgment

The data on which this research is based were drawn from four observational studies included in the InterLACE project. This acknowledgement covers all 11 studies in the paper by Zhu et al. [10] as the InterLACE collaborators contributed to the present paper. The Australian Longitudinal Study on Women's Health, the University of Newcastle, Australia, and the University of Queensland, Australia. We are grateful to the Australian Government Department of Health for funding and to the women who provided the survey data. The Melbourne Collaborative Cohort Study was supported by VicHealth and the Cancer Council, Victoria, Australia. The Danish Nurse Cohort Study was supported by the National Institute of Public Health, Copenhagen, Denmark. The Women's Lifestyle and Health Study was funded by a grant from the Swedish Research Council (Grant number 521-2011-2955). The MRC National Study of Health and Development and the National Child Development Study has core funding from

the UK Medical Research Council. The English Longitudinal Study of Ageing is funded by the National Institute on Aging (Grants 2RO1AG7644 and 2RO1AG017644-01A1) and a consortium of UK government departments. The UK Women's Cohort Study was funded by the World Cancer Research Fund. The Whitehall II study has been supported by grants from the Medical Research Council. The Seattle Middle Women's Health Study was supported by grants from the National Institute for Nursing Research.

The Study of Women Across the Nation has grant support from the National Institutes of Health (NIH), DHHS, through the National Institute on Aging (NIA), the National Institute of Nursing Research (NINR) and the NIH Office of Research on Women's Health (ORWH) (Grants U01NR004061; U01AG012505, U01AG012535, U01AG012531, U01AG012539, U01AG012546, U01AG012553, U01AG012554, U01AG012495). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIA, NINR, ORWH or the NIH. Clinical Centers: University of Michigan, Ann Arbor – Siobán Harlow, PI 2011 – present, MaryFran Sowers, PI 1994-2011; Massachusetts General Hospital, Boston, MA – Joel Finkelstein, PI 1999 – present; Robert Neer, PI 1994 – 1999; Rush University, Rush University Medical Center, Chicago, IL – Howard Kravitz, PI 2009 – present; Lynda Powell, PI 1994 – 2009; University of California, Davis/Kaiser – Ellen Gold, PI; University of California, Los Angeles – Gail Greendale, PI; Albert Einstein College of Medicine, Bronx, NY – Carol Derby, PI 2011 – present, Rachel Wildman, PI 2010 – 2011; Nanette Santoro, PI 2004 – 2010; University of Medicine and Dentistry – New Jersey Medical School, Newark – Gerson Weiss, PI 1994 – 2004; and the University of Pittsburgh, Pittsburgh, PA – Karen Matthews, PI.

NIH Program Office: National Institute on Aging, Bethesda, MD – Chhanda Dutta 2016 – present; Winifred Rossi 2012 – 2016; Sherry Sherman 1994 – 2012; Marcia Ory 1994 – 2001; National Institute of Nursing Research, Bethesda, MD – Program Officers.

Central Laboratory: University of Michigan, Ann Arbor – Daniel McConnell (Central Ligand Assay Satellite Services).

Coordinating Center: University of Pittsburgh, Pittsburgh, PA – Maria Mori Brooks, PI 2012 - present; Kim Sutton-Tyrrell, PI 2001 – 2012; New England Research Institutes, Watertown, MA - Sonja McKinlay, PI 1995 – 2001.

Steering Committee: Susan Johnson, Current Chair; Chris Gallagher, Former Chair

We thank Janet Cade, Nancy Woods, Ellen S. Mitchell, and Diana Kuh for providing the data used in this study. All studies would like to thank the participants for volunteering their time to be involved in the respective studies. The findings and views in this paper are not necessarily those of the original studies or their respective funding agencies.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2021.08.033.

Appendix 1. Mathematical methods

Notation. The data are denoted by y_{jk}, z_{jk} and c_{jkl}, where y is the outcome or response variable, z is the exposure or dose variable of interest, and c denotes the covariates; $j = 1, \ldots, J$ denotes the studies, $k = 1, \ldots, K_j$ denotes the observations in each study, and l denotes the covariates ($l = 1, \ldots, L$). For the sample data there are $J = 4$ studies with $K = 1500$ observations for each study and $L = 4$ covariates.

