Title: Estimating additive interaction in two-stage individual participant data meta-analysis

Authors: Maartje Basten[1-4], Lonneke A. van Tuijl[5,6], Kuan-Yu Pan[2,7,8], Adriaan W. Hoogendoorn[2,7], Femke Lamers[2,7], Adelita V. Ranchor[5], Joost Dekker[2,7], Philipp Frank[9], Henrike Galenkamp[10], Mirjam J. Knol[11], Nolwenn Noisel[12], Yves Payette[12], Erik R. Sund[13-15], Aeilko H. Zwinderman[16], Lützen Portengen[17], Mirjam I. Geerlings[4,18-20]

ORCiD IDs: Maartje Basten 0000-0001-9982-7048 Lonneke A. van Tuijl: 0000-0001-9031-0886 Kuan-Yu Pan: 0000-0003-0523-663X Adriaan W. Hoogendoorn: 0000-0003-4784-0712 Femke Lamers: 0000-0003-4344-5766 Joost Dekker: 0000-0003-2027-0027 Philipp Frank: 0000-0001-8824-8673 Nolwenn Noisel: 0000-0002-3143-6347 Erik R Sund: 0000-0002-0736-5139 Aeilko H Zwinderman: 0000-0003-0361-3139 Mirjam I Geerlings: 0000-0002-4037-036X

Correspondence Address: Maartje Basten, Department of Health Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. m.g.j.basten@vu.nl

Joint Authorship: N/A

Affiliations: [1] Department of Health Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [2] Amsterdam Public Health, Mental Health program, Amsterdam, the Netherlands. [3] Amsterdam Public Health, Health Behaviors and Chronic Diseases program, Amsterdam, the Netherlands. [4] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht and Utrecht University, Utrecht, the Netherlands. [5] Department of Health Psychology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [6] Department of Clinical Psychology, Utrecht University, Utrecht, The Netherlands. [7] Amsterdam UMC, location Vrije Universiteit Amsterdam, Department of Psychiatry, Boelelaan 1117, Amsterdam, the Netherlands. [8] Unit of Occupational Medicine, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [9] UCL Brain Sciences, Division of Psychiatry, University College London, London, UK. [10] Department of Public and Occupational Health, Amsterdam UMC, and Amsterdam Public Health Research Institute, University of Amsterdam, Amsterdam, the Netherlands. [11] Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. [12] CARTaGENE, CHU Sainte-Justine, 3175, Chemin de la Côte-Sainte-Catherine, Montréal, Québec, Canada. [13] Faculty of Nursing and Health Sciences, Nord University, Levanger, Norway. [14] HUNT Research Centre, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [15] Levanger hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway. [16] Amsterdam UMC, location University of Amsterdam, Clinical Epidemiology, Biostatistics and Bioinformatics, Meibergdreef 9, Amsterdam, the Netherlands. [17] Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands. [18] Amsterdam UMC, location University of Amsterdam, Department of General Practice, Meibergdreef 9, Amsterdam, the Netherlands [19] Amsterdam Public Health, Personalized Medicine, and Aging & Later Life, Amsterdam, the Netherlands. [20] Amsterdam Neuroscience, Neurodegeneration, and Mood, Anxiety, Psychosis, Stress, and Sleep, Amsterdam, the Netherlands.

Clinical trial registration number and website: N/A

Key words: additive interaction, effect-modification, individual participant data meta-analysis, relative excess risk due to interaction (RERI)

---

[1] Study investigators, conference presentations, preprint publication information, thanks.

Conflict of Interest: The authors declare no conflict of interests.

Disclaimer: HUNT: The study has used data from the Cancer Registry of Norway. The interpretation and reporting of these data are the sole responsibility of the authors, and no endorsement by the Cancer Registry of Norway is intended nor should be inferred

Data Availability Statement: The data that support the findings of this study are owned by participating cohort studies. Data may be shared upon reasonable request at each cohort depending on cohort-specific regulations. Further information is available from the corresponding author upon request.

