

# The consequences of neglected confounding and interactions in mixed-effects meta-regression: An illustrative example

Eric S. Knop<sup>1</sup>  | Markus Pauly<sup>1</sup>  | Tim Friede<sup>2,3</sup>  | Thilo Welz<sup>1</sup> 

<sup>1</sup>Department of Statistics, TU Dortmund University, Dortmund, Germany

<sup>2</sup>Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

<sup>3</sup>DZHK (German Center for Cardiovascular Research), Partner Site Göttingen, Göttingen, Germany

## Correspondence

Eric S. Knop, Department of Statistics, TU Dortmund University, Joseph-von-Fraunhofer-Straße 2-4, 44227 Dortmund, Germany.  
Email: [eric.knop@tu-dortmund.de](mailto:eric.knop@tu-dortmund.de)

## Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Numbers: FR 3070/3-1, PA-2409 7-1

## Abstract

Analysts seldom include interaction terms in their meta-regression model, which can introduce bias if an interaction is present. We illustrate this by re-analysing a meta-regression study in acute heart failure. Based on a total of 285 studies, the 1-year mortality rate related to acute heart failure is considered and the connection to the study-level covariates year of recruitment and average age of study participants are of interest. We show that neglecting a possibly confounding variable and an interaction term might lead to erroneous inference and conclusions. Based on our results and accompanying simulations, we recommend to include possible confounders and interaction terms, whenever they are plausible, in mixed-effects meta-regression models.

## KEYWORDS

acute heart failure, confounding, interactions, meta-regression

## Highlights

### What we know so far

- In meta-regression interaction terms of moderators are often not considered.
- Omitting important variables in a regression model can lead to biased estimates, whereas including redundant variables increases the variance of estimates.

### What is new

- Re-analysis of a meta-analysis on acute heart failure illustrates how neglecting possibly confounding variables and interaction terms can alter the results of a meta-regression model.
- A small simulation study is conducted, which shows how neglecting and unnecessarily including moderators and interaction effects confidence intervals estimates in meta-regression.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Research Synthesis Methods* published by John Wiley & Sons Ltd.

**Potential impact for RSM readers outside the authors' field**

- The results suggest to include possible confounding variables and interaction terms in meta-regression, whenever they are plausible, because otherwise wrong conclusions might be drawn.

**1 | INTRODUCTION**

In meta-regression, interactions reflecting effect moderators are often neglected. Certainly, in many situations, ignoring interactions is due to the small number of available studies.<sup>1</sup> But even if enough studies are available, interactions are often not considered.<sup>2,3</sup> As we know from ordinary least squares regression, this can lead to biased estimates, whereas including redundant moderators and interactions only increases the variance of the estimates (see Reference [4], pp. 99–101). A further common issue in statistical analyses that can lead to incorrect or misleading results is confounding. It arises when the relationship between an independent variable and a dependent variable is influenced by a third variable, which is ignored in the analysis. The latter is known as a confounding variable. A formal definition can be found elsewhere.<sup>5</sup> The topic is extremely important when distinguishing between mere associations and causal effects, see for example the work of Friedrich et al.<sup>6</sup> A discussion of confounding in meta-analysis can be found in the article of Durlak and Lipsey.<sup>7</sup>

In our work, we illustrate these points by a reanalysis of a recent meta-analysis, including mixed-effects meta-regression analyses: Kimmoun et al.<sup>3</sup> analysed mortality and readmission to hospital after acute heart failure in clinical studies. They found a statistically significant decline of mortality over calendar time. However, the average age of the patients also decreased over calendar time. This suggests that the observed trend might at least partially be explained by the confounding variable average age or a neglected interaction between the average age and the median recruitment year. Therefore, we conducted meta-regressions where we (i) included not only the recruitment year as a moderator but the average age as well and (ii) both variables and their interaction. We conclude with some practical recommendations, when possible confounding variables and interaction terms should be included.

