采用多层次结局缺失数据估计整群随机试验中的边际治疗效应

Chia-Rui Chang[1] 和 Rui Wang[1,2]

[1]哈佛大学陈曾熙公共卫生学院生物统计学系，马萨诸塞州，美国[2]哈佛朝圣者健康护理研究所及哈佛医学院人口医学系，马萨诸塞州，美国

arXiv:2304.05583v1 [stat.ME] 12 Apr 2023

摘要

整群随机试验（整群随机试验）的分析可能会因信息性结局数据缺失而变得复杂。已有研究提出了加权广义估计方- 程等方法，通过对每个簇中观测到的个体结局数据进行加权，来处理信息性缺失。这些现有方法主要针对个体层面发生缺失且每个簇的个体结局部分或完全观测到的设定。在存在缺失集群的情况下，例如由于集群脱落而导致某个簇的所有结局均缺失，这些方法实际上忽略了整群层面缺失；如果整群层面缺失具有信息性，就可能导致有偏推断。在具有多层次结构的整群随机试验中，也可能发生多个层面的信息性缺失，此时研究参与者嵌套于医疗服务提供者等子整群中，而子整群又嵌套于诊所等簇中。在本文中，我们基于加权广义估计方程，提出了用于估计整群随机试验中边际治疗效应的新估计量，并处理多个层面的结局数据缺失问题。我们证明，只要倾向评分模型中的一组模型被正确设定，所提出的多层次多重稳健估计量就是一致的，并且渐近服从正态分布。我们通过广泛的模拟评估了所提方法的性能，并以一项在马达加斯加农村地区评估疟疾风险降低干预措施的整群随机试验为例，说明了该方法的应用。

关键词—整群随机试验；多层次缺失数据；广义估计方程（GEE）；逆概率加权（IPW）；倾向评分；多重稳健；期望最大化（EM）算法

1 引言

整群随机试验（CRT）将同一集群中的所有个体随机分配接受相同的治疗，常用于生物医学研究中的干预评估（Hayes and Moulton, 2017）。由于同一集群内个体的结局可能存在相关性，整群随机试验的分析必须考虑集群内的这种依赖关系。广义估计方程（GEE）方法常用于估计整群随机试验中的边际治疗效应（Liang and Zeger, 1986）。与混合效应模型相比，GEE以总体边际效应参数为目标，且对结局分布的参数假设要求较少（Hubbard et al., 2010）。只要均值模型设定正确，该方法便能给出有效推断，并且对相关结构的设定错误具有稳健性。然而，存在信息性结局数据缺失时，基于完整数据的广义估计方程估计量可能会导致估计偏倚（例如，Prague 等人（2016））。

这里我们考虑这样一种设定：结局缺失取决于完全观测的基线协变量和暴露，这在整群随机试验数据的分析中很常见（例如，参见Hossain等人（2017a,b））。这种缺失过程被称为“协变量依赖型缺失（CDM）”，它是随机缺失（MAR）机制的更强版本，因为该缺失过程被假定不受已观测结局的影响。处理协变量依赖性缺失数据的两种常见方法包括多层多重插补（MMI）（Schafer和 Yucel，2002；DiazOrdaz等人，2016；Hossain等人，2017a,b）和逆概率加权（IPW）（Robins等人，1995）。我们采用逆概率加权框架，这避免了正确指定集群结局联合分布的必要性。

已有研究提出了用于处理整群随机试验中个体结局缺失的逆概率加权方法（Prague 等人，2016；Chen 等人，2020）。这些方法适用于缺失发生在个体层面，且每个集群的个体结局部分或全部可观测的设定。当存在缺失集群时，例如由于集群退出而导致该集群的所有结局均缺失，这些方法会忽略整群层面缺失；如果整群层面缺失具有信息性，则可能导致有偏推断（Giraudeau and Ravaud，2009）。

整群随机试验中出现缺失集群并不少见。两项针对整群随机试验的系统综述报告称，132项试验中有18%、86项试验中有31%存在缺失集群（Diaz-Ordaz 等，2014；Fiero 等，2016）。在具有多层次结构的整群随机试验中也可能出现多层缺失，其中研究参与者嵌套于家庭等子整群中，而家庭又嵌套于地区中。例如，在一项评估主动社区病例管理（pro-CCM）是否能减轻马达加斯加农村疟疾流行地区疟疾负担的研究中，22个福孔塔尼（最小行政单位）被随机分配接受 pro-CCM 或常规综合社区病例管理（iCCM）（Ratovoson 等，2022）。研究参与者嵌套于家庭中，而家庭则嵌套于各个福孔塔尼中。约24%的研究参与者和22%的家庭因搬迁、缺席、死亡或拒绝参与而失访。

本文基于广义估计方程框架，开发了新的估计量，用于估计存在多层次结局缺失的整群随机试验中的边际治疗效应。我们推导出多层权重，以处理（子）集群层面和个体层面的信息性缺失。我们进一步将多重稳健估计方法（Han，2014）应用于多层次结局缺失设定，并提出多层多重稳健广义估计方程（MMR-GEE）估计量。这种新型估计量允许分析者设定多组倾向评分（PS）模型；只要其中一组模型设定正确，就能对边际治疗效应进行一致估计。为处理集群层面缺失指示变量可能存在的错误分类问题（即所有个体结局均缺失的集群可能被误分类为整簇脱落），我们开发了期望最大化（EM）算法来估计倾向评分模型的参数（Dempster 等，1977）。

本文其余部分安排如下。第 2.1 节介绍多层缺失设定的符号，并回顾未加权广义估计方程和逆概率加权广义估计方程方法。第 2.2 节描述多层缺失过程以及全文所采用的假设。第 2.3 节和第 2.4 节分别介绍多层逆概率权重和我们提出的 MMR-GEE 估计量。第 2.5 节阐述 MMR-GEE 估计量的理论性质。第 2.6 节探讨集群层面观测到的缺失指示变量的错误分类问题。为简化符号，在第 2.1 至 2.6 节中，我们以两级整群随机试验为例展开讨论，其中信息性缺失可能发生在集群层面和个体层面。在第 2.7 节中，我们将方法推广至三级整群随机试验，其中信息性缺失可能发生在子集群层面和个体层面。第 3 节报告了广泛模拟研究的结果。第 4 节通过一项三级整群随机试验——“马达加斯加农村地区主动社区病例管理”研究（Ratovoson 等，2022年）——展示所提方法的应用。第 5 节总结全文，并讨论实际应用中的考量。

2 方法

2.1 符号与模型

我们考虑一个双臂平行整群随机试验，其结局为 Y_{ij}，集群层面基线协变量 P 的向量为 $\boldsymbol{Z}_i = (Z_i^1, Z_i^2, \ldots, Z_i^P)^{\mathrm{T}}$，个体级基线协变量 Q 的向量为 $\boldsymbol{X}_{ij} = (X_{ij}^1, X_{ij}^2, \ldots, X_{ij}^Q)^{\mathrm{T}}$，对应集群 $i = 1, ..., M$ 中的受试者 $j = 1, ..., n_i$。令 $A_i \in \{0, 1\}$ 表示集群 i 的二元治疗指标（治疗组为 $A_i = 1$，对照组为 $A_i = 0$）；治疗分配概率已知，为 $P(A_i = 1) = p_A$。假设集群层面协变量向量 $\boldsymbol{Z}_i$ 和个体层面协变量矩阵 $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots, \boldsymbol{X}_{in_i})^{\mathrm{T}}$ 在随机化前均被完整观测。这里，我们用 $\boldsymbol{R}_i = (R_{i1}, R_{i2}, \ldots, R_{in_i})^{\mathrm{T}}$ 表示个体层面缺失指示变量向量，并用 C_i 表示结局 $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^{\mathrm{T}}$的集群层面缺失指示变量。当观测到 Y_{ij}时，$R_{ij} = 1$；当 Y_{ij} 缺失时，$R_{ij} = 0$。如果集群 i 退出研究，导致该集群中没有任何

个体结局可以被观测到，则 $C_i = 1$；否则，

令 $s = \sum_{i=1}^{M} C_i$ 表示观察到的集群数量，$m_i = \sum_{j=1}^{n_i} R_{ij}$ 表示集群 i 中观察到的结局数量。不失一般性，令 $i = 1, ..., s$ 为观察到的集群的索引，$i = s+1, ..., M$ 为缺失集群的索引；对于观察到的集群 i 中的参与者，令 $j = 1, ..., m_i$ 为结局已观察到的参与者的索引，$j = m_i + 1, ..., n_i$ 为结局缺失的参与者的索引。表1展示了多层缺失设定下数据结构的示例。

表1：整群随机试验中多层缺失的数据结构，包括结局 Y_{ij}、集群层面缺失指示变量 C_i（若集群被观测到则为 $C_i = 1$，若集群缺失则为 $C_i = 0$），以及个体层面缺失指示变量 R_{ij}（若观测到 Y_{ij} 则为 $R_{ij} = 1$，若 Y_{ij} 缺失则为 $R_{ij} = 0$）

集群	Unit	Y_{ij}	R_{ij}	C_i
1	1	Y_{11}	1	1
1	⋮	⋮	1	1
1	m_1	Y_{1m_1}	1	1
1	m_1+1	$Y_{1(m_1+1)}$	0	1
1	⋮	⋮	0	1
1	n_1	Y_{1n_1}	0	1
⋮	⋮	⋮	⋮	1
s	1	Y_{s1}	1	1
s	⋮	⋮	1	1
s	m_s	Y_{sm_s}	1	1
s	m_s+1	$Y_{s(m_s+1)}$	0	1
s	⋮	⋮	0	1
s	n_s	Y_{sn_s}	0	1
s+1	1	$Y_{(s+1)1}$	0	0
s+1	⋮	⋮	0	0
s+1	n_{s+1}	$Y_{(s+1)n_{s+1}}$	0	0
⋮	⋮	⋮	0	0
M	1	Y_{M1}	0	0
M	⋮	⋮	0	0
M	n_M	Y_{Mn_M}	0	0

我们主要关注边际均值模型 $g(\boldsymbol{\mu}_{ij}(\boldsymbol{\beta}, A_i)) = g(\mathbb{E}[Y_{ij}|A_i]) = \beta_I + \beta_A A_i$，其链接函数为 $g(.)$，并据此对参数 $\boldsymbol{\beta} = (\beta_I, \beta_A)^{\mathrm{T}}$进行推断。对于连续结局和恒等链接函数，$\beta_A$ 对应于均值差边际治疗效应。没有缺失数据时，可以通过求解以下估计方程（Liang 和 Zeger，1986）得到 $\boldsymbol{\beta}$ 的估计量：

$$\sum_{i=1}^{M} \underbrace{\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i))}_{=U_i(\boldsymbol{Y}_i, A_i; \boldsymbol{\beta})} = 0. \tag{1}$$

$\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)}{\partial \boldsymbol{\beta}}$ 是设计矩阵，其中包含 $\boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i) = \{\boldsymbol{\mu}_{i1}(\boldsymbol{\beta}, A_i) \ldots, \boldsymbol{\mu}_{in_i}(\boldsymbol{\beta}, A_i)\}^{\mathrm{T}}$。 $\boldsymbol{V}_i = \boldsymbol{F}_i^{1/2}\boldsymbol{C}(\boldsymbol{\alpha})\boldsymbol{F}_i^{1/2}$ 是协方差矩阵，其中 $\boldsymbol{F}_i = \mathrm{diag}\{\mathrm{var}(y_{ij})\}$。 $\boldsymbol{C}(\boldsymbol{\alpha})$ 是工作相关矩阵，其索引为非对角元素 $\boldsymbol{\alpha}$。

当结局在协变量依赖性缺失（CDM）机制下缺失时，使用完整数据拟合模型（1）可能会导致对 $\boldsymbol{\beta}$ 的推断产生偏倚（ Prague 等人，2016 ）。若所有集群均可观测，即对于所有 i， $C_i = 1$ ，则可以尝试通过逆概率加权广义估计方程（IPW-GEE）校正偏倚（Robins 等人，1995）：

$$\sum_{i=1}^{M} \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \boldsymbol{W}_i(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\theta})(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)) = 0. \tag{2}$$

模型（2）通过依据权重矩阵 $\boldsymbol{W}_i(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = \mathrm{diag}\left[{}_{\pi_{ij}\ A_i\ \boldsymbol{Z}_i}^{R_{ij}}(.,.,\boldsymbol{X}_{ij};\boldsymbol{\theta})\right]_{j=1,\ldots,n_i}$ 对完整数据重新加权，从而恢复总体矩。给定条件下 Y_{ij} 被观测到的概率也称为倾向评分（PS），记为 $\pi_{ij}(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij}; \boldsymbol{\theta}) \overset{def}{=} P(R_{ij} = 1|A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij})$。在实践中，真实的倾向评分未知，需要进行估计。一种常见方法是假设一个逻辑回归模型，以处理指示变量和基线协变量为自变量，对缺失指标进行回归。若倾向评分模型设定正确，则可获得 $\boldsymbol{\beta}$ 的一致且渐近正态（CAN）估计量。