ABSTRACT

Individual participant data (IPD) meta-analysis provides important opportunities to study interaction and effect modification for which individual studies often lack power. While previous meta-analyses have commonly focused on multiplicative interaction, additive interaction holds greater relevance for public health and may in certain contexts better reflect biological interaction. Methodological literature on interaction in IPD meta-analysis does not cover additive interaction for models including binary or time-to-event outcomes. We aimed to describe how the Relative Excess Risk due to Interaction (RERI) and other measures of additive interaction or effect modification can be validly estimated within two-stage IPD meta-analysis. First, we explain why direct pooling of study-level RERI estimates may lead to invalid results. Next, we propose a three-step procedure to estimate additive interaction: 1) estimate effects of both exposures and their product term on the outcome within each individual study; 2) pool study-specific estimates using multivariate meta-analysis; 3) estimate an overall RERI and 95% confidence interval based on the pooled effect estimates. We illustrate this procedure by investigating interaction between depression and smoking and risk of smoking-related cancers using data from the PSYchosocial factors and Cancer (PSY-CA) consortium. We discuss implications of this procedure, including the application in meta-analysis based on published data.

Introduction

Individual participant data (IPD) meta-analyses are increasingly used in epidemiologic research. An IPD meta-analysis is a specific type of systematic review that involves the collection, checking and re-analysis of the original data for each participant in each study (1). One significant strength of IPD meta-analysis is the ability to yield greater statistical power for hypothesis testing compared with smaller-scale individual studies. Compared with meta-analysis based on summary data from previous research, IPD meta-analysis provides a greater control and harmonization in methods used across studies included. In addition, analyses can be conducted which were not previously done, including subgroup analyses or interaction analyses. Hence, utilizing IPD meta-analysis is considered the gold standard method of review (2). IPD meta-analysis is particularly interesting for studying interaction and effect modification, as much larger sample sizes are needed to identify interactions or effect modifications than to identify main effects (3-5). In interaction analysis, the main interest is in the combined effect of two exposures, while in effect modification the focus is on the possible different effects of one exposure within strata based on some other factor (6). Many studies that have been designed to examine main effects such as intervention effects or identification of risk factors, have lacked statistical power to study interaction or effect modification. Combining multiple studies in IPD meta-analysis allows the investigation of e.g., potential interacting risk factors in the development of disease, or identification of subgroups that benefit most from treatment or that are at highest risk for disease.

Several methods on how to study interaction or effect modification within IPD meta-analysis have been developed and compared with respect to power, potential for aggregation bias and susceptibility to confounding (7-12). These include interaction models for one-stage IPD meta-analyses, where all data are combined in one dataset that is used to build the statistical model of interest. Alternatively, the meta-analysis of interaction (MAOI) model for two-stage IPD meta-analysis is a method where the same statistical model is applied to each study and study-specific effect estimates are subsequently pooled in a meta-analysis. The method of choice may depend on data characteristics (11, 12). In practice, a two-stage procedure is often applied when participating studies for various reasons (e.g., legal restrictions) are unable to share their individual participant data. In this study we mainly focus on interaction and effect modification in two-stage IPD meta-analysis.

Interaction and effect modification can be studied on an additive and multiplicative scale (6). Positive *additive interaction* is present if the combined effect of two exposures is larger than the *sum* of the

individual effects of the two exposures. Positive *multiplicative* interaction is present if the combined effect of two exposures is larger than the *product* of the individual effects. Additive and multiplicative interaction fall along a continuum: when both exposures increase the risk of the outcome, the threshold for the combined effect to reach additive interaction is lower than the threshold for multiplicative interaction (13).

Unfortunately, whether multiplicative or additive interaction is being studied is often a matter of convenience, depending on whether the outcome of interest is continuous or binary. Interaction is most often examined by adding a product term of the exposures as an independent variable to the regression model. When performing regression models with a continuous outcome, the effect estimate for interaction is on the additive scale. However, in regression models with a binary or time-to-event outcome (e.g., logistic regression or Cox regression), the effect estimate for interaction is on the multiplicative scale. Current methods on studying interaction in IPD meta-analysis follow the same approach. Although multiplicative interaction is often the default in models with a binary or time-to-event outcome, additive interaction may better fit the objectives of the study.

When statistical interaction is examined to test potential biological interaction, the choice for studying multiplicative or additive interaction should be based on hypothesized underlying biological mechanisms (14, 15). When two exposures are hypothesized to each reinforce the same biological mechanism or to act at different stages in the same disease process, positive additive interaction may more likely be identified than multiplicative interaction. In addition, the additive scale is preferred from a public health perspective for which absolute risks are more important than relative risks (6, 14, 15). Additive interaction indicates e.g., which subgroups are at highest risk for disease or may benefit most from an intervention. By studying interaction on the additive scale, absolute risks due to interaction can be derived. These are informative in e.g., number needed to treat to prevent one additional bad outcome or in targeting interventions to subgroups which may benefit most when resources are limited.