**2 | METHODS**

A mixed-effects meta-regression model with two moderators and an interaction term is of the form

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}x_{2i}\beta_{12} + u_i + e_i, i = 1, \dots, k \quad (1)$$

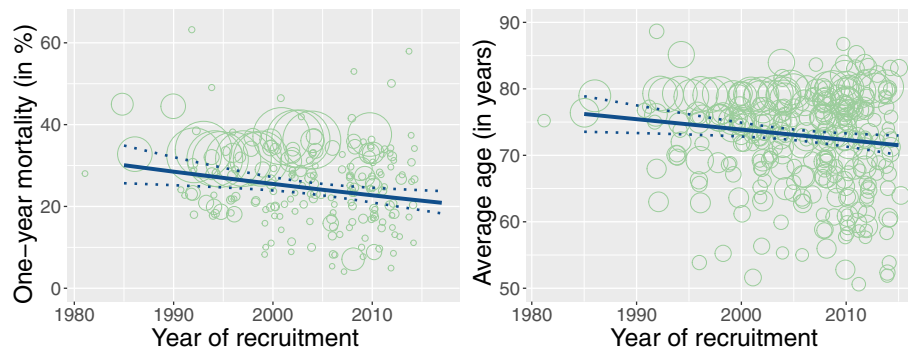
where  $k$  denotes the number of studies,  $y_i$  is a function of the effect measure,  $x_{1i}$  and  $x_{2i}$  denote the moderators,  $u_i$  the random effect and  $e_i$  the sampling error of the study  $i = 1, \dots, k$ . The  $u_i$  and  $e_i$  are usually assumed to be independent and follow normal distributions with  $u_i \sim \mathcal{N}(0, \tau^2)$  and  $e_i \sim \mathcal{N}(0, v_i)$ .<sup>8</sup> In the following analysis, the parameter vector  $\beta = (\beta_0, \beta_1, \beta_2, \beta_{12})^T$  is estimated via weighted least squares regression using inverse variance weights.<sup>9</sup> The variance of the random effect  $\tau^2$  is estimated via restricted maximum likelihood estimation (Reference [10], pp. 372–373) since it has been recommended as a good choice of estimator for continuous outcomes in other research.<sup>11</sup> The confidence intervals for the moderators are calculated using the Knapp-Hartung method,<sup>12</sup> because it was shown to be more accurate compared to other covariance estimators in simulation studies.<sup>13,14</sup> The models with only one and only two moderators in Section 3 can be seen as a special case of the model in Equation (1). All analyses were conducted in R 4.1.1<sup>15</sup> using the `metafor` package.<sup>16</sup> The R-Code will be made publicly available.

**3 | META-REGRESSIONS FOR ACUTE HEART FAILURE STUDIES**

The research synthesis by Kimmoun et al.<sup>3</sup> included 285 studies on acute heart failure, published between 1998 and 2017. Outcome measures were 30-day and 1-year readmission rates and 30-day and 1-year mortality in 108, 61, 148 and 204 of the studies, respectively. Furthermore, study characteristics like the median year of recruitment, the number of patients for the follow-up and the average age of the patients (in 260 studies) were reported. Other variables provided information about the medication and medical history of the patients. For more information, we refer to Kimmoun et al.<sup>3</sup>; the authors made their dataset publicly available.

**3.1 | Meta-regression analysis with a univariable model**

A major finding of the analysis of Kimmoun et al.<sup>3</sup> was a statistically significant decline in the 1-year mortality over calendar time. The decline is shown in the left panel