2.2 多层缺失过程与假设

我们考虑以下多层缺失过程：簇在随机化后退出研究或撤出研究，其中这种整群层面缺失由模型 $\lambda_i(A_i, \boldsymbol{Z}_i; \boldsymbol{\gamma}) \overset{def}{=} P(C_i = 1|A_i, \boldsymbol{Z}_i)$ 所诱导，参数为 $\boldsymbol{\gamma}$。对于在整个研究期间均未退出的簇，个体参与者的结局可能缺失，其中这种个体级缺失由另一个模型 $\phi_{ij}(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij}|C_i = 1; \boldsymbol{\eta}) \overset{def}{=} P(R_{ij} = 1|C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij})$ 所诱导，参数为 $\boldsymbol{\eta}$。在此类设定下，基于至少部分观测到个体结局的簇的数据所采用的标准逆概率加权广义估计方程方法忽略了簇级缺失过程，因此可能导致对 $\boldsymbol{\beta}$ 的估计有偏。本文始终作出以下假设：

1)非信息性簇大小：我们假设簇大小与结局无关，但可能会影响缺失过程。

2) 多层次协变量依赖缺失：给定基线协变量和处理后，多层缺失过程既不依赖于已观测结局，也不依赖于缺失结局：$(C_i, \boldsymbol{R}_i) \perp\!\!\!\perp \boldsymbol{Y}_i|A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i$。

3) $\boldsymbol{C}_i$ 中关于 $\boldsymbol{X}_i$的条件交换性：给定集群层面协变量和处理指示变量后，簇级缺失过程独立于个体层面协变量：$C_i \perp\!\!\!\perp \boldsymbol{X}_i \mid A_i, \boldsymbol{Z}_i$。请注意， $\boldsymbol{Z}_i$ 可以包含 $\boldsymbol{X}_i$ 的汇总测量（例如集群的平均年龄、男性/女性比例），我们允许 C_i 依赖于此类汇总测量，但不依赖于某个特定的 X_{ij}。该假设意味着 P ($C_i = 1 \mid A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i) = \mathrm{P}\ (C_i = 1 \mid A_i, \boldsymbol{Z}_i)$ 4) 无 c

个体级缺失过程中的协变量干扰：对于参与者 j in

在观测到的集群 i 中，在给定基线协变量和处理指示变量的条件下，其他参与者的个体层面协变量（$\boldsymbol{X}_{ij'}$，适用于 $j' \neq j$）不会影响参与者 j 的缺失过程：$R_{ij} \perp\!\!\!\perp \boldsymbol{X}_{ij'} \mid C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij}$，对所有 $j' \neq j$ 均成立。这意味着 $P(R_{ij} = 1 \mid C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i) = P(R_{ij} = 1 \mid C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij})$。

2.3 多层次逆概率加权广义估计方程（MIPW-GEE）

我们将纵向研究中失访情形下的加权方法（Robins 等人，1995；Mitani 等人，2022年）加以改编，以便在多层次协变量依赖缺失设定下估计 $\boldsymbol{\beta}$。观测到 Y_{ij} 的条件概率可表示为：

$$\begin{aligned} P(R_{ij}C_i = 1|A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i) &= P(R_{ij} = 1, C_i = 1|A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i) \\ &= P(R_{ij} = 1|C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i)P(C_i = 1|A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i) \\ &= P(R_{ij} = 1|C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij})P(C_i = 1|A_i, \boldsymbol{Z}_i), \end{aligned} \tag{3}$$

其中，$P(C_i = 1|A_i, \boldsymbol{Z}_i)$对应于由 $\lambda_i(A_i, \boldsymbol{Z}_i; \boldsymbol{\gamma})$引起的簇级缺失过程，而 $P(R_{ij} = 1|C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij})$对应于由$\phi_{ij}(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij} \mid C_i = 1; \boldsymbol{\eta})$引起的个体层面的缺失过程。通过修改式（2）的加权矩阵，我们提出如下多重逆概率加权广义估计方程（MIPW-GEE）估计量：

$$\boldsymbol{W}_i(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\gamma}, \boldsymbol{\eta}) = \operatorname{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij} \mid C_i = 1; \boldsymbol{\eta})\lambda_i(A_i, \boldsymbol{Z}_i; \boldsymbol{\gamma})}\right]_{j=1,\ldots,n_i}. \tag{4}$$

MIPW-GEE估计量的一致性要求两个PS模型均设定正确：即，对于某些 $\boldsymbol{\eta}$ 和 $\boldsymbol{\gamma}$，$\phi_{ij}(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij}|C_i = 1; \boldsymbol{\eta}) = P(R_{ij} = 1|C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij})$ 且 $\lambda_i(A_i, \boldsymbol{Z}_i; \boldsymbol{\gamma}) = P(C_i = 1|A_i, \boldsymbol{Z}_i)$。

2.4 多层多重稳健广义估计方程（多重稳健广义估计方程）

MIPW-GEE估计量的一致性取决于集群层面和个体层面倾向评分模型均设定正确。为防止倾向评分模型设定错误，我们基于经验似然理论（Owen，2001；Qin等，2009；Han，2014），提出了一个 $\boldsymbol{\beta}$的^多重稳健估计量，记为 $\boldsymbol{\beta}_{MR}$。所提出的多层多重稳健广义估计方程（MMR-GEE）估计量^允许分析人员指定多组倾向评分模型；只要集群层面和个体层面分别有一个倾向评分模型设定正确，$\boldsymbol{\beta}_{MR}$ 就是 $\boldsymbol{\beta}$ 的一致估计量。

将权重替换为多重稳健权重后，通过求解估计方程 (2) 可得到 MMR-GEE 估计量：

$$\boldsymbol{W}_i^{MR} = \operatorname{diag}\left[R_{ij}C_i w_{ij}^{MR}\right]_{j=1,\ldots,n_i}. \tag{5}$$

我们将 Han 和 Wang（2013）针对独立数据设定的方法扩展到可能在多个层级发生信息性缺失的聚类数据设定，从而推导出多重稳健权重。令 $\mathcal{P}_1 = \{\phi_{ij}^k(\boldsymbol{\eta}^k) : k = 1, ..., K\}$ 表示针对 $\phi_{ij}(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij} \mid C_i = 1; \boldsymbol{\eta})$ 的 K 个假设个体层面倾向评分模型集合，令 $\mathcal{P}_2 = \{\lambda_i^\ell(\boldsymbol{\gamma}^\ell) : \ell = 1, ..., L\}$ 表示针对 $\lambda_i(A_i, \boldsymbol{Z}_i; \gamma)$ 的 L 个假设集群层面倾向评分模型集合，其中 $\boldsymbol{\eta}^k$ 和 $\boldsymbol{\gamma}^\ell$ 分别为第 k 个和第 ℓ 个模型的参数向量。令 $\hat{}\boldsymbol{\eta}^k$ 和 $\hat{}\boldsymbol{\gamma}^\ell$ 分别为 $\boldsymbol{\eta}^k$ 和 $\boldsymbol{\gamma}^\ell$ 的估计量。现在定义 $\chi^{k,l}(\hat{}\boldsymbol{\eta}^k, \hat{}\boldsymbol{\gamma}^\ell) = \left(\sum_{i=1}^M n_i\right)^{-1} \sum_{i=1}^M \sum_{j=1}^{n_i} \phi_{ij}^k(\hat{}\boldsymbol{\eta}^k)\lambda_i^\ell(\hat{}\boldsymbol{\gamma}^\ell)$。对于结局可观测的个体，其多重稳健权重 $i = 1, ..., s$, $j = 1, ..., m_i$ 可通过求解一个约束优化问题得到：

$$\hat{w}_{ij}^{MR} = \underset{w_{ij}}{\operatorname{argmax}} \prod_{i=1}^{s} \prod_{j=1}^{m_i} w_{ij}, \tag{6}$$

需满足以下约束条件：

$$w_{ij} \geq 0, \quad \sum_{i=1}^{s} \sum_{j=1}^{m_i} w_{ij} = 1, \quad \sum_{i=1}^{s} \sum_{j=1}^{m_i} w_{ij} \phi_{ij}^k(\hat{\boldsymbol{\eta}}^k) \lambda_i^\ell(\hat{\boldsymbol{\gamma}}^\ell) = \chi^{k,l}(\hat{\boldsymbol{\eta}}^k, \hat{\boldsymbol{\gamma}}^\ell) \quad (k = 1, ..., K, \ell = 1, ..., L).$$

第一个约束要求权重为非负数。第二个约束要求权重之和为 1。第三个约束为每个在有偏样本上评估的假设模型赋予权重，以表示总体均值。对于 $i = 1, ..., M$, $j = 1, ..., n_i$，现在定义 $\hat{}\boldsymbol{\eta} = \{(\hat{}\boldsymbol{\eta}^1)^{\mathrm{T}}, ..., (\hat{}\boldsymbol{\eta}^K)^{\mathrm{T}}\}$, $\hat{}\boldsymbol{\gamma} = \{(\hat{}\boldsymbol{\gamma}^1)^{\mathrm{T}}, ..., (\hat{}\boldsymbol{\gamma}^L)^{\mathrm{T}}\}$, 和

$$\begin{aligned}
\hat{g}_{ij}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}}) = \{&\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1) - \chi^{1,1}(\hat{\boldsymbol{\eta}}^1, \hat{\boldsymbol{\gamma}}^1), ..., \phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^L(\hat{\boldsymbol{\gamma}}^L) - \chi^{1,L}(\hat{\boldsymbol{\eta}}^1, \hat{\boldsymbol{\gamma}}^L), \\
&\phi_{ij}^2(\hat{\boldsymbol{\eta}}^2)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1) - \chi^{2,1}(\hat{\boldsymbol{\eta}}^2, \hat{\boldsymbol{\gamma}}^1), ..., \phi_{ij}^2(\hat{\boldsymbol{\eta}}^2)\lambda_i^L(\hat{\boldsymbol{\gamma}}^L) - \chi^{2,L}(\hat{\boldsymbol{\eta}}^2, \hat{\boldsymbol{\gamma}}^L), \ldots, \\
&\phi_{ij}^K(\hat{\boldsymbol{\eta}}^K)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1) - \chi^{K,1}(\hat{\boldsymbol{\eta}}^K, \hat{\boldsymbol{\gamma}}^1), ..., \phi_{ij}^K(\hat{\boldsymbol{\eta}}^K)\lambda_i^L(\hat{\boldsymbol{\gamma}}^L) - \chi^{K,L}(\hat{\boldsymbol{\eta}}^K, \hat{\boldsymbol{\gamma}}^L)\}^{\mathrm{T}}.
\end{aligned}$$

约束优化问题 (6) 可以通过拉格朗日乘子法求解，得到：

$$\hat{w}_{ij}^{MR} = \left\{\frac{1}{1 + \hat{\boldsymbol{\rho}}^T \hat{g}_{ij}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}})}\right\} \Bigg/ \left\{\sum_{i=1}^{s} m_i\right\} \quad i = 1, ..., s, j = 1, ..., m_i, \tag{7}$$

其中，$\hat{}\boldsymbol{\rho}^T = (\hat{}\rho_{11}, \hat{}\rho_{12}, ..., \hat{}\rho_{KL})$ 是一个 $(KL) \times 1$ 向量，通过求解以下方程得到：

$$\sum_{i=1}^{s} \sum_{j=1}^{m_i} \frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}})}{1 + \boldsymbol{\rho}^T \hat{g}_{ij}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}})} = \mathbf{0}. \tag{8}$$

详细推导见网络附录 A。方程(8) 可能有多个根。我们采用 Han (2014) 的凸最小化方法来获得 $\hat{}\boldsymbol{\rho}$。

2.5 多重稳健广义估计方程的一致性与渐近正态性

在本节中，我们首先证明，MMR-GEE估计量具有多重稳健性，即当 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 均包含设定正确的倾向评分模型时，$\hat{}\boldsymbol{\beta}_{MR}$ 是 $\boldsymbol{\beta}$ 的一致估计量。随后，我们建立 $\boldsymbol{\beta}_{MR}$ 的渐近分布。我们考虑簇数 M 趋于无穷大且簇大小有上界的情形。为简化符号，我们考虑簇大小固定的设定。借助 Lindeberg-Feller 中心极限定理，结果可以推广到簇大小不等的情形。下文中，我们用下标星号表示概率极限，用 $(\boldsymbol{\eta}_0, \boldsymbol{\gamma}_0)$ 表示倾向评分模型的真实参数，并用 $\boldsymbol{\beta}_0$ 表示边际均值模型的真实参数。

$\hat{}$ 2.5.1 $\boldsymbol{\beta}_{MR}$ 的多重稳健性

$\hat{}$ 为证明 $\boldsymbol{\beta}_{MR}$ 的多重稳健性，我们首先说明：当 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 均包含正确设定的倾向评分模型时，公式（5）的多重稳健权重在渐近意义上等价于公式（4）的多层逆概率权重，前提是真实的正确模型已知。通过在假设正确的倾向评分模型已知的条件下，建立多重稳健权重与另一版本的经验似然权重之间的联系，可以确立这种渐近等价性；我们将在下文推导这一过程。

不失一般性，设 $\phi_{ij}^1(\boldsymbol{\eta}^1)$ 和 $\lambda_i^1(\boldsymbol{\gamma}^1)$ 分别是 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 中第一个正确设定的模型。此外，设 p_{ij} 为 $i = 1, ..., s, j = 1, ..., m_i$ 时，给定 $R_{ij}C_i = 1$ 条件下 $(Y_{ij}, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij})$ 的经验概率。通过采用与 2.4 中相同的拉格朗日乘数法求解（6）的约束优化问题的经验版本，可以得到 p_{ij} 的估计量，记为 $\hat{}p_{ij}$。经过一些代数运算，$\hat{}w_{ij}^{MR}$ 可表示为（推导见网络附录B）：

$$\hat{w}_{ij}^{MR} = \hat{p}_{ij} \frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1, \hat{\boldsymbol{\gamma}}^1\right)}{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)} = \frac{1 + o_p(1)}{\left(\sum_{i=1}^M n_i\right) \phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}, \quad i = 1, \ldots, s, j = 1, \ldots m_i. \tag{9}$$

