arXiv:2301.03244v1 [stat.ME] 9 Jan 2023

混合效应荟萃回归中被忽略的混杂和交互作用的影响

Eric S. Knop[1],Markus Pauly[1 2],Tim Friede[3 4] 和 Thilo Welz[1]

2023年1月10日

摘要 分析人员很少在荟萃回归模型中纳入交互项，而当交互作用存在时，这可能会引入偏倚。本文通过重新分析一项关于急性心力衰竭的研究实例来说明这一点；在该实例中，忽略交互作用可能导致错误的推断和结论。此外，我们基于这一实例开展了一项简短的模拟研究，以突出省略交互项或不必要地纳入交互项所造成的影响。根据研究结果，我们建议，只要交互作用具有合理性，就应始终在混合效应荟萃回归模型中纳入交互项。

关键词： 急性心力衰竭，交互作用，荟萃回归，模拟

1 引言

在荟萃回归中，反映效应调节变量的交互作用常常被忽略。当然，在许多情况下，忽略交互作用是因为可用研究数量较少 [Davey 等， 2011]。但即使有足够的研究可用，人们也常常不考虑交互作用（例如 Andreas 等[2019]；Kimmoun 等人 [2021]）。正如我们从普通最小二乘回归中所知，这会导致有偏估计，而纳入冗余调节变量和交互作用只会增加估计的方差（Greene [2019], 第 99–101 页）。

在我们的研究中，我们通过重新分析一项近期的荟萃分析来阐释这些观点，其中包括混合效应荟萃回归分析 [Kimmoun 等人， 2021]。Kimmoun 等人 [2021] 分析了临床研究中急性心力衰竭患者的死亡率和再住院情况。他们发现，死亡率随日历时间推移而显著下降。然而，患者的平均年龄也随日历时间推移而下降。这表明，观察到的趋势至少部分可能是

[1]统计系，多特蒙德工业大学[2]鲁尔大学联盟，可信数据科学与安全研究中心，多特蒙德[3]医学统计系，哥廷根大学[4]德国心血管研究中心（DZHK），哥廷根合作站点

由忽略平均年龄与中位招募年份之间的交互作用所致。因此，我们进行了荟萃回归，将中位招募年份、平均年龄及其交互作用纳入模型。此外，我们开展了一项模拟研究，模拟该数据示例，以评估模型设定错误的影响。在模拟中，我们分别假设仅将招募年份作为调节变量的模型，或包含两个调节变量及其交互作用的模型，是真实模型。在这两种情形下，我们分别拟合了包含（i）一个调节变量、（ii）两个调节变量，或（iii）两个调节变量及其交互作用的模型。我们根据经验覆盖概率、区间长度中位数以及系数估计量的经验偏倚，对调节变量的置信区间进行了比较。

2 方法

一个混合效应 meta 回归模型可表示为:

模型包含两个调节变量和一个交互项，其形式如下

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}x_{2i}\beta_{12} + u_i + e_i, i = 1, ..., k \tag{1}$$

其中， k 表示研究数量， y_i 是效应量的函数， x_{1i} 和 x_{2i} 表示调节变量， u_i 表示随机效应， e_i 表示研究 $i = 1$，...， k的抽样误差。通常假定 u_i 和 e_i 相互独立且服从正态分布，其中$u_i \sim$ N $(0, \tau^2)$， $e_i \sim$ N $(0, v_i)$ [Raudenbush，2009]。在以下分析中，参数向量 $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_{12})^T$ 通过加权最小二乘回归进行估计，并使用逆方差权重 [Berkey 等， 1995]。随机效应 τ^2 的方差采用限制性最大似然估计进行估计（Fahrmeir 等， [2013], 第 372–373 页），因为其他研究已推荐将其作为连续结局的一种良好估计量 [Veroniki 等， 2016]。调节变量的置信区间采用 Knapp-Hartung 方法计算 [Knapp 和 Hartung， 2003], ，因为模拟研究表明，与其他协方差估计量相比，该方法更准确（Viechtbauer 等， [2015]; Welz 和 Pauly [2020]）。第 3 节和第 4 节中仅包含一个和仅包含两个调节变量的模型，可视为方程 (1) 中模型的特殊情形。所有分析均使用 `R 4.1.1` [R 核心团队， 2021] 借助 `metafor`程序包 [Viechtbauer， 2010]完成。R 代码将公开发布。

3 急性心力衰竭研究的荟萃回归

Kimmoun 等人 [2021] 的研究综合纳入了285项关于急性心力衰竭的研究，这些研究发表于1998年至2017年间。结局指标包括30天和一年再入院率，以及30天