To study additive interaction for binary and time-to-event outcomes, several measures have been developed that can be derived from relative risks (RR), odds ratios (OR), or hazard ratios (HR). The relative excess risk due to interaction (RERI) is most often considered in the literature (6). Other measures are the proportion attributable to interaction of the combined effect of two exposures (AP), or the synergy index (S) (16, 17). Thus far, few studies used RERI or other additive interaction measures to study interaction within IPD meta-analyses. In one-stage IPD meta-analysis, additive interaction measures can be calculated from effect estimates obtained from the analyses on the

7

pooled data (18). In two-stage IPD meta-analyses different procedures have been used. Some studies calculated additive interaction measures within each study and subsequently pooled study-specific estimates (19-21). Other studies have first pooled the individual effect estimates of the exposures and product term and calculated measures of additive interaction based on these pooled estimates (22, 23). There is currently no in-depth discussion in the literature regarding the validity of these procedures.

The aim of this study was to describe how measures of additive interaction and effect modification for binary and time-to-event outcomes can be validly estimated within *two-stage* IPD meta-analyses. We propose a procedure to calculate measures of additive interaction based on the pooled effect estimates at stage two of the IPD meta-analysis. For simplicity, we will focus on interaction but our conclusions are also relevant for effect modification. For the same reason, we will focus on the RERI, which is most commonly used, but our approach can also be used to estimate AP or S. We will first explain the RERI and its features in more detail and discuss potential issues related to direct comparisons and pooling of study-level RERI estimates. Second, we provide a three-step procedure on how to obtain a valid estimate of additive interaction within two-stage IPD meta-analyses. We illustrate our procedure using an example from the psychosocial factors and cancer (PSY-CA) study (24-26). Finally, we discuss for which situations this procedure can be applied, including meta-analysis based on summary data of previous research.

Relative Excess Risk due to Interaction (RERI)

The RERI is a measure of additive interaction or effect modification for models with a binary outcome. For two dichotomous exposures A and B, RRs can be calculated for when both exposures are present ($A+B+$), exposure A is present and B is absent ($A+B-$), and exposure B is present and A is absent ($A-B+$), with the absence of both exposures ($A-B-$) as the reference group. Subsequently, the RERI estimate is calculated as follows:

$$RERI = RR_{A+B+} - RR_{A-B+} - RR_{A+B-} + 1 \qquad\qquad (1)$$

The RERI can range from minus infinity to infinity. RERI = 0 reflects no interaction; RERI > 0 reflects positive interaction; and a RERI < 0 reflects negative interaction, i.e., the combined effect of two exposures is *smaller* than the sum of the individual effects of the two exposures. By multiplying the RERI with the absolute background risk, the absolute risk due to interaction is estimated (27). The RERI formula can also be applied to ORs derived from logistic regression models or HRs from models

8

with time-to-event outcomes when ORs and HRs approximate RRs (6, 28). A more general formula that can be applied when either one or both exposures are continuous is as follows (6, 27):

$$\text{RERI} = e^{(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3)} - e^{(\hat{\beta}_1)} - e^{(\hat{\beta}_2)} + 1 \qquad\qquad (2)$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ refer to the estimated regression coefficients of exposure 1 and exposure 2 and where $\hat{\beta}_3$ refers to their product term.

Several methods have been described to calculate a 95% confidence interval (CI), including the Delta method (29), four-by-two table method (30), or bootstrapping. A detailed discussion of these methods for confidence interval estimation is provided by VanderWeele and Knol (6).

Importantly, RERI, but also AP and S, should only be applied to exposures that increase the risk of the outcome, but not to exposures decreasing the risk, i.e., preventive factors (31). Using preventive exposures can give inconsistent results. This is because these measures are calculated from ratio measures (RRs, ORs or HRs) which have an asymmetric scale: preventive effects are limited between 0 and 1 while risk effects may vary between 1 and infinity. Therefore, preventive factors should be recoded so that the category with the lowest risk is the reference category (31).

Comparing and pooling study-level RERI estimates: potential issues

When additive interaction is examined in two-stage IPD meta-analysis but also in meta-analysis based on summary data from previous research, this is typically done by calculating estimates of additive interaction for each individual study and subsequent pooling of these estimates (19-21, 32, 33). Here we will discuss potential issues related to direct comparisons and pooling of study-level RERI estimates.