将 $\hat{}w_{ij}^{MR}$ 代入式 (5) 的加权矩阵中，我们得到以下关系：

$$\hat{\boldsymbol{W}}_i^{MR} = \operatorname{diag}\left[R_{ij}C_i\hat{w}_{ij}^{MR}\right]_{j=1,\ldots,n_i} = \operatorname{diag}\left[\frac{R_{ij}C_i + o_p(1)}{\left(\sum_{i=1}^M n_i\right) \phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\right]_{j=1,\ldots,n_i}, \tag{10}$$

即 渐近地与正确设定的多重逆概率加权广义估计方程估计量的权重成比例 of

(4). 当簇的数量 M 趋于无穷时，我们有：

$$\begin{aligned}
&\left(\frac{\sum_{i=1}^{M} n_i}{M}\right) \sum_{i=1}^{M} \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \hat{\boldsymbol{W}}_i^{MR} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \\
&= \frac{1}{M} \sum_{i=1}^{M} \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \operatorname{diag}\left[\frac{R_{ij} C_i + o_p(1)}{\phi_{ij}^1(\boldsymbol{\eta}_0) \lambda_i^1(\boldsymbol{\gamma}_0)}\right] (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \\
&\xrightarrow{p} E\left\{\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \operatorname{diag}\left[\frac{R_{ij} C_i}{\phi_{ij}^1(\boldsymbol{\eta}_0) \lambda_i^1(\boldsymbol{\gamma}_0)}\right] (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i))\right\} = 0, \qquad (11)
\end{aligned}$$

$\hat{}$ 这证明了 $\boldsymbol{\beta}_{MR}$ 的一致性。结果总结如下：

定理 1 当 $\mathcal{P}_1$ 包含一个关于 $\phi_{ij}(\boldsymbol{\eta})$ 的正确模型，且 $\mathcal{P}_2$ 包含一个关于 $\lambda_i(\gamma)$ 的正确模型时，随着 $M \to \infty$，$\hat{}\boldsymbol{\beta}_{MR} \xrightarrow{P} \boldsymbol{\beta}_0$。

证明见网络附录B。

2.5.2 渐近分布

$\hat{}$ 假设正确模型已知，我们沿用 Han（2014）的定理2中的方法，推导 $\boldsymbol{\beta}_{MR}$ 的渐近分布。不失一般性，设 $\phi_{ij}^1(\boldsymbol{\eta}^1)$ 和 $\lambda_i^1(\boldsymbol{\gamma}^1)$ 分别是关于 $\phi_{ij}(\boldsymbol{\eta})$ 和 $\lambda_i(\boldsymbol{\gamma})$ 的正确设定的模型。$\boldsymbol{\eta}^1$ 和 $\boldsymbol{\gamma}^1$ 的得分函数分别记为 $\boldsymbol{S}_{1,\ i}(\boldsymbol{\eta}^1)$ 和 $\boldsymbol{S}_{2,\ i}(\boldsymbol{\gamma}^1)$，如下：

$$\boldsymbol{S}_{1,i}(\boldsymbol{\eta}^1) = \sum_{j=1}^{n_i} \frac{C_i[R_{ij} - \phi_{ij}^1(\boldsymbol{\eta}^1)]}{\phi_{ij}^1(\boldsymbol{\eta}^1)\{1 - \phi_{ij}^1(\boldsymbol{\eta}^1)\}} \frac{\partial \phi_{ij}^1(\boldsymbol{\eta}^1)}{\partial \boldsymbol{\eta}^1},\ \boldsymbol{S}_{2,i}(\boldsymbol{\gamma}^1) \quad = \frac{C_i - \lambda_i^1(\boldsymbol{\gamma}^1)}{\lambda_i^1(\boldsymbol{\gamma}^1)\{1 - \lambda_i^1(\boldsymbol{\gamma}^1)\}} \frac{\partial \lambda_i^1(\boldsymbol{\gamma}^1)}{\partial \boldsymbol{\gamma}^1}.$$

Let

$$\begin{aligned}
\boldsymbol{L} &= \mathbb{E}\left[\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}(\boldsymbol{\eta}_*, \boldsymbol{\gamma}_*)}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\right\} \boldsymbol{U}_i(\boldsymbol{\beta})\right], \\
\boldsymbol{G} &= \mathbb{I}^{\mathrm{T}} \mathbb{E}\left[\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}(\boldsymbol{\eta}_*, \boldsymbol{\gamma}_*)^{\otimes 2}}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\right\}\right] \mathbb{I}, \qquad (12) \\
\boldsymbol{Q}_i(\boldsymbol{\gamma}^1, \boldsymbol{\eta}^1) &= \operatorname{diag}\left\{\frac{R_{ij} C_i}{\phi_{ij}^1(\boldsymbol{\eta}^1)\lambda_i^1(\boldsymbol{\gamma}^1)}\right\} \boldsymbol{U}_i(\boldsymbol{\beta}) - \mathbf{L}\mathbf{G}^{-1} \mathbb{I}^{\mathrm{T}} \operatorname{diag}\left\{\frac{R_{ij} C_i - \phi_{ij}^1(\boldsymbol{\eta}^1)\lambda_i^1(\boldsymbol{\gamma}^1)}{\phi_{ij}^1(\boldsymbol{\eta}^1)\lambda_i^1(\boldsymbol{\gamma}^1)} \boldsymbol{g}_{ij}(\boldsymbol{\eta}_*, \boldsymbol{\gamma}_*)\right\} \mathbb{I},
\end{aligned}$$

其中，$\mathbb{I} = (1, 1, \ldots, 1)^{\mathrm{T}}$，且对于任意矩阵 $\boldsymbol{C}$，$\boldsymbol{C}^{\otimes 2} = \boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}$。此外，记 $\boldsymbol{S}_{1,\ i} = \boldsymbol{S}_{1,\ i}(\boldsymbol{\eta}_0)$、$\hat{}\boldsymbol{S}_{2,\ i} = \boldsymbol{S}_{2,\ i}(\boldsymbol{\gamma}_0)$ 和 $\boldsymbol{Q}_i = \boldsymbol{Q}_i(\boldsymbol{\gamma}_0, \boldsymbol{\eta}_0)$。下述定理给出了 $\boldsymbol{\beta}_{MR}$ 的渐近分布。**定理 2** 当 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 分别包含关于 $\phi_{ij}(\boldsymbol{\eta})$ 和 $\lambda_i(\boldsymbol{\gamma})$ 的正确设定的模型时，$\sqrt{M}(\hat{}\boldsymbol{\beta}_{MR} - \boldsymbol{\beta}_0)$ 服从均值为 0、方差为 $\operatorname{var}(\boldsymbol{Z}_i)$ 的渐近正态分布，其中

$$\boldsymbol{Z}_i = \left[E\left\{\frac{\partial \boldsymbol{U}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\right\}\right]^{-1} \left[\boldsymbol{Q}_i - \left\{\mathbb{E}\left(\boldsymbol{Q}_i \boldsymbol{S}_{1,i}^{\mathrm{T}}\right)\right\}\left\{\mathbb{E}\left(\boldsymbol{S}_{1,i}^{\otimes 2}\right)\right\}^{-1} \boldsymbol{S}_{1,i} - \left\{\mathbb{E}\left(\boldsymbol{Q}_i \boldsymbol{S}_{2,i}^{\mathrm{T}}\right)\right\}\left\{\mathbb{E}\left(\boldsymbol{S}_{2,i}^{\otimes 2}\right)\right\}^{-1} \boldsymbol{S}_{2,i}\right].$$

详细的推导和证明请参见网络附录C。

MMR-GEE估计量的渐近方差需要知道正确的PS模型，而研究者通常无法获得这些模型。因此，渐近方差公式无法轻易用于获得标准误估计。我们建议在推断中使用非参数自助法（Davison andHinkley，1997）。聚类数据中常用的两种自助法策略包括“聚类自助法”和“个体自助法”。“聚类自助法”方法有放回地抽样 M 簇，并将重抽样簇中的所有个体纳入自助法样本（FieldandWelsh，2007）。“个体自助法”方法在原始簇内有放回地抽样 n_i 个体，同时在所有自助法重复中保持总样本量不变（RobertsandFan，2004）。在当前设定中，由于整个簇的结局可能缺失，仅将“个体自助法”方法应用于具有观测数据的簇，无法反映簇级PS模型估计中的不确定性，并可能导致标准误低估。因此，我们建议使用“簇自助法”方法。

2.6 使用 EM 算法处理集群层面缺失指标的错误分类

所提出的 MMR-GEE 估计量的一致性要求正确设定的 PS模型中的参数能够得到一致估计，这通常可以通过例如基于矩的估计量或基于似然的估计量来实现。在观测数据中，当某个集群没有任何个体结局可用时，该集群的结局数据可能因簇级缺失过程而缺失（即真实的集群层面缺失指示变量 $C_i = 0$）；也可能是该集群仍留在研究中，但该集群的所有个体结局均缺失，尤其是在簇大小较小时（即 $C_i = 1$，但对于所有 $j = 1$，.....，n_i，$R_{ij} = 0$）。令 C_i^O 表示观测簇级缺失指示变量。在这两种情况下，我们观测到 $C_i^O = I\left(\sum_{j=1}^{n_i} R_{ij} > 0\right) = 0$，但 C_i 可以为 0 或 1。由于对 PS模型中的参数进行一致估计需要知道 C_i，若直接指定 $C_i = C_i^O$，就可能发生错误分类。我们总结 (C_i, C_i^O) 的所有可能模式，包括:

1. $C_i = 0, C_i^O = 0$: 当一个集群在随机化后退出时，该集群中所有参与者的结局均无法观测，因此 C_i^O为 0。

2. $C_i = 1, C_i^O = 0$: 当集群未退出研究时，如果集群内所有个体结局均缺失，我们仍可能观察到 $C_i^O = 0$。这种情况更可能发生在簇大小较小时。

3. $C_i = 1, C_i^O = 1$: 当集群未退出研究时，如果集群 i 中至少一名参与者的结局被观测到，则观测到的集群层面缺失指示变量就是真实的集群层面缺失指示变量。

这些模式也总结于网络附录D.1中的表 S1。在模式（2）下，观测到的C_i^O将真实的 C_i错误分类，从而导致PS模型估计参数产生偏差。更具体地说，假设PS模型为：

$$\operatorname{logit}\left\{\lambda_i(A_i, \boldsymbol{Z}_i; \boldsymbol{\gamma})\right\} = \boldsymbol{Z}_i^* \boldsymbol{\gamma}, \quad \operatorname{logit}\left\{\phi_{ij}(A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij} \mid C_i = 1; \boldsymbol{\eta})\right\} = \boldsymbol{X}_{ij}^* \boldsymbol{\eta}, \tag{13}$$

其中 $\boldsymbol{Z}_i^* = \begin{pmatrix} 1 & A_i & \boldsymbol{Z}_i^{\mathrm{T}} \end{pmatrix}$、$\boldsymbol{X}_{ij}^* = \begin{pmatrix} 1 & A_i & \boldsymbol{Z}_i^{\mathrm{T}} & \boldsymbol{X}_{ij}^{\mathrm{T}} \end{pmatrix}$、$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_I & \gamma_A & \boldsymbol{\gamma}_{\boldsymbol{Z}}^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}}$和 $\boldsymbol{\eta} = \begin{pmatrix} \eta_I & \eta_A & \boldsymbol{\eta}_{\boldsymbol{Z}}^{\mathrm{T}} & \boldsymbol{\eta}_{\boldsymbol{X}}^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}}$。由于 $P\left(R_{ij}=1 | C_i=1、A_i、\boldsymbol{Z}_i、\boldsymbol{X}_{ij}\right) \neq P\left(R_{ij}=1 | C_i^O=1、A_i、\boldsymbol{Z}_i、\boldsymbol{X}_{ij}\right)$ 且 $P\left(C_i=1 | A_i、\boldsymbol{Z}_i\right) \neq P\left(C_i^O=1 | A_i、\boldsymbol{Z}_i\right)$，即使PS模型设定正确，基于 C_i^O的 $\boldsymbol{\gamma}$和 $\boldsymbol{\eta}$的估计量也可能存在偏差。为解决这一潜在的错误分类问题，我们将 C_i 视为部分观测数据，并提出一种期望最大化算法（Dempster 等，1977）来估计PS模型中的参数。

在当前设定中，“完整”数据为 $\{C_i, \boldsymbol{R}_i, \boldsymbol{Y}_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i\}_{i=1}^M$，为简化记号，记作 $(\boldsymbol{C}, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X})$。完整数据对数似然为：

$$\begin{aligned}
&\ell_c(\boldsymbol{C}, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\gamma}, \boldsymbol{\eta}) \\
&= \sum_{i=1}^{M} C_i \left\{ \log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \gamma\right)\right) + \sum_{j=1}^{n_i} R_{ij} \log\left(\operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) + \left(1 - R_{ij}\right) \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right\} + \\
&\left(1 - C_i\right) \log\left(1 - \operatorname{expit}\left(\boldsymbol{Z}_i^* \gamma\right)\right) + C_i \left(\log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \gamma\right)\right) + \sum_{j=1}^{n_i} \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right).
\end{aligned} \tag{14}$$