和一年死亡率，分别见于108、61、148和204项研究。此外，研究还报告了研究特征，例如招募年份中位数、随访患者人数，以及患者的平均年龄（见于260项研究）。其他变量提供了患者用药和病史的信息。更多信息请参阅Kimmoun 等人 [2021]的研究；作者已将其数据集公开。

3.1 使用单变量模型的荟萃回归分析

Kimmoun 等人 [2021] 的分析有一项重要发现：一年死亡率随日历时间推移而显著下降。图1左图展示了这一下降趋势（改编自 Kimmoun 等人 [2021]的图2B和图1A）。研究人员针对 $k = 204$ 项研究进行了荟萃回归，以经 logit 变换的一年死亡率作为结果变量 y_i （见方程（1）），并仅将招募年份作为解释变量 x_{1i}。 β_0 和 β_1 的估计值分别为$\hat{}\hat{}\beta_0 = 29.432$ （95%-CI：[7.328，51.537]）和 $\beta_1 = -0.015$ （95%-CI：[-0.263，-0.042]）。然而，平均年龄也随着招募年份的增加而降低。图1右图显示，参与者的平均年龄每十年下降.56岁。因此，我们有必要考察在模型中同时纳入平均年龄及其交互作用后，招募年份的效应会如何变化。

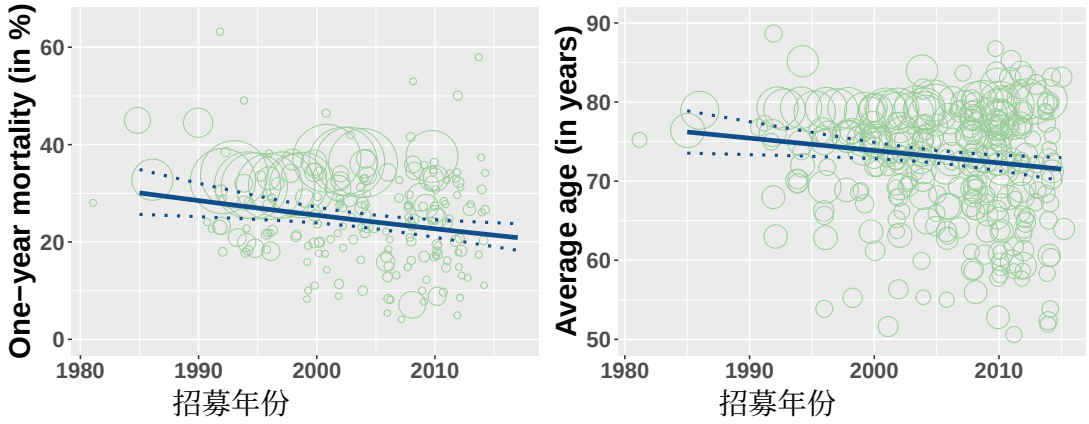


图1：Kimmoun 等人 [2021] 分析中的两个模型：一年死亡率的招募年份荟萃回归模型（左），以及以研究规模的对数为权重、针对平均年龄的招募年份加权线性模型（右）。点的大小与相应研究规模成正比。实线表示模型预测，虚线表示相应的95%置信区间。

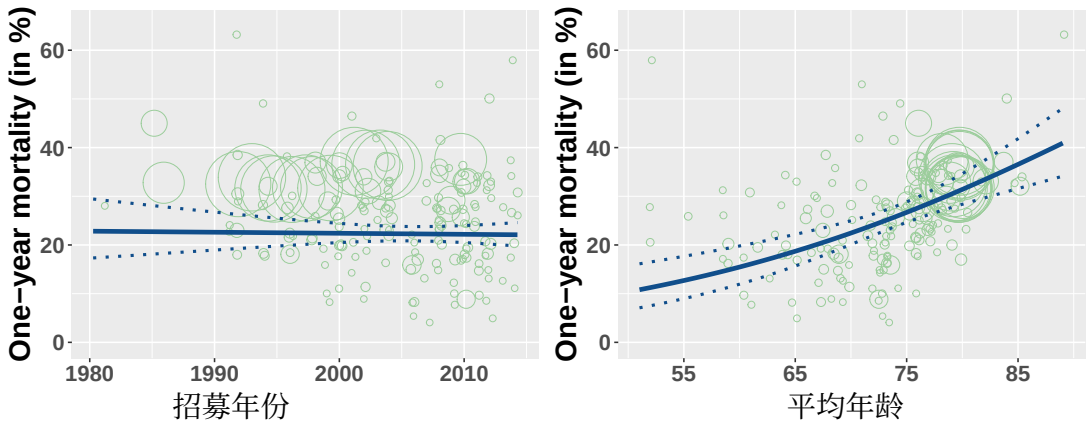


图2：在包含这两个调节变量及其交互作用的混合效应荟萃回归模型中，招募年份（左）和平均年龄（右）对一年死亡率的影响趋势。点的大小与相应研究规模成正比。实线表示另一个解释变量取其中位数时模型的预测值，虚线表示相应的95%置信区间。