Inconsistent direction of exposure effects across studies
When performing a two-stage IPD meta-analysis, the direction of the effect for an exposure may differ across studies: an exposure may be related to an increased risk in study *A*, while a decreased risk may be found in study *B*. As described above, a RERI estimate can only be validly calculated for exposures that are related to increased risk of the outcome. As such, different reference categories should be used in study *A* and *B* in order to obtain a valid RERI estimate in each study. When changing the reference category for part of the studies, interpretation of RERI will differ across

9

studies and cannot be compared. In this situation, pooling RERI estimates to derive one interpretable effect is impossible.

<u>Inconsistency with exposure effect estimates due to asymmetry of RERI components</u>

If both exposures are consistently associated with increased risk of the outcome across studies, pooling of study-level RERI estimates might seem valid given its continuous scale and its potential range from minus infinity to infinity. However, the RERI estimate is a function of multiple ratio measures, either formulated in terms of RRs, ORs or HRs. Due to the asymmetry of these components, the distribution of RERI estimates may deviate significantly from a normal distribution. This complicates direct pooling of RERI estimates across studies using classical methods for meta-analysis. This issue is similar to that when pooling ratio measures like RRs, ORs and HRs, that are usually log transformed to obtain a symmetric scale (34). However, a log transformation of RERI is less straightforward as it is a sum of multiple ratio measures. Direct pooling of study-level RERI estimates in two-stage IPD meta-analysis is likely to result in an estimated overall RERI that does not match the pooled effect estimates for both exposures and their product term, i.e. the multiplicative interaction effect. We illustrate this potential deviation using real-world data in a supplementary data file (Supplementary file, Table S1, Table S2).

Recommended procedure: calculate overall RERI on pooled effect estimates

To bypass the issues described above, we propose a procedure to examine additive interaction using the RERI within a two-stage IPD meta-analysis. This procedure consists of three steps:
1) Study-level effect estimation of exposures and their product term on the outcome
2) Pooling effect estimates in multivariate meta-analysis
3) Estimation of overall RERI and 95% confidence interval based on pooled estimates
Step 1 is conducted during stage 1 of the IPD meta-analysis, while steps 2 and 3 are conducted during stage 2. The steps are described into detail below:

<u>1. Study-level effect estimation</u>

To test interaction for dichotomous and/or continuous exposures, both exposures and their product term are entered as independent variables in the regression model of interest within each study. Alternatively, if both exposures are dichotomous, a categorical variable could be created including four categories: both exposures present (*A+B+*), only exposure *A* (*A+B-*), only exposure *B* (*A-B+*), and absence of both exposures (*A-B-*). This variable could be entered in the regression model, with the category of (*A-B-*) as reference group. Next, covariates can be entered into the model to control for

10

potential confounding in the associations between both exposures and the outcome. From this model, beta coefficients of the exposures and the product term, together with their standard errors, variances and covariances need to be extracted as input for stage 2.

2. Pooling of effect estimates

After interaction analyses are performed in each study, effect estimates for both exposures and the product term of all studies are pooled at stage two. Effect estimates can be pooled in a fixed-effects model, assuming that the true effect is the same in each study, or in a random-effects model, which allows for between-study heterogeneity in the true effect. In most meta-analyses, univariate meta-analysis models are used, pooling one effect estimate at a time. Alternatively, in multivariate meta-analysis, multiple effect estimates are pooled at the same time, taking into account correlations between effects. We recommend to use multivariate meta-analyses to pool the effect estimates of the exposures and the product term and their variances and covariances all at once. By using this method the correlations between the regression coefficients of the exposures and their product term are accounted for. Moreover, pooled variances and covariances of effects are automatically estimated, which are required for estimation of the 95% confidence interval of the overall RERI. More detailed information on using multivariate meta-analyses can be found elsewhere (35).

3. Overall RERI estimation

The pooled regression coefficients from the multivariate meta-analysis can then be entered in the RERI formula (1) or (2) to obtain an estimate for additive interaction. The 95% confidence interval of the overall RERI can be estimated based on the pooled variances and covariances of the beta coefficients.

As described earlier, RERI should only be applied to exposures increasing the risk of the outcome but not to exposures decreasing the risk. This means that the *pooled* regression coefficients for both exposures should be positive. If the *pooled* regression coefficient for one of the exposure variables is negative, it is invalid to use this regression coefficient to estimate an overall RERI. This can be solved by reversely coding the preventive exposure variable in *all studies*. Subsequently, steps 1 and 2 have to be repeated and the overall RERI can be calculated. Alternatively, it is also possible to reverse in all studies the beta coefficients of the preventive exposure and the product term and their covariances with the second exposure. This reversal should be incorporated in interpreting the direction of the RERI estimate.