Fractional polynomials. A fractional polynomial of a variable x is defined as $f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m x^{p_m}$ with degree M, the number of terms, and powers p_m [8]. Usually, $M = 2$ and the values p_m are selected from -2, -1, -0.5, 0, 0.5, 1, 2 and 3; x^0 is defined to be $ln(x)$ and if a power p is repeated the terms are $x^p + x^p ln(x)$.

Fitting models for each study separately. For the sample data we fitted models $f(y_{jk}) = \beta_0 + \sum_{m=1}^{M} \beta_m z_{jk}^{p_m} + \sum_{l=1}^{L} \gamma_{jkl} c_{jkl}$ for each study separately using multivariable linear regression. The fractional polynomial and covariate terms were all fitted at the same time (using the Stata command *fp*. The powers p_1 and p_2 for each study were selected based on the model with the lowest deviance (the default method in Stata). The continuous covariate baseline age was fitted as a linear term. In this case, for simplicity, none of the variables was transformed (e.g., centred or scaled).

Predicted values and standard errors of predicted values. For each model predicted values and their standard errors need to be calculated for every participant, not only those in that particular study. To do this it may be convenient to work with the data in "long form", that is with the observations of Study 1 stacked on those for Study 2, and so on. We use the index i for each of the $J \times K$ rows and retain the index j for studies. With this change in notation, each row includes the predicted values ψ_{ij} and their standard errors s_{ij} for each of the J studies. Fig. 2 shows the $J \times K$ predicted values and their 95% confidence intervals ($\psi_{ij} \pm 1.96\ s_{ij}$) plotted for each study separately.

Meta-analysis. Standard metaanalysis methods are now used to average the predicted values ψ_{ij} across each row using inverse variance weights calculated from the standard errors s_{ij}. Using similar notation to Sauerbrei and Royston [6], let $v_{ij} = s_{ij}^2$ and $R_i = \sum_{j=1}^{J} (1/v_{ij})$. For the fixed effects estimate, the weights are given by $w_{ij} = 1/(v_{ij} R_i)$ and the estimate is

$\psi_i^{FE} = \sum_{j=1}^{J} w_{ij}\psi_{ij}$, with variance given by $var_i^{FE} = 1/R_i$

for row i of the stacked data. For the random effects estimate, first calculate $Q_i = \sum_{j=1}^{J}[(\psi_{ij} - \psi_i^{FE})^2/v_{ij}]$, $D_i = R_i - [\sum_{j=1}^{J}(1/v_{ij}^2)/R_i]$, $\tau_i^2 = max\{0, [Q_i - (J-1)]/(R_i - D_i)\}$ and $W_i = \sum_{j=1}^{J} 1/(v_{ij} + \tau_i^2)$. Then the weights are given by $u_{ij} = 1/\{(v_{ij} + \tau_i^2)W_i\}$ and the random effects estimate is

$\psi_i^{RE} = \sum_{j=1}^{J} u_{ij}\psi_{ij}$, with variance given by $var_i^{RE} = 1/W_i$.

Fig. 3 shows the fixed effects weights w_{ij} and the random effects weights u_{ij} for each study, plotted for all i and Fig. 4 shows the results of the meta-analysis using either fixed or random effects analysis.

Appendix 2. Segments of Stata and R code and output

——————— Stata code for Table 1 ———————

```
by study, sort : summarize meno bmi age
tabulate smoke study, chi2 column
tabulate educ study, chi2 column
tabulate child study, chi2 column
```

——————— Stata code for Fig. 1 ———————

Plot of scatter plots and lowess curves of age at natural menopause against BMI for each study, for example, for Study 1 as shown in Fig. 1

```
twoway (scatter meno bmi if study==1, msize(tiny)) (lowess meno bmi if study==1, noweight
bwidth(0.5) lwidth(thick)), ytitle(Age at menopause) ytitle(, margin(medium)) ylabel(35(10)65)
xtitle(Body mass index) xlabel(15(10)50) title(Study 1) legend(off)
```