3.2 含交互作用的元回归分析

为此，我们考虑了一个包含两个调节变量及其交互作用的模型，并使用中心化的解释变量，以减小多重共线性的影响（Aiken等人 [1991], 第28-31页）。在该模型中，我们以181项研究为基础，将（经 logit 变换的）一年死亡率对招募年份（方程（1）中的x_{1i}）、平均年龄（x_{2i}）及二者的交互作用进行回归。平均年龄对一年死亡率有显著影响（$\hat{}\beta_2 = 0.0333$，95%-CI：$[0.0208, 0.0457]$），但在包含两个调节变量及其交互作用的荟萃回归模型中，招募年份的效应并不显著（$\hat{}\beta_1 = -0.0066$，95%-CI：$[-0.0185, 0.0052]$）。图2展示了当另一变量处于其中位数水平时，各解释变量的预测效应。交互项（$\hat{}\beta_{12} = -0.0018$，95%-CI：$[-0.0035, -0.0001]$）和截距（$\hat{}\beta_0 = -1.1477$，95%-CI：$[-1.2271, -1.0684]$）在0.05显著性水平上也显著。因此，拟合包含交互作用的混合效应荟萃回归模型后可见，表面上的时间趋势实质上是由日历时间推移过程中年龄的变化所解释的。这一点从图2中已可看出，而且与 Kimmoun 等人 [2021]得出的结论相反。他们观察到的时间趋势似乎源于忽略了平均年龄与招募年份之间的相关性；他们的荟萃回归模型仅纳入了一个调节变量，未考虑这种相关性。

4 模拟

为说明忽略交互项会对模型产生多大影响，我们开展了一项小型模拟研究，模拟 Kimmoun 等人 [2021] 中的情境。我们假设含交互作用模型中的估计参数就是真实参数，即假设$\beta = (-1.1477,\ -0.0066,\ 0。0333,\ -0.0018)^T$ 为回归系数，$k = 181$ 为研究数量，且 $\tau^2 = 0.2484$。研究规模 $n_i,\ i = 1,\ .....,\ k$、中心化招募年份、平均年龄以及研究 i 的交互作用，也假设与数据示例相同。通过方程 $\theta_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}x_{2i}\beta_{12} + u_i$，我们生成真实的 logit 变换死亡率$\theta_i$，其中 u_i 取自均值为 0、方差为 τ^2的正态分布。因此，假设研究 i 中一年内死亡的概率为 $p_i = \exp(\theta_i)/(1 + \exp(\theta_i))$。研究 i 中的死亡人数 d_i 是从参数为 n_i 和 p_i的二项分布中抽取的。此外，观测到的 logit 变换死亡率 y_i 设为

$$y_i = \log\left(\frac{d_i + 0.5}{n_i - d_i + 0.5}\right),$$

其抽样方差为 $v_i = (d_i + 0.5)^{-1} + (n_i - d_i + 0.5)^{-1}$，参见 Sutton 等人，第 18 页。请注意，$d_i/(n_i - d_i)$ 等同于率之比。在两个方程中都加上了 0.5，以避免除以 0，参见 Viechtbauer [2007]。

4.1 忽略调节变量及交互作用的影响

根据这一流程，我们生成了10,000个数据集，并分别拟合了以下模型：(i) 包含两个调节变量及其交互作用的模型；(ii) 包含两个调节变量但不含交互作用的模型；(iii) 仅包含一个调节变量的模型。对于所有模型，我们还计算了估计系数的95%置信区间。相应的经验覆盖率、区间长度中位数和偏差见表1。仅包含一个调节变量的模型对两个参数的覆盖率都较低，尤其是 β_1，真实参数仅在77.2%的情况下落入区间。对于包含两个调节变量的模型，β_0 的覆盖率为89.8%，β_2 的覆盖率为90.6%，也低于名义置信水平。在包含交互作用的模型中，所有系数在约95%的情况下均落入区间。对于 β_0 和 β_1，仅包含一个调节变量的模型区间最长。包含两个调节变量的模型所得区间通常略短于包含交互作用的模型。此外，不含交互作用的模型绝对偏差更大。对于 β_1，包含两个调节变量的模型中的偏差绝对值几乎是包含交互作用的模型中的三倍。仅包含一个调节变量的模型中的偏差则再次超过包含两个调节变量的模型中偏差的五倍。