Example: depression, smoking and smoking-related cancers in PSY-CA study

We will now illustrate the proposed three-step procedure with an example from the PSYchosocial factors and Cancer (PSY-CA) study, including a time-to-event outcome. The PSY-CA consortium, involving 18 prospective cohort studies, was established to perform IPD meta-analyses on the association of depression, anxiety and other psychosocial factors with cancer incidence (24-26). A detailed description of the PSY-CA study, including study protocol and power calculations, has been published previously (24). Of note, the power analysis in PSY-CA concerned univariate instead of multivariate meta-analysis and multiplicative instead of additive interaction was considered. Notice that our power calculation is conservative, since power is larger for multivariate meta-analysis than for univariate meta-analysis (36) and larger for testing additive interaction than for multiplicative interaction (5). In this example, we examined potential interaction between depression and smoking at baseline on smoking-related cancers incidence during follow-up, which involves an adapted analysis previously published in Basten et al. (26). We used IPD of six cohorts: CARTaGENE (37-39), the English Longitudinal Study of Aging (ELSA) (40), the Healthy Life in an Urban Setting (HELIUS) study (41, 42), HUNT 2 (43), Lifelines (44, 45) and the Netherlands Study of Depression and Anxiety (NESDA) (46). To ease interpretation of results, dichotomous measures of depression and smoking were used. Depression (yes/no) was based on clinical interviews or, if not available, on symptom questionnaires using validated clinical cut-offs. Smoking was defined as currently smoking (yes/no). Cancer cases during follow-up, including cancer type and date of diagnosis, were identified through linkage to national or regional cancer registries. This analysis included a total of 252,686 participants, 7,401 incident smoking-related cancer diagnoses, and 2,536,124 person years of follow-up. All analyses were performed using R Statistical Software (47).

In step one, an interaction model using Cox regression analysis was estimated in each cohort study. We included depression, smoking and the product term of depression and smoking as independent variables into the model and smoking-related cancer incidence as dependent variable. Entry age (age at baseline) and exit age (age at diagnosis, death, loss to follow-up or study end) were entered in the model as time variables. The model was adjusted for potential confounders of the associations between both depression and smoking with cancer incidence. For simplicity, we present a basic model adjusted for confounders that were available in all cohorts: sex, education, country of birth and birth year. Effect estimates for depression, smoking and the product term in each cohort study are presented in Table 1.

In step two, results from all participating cohorts were pooled using random-effects multivariate meta-analysis. Analyses were performed using R package 'mvmeta' (48). We used a random-effects model to account for potential varying effects across cohorts, as cohorts within PSY-CA included both

12

population-based and clinical samples and varied substantially in age and time of follow-up. The regression coefficients, variances and covariances of depression, smoking and the product term of each cohort were entered in the multivariate meta-analysis. The pooled effects showed that depression (B: 0.18, 95%CI: 0.06 to 0.31; HR: 1.20, 95%CI: 1.06 to 1.36) and smoking (B: 0.72, 95%CI: 0.62 to 0.81; HR: 2.05, 95%CI: 1.86 to 2.25) were both associated with an increased risk of smoking-related cancer (Table 1). The product term, representing multiplicative interaction, showed that the combined effect of depression and smoking was not higher than *the product* of the individual effects ($B_{product}$: 0.01; 95% CI: -0.18 to 0.20; HR: 1.01, 95%CI: 0.84 to 1.22).

As both depression and smoking were associated with an increased risk (i.e., no preventive effects), we proceeded with step three. We entered the pooled effect estimates (Table 1) in the RERI formula (2): $RERI = e^{(0.18 + 0.72 + 0.01)} - e^{(0.18)} - e^{(0.72)} + 1 = 0.23$. The overall RERI was larger than 0, suggesting that the combined effect of depression and smoking is larger than the sum of the individual effects. The hazard ratio of having depression and being a smoker is 0.23 more than if there were no interaction between depression and smoking. We calculated the 95% CI based on the pooled variances and covariances using the delta method. This was done using the function 'deltamethod' in R package 'msm' (49). This resulted in a 95% confidence interval ranging from -0.14 to 0.61, indicating that there is no strong evidence for the presence of an interaction effect and a large interaction effect seems unlikely.