在第 ν 次迭代中，给定观测数据（$\boldsymbol{C}^O, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}$），期望步骤（E步）的条件期望为：

$$\begin{aligned}
&Q\left(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\gamma}^{(\nu)}, \boldsymbol{\eta}^{(\nu)}, \boldsymbol{C}^O, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}\right) \\
&= \mathbb{E}_{\boldsymbol{\gamma}^{(\nu)}, \boldsymbol{\eta}^{(\nu)}}\left\{\ell(\boldsymbol{C}, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\gamma}, \boldsymbol{\eta}) \mid \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}\right\} \\
&= \sum_{i=1}^{M} C_i^O \left\{ \log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \gamma\right)\right) + \sum_{j=1}^{n_i} R_{ij} \log\left(\operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) + \left(1 - R_{ij}\right) \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right\} + \\
&\qquad \left(1 - w_i^{(\nu)}\right)\left(1 - C_i^O\right) \log\left(1 - \operatorname{expit}\left(\boldsymbol{Z}_i^* \gamma\right)\right) + \\
&\qquad w_i^{(\nu)}\left(1 - C_i^O\right) \left(\log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \gamma\right)\right) + \sum_{j=1}^{n_i} \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right), \\
&\text{where } w_i^{(\nu)} = C_i^O + \left(1 - C_i^O\right) \times \left\{ \frac{\prod_{j=1}^{n_i} \left(1 - \operatorname{expit}(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}^{(\nu)})\right) \operatorname{expit}(\boldsymbol{Z}_i^* \boldsymbol{\gamma}^{(\nu)})}{1 - \operatorname{expit}(\boldsymbol{Z}_i^* \boldsymbol{\gamma}^{(\nu)}) \left[1 - \prod_{j=1}^{n_i} \left(1 - \operatorname{expit}(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}^{(\nu)})\right)\right]} \right\}.
\end{aligned} \tag{15}$$

对于最大化步骤（M步），我们建议使用优化软件（例如 R 中的 Optim 函数）（Nash, 2019；R Core Team, 2021）来最大化完整数据似然。迭代执行 E步和 M步后，算法收敛时即可得到 EM 估计量 $\hat{}\boldsymbol{\gamma}^{EM}$ 和 $\hat{}\boldsymbol{\eta}^{EM}$。当 PS模型设定正确时，尽管簇级缺失指示变量存在错误分类，$\hat{}\boldsymbol{\gamma}^{EM}$ 和 $\hat{}\boldsymbol{\eta}^{EM}$ 仍是

PS模型真实参数的一致估计量，即 $\hat{\boldsymbol{\eta}}^{EM} \xrightarrow{P} \boldsymbol{\eta}_0$ 和 $\hat{\boldsymbol{\gamma}}^{EM} \xrightarrow{P} \boldsymbol{\gamma}_0$。完整数据似然和 E步的详细推导以及该算法的伪代码见网络附录D。

2.7 扩展至三级整群随机试验

在三级整群随机试验中，研究参与者嵌套于家庭或医疗服务提供者等子整群中，而子整群又嵌套于地区或诊所等簇中。在本节中，我们将介绍如何修改所提出的方法，以处理子整群和个体层面的信息性结局缺失，正如引言中所述的 Pro-CCM 研究（Ratovoson 等，2022）。

设 Y_{ijk} 为结局，$\boldsymbol{X}_{ijk} = (X^1_{ijk}, X^2_{ijk}, \ldots)^{\mathrm{T}}$ 为参与者 $k = 1, ..., n_{cs}$ 在集群 $i = 1, ..., n_c$ 中的子集群 $j = 1, ..., n_s$ 的基线协变量向量。这里，基线协变量 $\boldsymbol{X}_{ijk}$ 可以包含个体、子集群和集群层面的信息。与之前类似，我们考虑采用相同二元治疗指标记法 A_i 的两臂平行整群随机试验；我们还假设所有协变量在随机化前均已完整观测。对于多层缺失过程，$\boldsymbol{R}_{ij} = (R_{ij1}, R_{ij2}, \ldots)^{\mathrm{T}}$ 用于表示个体层面缺失指示变量的向量，C_{ij} 用于表示结局 $\boldsymbol{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \ldots)^{\mathrm{T}}$ 的子集群层面缺失指示变量。当 Y_{ijk} 被观测到时，$R_{ijk} = 1$；当 Y_{ijk} 缺失时，$R_{ijk} = 0$。当子集群 j 中所有参与者的结局均缺失时，$C_{ij} = 0$；否则为 $C_{ij} = 1$。本质上，表1表示一个集群的数据结构，而三级整群随机试验的数据结构是所有集群的串接。

对于三级整群随机试验中的多层缺失设定，MIPW-GEE 和 MMR-GEE 的估计方程可修改如下：

$$\sum_{i=1}^{n_c} \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \boldsymbol{W}_i(A_i, \boldsymbol{X}_i; \boldsymbol{\gamma}, \boldsymbol{\beta})(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)) = 0. \tag{16}$$

$\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)}{\partial \boldsymbol{\beta}}$ 是设计矩阵，其中 $\boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i) = \{\{\boldsymbol{\mu}_{ijk}(\boldsymbol{\beta}, A_i)\}_{k=1}^{n_{cs}}\}_{j=1}^{n_s}$。$\boldsymbol{V}_i = \boldsymbol{F}_i^{1/2} \boldsymbol{C}(\boldsymbol{\alpha}) \boldsymbol{F}_i^{1/2}$ 是协方差矩阵，其中 $\boldsymbol{F}_i = \mathrm{diag}\{\mathrm{diag}[\mathrm{var}(y_{ijk})]_{k=1}^{n_{cs}}\}_{j=1}^{n_s}$。在三级整群随机试验中，可交换相关结构和区块可交换相关结构是指定 $\boldsymbol{C}(\boldsymbol{\alpha})$ 的常见选择。为考虑信息性缺失子簇，多层加权矩阵的形式如下：

$$\begin{aligned}
\boldsymbol{W}_i(A_i, \boldsymbol{X}_i; \boldsymbol{\gamma}, \boldsymbol{\eta}) &= \mathrm{diag}\left\{\boldsymbol{W}_{ij}(A_i, \boldsymbol{X}_{ij}; \boldsymbol{\gamma}, \boldsymbol{\eta})\right\}_{j=1}^{n_s} \\
&= \mathrm{diag}\left\{\mathrm{diag}\left[\boldsymbol{W}_{ijk}(A_i, \boldsymbol{X}_{ijk}; \boldsymbol{\gamma}, \boldsymbol{\eta})\right]_{k=1}^{n_{cs}}\right\}_{j=1}^{n_s} \\
&= \mathrm{diag}\left\{\mathrm{diag}\left[\frac{R_{ijk} C_{ij}}{\phi_{ijk}(A_i, \boldsymbol{X}_{ijk} \mid C_{ij} = 1; \boldsymbol{\eta}) \lambda_{ij}(A_i, \boldsymbol{X}_{ij}; \boldsymbol{\gamma})}\right]_{k=1}^{n_{cs}}\right\}_{j=1}^{n_s},
\end{aligned} \tag{17}$$

其中，$\phi_{ijk}(A_i, \boldsymbol{X}_{ijk} | C_{ij} = 1; \boldsymbol{\eta})$ 是 $P(R_{ijk} = 1 | C_{ij} = 1, A_i, \boldsymbol{X}_{ij})$ 的个体层面的缺失过程，而 $\lambda_{ij}(A_i, \boldsymbol{X}_{ij}; \boldsymbol{\gamma})$ 是 $P(C_{ij} = 1 | A_i, \boldsymbol{X}_{ij})$ 的子簇层面缺失过程。其扩展

the MMR-GEE估计量可通过替换公式（16）中的加权矩阵 b 得到 y:

$$\boldsymbol{W}_i^{MR} = \mathrm{diag}\{\mathrm{diag}\left[R_{ijk}C_{ij}w_{ijk}^{MR}\right]_{k=1}^{n_{cs}}\}_{j=1}^{n_s}. \tag{18}$$

w_{ijk}^{MR}的估计采用与第 2.4 节相同的策略。

3 模拟研究：疟疾流行的加纳贫血治疗

我们开展了一项模拟研究，以考察在不同簇大小和倾向评分模型设定下，我们提出的 MMR-GEE 估计量在有限样本中的表现，以及期望最大化算法的运行特性。我们基于在疟疾流行的加纳开展的缺铁性贫血治疗研究设计了这项模拟研究。该整群随机试验评估了铁补充剂对降低西非加纳婴儿和幼儿疟疾发病率的作用（Zlotkin 等，2013）。研究按集群对 6 至 35 个月大的儿童进行随机分组；集群定义为由一个或多个家庭组成的院落。治疗组的 780 个集群（967 名儿童）每日接受含铁微量营养素粉末（MNP）；对照组的 772 个集群（991 名儿童）每日接受不含铁的 MNP。研究结束时，铁组的 12 个集群（25 名儿童）和无铁组的 19 个集群（29 名儿童）失访。主要结局是研究期间的疟疾发病率。我们重点关注干预后 1 个月期间疟疾的累积发病率。根据 Zlotkin 等（2013）提取的结局和基线协变量描述性汇总结果见网络附录 E.1 的表 S2。

3.1 数据生成过程

我们生成了与表 S2 汇总统计量相匹配的数据。我们将主要结局 Y_{ij}（每 100 儿童年的疟疾发病率）视为连续变量，其生成方式为

$$Y_{ij} = \beta_I^* + \beta_A^* A_i + \boldsymbol{Z}_i^O\boldsymbol{\beta}_Z + \boldsymbol{X}_{ij}^O\boldsymbol{\beta}_X + A_i\boldsymbol{Z}_i^O\boldsymbol{\beta}_{AZ} + A_i\boldsymbol{X}_{ij}^O\boldsymbol{\beta}_{AX} + \delta_i + \varepsilon_i, \tag{19}$$

其中治疗分配 (A_i) 根据概率为$p = 0.5$的伯努利分布进行模拟， $\boldsymbol{Z}_i^O$包含家庭规模、家庭教育和财富。 $\boldsymbol{X}_{ij}^O$ 包含年龄、消瘦 z分数和发育迟缓 z分数。 $\delta_i \sim N(0, \sigma_\delta^2)$ 是簇随机截距， $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ 是残差标准误。

通过对 $(\boldsymbol{Z}_i^O, \boldsymbol{X}_{ij}^O, A_i\boldsymbol{Z}_i^O, A_i\boldsymbol{X}_{ij}^O)$、随机截距和残差标准误进行边际化，得到的真实边际均值模型为 $\mathbb{E}[Y_{ij}|A_i] = \beta_I + \beta_A A_i$。研究得出结论，干预后期间每日使用含铁 MNP 并未增加疟疾风险，因此

选择 $(\beta_I^*, \beta_A^*, \boldsymbol{\beta}_Z^{\mathrm{T}}, \boldsymbol{\beta}_X^{\mathrm{T}}, \boldsymbol{\beta}_{AZ}^{\mathrm{T}}, \boldsymbol{\beta}_{AX}^{\mathrm{T}})$，使截距 (β_I) 为 63.5，且边际治疗效应 (β_A) 为 0。

通过以下逻辑回归模型诱导结局缺失数据过程：

$$\operatorname{logit}\left(\lambda_i\left(A_i, \boldsymbol{Z}_i^C; \boldsymbol{\gamma}\right)\right) = \gamma_I + \gamma_A A_i + \boldsymbol{Z}_i^C \boldsymbol{\gamma}_{\boldsymbol{Z}} + A_i \boldsymbol{Z}_i^C \boldsymbol{\gamma}_{A\boldsymbol{Z}}, \tag{20}$$

$$\operatorname{logit}\left(\phi_{ij}\left(A_i, Z_i^I, \boldsymbol{X}_{ij}^I \mid C_i = 1; \boldsymbol{\eta}\right)\right) = \eta_I + \eta_A A_i + \eta_Z Z_i^I + \boldsymbol{X}_{ij}^I \boldsymbol{\eta}_X + A_i \boldsymbol{X}_{ij}^I \boldsymbol{\eta}_{AX}. \tag{21}$$

$\boldsymbol{Z}_i^C$包括家庭规模、家庭教育和财富。 Z_i^I包括辅食 ≤ 6 个月。$\boldsymbol{X}_{ij}^I$ 包括消瘦 z 分数和发育迟缓 z 分数。原始研究中约有 2% 的簇和 2.8% 的参与者缺失。为便于说明，我们通过选择$\boldsymbol{\gamma}$ 和 $\boldsymbol{\eta}$ 来增加缺失率，使 12% 的簇缺失，且总体参与者中有 30% 缺失结局。我们通过改变边际效应参数 $\beta_A = (0, 1.5)$ 、簇大小和组内相关系数，考察了一系列设定。参数值的详细说明见网络附录 E.2。