——————— Stata code and output for fractional polynomials ———————

Fit linear regression models for age at natural menopause as a function of a fractional polynomial for BMI and linear terms for the other covariates using the Stata function *fp*, and the default criteria for choosing the powers, for example, for Study 1.

```
fp <bmi>, replace all: regress meno <bmi> i.smoke i.educ i.child age if study==1
```

The output (summarized in Table 2) is as follows

```
Fractional polynomial comparisons:
------------------------------------------------------------------------------
          bmi |   df    Deviance  Res. s.d.   Dev. dif.   P(*)   Powers
--------------+---------------------------------------------------------------
      omitted |    0   7273.778      2.742      5.470    0.246
       linear |    1   7273.030      2.742      4.723    0.196   1
        m = 1 |    2   7271.262      2.741      2.954    0.231   -2
        m = 2 |    4   7268.307      2.739      0.000       --   -2 0
------------------------------------------------------------------------------
(*) P = sig. level of model with m = 2 based on F with 1487 denominator dof.
      Source |       SS           df       MS      Number of obs   =     1,500
-------------+----------------------------------   F(10, 1489)     =      5.70
       Model |  427.367172         10  42.7367172   Prob > F        =    0.0000
    Residual |  11168.8228     1,489   7.5008884   R-squared       =    0.0369
-------------+----------------------------------   Adj R-squared   =    0.0304
       Total |    11596.19     1,499  7.73595063   Root MSE        =    2.7388
------------------------------------------------------------------------------
        meno |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       bmi_1 |  -1134.302   555.9948    -2.04   0.042    -2224.918   -43.68532
       bmi_2 |  -2.933602   1.712177    -1.71   0.087    -6.292136    .4249326
             |
```

```
          smoke |
             1  |    .194115    .163553     1.19   0.235    -.1267038    .5149337
             2  |  -.0934166   .2111038    -0.44   0.658    -.5075091     .320676
                |
           educ |
             2  |    .097636   .2009809     0.49   0.627    -.2965998    .4918719
             3  |   .3545823   .1588956     2.23   0.026     .0428992    .6662654
                |
          child |
             1  |  -.2449392   .3479638    -0.70   0.482    -.9274906    .4376121
             2  |   .1256801   .2785451     0.45   0.652    -.4207024    .6720626
             3  |    -.19038   .2781773    -0.68   0.494     -.736041    .3552811
                |
            age |   .3028138   .0499079     6.07   0.000     .2049166     .400711
          _cons |   49.38168   6.878085     7.18   0.000     35.88992    62.87345
------------------------------------------------------------------------------
```

———— R code and output for fractional polynomials ————

```
library(mfp)
> d1 <- subset(dat, study==1); d2 <- subset(dat, study==2); d3 <- subset(dat, study==3);
d4 <- subset(dat, study==4)
> f1 <- mfp(meno ~ fp(bmi, df=4, select=NA, alpha=NA, scale=F) + factor(smoke)+factor(educ)
+factor(child)+age, family=gaussian, data=d1)
> summary(f1)
> f1
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    37.51567    2.38370  15.738  < 2e-16 ***
age             0.30581    0.04996   6.122 1.18e-09 ***
factor(educ)2   0.10066    0.20122   0.500   0.6170
factor(educ)3   0.35184    0.15908   2.212   0.0271 *
factor(child)1 -0.22827    0.34832  -0.655   0.5123
factor(child)2  0.16367    0.27832   0.588   0.5566
factor(child)3 -0.14579    0.27776  -0.525   0.5997
factor(smoke)1  0.19717    0.16375   1.204   0.2287
factor(smoke)2 -0.10913    0.21124  -0.517   0.6055
I(bmi^1)        0.01281    0.01487   0.862   0.3889
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for gaussian family taken to be 7.519492)
     Null deviance: 11596  on 1499  degrees of freedom
Residual deviance: 11204  on 1490  degrees of freedom
AIC: 7295
Number of Fisher Scoring iterations: 2
Call:
mfp(formula = meno ~ fp(bmi, df = 4, select = NA, alpha = NA,
     scale = F) + factor(smoke) + factor(educ) + factor(child) +
     age, data = d1, family = gaussian)
Deviance table:
               Resid. Dev
Null model    11596.19
Linear model  11204.04
Final model   11204.04
Fractional polynomials:
                df.initial select alpha df.final power1 power2
age                      1      1  0.05        1      1      .
factor(educ)2            1      1  0.05        1      1      .
```

```
factor(educ)3          1      1  0.05        1      1      .
factor(child)1         1      1  0.05        1      1      .
factor(child)2         1      1  0.05        1      1      .
factor(child)3         1      1  0.05        1      1      .
factor(smoke)1         1      1  0.05        1      1      .
factor(smoke)2         1      1  0.05        1      1      .
bmi                    4      1  0.05        1      1      .
```