4.2 冗余调节变量和交互作用的影响

上述结果表明，当潜在调节效应和交互效应确实存在时，将其纳入模型十分重要。接下来，我们将考察在仅存在单一调节效应的情形下，采用更复杂建模可能带来的缺点。为此，我们生成了10,000个单变量模型，并分别采用包含一个调节变量、两个调节变量，或两个调节变量及其交互作用的模型进行估计。为了使本次模拟能够与第4.1节中的模拟相比较，我们使用中心化调节变量，并仅采用纳入交互作用模型的181项研究来获得参数估计。因此，我们设定 $\tau^2 = 0.2935$ 和$\beta = (-1.1497, -0.0148, 0, 0)^T$，即模型中仅包含截距和一个调节变量。所有估计模型的经验覆盖率、区间长度中位数和偏差见表2。这里，包含交互作用的模型的覆盖率仅略低于真实模型（该模型仅包含一个调节变量）。对于 β_0 和 β_1 ，与仅包含一个调节变量的模型相比，包含交互作用的模型的区间长度中位数略大。包含交互作用的模型中 β_0 的偏差也略大。至于 β_1 的偏差，两个模型并无差异。

5 讨论

受一项关于急性心力衰竭患者死亡率时间趋势的荟萃回归启发[Kimmoun 等人，2021], 我们考察了在

表1：不同模型的覆盖率（名义水平为0.95）、区间长度中位数和偏差（真实模型为含交互作用的模型）

调节变量/模型		one	two	两个及交互作用
覆盖率	β_0	0.9159	0.8979	0.9447
	β_1	0.7718	0.9547	0.9495
	β_2	-	0.9062	0.9527
	β_{12}	-	-	0.9502
区间长度	β_0	0.1615	0.1517	0.1535
	β_1	0.0240	0.0231	0.0230
	β_2	-	0.0239	0.0242
	β_{12}	-	-	0.0034
Bias	β_0	0.0282	0.0264	0.0077
	β_1	-0.0078	-0.0012	0.0001
	β_2	-	-0.0038	0.0000
	β_{12}	-	-	-0.00004

表2：当仅含一个调节变量的模型为真实模型时，不同模型的覆盖率（名义水平为0.95）、区间长度中位数和偏差

调节变量/模型		one	two	两个调节变量和交互作用
覆盖率	β_0	0.9431	0.9495	0.9504
	β_1	0.9495	0.9481	0.9501
	β_2	-	0.9544	0.9539
	β_{12}	-	-	0.9518
区间长度	β_0	0.1619	0.1618	0.1663
	β_1	0.0240	0.0247	0.0249
	β_2	-	0.0253	0.0262
	β_{12}	-	-	0.0037
Bias	β_0	0.0087	0.0092	0.0071
	β_1	0.0001	0.0003	0.0004
	β_2	-	-0.0002	-0.0003
	β_{12}	-	-	0.0000

荟萃回归中遗漏交互作用和不必要地拟合交互作用的影响。重新分析 Kimmoun 等人的研究 [2021]，后，我们发现，该研究中观察到的时间趋势似乎是由忽略了与患者年龄的交互作用所致。通过一项小型模拟研究，我们发现，拟合更复杂的模型会导致置信区间更宽；而忽略交互作用则可能导致估计产生较大偏倚，并使覆盖率过高。因此，我们建议，只要拿不准，就应始终在模型中纳入交互作用。否则，可能会得出错误的结论。

本文所展示的模拟受到所讨论示例的启发，因此范围有限。尽管如此，似乎可以合理推断，对于设定不同的模型也能得出相同结论，因为模型误设的影响与普通最小二乘回归中已知的影响相似。为验证这一研究假设，有必要开展一项参数和模型设定各异的扩展模拟研究，未来将进行这项研究。

财务披露

本研究得到德国研究基金会项目的资助（资助编号：FR 3070/3-1（Tim Friede）和 PA-2409 7-1（Markus Pauly））。

利益冲突

作者声明不存在潜在利益冲突。

数据可用性声明

用于分析数据示例和模拟研究的 R脚本将在 osf.io 上公开发布，网址为 https://osf.io/t5zsc/。Kimmoun 等人 [2021]的数据集由作者本人公开发布，网址为 https://osf.io/cxv5k/。