3.2 分析方法

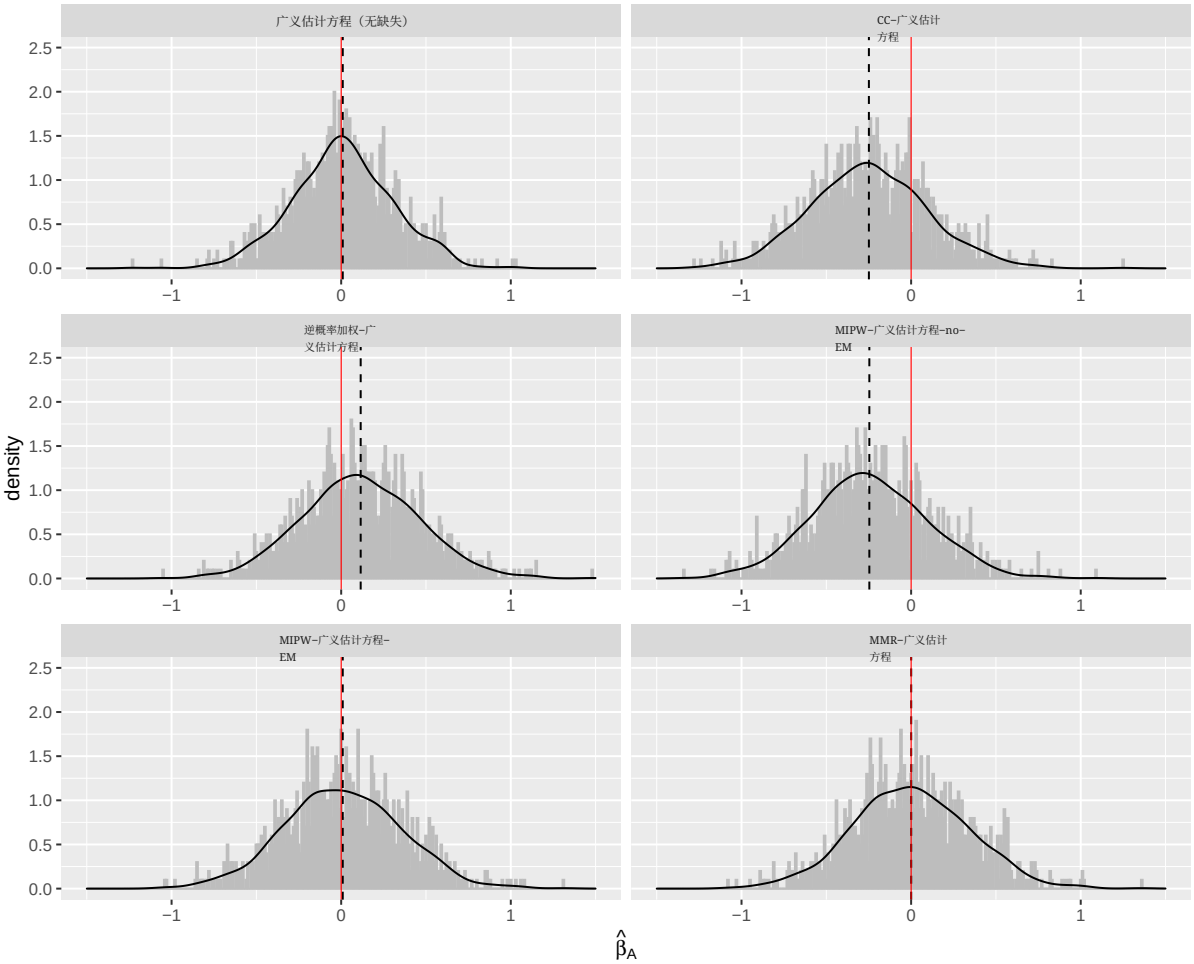
为说明纠正信息性多层结局缺失可能导致的偏倚的重要性，我们比较了以下四种方法。首先，我们基于模型（1）进行了未加权的完全病例广义估计方程分析。其次，我们采用基于模型（2）的逆概率加权广义估计方程方法，其中倾向评分通过无条件逻辑回归模型进行估计，该模型的函数形式与模型（21）相同，但忽略了整群层面缺失：

$$\operatorname{logit}\left(\pi_{ij}\left(A_i, Z_i^I, \boldsymbol{X}_{ij}^I; \boldsymbol{\theta}\right)\right) = \theta_I + \theta_A A_i + \theta_Z Z_i^I + \boldsymbol{X}_{ij}^I \boldsymbol{\theta}_X + A_i \boldsymbol{X}_{ij}^I \boldsymbol{\theta}_{AX}. \tag{22}$$

第三，我们采用了 MIPW-GEE 方法，其中集群层面和个体层面的 PS 模型均设定正确。为估计 PS 模型中的参数，我们基于 C_i^O 拟合标准逻辑回归模型（记为 MIPW-GEE-no-EM），并应用期望最大化算法（记为 MIPW-GEE-EM）。最后，我们通过指定$\mathcal{P}_1 = \{\lambda_i^k(\boldsymbol{\gamma}^k)\ k = 1、\ 2\}$和 $\mathcal{P}_2 = \{\phi_{ij}^\ell(\boldsymbol{\eta}^\ell)\ \ell = 1、\ 2\}$，实施了所提出的 MMR-GEE 估计量，所有参数均通过期望最大化算法估计。 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 均包含一个设定正确的模型和一个设定错误的模型。PS 模型设定的详情见网络附录 E.3。对于所有方法，我们均采用可交换工作相关结构，并使用“簇自助法”计算标准误估计。

3.3 模拟结果

图1展示了在零假设下（即$\beta_A = 0$），当 $M = 1552$ $n_i \sim DU(1, 5)$、 $ICC = 0.0804$ 时，估计的边际处理效应的经验分布。红线表示真实的边际效应，并



ˆ 图1：在 $n_i \sim DU(1,5)$ $M = 1552, ICC = 0.08$ 条件下，基于 1,000次重复 得到的估计的边际处理效应 (β_A) 的经验分布。红线表示真实边际处理效应 (0)。黑色虚线表示估计的边际处理效应的经验均值。

黑色虚线表示 1,000 个效应估计值的经验均值。总体而言，所提出的多重稳健广义估计方程和 EM多重逆概率加权广义估计方程估计量所得估计值均以真实的 β_A 为中心。由于完全病例广义估计方程忽略了信息性缺失数据，而逆概率加权广义估计方程未能考虑整群层面缺失，因此两者均导致有偏估计。多重逆概率加权广义估计方程-无期望最大化试图对多层缺失过程进行调整。然而，由于未使用期望最大化算法校正 C_i 中的错误分类，多重逆概率加权广义估计方程-无期望最大化所得估计值存在偏差。另一方面，EM多重逆概率加权广义估计方程恰当地处理了 C_i 中的错误分类，因此偏差消失了。其他设定下的相应图见网络附录 E.4。

表 2 汇总了经验偏差、经验标准误、平均标准误估计值和覆盖概率。在所有情景下，完全病例广义估计方程和逆概率加权广义估计方程所得边际治疗效应估计值均有偏（完全病例广义估计方程的偏差范围为 -0.262 至 0.599，逆概率加权广义估计方程为 0.105 至 0.330），而多重逆概率加权广义估计方程-无期望最大化和多重稳健广义估计方程的估计值均值非常接近真实参数值。当簇大小较小时，多重逆概率加权广义估计方程-无期望最大化的结果有偏，但随着 n_i 增加至 $DU(30, 50)$，偏差消失。在所有情景下，基于自助法的标准误均值都接近经验标准误。当边际治疗效应得到一致估计时（即多重稳健广义估计方程和 EM多重逆概率加权广义估计方程，以及 $n_i \sim DU(30, 50)$ 时的多重逆概率加权广义估计方程-无期望最大化），覆盖真实参数值的 95% 置信区间比例接近 95%。另一方面，完全病例广义估计方程和逆概率加权广义估计方程的经验覆盖率可能远低于名义水平（例如，完全病例广义估计方程为 <40%，逆概率加权广义估计方程为 <70%）。

进一步比较了在正确设定的 PS模型中使用和不使用 EM算法估计参数的策略，并在表 3 中呈现了经验绝对偏差。在所有情形下，$\hat{\boldsymbol{\gamma}}^{EM}$ 和$\hat{\boldsymbol{\eta}}^{EM}$ 均以真实值为中心，而使用观测到的 C_i^O通过标准逻辑回归估计的参数，对于 $n_i \sim DU(1, 4)$、$n_i \sim DU(1, 5)$ 和 $n_i = 3$可能存在较大偏差，但对于 $n_i \sim DU(30, 50)$，偏差消失了。如第 2.6 节所述，由于集群中所有个体结局均缺失，C_i 中的错误分类更有可能发生在簇大小较小时。随着簇大小增加，C_i 中的错误分类发生的可能性降低，因为集群内所有个体结局均缺失的概率 $\prod_{j=1}^{n_i} P(R_{ij} = 0|A_i, \boldsymbol{Z}_i, \boldsymbol{X}_{ij})$ 将非常小。因此，较大的簇大小无需应用 EM算法。还需注意的是，即使不需要使用 EM估计量，应用 EM算法也不会影响估计参数，因为我们可以看到，对于 $n_i \sim DU(30, 50)$，MIPW-GEE-EM估计量与 MIPW-GEE-no-EM估计量的表现几乎完全相同。

表2：经验偏差、经验标准误、使用“簇自助法”估计的自助法标准误的均值，以及基于 1,000次重复和100次自助重抽样计算的覆盖概率。覆盖概率是指真实β_A 落在根据自助法标准误构建的 95% 置信区间内的百分比。

			簇自助法	
	经验偏倚	经验标准误。	估计值 标准误	协变量概率（%）
零假设（$\beta_A = 0$）				
	$M = 1552, n_i \sim DU(1, 5)$ $ICC = 0.0804, P(C_i = C_i^O) = 0.87$			
广义估计方程（无缺失）	0.009	0.296	0.293	94.4
完全病例广义估计方程	-0.249	0.345	0.350	89.3
逆概率加权广义估计方程	0.115	0.345	0.349	93.4
多重逆概率加权广义估计方程-无期望最大化	-0.246	0.343	0.346	88.8
EM多重逆概率加权广义估计方程	0.009	0.344	0.348	94.9
多重稳健广义估计方程	< 0.001	0.343	0.346	94.5
	$M = 1552, n_i = 3, ICC = 0.0804, P(C_i = C_i^O) = 0.99$			
广义估计方程（无缺失）	-0.010	0.205	0.202	94.0
完全病例广义估计方程	-0.262	0.237	0.237	80.0
逆概率加权广义估计方程	0.105	0.237	0.236	90.4
多重逆概率加权广义估计方程-无期望最大化	-0.025	0.239	0.235	94.1
EM多重逆概率加权广义估计方程	-0.007	0.240	0.236	94.0
多重稳健广义估计方程	-0.009	0.239	0.234	94.1
	$M = 300, n_i \sim DU(30, 50)$ $ICC = 0.2, P(C_i = C_i^O) = 1$			
广义估计方程（无缺失）	0.007	0.266	0.282	94.6
完全病例广义估计方程	-0.247	0.291	0.308	86.6
逆概率加权广义估计方程	0.121	0.292	0.309	92.8
多重逆概率加权广义估计方程-无期望最大化	0.009	0.282	0.299	93.9
EM多重逆概率加权广义估计方程	0.008	0.281	0.299	93.9
多重稳健广义估计方程	0.013	0.282	0.299	94.1
备择假设下的数据 ($\beta_A = 1.5$)				
	$M = 1552, n_i \sim DU(1, 4)$ $ICC = 0.0804, P(C_i = C_i^O) = 0.94$			
完整数据	0.001	0.179	0.188	95.6
完整案例	0.483	0.204	0.211	37.8
逆概率加权广义估计方程	0.270	0.201	0.209	73.8
多重逆概率加权广义估计方程-无期望最大化	0.072	0.211	0.216	93.0
EM多重逆概率加权广义估计方程	0.002	0.212	0.217	95.1
多重稳健广义估计方程	0.009	0.212	0.214	95.1
	$M = 1552, n_i = 3, ICC = 0.2, P(C_i = C_i^O) = 0.99$			
完整数据	-0.002	0.174	0.177	94.4
完整案例	0.511	0.198	0.197	27.0
逆概率加权广义估计方程	0.279	0.195	0.196	69.9
多重逆概率加权广义估计方程-无期望最大化	0.034	0.202	0.200	93.5
EM多重逆概率加权广义估计方程	< 0.001	0.202	0.200	94.8
多重稳健广义估计方程	0.006	0.196	0.196	95.1
	$M = 300, n_i \sim DU(30, 50)$ $ICC = 0.2, P(C_i = C_i^O) = 1$			
完整数据	0.009	0.308	0.314	95.1
完整案例	0.599	0.339	0.343	56.9
逆概率加权广义估计方程	0.330	0.335	0.341	85.2
多重逆概率加权广义估计方程-无期望最大化	0.013	0.327	0.328	94.9
EM多重逆概率加权广义估计方程	0.013	0.327	0.328	94.9
多重稳健广义估计方程	0.011	0.323	0.329	95.3

表3：基于1,000次重复，PS模型中参数估计值的经验绝对偏差。所有参数均由设定正确的集群和个体层面PS模型估计，但采用了不同的估计方法：一种采用所提出的EM算法，另一种基于 C_i^O拟合标准逻辑回归。