This example included two categorical exposures. We could have also tested interaction using a categorical independent variable including four categories (2 [yes/no depression] ×2 [yes/no smoking]) in step 1 and calculate an overall RERI at step 3, using formula 1. This procedure would have resulted in the same overall RERI estimate.

DISCUSSION

IPD meta-analyses provide important opportunities to study interaction and effect modification for which individual studies often lack power. This paper describes a procedure to study additive interaction for binary and time-to-event outcomes within two-stage IPD meta-analyses. Since additive interaction may more closely reflect hypotheses on biological interaction and is more relevant from a public health perspective, this paper provides an important addition to previous methodological studies on interaction within IPD meta-analysis (7-12). The procedure was illustrated using an example from the PSY-CA study. We used this procedure previously to examine whether various psychosocial factors, including depression and anxiety, interact with or modify the effects of

13

health behaviors, such as smoking and alcohol use, in relation to cancer incidence (26). To date, few previous studies have used a similar procedure as presented in this paper (23).

This paper described a procedure to study additive interaction using the RERI in two-stage IPD meta-analysis, but this procedure has wider implications. First, the same procedure can also be applied for testing effect modification, that is when the effect of only one exposure on the outcome is of interest across subgroups. A main difference in testing effect modification is that only confounders for the exposure are included but not for the effect modifier (50). Second, the proposed procedure can also be used to estimate other measures of additive interaction or effect modification, including the attributable proportion due to interaction (AP) and the synergy index (S). Overall estimates of AP and S can be calculated based on the same pooled regression coefficients of the exposures and their product term in step three. Third, the example from the PSY-CA project involved two dichotomous exposures and time-to-event data which were analyzed using Cox regression. The procedure is also applicable to models including one or more continuous exposures and to logistic regression models. Fourth, similar steps may also be used to study additive interaction in meta-analysis based on published studies. If individual studies provide effect estimates for the exposures and the product term, these can be pooled and subsequently used to calculate an overall RERI estimate. However, a major limitation with published papers is that variances and covariances of the effect estimates, which are needed for estimating a multivariate meta-analysis and the final confidence interval of the overall RERI, are often not reported. Possible solutions to deal with missing variances and covariances in multivariate meta-analysis have been described elsewhere (35). Finally, additive interaction may also be estimated in one-stage IPD meta-analysis. The effect estimates for the exposures and their product term obtained in one-stage IPD meta-analysis can be used to estimate the RERI or other measures of additive interaction. The practical implications of applying one-stage or two-stage IPD meta-analysis are described elsewhere (8,12).

The choice for studying multiplicative or additive interaction should depend on the objectives of the study and not be just a matter of convenience. Knol and VanderWeele provide recommendations for presenting analyses of effect modification and interaction (50). They propose to present measures of interaction on additive and multiplicative scales. In this way sufficient information is provided to the reader to interpret interaction on whatever scale is preferred.

IPD meta-analyses are well suited to study interaction and effect modification. The current paper describes a procedure to study additive interaction for binary and time-to-event outcomes within two-stage IPD meta-analysis, illustrated with an empirical example. Our procedure will hopefully

encourage researchers to apply additive interaction analysis within ongoing IPD meta-analyses projects. Furthermore, this procedure may encourage researchers of smaller studies who aim to study interaction to set up an IPD meta-analysis project.

15

REFERENCES

1.   Tierney JF, Stewart LA, Clarke M. Chapter 26: Individual participant data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. 6.4: Cochrane; 2023.

2.   Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof. 2002;25(1):76-97.

3.   Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol. 2004;57(3):229-36.

4.   McClelland GH, Judd CM. Statistical difficulties of detecting interactions and moderator effects. Psychol Bull. 1993;114(2):376-90.

5.   VanderWeele TJ. Sample Size and Power Calculations for Additive Interactions. Epidemiol Methods. 2012;1(1):159-88.

6.   VanderWeele TJ, Knol MJ. A Tutorial on Interaction. Epidemiologic Methods. 2014;3(1):33-72.

7.   Fisher DJ, Copas AJ, Tierney JF, Parmar MK. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. J Clin Epidemiol. 2011;64(9):949-67.

8.   Riley RD, Fisher DJ. Using IPD Meta-Analysis to Examine Interactions between Treatment Effect and Participant-level Covariates. In: Riley RD, Tierney JF, Stewart LA, editors. Individual participant data meta-analysis: a handbook for healthcare research: John Wiley & Sons Ltd; 2021. p. 163-98.