**FIGURE 1** Two models from the analysis by Kimmoun et al.<sup>3</sup>: Meta-regression model of year of recruitment for the 1-year mortality (left) and weighted linear model of the year of recruitment for the average age with the logarithmic study size as weights (right). Point sizes are proportional to the respective study sizes. Straight lines are the models' prediction, dotted lines the corresponding 95% CIs. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

of Figure 1 (adopted from figures 1A and 2B in Reference [3]). A meta-regression with the logit transformed one-year mortality as the outcome variable  $y_i$  from Equation (1) and the year of recruitment as the only explanatory variable  $x_{1i}$  was conducted for  $k = 204$  studies. The estimates of  $\beta_0$  and  $\beta_1$  are  $\hat{\beta}_0 = 29.432$  (95%-CI: [7.328, 51.537]) and  $\hat{\beta}_1 = -0.015$  (95%-CI: [-0.263, -0.042]), so a significant decline in mortality over calendar time was observed. However, the average age is decreasing in the recruitment year as well. In the right plot of Figure 1, it is shown that the average age of the participants decreased by 1.56 years every 10 years. Therefore, it is of interest whether the observed effect of the recruitment year is confounded by the average age or if there is an interaction between those two variables.

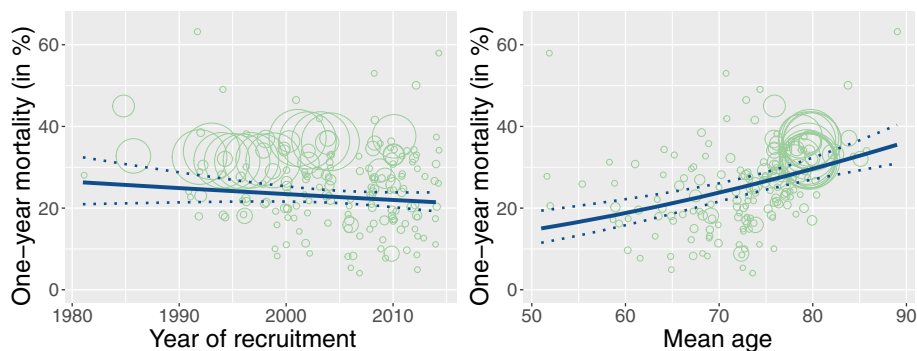
### 3.2 | Meta-regression analysis with a two-variable model

In order to check for possible confounding, we considered a model with two moderators. In this model, the (logit transformed) one-year mortality was regressed on the year of recruitment ( $x_{1i}$  in Equation (1)) and the average age ( $x_{2i}$ ) for 181 studies. The estimated intercept ( $\hat{\beta}_0 = 12.9584$ , 95%-CI: [-11.0633, 36.980]) and effect of the recruitment year turned out to be not significant ( $\hat{\beta}_1 = -0.0081$ , 95%-CI: [-0.0200, 0.0038]), while the estimated effect of the average age was significant ( $\hat{\beta}_2 = 0.0299$ , 95%-CI: [0.0178, 0.0420]). This suggests, that in the univariable model the effect of the recruitment year was in fact confounded by the variable average age. This is also visible in Figure 2, which shows the predicted effect of each explanatory variable at the median level of the respective other variable. Hence, fitting a mixed-effects meta-regression model which also includes the

average age as a moderator reveals that the apparent time trend is essentially captured by the changes in age over calendar time. Their observed time trend seems to stem from the neglected confounding variable average age, which was not considered in the meta-regression model of Kimmoun et al.<sup>3</sup> with one moderator only. However, it could be that the recruitment year influences the mortality not directly but via an interaction between the recruitment year and the average age.