要点

目前我们已知：

- 在荟萃回归中，调节变量的交互项往往未被考虑。
- 在回归模型中遗漏重要变量可能导致有偏估计，而纳入冗余变量则会增加估计量方差。

创新之处：

- 通过重新分析一项关于急性心力衰竭的荟萃分析，本研究说明在荟萃回归模型中忽略交互作用会如何改变结果。
- 研究开展了一项小规模模拟研究，展示在荟萃回归中，忽略调节变量和交互效应，以及不必要地纳入它们，会如何影响置信区间估计。

对作者所在领域之外的 RSM 读者可能产生的影响：

- 结果表明，当交互作用具有合理可能性时，荟萃回归应始终纳入交互项，因为否则参数估计的置信区间可能覆盖率不佳。

参考文献

L.S. Aiken、S.G. West 和 R.R. Reno。《多元回归：交互作用的检验与解释》。SAGE Publications，纽伯里帕克，加利福尼亚州，纽伯里帕克，加利福尼亚州，1991 年。ISBN 9780761907121。

Stefan Andreas、Christian Röver、Judith Heinz、Sebastian Straube、Henrik Watz 和 Tim Friede。二十年间临床试验中慢性阻塞性肺疾病急性加重的下降——系统综述与荟萃回归。《呼吸研究》，20(1):1–11，2019年。doi：10.1186/s12931-019-1163-2。

Catherine Berkey、David Hoaglin、F Mosteller 和 Graham Colditz。用于荟萃分析的随机效应回归模型。《医学统计学》，14:395–411，1995年2月。doi：10.1002/sim.4780140406。

Jonathan Davey、Rebecca M Turner、Mike J Clarke 和 Julian Higgins。Cochrane系统综述数据库中荟萃分析及其组成研究的特征：一项横断面描述性分析。《*BMC*医学研究方法学》，11(1):1–11，2011。doi: 10. 1186/1471-2288-11-160。

Ludwig Fahrmeir、Thomas Kneib、Stefan Lang 和 Brian Marx。回归：模型、方法与应用。施普林格出版社，柏林，2013年01月。ISBN 978-3-642-34332-2。doi: 10.1007/ 978-3-642-34333-9。

William H. Greene。计量经济学分析，电子书，全球版。培生德国，2019年。URL https://elibrary.pearson.de/book/99.150005/9781292231150。

Antoine Kimmoun、Koji Takagi、Emmanuel Gall、Shiro Ishihara、Pierre Hammoum、Nathan El Bèze、Alexandre Bourgeois、Guillaume Chassard、Hugo Pegorer-Sfes、Etienne Gayat 等。过去四十年急性心力衰竭患者死亡率和再入院率的时间趋势：系统综述与荟萃回归。《欧洲心力衰竭杂志》，23，2021。doi: 10.1002/ ejhf.2103。

Guido Knapp 和 Joachim Hartung。针对含单协变量的随机效应荟萃回归的改进检验。《医学统计学》，22(17):2693 – 2710，2003。doi: 10.1002/sim.1482。URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.1482。

R 核心团队。*R*：一种用于统计计算的语言与环境。R统计计算基金会，奥地利维也纳，2021。URL https://www.R-project.org/。

S.W. Raudenbush。《分析效应量：随机效应模型》，第 295–315 页。罗素·塞奇基金会，纽约州纽约市，2009 年 1 月。

Alex J Sutton、Keith R Abrams、David R Jones、David R Jones、Trevor A Sheldon 和 Fujian Song。《医学研究中荟萃分析的方法》。

Areti Angeliki Veroniki、Dan Jackson、Wolfgang Viechtbauer、Ralf Bender、Jack Bowden、Guido Knapp、Oliver Kuss、Julian PT Higgins、Dean Langan 和 Georgia Salanti。荟萃分析中研究间方差及其不确定性的估计方法。《研究综合方法》，7（1）：55–79，2016。doi：10.1002/jrsm.1164。URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1164。

Wolfgang Viechtbauer。《荟萃分析中异质性程度的置信区间》。《医学统计学》，26(1):37–52，2007。doi: 10.1002/sim.2514。

Wolfgang Viechtbauer。使用 metafor 软件包在 R 中进行荟萃分析。《统计软件杂志》，36(3):1–48，2010。URL https://www.jstatsoft.org/v36/i03/。

Wolfgang Viechtbauer、José López-López、Julio Sanchez-Meca 和 Fulgencio Marín-Martínez。混合效应荟萃回归模型中调节变量检验程序的比较。《心理学方法》，20，2015年8月。doi：10.1037/met0000023。

Thilo Welz 和 Markus Pauly。一项比较线性混合效应荟萃回归中稳健检验的模拟研究。《研究综合方法》，11(3):331–342，2020年。doi：10.1002/jrsm.1388。URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1388。