	使用 EM	不使用 EM	使用 EM	不含 EM	使用 EM	不使用 EM
$\beta_A = 0$	$n_i \sim DU(1, 5)$		$n_i = 3$		$n_i \sim DU(30, 50)$	
γ_I	0.01	0.41	0.01	< 0.01	0.02	0.02
γ_A	0.03	0.03	0.02	0.08	0.03	0.03
γ_{Z_1}	< 0.01	0.09	< 0.01	< 0.01	0.01	0.01
γ_{Z_2}	0.02	0.14	0.01	< 0.01	0.03	0.03
γ_{Z_3}	< 0.01	0.10	< 0.01	< 0.01	0.01	0.01
γ_{AZ_1}	< 0.01	0.25	< 0.01	0.08	0.03	0.03
γ_{AZ_2}	0.04	0.41	0.03	0.10	0.04	0.04
γ_{AZ_3}	< 0.01	0.29	0.02	0.01	0.07	0.03
η_I	< 0.01	1.18	0.01	0.04	< 0.01	< 0.01
η_A	0.01	0.23	< 0.01	0.04	< 0.01	< 0.01
η_{X_4}	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
η_{X_5}	< 0.01	0.01	< 0.01	< 0.01	< 0.01	< 0.01
η_{Z_4}	0.01	0.04	0.01	0.01	< 0.01	< 0.01
η_{AX_4}	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
η_{AX_5}	< 0.01	0.01	< 0.01	< 0.01	< 0.01	< 0.01
$\beta_A = 1.5$	$n_i \sim DU(1, 4)$		$n_i = 3$		$n_i \sim DU(30, 50)$	
γ_I	0.04	0.36	0.02	0.01	0.06	0.06
γ_A	0.01	0.40	< 0.01	0.20	0.03	0.03
γ_{Z_3}	0.01	0.08	< 0.01	0.02	0.03	0.03
γ_{Z_4}	0.01	0.09	< 0.01	0.01	0.02	0.02
γ_{AZ_3}	0.01	0.03	< 0.01	0.03	0.04	0.04
γ_{AZ_4}	< 0.01	0.13	< 0.01	0.07	0.01	0.01
η_I	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
η_A	< 0.01	0.01	0.01	0.01	0.01	0.01
η_{Z_3}	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
η_{X_1}	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
η_{X_2}	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
η_{X_3}	0.01	0.01	0.01	0.01	< 0.01	< 0.01
η_{X_4}	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
η_{AZ_3}	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
η_{AX_1}	< 0.01	< 0.01	0.01	0.01	< 0.01	< 0.01
η_{AX_2}	< 0.01	< 0.01	0.01	< 0.01	< 0.01	< 0.01
η_{AX_3}	0.01	0.01	0.01	0.01	< 0.01	< 0.01
η_{AX_4}	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

4 应用

我们使用引言和第 2.7 节中介绍的 Pro-CCM 研究来说明所提出的方法。Ratovoson等（2022年）研究了 pro-CCM 干预在马达加斯加 Mananjary 区降低疟疾患病率方面的效果。22 个集群（即村级行政区）按随机方式分配

至 pro-CCM（治疗组）或 iCCM（对照组），比例为 1:1。研究参与者嵌套于家庭中，而家庭又嵌套于各个村级行政区中。研究参与者的疾病状态通过基线和终点评估时的快速诊断检测（RDT）进行判定。此处我们关注终点评估时个体层面的 RDT 结果（RDT $= 1$ 若为阳性则记为 1，若为阴性则记为 0）。该数据集包含 29,683 名参与者、七个个体级基线协变量（男性指示变量 $X_{ijk,\ 1}$、年龄 $X_{ijk,\ 2}$、小学教育指示变量 $X_{ijk,\ 3}$、中学教育指示变量 $X_{ijk,\ 4}$、高等教育指示变量 $X_{ijk,\ 5}$、睡觉时使用蚊帐指示变量 $X_{ijk,\ 6}$、在院子里睡觉指示变量 $X_{ijk,\ 7}$）以及四个家庭层面的基线协变量（家庭规模$Z_{ij,\ 1}$、男性比例 $Z_{ij,\ 2}$、最高教育程度 $Z_{ij,\ 3}$、室内滞留喷洒指示变量 $Z_{ij,\ 4}$）。

终点时个体层面结局的总体缺失率为31%，相当于22.3%的家庭存在缺失。基于调整社会人口学特征的混合效应模型的结果表明，干预组与对照组参与者在终点时的快速诊断检测阳性率无统计学差异（OR $= 0.71$；95% CI：0.36–1.43）（Ratovoson 等，2022）。在假设结局存在协变量依赖型缺失的情况下，我们使用广义估计方程方法重新分析了该数据集，以估计边际治疗效应。我们基于赤池信息准则应用后退逐步法，为倾向评分选择协变量，得到以下模型：

$$\begin{aligned}\operatorname{logit}\left(\pi_{ijk};\boldsymbol{\theta}\right) =&\theta_I + \theta_A A_i + \sum_{q\in\{1,2,3,4\}} \theta_X^{(q)} X_{ijk,q} + \sum_{q\in\{1,2,3,4\}} \theta_Z^{(q)} Z_{ij,q} \\ &+ A_i \sum_{q\in\{1,2,3,4\}} \theta_{AX}^{(q)} X_{ijk,q} + A_i \sum_{q\in\{1,2,4\}} \theta_{AZ}^{(q)} Z_{ij,q}\end{aligned} \tag{23}$$

$$\operatorname{logit}\left(\lambda_{ij};\boldsymbol{\gamma}\right) =\gamma_I + \gamma_A A_i + \sum_{q\in\{1,3,4\}} \gamma_Z^{(q)} Z_{ij,q} + A_i \sum_{q\in\{4\}} \gamma_{AZ}^{(q)} Z_{ij,q} \tag{24}$$

$$\operatorname{logit}\left(\phi_{ijk};\boldsymbol{\eta}\right) =\eta_I + \eta_A A_i + \sum_{q\in\{1,2,3,4\}} \eta_X^{(q)} X_{ijk,q} + \sum_{q\in\{1,2,3,4\}} \eta_Z^{(q)} Z_{ij,q} A_i \sum_{q\in\{1,2,3\}} \eta_{AZ}^{(q)} Z_{ij,q} \tag{25}$$

模型拟合结果见网络附录F 中的表 S3。我们进行了以下四项分析：基于终点评估时具有已观测快速诊断检测结果的参与者的完全病例广义估计方程、使用模型 (23) 估计倾向评分的逆概率加权广义估计方程、使用模型 (24)和(25) 估计子集群和个体层面倾向评分的多重逆概率加权广义估计方程，以及使用两组 PS模型 $\mathcal{P}_1 = \{\lambda_{ij}^k(\boldsymbol{\gamma}^k)\ k = 1, 2\}$ 和 $\mathcal{P}_2 = \{\phi_{ijk}^\ell(\boldsymbol{\eta}^\ell)\ \ell = 1, 2\}$的多重稳健广义估计方程。$\mathcal{P}_1$ 包含模型 (24)以及另一个纳入 (A, AZ_1, AZ_3) 的模型； $\mathcal{P}_2$ 包含模型 (25) 以及另一个纳入 (A, X_2, $(X_2)^2$, X_3, X_5, Z_4, AZ_4) 的模型。多重逆概率加权广义估计方程和多重稳健广义估计方程中倾向评分模型的参数，均使用第 2.6 节提出的期望最大化算法进行估计。

完全病例广义估计方程得出的边际效应估计值（$OR = 0.76$，95%置信区间：0.46-1.23）与原始结果相似。纳入潜在信息性结局缺失的分析方法得出的效应估计值略微接近零假设（逆概率加权广义估计方程 $OR = 0.81$，95%置信区间：0.49-1.33；多重逆概率加权广义估计方程 $OR = 0.84$，95%置信区间：0.51-1.38；MMR-GEE $OR = 0.82$，95%置信区间：0.50-1.34）。不过，结论保持不变，因为所有方法所得的置信区间都包含零假设。未明确对子整群层面缺失进行建模的逆概率加权广义估计方程估计量，得出的效应估计值与我们提出的多重逆概率加权广义估计方程

和MMR-GEE估计量相似。这种相似性表明，子整群的缺失可能完全随机。事实上，尽管有22.3%的家庭缺失，但子整群层面缺失的估计概率均接近1（即，均值为 $P(C_{ij}=1|A_i, \boldsymbol{Z}_i; \hat{\boldsymbol{\gamma}}^{EM}) = 0.99$，范围为0.96-1.00）。在这个具体例子中，所有方法得出的结论相同；不过，所提出方法的应用使我们能够在对多层面结局缺失机制的一系列假设下，评估潜在信息性缺失对效应估计值的影响（Little et al., 2012）。

5 讨论

基于经验似然理论，本文提出了一种新的估计程序和推断方法，用于估计存在多层次结局缺失的整群随机试验中的边际治疗效应，并能防范PS模型的部分设定错误。为处理集群层面和个体层面的信息性缺失，我们推导了多层逆概率权重，并应用期望最大化算法来校正集群层面缺失指示变量可能存在的错误分类。所提出的MMR-GEE估计量允许分析者设定多组PS模型；只要 $\mathcal{P}_1$ 中的一个模型和 $\mathcal{P}_2$ 中的一个模型设定正确，且PS模型中的参数得到一致估计，就能获得一致的治疗效应估计值。

我们考察了多层结局数据缺失时的几种分析策略，并制作了流程图，以帮助选择这些方法（见图2）。首先，我们提出的方法针对多层次协变量依赖缺失设定。如果认为簇的缺失完全随机，那么采用逆概率加权广义估计方程方法来处理个体层面的信息性结局缺失或许就足够了。其次，尽管多重稳健广义估计方程允许设定多组PS模型，但如果分析者对真实的多层缺失过程有充分了解，也可以采用多重逆概率加权广义估计方程方法。最后，期望最大化算法旨在应对因 C_i中的错误分类而导致的PS模型参数估计难题；簇大小较小时，这种情况更有可能发生。当簇大小较大时，簇内所有个体结局均缺失的概率会变得非常小，因此 C_i 中的这类错误分类几乎不可能发生，此时无需使用期望最大化算法。尽管如此，当簇的大小不一且无法确定这种错误分类发生的可能性时，我们仍建议使用期望最大化算法。若 C_i中不存在错误分类，期望最大化算法将立即收敛，因此额外的计算负担极小。

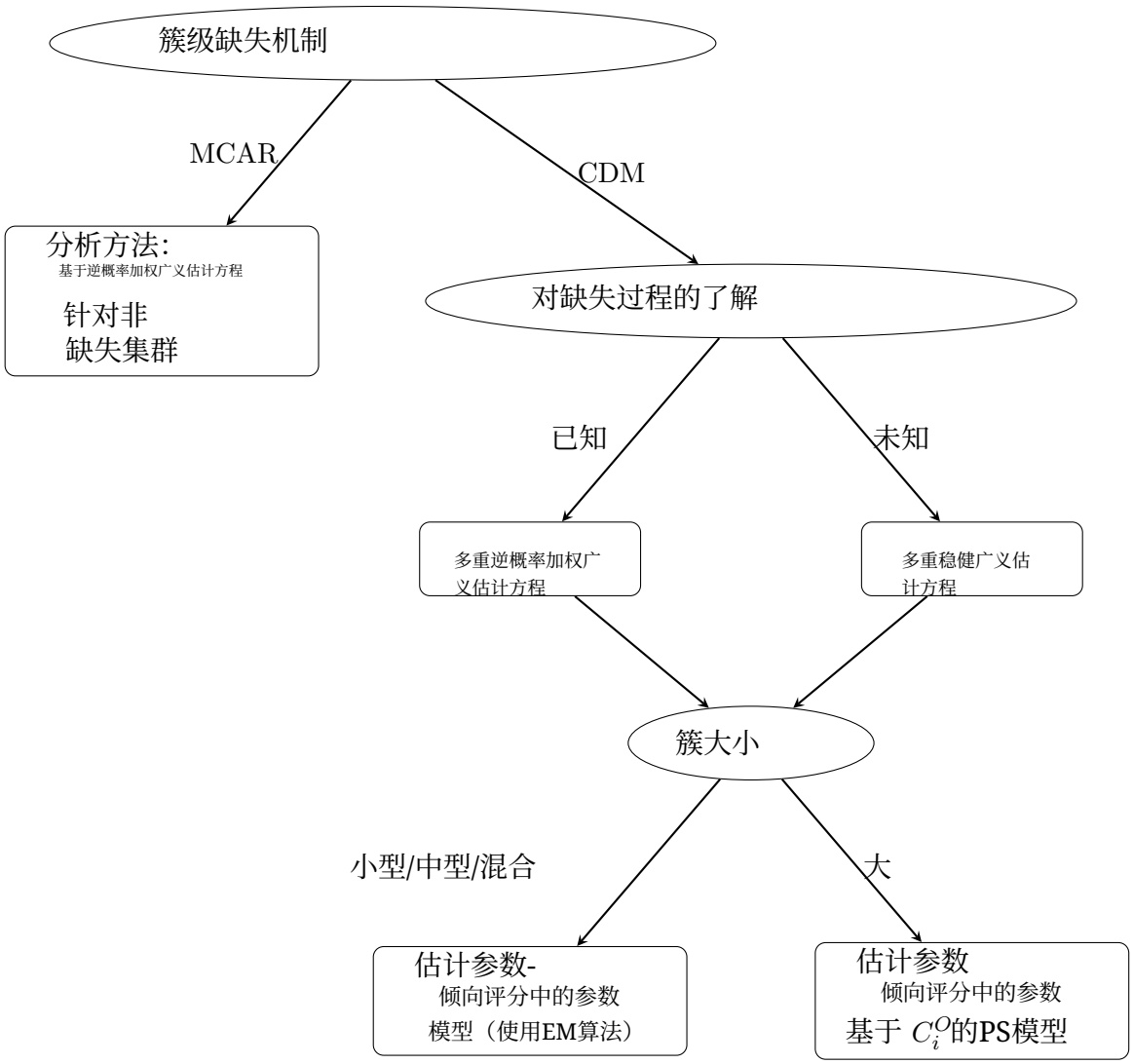


图2：不同缺失机制和情景下的建模与估计策略。

数据可用性声明

支持本文研究发现的数据可在 https://doi.org/10.7910/ DVN/IIDE2B 公开获取。

致谢

本文的研究部分由美国国立卫生研究院下属的美国国家过敏症和传染病研究所资助（R01 AI136947）。本文内容完全由作者负责，不一定代表美国国立卫生研究院的官方观点。