Note that the default setting for the R function *mfp* selected a linear model for Study 1 (and also different linear models for all other Studies). In contrast, the default settings for the Stata function *fp* selected the more complicated models shown in Table 2. However, as can be seen below, the estimated values ψ_{ij} and their standard errors s_{ij} are very similar. This suggests that the Stata models are over-fitted.

———— Stata code and output for predicted values and standard errors ————

Note because the data are already in long format with Study 1 data first, followed by Study 2 data (and so on) the following commands provide estimates of predicted values, and their standard errors for all 4×1500 observations.

```
predict y1, xb
predict se1, stdp
```

Similarly for all studies. The first 3 rows of the data now include these values.

```
list meno y1 se1 y2 se2 y3 se3 y4 se4 in 1/3
```

	meno	y1	se1	y2	se2	y3	se3	y4	se4
1	53	53.18607	.2443321	51.4018	.2367986	53.12684	.1990782	51.84632	.1531551
2	55	51.56691	.2874484	49.22062	.2947484	50.9301	.217362	49.36789	.1938314
3	56	51.85093	.2149087	50.27079	.1835253	51.26583	.1570621	49.37421	.1321854

For the plots in Fig. 2, for example, for Study 1.

```
gen ucl1 = y1 + 1.96*se1
gen lcl1 = y1 - 1.96*se1
twoway (lowess y1 bmi, lcolor(black) lpattern(solid)) (lowess ucl1 bmi, lcolor(black)
lpattern(dash)) (lowess lcl1 bmi, lcolor(black) lpattern(dash)), ytitle(Predicted age at
menopause) ylabel(40(5)55) xtitle(Body mass index) xlabel(15(5)50) title(Study 1) legend(off)
```

———— R code and output for predicted values and standard errors ————

For Study 1

```
s1p <- predict(f1, se.fit=T, newdata=dat)
y1 <- s1p$fit; se1 <- s1p$se.fit; v1 <- se1^2
lcl1 <- y1-1.96*se1
ucl1 <- y1+1.96*se1
```

	meno	y1	se1	y2	se2	y3	se3	y4	se4
1	53	53.13694	0.2433305	51.40186	0.2365143	53.08190	0.1944257	51.82207	0.1527804
2	55	51.46335	0.2836528	49.22723	0.2946102	50.87636	0.2104466	49.30538	0.1920844
3	56	51.74361	0.2093062	50.27884	0.1830635	51.23898	0.1541292	49.30062	0.1283526

Similarly for all other studies

———— Stata code and output for fixed effects and random effects meta-analysis ——-

Fixed Effects Meta Analysis

```
gen v1=se1^2
gen v2=se2^2
gen v3=se3^2
gen v4=se4^2
gen suminv = 1/v1 + 1/v2 + 1/v3 + 1/v4
gen w1 = (1/v1)/suminv
gen w2 = (1/v2)/suminv
gen w3 = (1/v3)/suminv
gen w4 = (1/v4)/suminv
gen phiFE = w1*y1 + w2*y2 + w3*y3 + w4*y4
gen varphiFE = 1/suminv
```

Random Effects Meta Analysis

```
  gen Q =  (y1 - phiFE)^2*1/v1 + (y2 - phiFE)^2*1/v2 + (y3 - phiFE)^2*1/v3 + (y4 -
phiFE)^2*1/v4
  gen denom = suminv - ((1/v1)^2 + (1/v2)^2 + (1/v3)^2 + (1/v4)^2)/suminv
  gen tausq = max(0, ((Q - (4-1))/denom))
  gen wran1 = 1/(v1 + tausq)
  gen wran2 = 1/(v2 + tausq)
  gen wran3 = 1/(v3 + tausq)
  gen wran4 = 1/(v4 + tausq)
  gen wransum = wran1 + wran2 + wran3 + wran4
  gen w1std = wran1/wransum
  gen w2std = wran2/wransum
  gen w3std = wran3/wransum
  gen w4std = wran4/wransum
  gen phiRE =  w1std*y1 + w2std*y2 + w3std*y3 + w4std*y4
  gen varphiRE = 1/wransum
```