9.   Freeman SC, Fisher D, Tierney JF, Carpenter JR. A framework for identifying treatment-covariate interactions in individual participant data network meta-analysis. Res Synth Methods. 2018;9(3):393-407.

10.   Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Anti-Lymphocyte Antibody Induction Therapy Study G. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med. 2002;21(3):371-87.

11.   Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Stat Med. 2007;26(15):2982-99.

12.   Walker R, Stewart L, Simmonds M. Estimating interactions in individual participant data meta-analysis: a comparison of methods in practice. Syst Rev. 2022;11(1):211.

13.   Mehta NK, Zheng H, Myrskyla M. How do age and major risk factors for mortality interact over the life-course? Implications for health disparities research and public health policy. SSM Popul Health. 2019;8:100438.

14. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. Am J Epidemiol. 1980;112(4):467-70.

15. Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. Int J Epidemiol. 1981;10(4):383-7.

16. Rothman KJ. Modern Epidemiology. Boston, MA: Little, Brown and Company; 1986.

17. Rothman KJ. Epidemiology: An introduction. New York: Oxford University Press; 2002.

18. Olsson A, Bouaoun L, Schuz J, Vermeulen R, Behrens T, Ge C, et al. Lung Cancer Risks Associated with Occupational Exposure to Pairs of Five Lung Carcinogens: Results from a Pooled Analysis of Case-Control Studies (SYNERGY). Environ Health Perspect. 2024;132(1):17005.

19. Jabbarpoor Bonyadi MH, Yaseri M, Soheilian M. Association of combined complement factor H Y402H and ARMS2/LOC387715 A69S polymorphisms with age-related macular degeneration: an updated meta-analysis. Ophthalmic Genet. 2020;41(4):301-7.

20. Lee W, Choi HM, Kim D, Honda Y, Leon Guo YL, Kim H. Synergic effect between high temperature and air pollution on mortality in Northeast Asia. Environ Res. 2019;178:108735.

21. Fontvieille E, Viallon V, Recalde M, Cordova R, Jansana A, Peruchet-Noray L, et al. Body mass index and cancer risk among adults with and without cardiometabolic diseases: evidence from the EPIC and UK Biobank prospective cohort studies. BMC Med. 2023;21(1):418.

22. Fang W, Li Z, Gao J, Meng R, He G, Hou Z, et al. The joint and interaction effect of high temperature and humidity on mortality in China. Environ Int. 2023;171:107669.

23. Oze I, Charvat H, Matsuo K, Ito H, Tamakoshi A, Nagata C, et al. Revisit of an unanswered question by pooled analysis of eight cohort studies in Japan: Does cigarette smoking and alcohol drinking have interaction for the risk of esophageal cancer? Cancer Med. 2019;8(14):6414-25.

24. van Tuijl LA, Voogd AC, de Graeff A, Hoogendoorn AW, Ranchor AV, Pan KY, et al. Psychosocial factors and cancer incidence (PSY-CA): Protocol for individual participant data meta-analyses. Brain Behav. 2021;11(10):e2340.

25. van Tuijl LA, Basten M, Pan KY, Vermeulen R, Portengen L, de Graeff A, et al. Depression, anxiety, and the risk of cancer: An individual participant data meta-analysis. Cancer. 2023.

26. Basten M, Pan KY, van Tuijl LA, de Graeff A, Dekker J, Hoogendoorn AW, et al. Psychosocial factors, health behaviors and risk of cancer incidence: Testing interaction and effect modification in an individual participant data meta-analysis. Int J Cancer. 2024.

27. Knol MJ, van der Tweel I, Grobbee DE, Numans ME, Geerlings MI. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int J Epidemiol. 2007;36(5):1111-8.

28. Li R, Chambless L. Test for additive interaction in proportional hazards models. Ann Epidemiol. 2007;17(3):227-36.

29.   Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. Epidemiology. 1992;3(5):452-6.

30.   Zou GY. On the estimation of additive interaction by use of the four-by-two table and beyond. Am J Epidemiol. 2008;168(2):212-24.

31.   Knol MJ, VanderWeele TJ, Groenwold RH, Klungel OH, Rovers MM, Grobbee DE. Estimating measures of interaction on an additive scale for preventive exposures. Eur J Epidemiol. 2011;26(6):433-8.

32.   Keijzer MB, Borm GF, Blom HJ, Bos GM, Rosendaal FR, den Heijer M. No interaction between factor V Leiden and hyperhomocysteinemia or MTHFR 677TT genotype in venous thrombosis. Results of a meta-analysis of published studies and a large case-only study. Thromb Haemost. 2007;97(1):32-7.