补充材料

第 2.4、2.5、2.6、3 和 4 节中引用的网络附录、表格和图均收录于补充材料中。

参考文献

Tom Chen、Eric JTchetgenTchetgen 和 RuiWang。用于估计关联参数的随机二阶广义估计方程方法。《计算与图形统计学杂志》，29(3)：547–561，2020 年。

Anthony Christopher Davison 和DavidVictorHinkley。《自助法及其应用》。第 1 号。剑桥大学出版社，1997 年。

Arthur P Dempster、NanM Laird 和 DonaldB Rubin。通过期望最大化算法从不完整数据中进行最大似然估计。英国皇家统计学会杂志：B辑，39(1)：1–22，1977。

KarlaDiaz-Ordaz、Michael G Kenward、Abie Cohen、Claire LColeman 和 SandraEldridge。整群随机试验中的缺失数据是否得到了充分处理？一项系统综述和指南。《临床试验》，11(5)：590–600，2014 年。

K DiazOrdaz、MG Kenward、M Gomes 和 R Grieve。整群随机试验中双变量结局的多重插补方法。《医学统计学》，35(20):3482–3496，2016 年。

ChristopherAField 和AlanH Welsh。对聚类数据进行自助法。英国皇家统计学会杂志：B辑，69(3)：369–390，2007。

Mallorie H Fiero、Shuang Huang、Eyal Oren 和 Melanie L Bell。整群随机试验中的统计分析与缺失数据处理：系统综述。《Trials》，17(1):72，2016。

Bruno Giraudeau 和 Philippe Ravaud。防止整群随机试验中的偏倚。《PLoS Medicine》，6 (5):e1000065，2009。Peisong Han。缺失数据回归分析中的多重稳健估计。《美国统计学会会刊》，109(507):1159–1173，2014。Peisong Han 和 Lu Wang。缺失数据估计：超越双重稳健性。《Biometrika》，100 (2):417–430，2013。Richard J Hayes 和 Lawrence H Moulton。整群随机试验。Chapman and Hall/CRC，2017。Anower Hossain、Karla Diaz-Ordaz 和 Jonathan W Bartlett。整群随机试验中协变量依赖性缺失下的二元结局缺失。《医学统计学》，36(19):3092–3109，2017a。Anower Hossain、Karla Diaz-Ordaz 和 Jonathan W Bartlett。整群随机试验中协变量依赖性缺失下的连续结局缺失。《Statistical Methods in Medical Research》，26(3):1543–1562，2017b。

Alan E Hubbard、Jennifer Ahern、Nancy L Fleischer、Mark Van der Laan、Sheri A Satariano、Nicholas Jewell、Tim Bruckner 和 William A Satariano。是使用广义估计方程还是不使用广义估计方程：比较总体平均模型与混合模型，以估计邻里风险因素与健康之间的关联。Epidemiology，21：467–474，2010。

Kung-Yee Liang 和 Scott L Zeger。使用广义线性模型分析纵向数据。Biometrika，73(1)：13–22，1986。

Roderick J Little、Ralph D'Agostino、Michael L Cohen、Kay Dickersin、Scott S Emerson、John T Farrar、Constantine Frangakis、Joseph W Hogan、Geert Molenberghs、Susan A Murphy 等。临床试验中缺失数据的预防与处理。New England Journal of Medicine，367(14)：1355–1360，2012。

Aya A Mitani、Elizabeth K Kaye 和 Kerrie P Nelson。利用逆概率删失权重处理具有信息性簇大小的纵向聚类数据中的失访。《应用统计年刊》，16(1):596–611，2022年。John C Nash。optimr：对"optim"函数的替代与扩展，2019年。网址：https: //CRAN.R-project.org/package=optimr。Art B Owen。经验似然。Chapman and Hall/CRC，2001年。Melanie Prague、Rui Wang、Alisa Stephens、Eric Tchetgen Tchetgen 和 Victor DeGruttola。考虑交互作用和复杂的个体间依赖性，以估计存在结局缺失的集群随机试验中的治疗效应。《生物计量学》，72(4):1066–1077，2016年。

Jing Qin、Biao Zhang 和 Denis HY Leung。《缺失数据问题中的经验似然》。美国统计学会会刊，104(488)：1492–1503，2009 年。

R 核心团队。R：用于统计计算的语言和环境。R 统计计算基金会，奥地利维也纳，2021。网址：https://www.R-project.org/。

Rila Ratovoson、Andres Garchitorena、Daouda Kassie、Jemima A Ravelonarivo、Voahangy Andrianaran- jaka、Seheno Razanatsiorimalala、Avotra Razafimandimby、Fanjasoa Rakotomanana、LaurieOhlstein、RezikyMangahasimbola等。主动社区病例管理降低了马达加斯加农村地区的疟疾患病率：一项整群随机试验的结果。BMC Medicine，20(1):1–15，2022年。

J Kyle Roberts 和 Xitao Fan。在多层/层级线性建模框架中进行自助法：使用 SAS 和 S-PLUS 的入门指南。《多元线性回归观点》，30(1):23–34，2004。James M Robins、Andrea Rotnitzky 和 Lue Ping Zhao。缺失数据情形下重复结局半参数回归模型的分析。《美国统计学会会刊》，90(429):106–121，1995。Joseph L Schafer 和 Recai M Yucel。含缺失值的多变量线性混合效应模型的计算策略。《计算与图形统计学杂志》，11(2):437–457，2002。Stanley Zlotkin、Samuel Newton、Ashley M Aimone、Irene Azindow、Seeba Amenga-Etego、Kofi Tchum、Emmanuel Mahama、Kevin E Thorpe 和 Seth Owusu-Agyei。铁强化对加纳婴儿和幼儿疟疾发病率的影响：一项随机试验。《美国医学会杂志》，310(9):938–947，2013。

arXiv:2304.05583v1 [stat.ME] 12 Apr 2023

《基于多层次结局缺失的整群随机试验中边际治疗效应的估计》的补充材料

网络附录 A 约束优化问题的解答

回顾一下，结局已观测个体的多重稳健权重为 e:

$$\hat{w}_{ij}^{MR} = \underset{w_{ij}}{\operatorname{argmax}} \prod_{i=1}^{s}\prod_{j=1}^{m_i} w_{ij}$$

满足以下约束：

$$w_{ij} \geq 0 \quad (i = 1, ..., s, j = 1, ..., m_i), \tag{a.1}$$

$$\sum_{i=1}^{s}\sum_{j=1}^{m_i} w_{ij} = 1, \tag{a.2}$$

$$\sum_{i=1}^{s}\sum_{j=1}^{m_i} w_{ij}\phi_{ij}^{k}(\hat{\boldsymbol{\eta}}^k)\lambda_i^{\ell}(\hat{\boldsymbol{\gamma}}^{\ell}) = \hat{\chi}^{k,\ell} \quad (k = 1, ..., K, \ell = 1, ..., L), \tag{a.3}$$

其中 $\hat{\chi}^{k,l} = \left(\sum_{i=1}^{M} n_i\right)^{-1} \sum_{i=1}^{M}\sum_{j=1}^{n_i} \phi_{ij}^{k}(\hat{\boldsymbol{\eta}}^k)\lambda_i^{\ell}(\hat{\boldsymbol{\gamma}}^{\ell})$。根据约束条件 (a.2)，(a.3) 可写为

$$\begin{aligned}
0 &= \sum_{i=1}^{s}\sum_{j=1}^{m_i} w_{ij}\phi_{ij}^{k}(\hat{\boldsymbol{\eta}}^k)\lambda_i^{\ell}(\hat{\boldsymbol{\gamma}}^{\ell}) - \hat{\chi}^{k,l} \\
&= \sum_{i=1}^{s}\sum_{j=1}^{m_i} w_{ij}\phi_{ij}^{k}(\hat{\boldsymbol{\eta}}^k)\lambda_i^{\ell}(\hat{\boldsymbol{\gamma}}^{\ell}) - \hat{\chi}^{k,l}\underbrace{\sum_{i=1}^{s}\sum_{j=1}^{m_i} w_{ij}}_{=1} \\
&= \sum_{i=1}^{s}\sum_{j=1}^{m_i} w_{ij}[\phi_{ij}^{k}(\hat{\boldsymbol{\eta}}^k)\lambda_i^{\ell}(\hat{\boldsymbol{\gamma}}^{\ell}) - \hat{\chi}^{k,l}] \quad (k = 1, ..., K, \ell = 1, ..., L).
\end{aligned} \tag{a.4}$$

回忆一下，$\hat{g}_{ij}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}}) = \{\phi_{ij}^{1}(\hat{\boldsymbol{\eta}}^1)\lambda_i^{1}(\hat{\boldsymbol{\gamma}}^1) - \hat{\chi}^{1,1}, ..., \phi_{ij}^{K}(\hat{\boldsymbol{\eta}}^K)\lambda_i^{L}(\hat{\boldsymbol{\gamma}}^L) - \hat{\chi}^{K,L}\}^{\mathrm{T}}$。因此，（a.4）的向量形式等价于：

$$\begin{aligned}
\mathbf{0} &= \sum_{i=1}^{s}\sum_{j=1}^{m_i} w_{ij}\left\{\phi_{ij}^{1}(\hat{\boldsymbol{\eta}}^1)\lambda_i^{1}(\hat{\boldsymbol{\gamma}}^1) - \hat{\chi}^{1,1}, ..., \phi_{ij}^{K}(\hat{\boldsymbol{\eta}}^K)\lambda_i^{L}(\hat{\boldsymbol{\gamma}}^L) - \hat{\chi}^{K,L}\right\}^{\mathrm{T}} \\
&= \sum_{i=1}^{s}\sum_{j=1}^{m_i} w_{ij}\hat{g}_{ij}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}}).
\end{aligned} \tag{a.5}$$

注意，最大化 $\prod_{i=1}^{s}\prod_{j=1}^{m_i} w_{ij}$ 相当于在约束条件 (a.1)-(a.3) 下最大化 $\sum_{i=1}^{s}\sum_{j=1}^{m_i}\log w_{ij}$。我们使用拉格朗日乘数法来求解这个优化问- 题。令

$$Q=\sum_{i=1}^{s}\sum_{j=1}^{m_i}\log w_{ij}+S\left(\sum_{i=1}^{s}\sum_{j=1}^{m_i}w_{ij}-1\right)-\left(\sum_{i=1}^{s}m_i\right)\boldsymbol{\rho}^{\mathrm{T}}\sum_{i=1}^{s}\sum_{j=1}^{m_i}w_{ij}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}}), \tag{a.6}$$

其中，S 是约束 (a.2) 的拉格朗日乘子，$\boldsymbol{\rho}^{\mathrm{T}}=(\rho_{11},\rho_{12},...,\rho_{KL})$ 是约束 (a.5) 的拉格朗日乘子。为方便起见，我们还引入常数项 $\sum_{i=1}^{s}m_i$。对 Q 分别关于 w_{ij}、S、$\boldsymbol{\rho}^{\mathrm{T}}$ 求偏导，并令其等于 0，可得

$$\frac{\partial Q}{\partial w_{ij}}=w_{ij}^{-1}+S-\left(\sum_{i=1}^{s}m_i\right)\boldsymbol{\rho}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})\overset{\text{set}}{=}0\ , \tag{a.7}$$

$$\frac{\partial Q}{\partial S}=\sum_{i=1}^{s}\sum_{i=1}^{m_i}w_{ij}-1\overset{\text{set}}{=}0, \tag{a.8}$$

$$\frac{\partial Q}{\partial \boldsymbol{\rho}^{\mathrm{T}}}=-\left(\sum_{i=1}^{s}m_i\right)\sum_{i=1}^{s}\sum_{j=1}^{m_i}w_{ij}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})\overset{\text{set}}{=}\mathbf{0}. \tag{a.9}$$

将 (a.7) 乘以 w_{ij}，我们得到

$$\begin{aligned}
0&=w_{ij}\frac{\partial Q}{\partial w_{ij}}\\
&=\sum_{i=1}^{s}\sum_{j=1}^{m_i}w_{ij}\frac{\partial Q}{\partial w_{ij}}\\
&=\sum_{i=1}^{s}\sum_{j=1}^{m_i}\left[1+Sw_{ij}-\left(\sum_{i=1}^{s}m_i\right)\boldsymbol{\rho}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})w_{ij}\right]\\
&=\sum_{i=1}^{s}m_i+S\underbrace{\sum_{i=1}^{s}\sum_{j=1}^{m_i}w_{ij}}_{=1}-\left(\sum_{i=1}^{s}m_i\right)\boldsymbol{\rho}^{\mathrm{T}}\underbrace{\sum_{i=1}^{s}\sum_{j=1}^{m_i}w_{ij}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}_{=0}\\
&=\sum_{i=1}^{s}m_i+S\\
&\Longrightarrow \hat{S}=-\sum_{i=1}^{s}m_i
\end{aligned}$$