### 3.3 | Meta-regression analysis with interaction

To check this, we considered a model with two moderators and their interaction, where we used centred explanatory variables to reduce the impact of multicollinearity (Reference [17], pp. 28–31). In the model the (logit transformed) one-year mortality was regressed on the year of recruitment ( $x_{1i}$  in Equation (1)), the average age ( $x_{2i}$ ) and their interaction for 181 studies. While the average age has a significant effect on the one-year mortality ( $\hat{\beta}_2 = 0.0333$ , 95%-CI: [0.0208, 0.0457]) the year of recruitment turned out to be not significant ( $\hat{\beta}_1 = -0.0066$ , 95%-CI: [-0.0185, 0.0052]) in the meta-regression model with two moderators and their interaction. However, the interaction term ( $\hat{\beta}_{12} = -0.0018$ , 95%-CI: [-0.0035, -0.0001]) and intercept ( $\hat{\beta}_0 = -1.1477$ , 95%-CI: [-1.2271, -1.0684]) are significant at level 0.05 as well. This suggests, that the recruitment year does not influence the one-year mortality directly, but via an interaction with the average age. In Figure 3, it is shown how this interaction influences the predicted one-year mortality. In the left panel, the predicted time trend in mortality is shown for a comparatively low average age of 60.5, while in the left panel it is shown for a higher average age of 79.5. The average ages



**FIGURE 2** Trends of year of recruitment (left) and average age (right) on the 1-year mortality in a mixed-effects meta-regression model with these two moderators. Point sizes are proportional to the respective study sizes. Straight lines are the models' prediction at the median level of the respective other explanatory variable, dotted lines the corresponding 95% CIs. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 3** Trends of year of recruitment for an average age of 60.5 (left) and 79.5 (right) years in a mixed-effects meta-regression model with two moderators and their interaction. Point sizes are proportional to the respective study sizes. Straight lines are the models' prediction, dotted lines the corresponding 95% CIs. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

60.5 and 79.5 correspond to the lower and upper quartile among all average ages of the included studies, respectively. While the time trend is increasing for studies with a lower average age, it is decreasing for studies with a higher average age. Hence, this model leads to the conclusion that there is a time trend present in the data, which depends on the age of the patients incorporated in the study.

## 4 | DISCUSSION

Motivated by a meta-regression on a temporal trend in mortality of acute heart failure patients,<sup>3</sup> we considered the impact of a neglected confounding variable and interaction in a meta-regression. Re-analysing the study of Kimmoun et al.,<sup>3</sup> we showed that the time trend found in the study is questionable. Considering a model, where not only the median year of recruitment but also the average age is included as a moderator, leads to the conclusion that the time trend is confounded by the effect of the patients' age. However, when the interaction of those variables is included as well, the results indicate that there is in fact a time trend, but the time trend depends on the age of the patients. This shows the importance of

including possible confounding variables and their interactions in meta-regression models. In the supplement to this paper we also conducted a small simulation study, where we analysed the impact of omitted and redundantly included interactions more generally. The simulation settings are based on the acute heart failure data above. The results are analogous to insights from multiple-regression literature, that is omitting regressors leads to biased estimates, while including unnecessary regressors causes more variance (see e.g. Reference [4], pp. 99–101). Finally, we would like to stress that limitations in software<sup>18</sup> can be another potential reason why most meta-regressions studies only include one-covariate at a time.

A limitation of this paper is that the conducted meta-regressions include 204 and 181 studies, respectively. This is much more than is common in many meta analyses.<sup>1</sup> We conducted another simulation study that showed that with a subsample of only 30 studies, the significant impact of the interaction term often cannot be detected. For this simulation, we refer to the supplement as well. Nevertheless, based on our results, we suggest to always include possible confounders and interactions in meta-regression models. Although the ideal modelling approach remains unclear, this work has shown the dangers of using overly

simple univariable meta-regression models for statistical inference. Even though many applications will have a much smaller sample size than in the analysis above, considering possible confounding variables and interactions may generally improve insights into the underlying data.

## AUTHOR CONTRIBUTIONS

**Eric Samuel Knop:** Data curation; formal analysis; methodology; software; visualization; writing – original draft; writing – review and editing. **Markus Pauly:** Conceptualization; funding acquisition; project administration; supervision; writing – review and editing. **Tim Friede:** Conceptualization; funding acquisition; supervision; writing – review and editing. **Thilo Welz:** Formal analysis; methodology; supervision; writing – original draft; writing – review and editing.

## ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

## FUNDING INFORMATION

This work was supported by the German Research Foundation projects, Grant no. FR 3070/3-1 (Tim Friede) and PA-2409 7-1 (Markus Pauly).

## CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

## DATA AVAILABILITY STATEMENT

The R-script used for the analysis of the data example and the simulation study will be made publicly available on [osf.io](https://osf.io/t5zsc/) under <https://osf.io/t5zsc/>. The dataset from Kimmoun et al.<sup>3</sup> was made publicly available by the authors themselves under <https://osf.io/cxv5k/>.

## ORCID

Eric S. Knop  <https://orcid.org/0000-0001-6134-1624>

Markus Pauly  <https://orcid.org/0000-0002-0976-7190>

Tim Friede  <https://orcid.org/0000-0001-5347-7441>

Thilo Welz  <https://orcid.org/0000-0001-6223-5698>

## REFERENCES

- Davey J, Turner RM, Clarke MJ, Higgins J. Characteristics of meta-analyses and their component studies in the cochrane database of systematic reviews: a cross-sectional, descriptive analysis. *BMC Med Res Methodol*. 2011;11(1):1-11. doi:10.1186/1471-2288-11-160
- Andreas S, Röver C, Heinz J, Straube S, Watz H, Friede T. Decline of COPD exacerbations in clinical trials over two decades—a systematic review and meta-regression. *Respir Res*. 2019;20(1):1-11. doi:10.1186/s12931-019-1163-2
- Kimmoun A, Takagi K, Gall E, et al. Temporal trends in mortality and readmission after acute heart failure: a systematic

review and meta-regression in the past four decades. *Eur J Heart Fail*. 2021;23:420-431. doi:10.1002/ejhf.2103

- Greene WH. *Econometric Analysis, Ebook*. global ed. Pearson Deutschland; 2019 <https://elibrary.pearson.de/book/99.150005/9781292231150>
- Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*. 1986;15(3):413-419.
- Friedrich S, Antes G, Behr S, et al. Is there a role for statistics in artificial intelligence? *Advances in Data Analysis and Classification*. 2022;16(4):823-846.
- Durlak JA, Lipsey MW. A practitioner's guide to meta-analysis. *Am J Community Psychol*. 1991;19:291-332.
- Raudenbush S. Analyzing effect sizes: random-effects models. In: Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation; 2009:295-315.
- Berkey C, Hoaglin D, Mosteller F, Colditz G. A random-effects regression model for meta-analysis. *Stat Med*. 1995;14:395-411. doi:10.1002/sim.4780140406
- Fahrmeir L, Kneib T, Lang S, Marx B. *Regression: Models*. Springer-Verlag, Berlin; 2013. doi:10.1007/978-3-642-34333-9
- Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods*. 2016;7(1):55-79. doi:10.1002/jrsm.1164
- Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med*. 2003;22(17):2693-2710. doi:10.1002/sim.1482
- Viechtbauer W, López-López J, Sanchez-Meca J, Marin-Martínez F. A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychol Methods*. 2015;20:360-374. doi:10.1037/met0000023
- Welz T, Pauly M. A simulation study to compare robust tests for linear mixed-effects meta-regression. *Res Synth Methods*. 2020;11(3):331-342. doi:10.1002/jrsm.1388
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021 <https://www.R-project.org/>
- Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1-48. <https://www.jstatsoft.org/v36/i03/>
- Aiken L, West S, Reno R. *Multiple Regression: Testing and Interpreting Interactions*. SAGE Publications; 1991.
- Tipton E, Pustejovsky JE, Ahmadi H. Current practices in meta-regression in psychology, education, and medicine. *Res Synth Methods*. 2019;10(2):180-194.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Knop ES, Pauly M, Friede T, Welz T. The consequences of neglected confounding and interactions in mixed-effects meta-regression: An illustrative example. *Res Synth Meth*. 2023;14(4):647-651. doi:10.1002/jrsm.1643