33.   Yu IT, Tse LA. Exploring the joint effects of silicosis and smoking on lung cancer risks. Int J Cancer. 2007;120(1):133-9.

34.   Higgins JPT, Li T, Deeks JJe. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022): Cochrane; 2022.

35.   Riley RD, Jackson D, White IR. Multivariate meta-analysis using IPD. In: Riley RD, Tierney JF, Stewart LA, editors. Individual participant data meta-analysis: a handbook for healthcare research: John Wiley & Sons Ltd; 2021. p. 311-46.

36.   Hattle M, Burke DL, Trikalinos T, Schmid CH, Chen Y, Jackson D, et al. Multivariate meta-analysis of multiple outcomes: characteristics and predictors of borrowing of strength from Cochrane reviews. Syst Rev. 2022;11(1):149.

37.   Awadalla P, Boileau C, Payette Y, Idaghdour Y, Goulet JP, Knoppers B, et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. Int J Epidemiol. 2013;42(5):1285-99.

38.   Borugian MJ, Robson P, Fortier I, Parker L, McLaughlin J, Knoppers BM, et al. The Canadian Partnership for Tomorrow Project: building a pan-Canadian research platform for disease prevention. CMAJ. 2010;182(11):1197-201.

39.   Dummer TJB, Awadalla P, Boileau C, Craig C, Fortier I, Goel V, et al. The Canadian Partnership for Tomorrow Project: a pan-Canadian platform for research on chronic disease prevention. CMAJ. 2018;190(23):E710-E7.

40.   Steptoe A, Breeze E, Banks J, Nazroo J. Cohort profile: the English longitudinal study of ageing. Int J Epidemiol. 2013;42(6):1640-8.

41. Snijder MB, Galenkamp H, Prins M, Derks EM, Peters RJG, Zwinderman AH, et al. Cohort profile: the Healthy Life in an Urban Setting (HELIUS) study in Amsterdam, The Netherlands. BMJ Open. 2017;7(12):e017873.

42. Stronks K, Snijder MB, Peters RJ, Prins M, Schene AH, Zwinderman AH. Unravelling the impact of ethnicity on health in Europe: the HELIUS study. BMC Public Health. 2013;13:402.

43. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, et al. Cohort Profile: the HUNT Study, Norway. Int J Epidemiol. 2013;42(4):968-77.

44. Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol. 2015;44(4):1172-80.

45. Sijtsma A, Rienks J, van der Harst P, Navis G, Rosmalen JGM, Dotinga A. Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. Int J Epidemiol. 2022;51(5):e295-e302.

46. Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P, et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int J Methods Psychiatr Res. 2008;17(3):121-40.

47. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.

48. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Stat Med. 2012;31(29):3821-39.

49. Jackson CH. Multi-State Models for Panel Data: The msm Package for R. J Stat Softw. 2011;38(8):1-28.

50. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol. 2012;41(2):514-20.

Table 1 Depression, smoking and their interaction on smoking-related cancer incidence, an example from the PSY-CA consortium

Cohorts	PY (events)	Depression B (95% CI)	Smoking B (95% CI)	Product term B (95% CI)
CARTaGENE	236,496 (1334)	0.28 (-0.02 to 0.58)	0.59 (0.46 to 0.72)	0.02 (-0.44 to 0.48)
ELSA	106,877 (705)	0.23 (0.03 to 0.44)	0.87 (0.67 to 1.07)	-0.08 (-0.43 to 0.28)
HELIUS	100,986 (142)	0.09 (-0.55 to 0.73)	0.76 (0.38 to 1.14)	0.19 (-0.72 to 1.11)
HUNT 2	1,038,540 (3607)	0.15 (-0.09 to 0.38)	0.76 (0.69 to 0.83)	-0.07 (-0.41 to 0.26)
Lifelines	1,023,663 (1528)	0.14 (-0.23 to 0.51)	0.64 (0.52 to 0.76)	0.00 (-0.55 to 0.56)
NESDA	29,561 (85)	0.09 (-0.61 to 0.79)	0.95 (0.38 to 1.51)	0.18 (-0.72 to 1.08)
Pooled estimates	2,536,124 (7401)	0.18 (0.06 to 0.31)	0.72 (0.62 to 0.81)	0.01 (-0.18 to 0.20)

Note. Adapted analysis previously published in Basten et al. (26). PY= person years.