The FE and RE averages of the predicted values y of meno for each study are shown here for the first few rows

```
. list meno y1 y2 y3 y4 phiFE phiRE  in 1/3
```

	meno	y1	y2	y3	y4	phiFE	phiRE
1.	53	53.18607	51.4018	53.12684	51.84632	52.30338	52.3873
2.	55	51.56691	49.22062	50.9301	49.36789	50.18005	50.26868
3.	56	51.85093	50.27079	51.26583	49.37421	50.42663	50.68446

The corresponding standard errors are

```
  gen sephiFE = sqrt(varphiFE)
  gen sephiRE = sqrt(varphiRE)
  list se1 se2 se3 se4 sephiFE sephiRE in 1/3
```

	se1	se2	se3	se4	sephiFE	sephiRE
1.	.2443321	.2367986	.1990782	.1531551	.0987976	.4302353
2.	.2874484	.2947484	.217362	.1938314	.118349	.5566828
3.	.2149087	.1835253	.1570621	.1321854	.0818929	.5652525

For graphs of the weights (Fig. 3) the code for fixed effects is here (and similarly for the random effects)

```
  twoway (lowess w1 bmi, lcolor(black) lpattern(solid)) (lowess w2 bmi, lcolor(black)
lpattern(dash)) (lowess w3 bmi, lcolor(black) lpattern(longdash)) (lowess w4 bmi, lcolor(black)
lpattern(shortdash)), ytitle(Fixed effect weights) ytitle(, margin(medium)) ylabel(0(0.2)1)
xtitle(Body mass index) xlabel(15(5)50) title(Fixed effect weights: lowess plots) legend
(label(1 "study 1″) label(2 "study 2″) label(3 "study 3″) label(4 "study 4″))
```

For graphs of the fixed effects estimates (Fig. 4) the code is here (and similarly for the random effects)

```
  gen lclFE = phiFE - 1.96* sephiFE
  gen uclFE = phiFE + 1.96* sephiFE
  gen lclRE = phiRE - 1.96* sephiRE
  gen uclRE = phiRE + 1.96* sephiRE
  twoway (lowess uclFE bmi, lcolor(black) lpattern(dash)) (lowess lclFE bmi, lcolor(black)
lpattern(dash)) (lowess phiFE bmi, lcolor(black) lpattern(solid)) , ytitle(Age at natural
menopause) ylabel(45(5)55) xtitle(Body mass index) xlabel(15(5)50) title(Meta-analysis: Fixed
effects) legend( order(3 1 ) label(3 "Estimate") label(1 "95% CI")  )
```

———— R code and output for fixed effects and random effects meta-analysis ————

```
  #Pooling the functional forms across studies
  suminv = 1/v1 + 1/v2 + 1/v3 + 1/v4
  #Standardised fixed effect weights
  w1 = (1/v1)/suminv
  w2 = (1/v2)/suminv
```

```
  w3 = (1/v3)/suminv
  w4 = (1/v4)/suminv
  #Overall fixed effect estimate and the variance
  phiFE = w1*y1 + w2*y2 + w3*y3 + w4*y4
  varphiFE = 1/suminv
  sephiFE = sqrt(varphiFE)
  lclFE = phiFE - 1.96*sephiFE
  uclFE = phiFE + 1.96*sephiFE
  ####Calculation random effect weights#########
  Q <- ((y1 - phiFE)^2*(1/v1)) + ((y2 - phiFE)^2*(1/v2)) + ((y3 - phiFE)^2*(1/v3)) + ((y4 -
phiFE)^2*(1/v4))
  denom = suminv - ((1/v1)^2 + (1/v2)^2 + (1/v3)^2 + (1/v4)^2)/suminv
  #S Squared
  #tausq = max(0, ((Q - (4-1))/denom))
  tausq1 = ((Q - (4-1))/denom)
  tausq <- ifelse(tausq1<0,0,tausq1)
  #random-effect weights
  wran1 = 1/(v1 + tausq)
  wran2 = 1/(v2 + tausq)
  wran3 = 1/(v3 + tausq)
  wran4 = 1/(v4 + tausq)
  wransum = wran1 + wran2 + wran3 + wran4
  w1std = wran1/wransum
  w2std = wran2/wransum
  w3std = wran3/wransum
  w4std = wran4/wransum
  #Overall random-effect estimate and the variance
  phiRE =  w1std*y1 + w2std*y2 + w3std*y3 + w4std*y4
  varphiRE = 1/wransum
  sephiRE <- sqrt(varphiRE)
  lclRE = phiRE - 1.96*sephiRE
  uclRE = phiRE + 1.96*sephiRE
```

	meno	phiFE	sephiFE	phiRE	sephiRE
1	53	52.28356	0.09802917	52.35814	0.4211312
2	55	50.13915	0.11653731	50.21595	0.5449056
3	56	50.37194	0.08020266	50.63507	0.5703076

References

[1] Greenland S, Longnecker MP. Methods for trend estimation from summarized dose–response data, with applications to meta-analysis. Am J Epidemiol 1992;135:1301–9.

[2] Liu Q, Cook NR, Bergström A, Hsieh CC. A two-stage hierarchical regression model for meta-analysis of epidemiologic nonlinear dose–response data. Computational Statistics & Data Analysis 2009;53:4157–67.

[3] Rota M, Bellocco R, Scotti L, Tramacere I, Jenab M, Corrao G, et al. Random-effects meta-regression models for studying nonlinear dose–response relationship, with an application to alcohol and esophageal squamous cell carcinoma. Statist. Med. 2010;29:2679–87. doi:10.1002/sim.4041.

[4] Debray TP, Moons KG, van Valkenhoef G, Efthimiou O, Hummel N, Groenwold RHH, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods 2015;6:293–309. doi:10.1002/jrsm.1160.

[5] Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Statist. Med. 2006;25:127–41.

[6] Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. Statist. Med. 2011;30:3341–60.

[7] White IR, Kaptoge S, Royston P, Sauerbrei W. Meta-analysis of non-linear exposure-outcome relationships using individual participant data: a comparison of two methods. Statist. Med. 2019;38:326–38.

[8] Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. J Royal Statistical Soc. Series C (Applied Statistics) 1994;43:429–67.

[9] Mishra GD, Anderson D, Schoenaker DAJM, Adami H-O, Avis NE, Brown D, et al. InterLACE: a new international collaboration for a life course approach to women's reproductive health and chronic disease events. Maturitas 2013;74:235–40.

[10] Zhu D, Chung HS, Pandeya N, Dobson AJ, Kuh D, Crawford SL, et al. Body mass index and age at natural menopause: an international pooled analysis of 11 prospective women's health studies. Eur J Epidemiol 2018;33:699–710.

[11] Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. Statist Med 2017;36:855–75.

[12] Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Statist. Med. 2012;31:3821–39.

[13] White IR. Multivariate random-effects meta-analysis. Stata J 2009;9:40–56.

[14] Austin PC. A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality. MultivarBehav Res 2011;46:119–51.

[15] Bagnardi V, Zambon A, Quatto P, Corrao G. Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. Am J Epidemiol 2004;159:1077–86.

[16] Li R, Spiegelman D (2010). "The SAS %metadose Macro." http://www.hsph.harvard. edu/donna-spiegelman/software/metadose/.

[17] Crippa A, Orsini N. Multivariate dose–response meta-analysis: the dosresmeta R package. J Stat Software 2016;72:1–15.

[18] Cordero CP, Dans AL. Key concepts in clinical epidemiology: detecting and dealing with heterogeneity in meta-analyses. J Clin Epidemiol 2021;130:149–51.

[19] Smith CT, Williamson PR, Marson A. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Statist. Med. 2005;24:1307–19.

[20] Steyerberg EW, Nieboer D, Debray TPA, van Houwelingen HC. Assessment of heterogeneity in an individual participant data meta–analysis of prediction models: An overview and illustration. Statist. Med. 2019;38:4290–309.