ˆ将 S 代回(a.7)，我们得到

$$\begin{aligned}
\hat{w}_{ij}(\boldsymbol{\rho})&=\left[\sum_{i=1}^{s}m_i+\left(\sum_{i=1}^{s}m_i\right)\boldsymbol{\rho}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})\right]^{-1}\\
&=\frac{1}{\sum_{i=1}^{s}m_i}\frac{1}{1+\boldsymbol{\rho}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}.
\end{aligned} \tag{a.10}$$

ˆ$\boldsymbol{\rho}^{\mathrm{T}}$ 可通过将 ˆ$w_{ij}(\boldsymbol{\rho})$ 代入 (a.9) 并求解得到。

$$-\left(\sum_{i=1}^{s}m_i\right)\sum_{i=1}^{s}\sum_{j=1}^{m_i}\hat{w}_{ij}(\boldsymbol{\rho})\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})=-\sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}{1+\boldsymbol{\rho}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}=\mathbf{0}$$

$$\Longrightarrow \sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}{1+\boldsymbol{\rho}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}=\mathbf{0} \tag{a.11}$$

因此，多重稳健权重是

$$\hat{w}_{ij}^{MR}=\frac{1}{\sum_{i=1}^{s}m_i}\frac{1}{1+\hat{\boldsymbol{\rho}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}. \tag{a.12}$$

ˆ 网络附录B $\boldsymbol{\beta}_{MR}$的多重稳健性

本节证明了我们提出的MMR-GEE估计量的多重稳健性。在此，我们考虑 $n_i=n$ 和 $M\to\infty$的设定。此外，我们用下标星号表示概率极限，用（$\boldsymbol{\eta}_0,\boldsymbol{\gamma}_0$）表示PS模型的真实参数，并用 $\boldsymbol{\beta}_0$ 表示边际模型的真实参数。

定理 1。当 $\mathcal{P}_1$ 包含 $\phi_{ij}(\boldsymbol{\eta})$ 的正确模型且 $\mathcal{P}_2$ 包含 $\lambda_i(\gamma)$ 的正确模型时，作为 $M\to\infty$，ˆ $\boldsymbol{\beta}_{MR}\xrightarrow{P}\boldsymbol{\beta}_0$。

证明。不失一般性地，令 $\phi_{ij}^1(\boldsymbol{\eta}^1)$ 和 $\lambda_i^1(\boldsymbol{\gamma}^1)$ 分别为 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 中的第一个模型，且为正确设定的模型。我们首先证明：

$$\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)\xrightarrow{P}\mathbb{E}\left[\phi_{ij}^1\left(\boldsymbol{\eta}_*^1\right)\lambda_i^1\left(\boldsymbol{\gamma}_*^1\right)\right]\overset{\text{def}}{=}\chi_*^{1,1}. \tag{b.1}$$

详见网络附录 B.1，了解详细证明。接下来，我们建立多重稳健权重与p_{ij}之间的联系，其中p_{ij}是（Y_{ij}、 A_i、 $\boldsymbol{Z}_i$、 $\boldsymbol{X}_{ij}$）在$R_{ij}C_i=1$条件下对于 $i=1$、...、 s、 $j=1$、...、 m_i的经验概率。p_{ij}的估计量可以通过求解如下约束优化问题的另一种形式获得：

$$\hat{p}_{ij}=\underset{p_{ij}}{\operatorname{argmax}}\prod_{i=1}^{s}\prod_{j=1}^{m_i}p_{ij}, \tag{b.2}$$

满足以下约束：

$$p_{ij}\geq 0,\sum_{i=1}^{s}\sum_{j=1}^{m_i}p_{ij}=1,\sum_{i=1}^{s}\sum_{j=1}^{m_i}p_{ij}\frac{\left\{\phi_{ij}^k(\hat{\boldsymbol{\eta}}^k)\lambda_i^\ell(\hat{\boldsymbol{\gamma}}^\ell)-\chi^{k,l}(\hat{\boldsymbol{\eta}}^k,\hat{\boldsymbol{\gamma}}^\ell)\right\}}{\phi_{ij}^k(\hat{\boldsymbol{\eta}}^k)\lambda_i^\ell(\hat{\boldsymbol{\gamma}}^\ell)}=0\ \ (k=1,...,K,\ell=1,...,L). \tag{b.3}$$

这里，前两个约束确保p_{ij} 是适当的经验概率，第三个约束则是 (a.3) 的经验版本。应用网络附录 A 中相同的拉格朗日乘数法，我们得到：

$$\hat{p}_{ij}=\frac{1}{\sum_{i=1}^{s}m_i}\frac{1}{1+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}}, \tag{b.4}$$

w其中，ˆ$\boldsymbol{\varepsilon}^{\mathrm{T}}=$(ˆ$\varepsilon_{11}$, ˆ$\varepsilon_{12}$, ..., ˆ$\varepsilon_{KL}$) 是一个 $(KL)\times 1$ 向量，可解以下方程：

$$\sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}}{1+\boldsymbol{\varepsilon}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}}=\mathbf{0}. \tag{b.5}$$

经过简单的代数运算，ˆ$\boldsymbol{\rho}^{\mathrm{T}}$ 可表示为ˆ$\boldsymbol{\varepsilon}$：

$$\hat{\boldsymbol{\rho}}^{\mathrm{T}}=\left\{\frac{\hat{\varepsilon}_{11}+1}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)},\frac{\hat{\varepsilon}_{12}}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)},...,\frac{\hat{\varepsilon}_{KL}}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}\right\}, \tag{b.6}$$

据此可将$\hat{}w_{ij}^{MR}$写为:

$$\hat{w}_{ij}^{MR} = \hat{p}_{ij} \frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1, \hat{\boldsymbol{\gamma}}^1\right)}{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)}. \tag{b.7}$$

关于 $\hat{}w_{ij}$ 和 $\hat{}p_{ij}$ 之间关系的详细证明见网络附录B.2。最后，在假设 $\phi_{ij}^1(\boldsymbol{\eta}^1)$ 和 $\lambda_i^1(\boldsymbol{\gamma}^1)$ 是正确设定的模型的条件下，观测到的个体比例依概率收敛于 $\chi^{1,}{}_*{}^1$。也就是说，

$$\frac{\sum_{i=1}^s m_i}{\sum_{i=1}^M n_i} \xrightarrow{P} \chi_*^{1,1}, \tag{b.8}$$

证明见网络附录B.3。

利用（1）方程（b.7）中 $\hat{}w_{ij}^{MR}$和 $\hat{}p_{ij}$ 之间的关系，（2）方程（b.8）中观测到的个体比例的收敛性，以及（3）基于经验似然理论的 $\hat{}\boldsymbol{\varepsilon} \xrightarrow{P} 0$这一事实（Han，2014），我们证明多重稳健权重等价于（推导见网络附录B.4）：

$$\hat{w}_{ij}^{MR} = \frac{1 + o_p(1)}{\left(\sum_{i=1}^M n_i\right)\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}, \quad i = 1, \ldots, s, j = 1, \ldots m_i, \tag{b.9}$$

由此可建立如下关系:

$$\hat{\boldsymbol{W}}_i^{MR} = \operatorname{diag}\left[R_{ij}C_i\hat{w}_{ij}^{MR}\right]_{j=1,\ldots,n_i} = \operatorname{diag}\left[\frac{R_{ij} + o_p(1)}{\left(\sum_{i=1}^M n_i\right)\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\right]_{j=1,\ldots,n_i}. \tag{b.10}$$

方程（b.10）表明，多重稳健权重渐近正比于正确设定的 MIPW-GEE 估计量的多层权重。随着集群数 M 趋于无穷大，我们有:

$$\begin{aligned}
&\left(\frac{\sum_{i=1}^M n_i}{M}\right)\sum_{i=1}^M \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1}\hat{\boldsymbol{W}}_i^{MR}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \\
&= \frac{1}{M}\sum_{i=1}^M \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1}\operatorname{diag}\left[\frac{R_{ij}C_i + o_p(1)}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\right](\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \\
&\xrightarrow{p} E\left\{\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1}\operatorname{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\right](\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i))\right\} = 0.
\end{aligned} \tag{b.11}$$

证明见网络附录B.5。 □

B.1 证明 $\chi^{1,1}(\hat{}\boldsymbol{\eta}^1, \hat{}\boldsymbol{\gamma}^1) \xrightarrow{P} \mathbb{E}\left[\phi_{ij}^1(\boldsymbol{\eta}_*^1)\ \lambda_i^1(\boldsymbol{\gamma}_*^1)\right] \stackrel{\text{def}}{=} \chi^{1,}{}_*{}^1$

根据连续映射定理，$\hat{}\boldsymbol{\eta}^1 \xrightarrow{P} \boldsymbol{\eta}_*^1$ 和 $\hat{}\boldsymbol{\gamma}^1 \xrightarrow{P} \boldsymbol{\gamma}_*^1$ 意味着 $\phi_{ij}(\hat{}\boldsymbol{\eta}^1) \xrightarrow{P} \phi_{ij}(\boldsymbol{\eta}_*^1)$ 和$\lambda_i(\hat{}\boldsymbol{\gamma}^1) \xrightarrow{P} \lambda_i(\boldsymbol{\gamma}_*^1)$。再次应用连续映射定理，我们有 $\phi_{ij}(\hat{}\boldsymbol{\eta}^1)\lambda_i(\hat{}\boldsymbol{\gamma}^1) \xrightarrow{P}\phi_{ij}(\boldsymbol{\eta}_*^1)\lambda_i(\boldsymbol{\gamma}_*^1)$】【。

$$\begin{aligned}
\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1, \hat{\boldsymbol{\gamma}}^1\right) &= \left(\sum_{i=1}^M n_i\right)^{-1}\sum_{i=1}^M\sum_{j=1}^{n_i}\phi_{ij}^k(\hat{\boldsymbol{\eta}}^k)\lambda_i^\ell(\hat{\boldsymbol{\gamma}}^\ell) \\
&= \frac{1}{M}\sum_{i=1}^M\left(\frac{M}{\sum_{i=1}^M n_i}\right)\sum_{j=1}^{n_i}\phi_{ij}^k(\hat{\boldsymbol{\eta}}^k)\lambda_i^\ell(\hat{\boldsymbol{\gamma}}^\ell)
\end{aligned}$$

$$
\begin{aligned}
&\xrightarrow{P} \lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}\left[\left(\frac{M}{\sum_{i=1}^{M} n_i}\right)\sum_{j=1}^{n_i}\phi_{ij}^{k}(\boldsymbol{\eta}_*^k)\lambda_i^{\ell}(\boldsymbol{\gamma}_*^{\ell})\right] \\
&= \lim_{M\to\infty} \frac{\sum_{i=1}^{M}\sum_{j=1}^{n_i}\mathbb{E}\left[\phi_{ij}^{k}(\boldsymbol{\eta}_*^k)\lambda_i^{\ell}(\boldsymbol{\gamma}_*^{\ell})\right]}{\sum_{i=1}^{M} n_i} \\
&= \chi_*^{1,1}
\end{aligned}
$$

B.2 证明 $\hat{w}_{ij}^{MR} = \hat{p}_{ij}\chi^{1,1}(\hat{\boldsymbol{\eta}}^1, \hat{\boldsymbol{\gamma}}^1) / \{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}$

首先，我们证明，求解多重稳健权重的 $\hat{\boldsymbol{\rho}}$ 等价于求解经验概率的 $\hat{\boldsymbol{\varepsilon}}$ 。回顾一下，$\hat{\boldsymbol{\varepsilon}}$ 是通过求解以下方程得到的:

$$
\sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}}{1+\boldsymbol{\varepsilon}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}} = \mathbf{0}.
$$

重新整理该方程，可得

$$
\begin{aligned}
\mathbf{0} &= \sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}}{1+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}} \\
&= \sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})} \\
&= \frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}\sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})} \\
&= \frac{1}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}\sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}{\frac{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}+\frac{\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}} \\
&= \frac{1}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}\sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}{1+\frac{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)-\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}+\frac{\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})} \\
&= \frac{1}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}\sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}{1+\left\{\frac{\hat{\varepsilon}_{11}+1}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)},\frac{\hat{\varepsilon}_{12}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)},...,\frac{\hat{\varepsilon}_{KL}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}\right\}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}
\end{aligned}
$$

因此，我们得出如下关系：方程 (b.5) 的解等价于

$$
\hat{\boldsymbol{\rho}}^{\mathrm{T}} = \left\{\frac{\hat{\varepsilon}_{11}+1}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)},\frac{\hat{\varepsilon}_{12}}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)},...,\frac{\hat{\varepsilon}_{KL}}{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}\right\}.
$$

将 $\hat{\boldsymbol{\rho}}^{\mathrm{T}}$ 的这一备择形式代入多重稳健权重，我们得到

$$
\begin{aligned}
\hat{w}_{ij}^{MR} &= \left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\left(\frac{1}{1+\hat{\boldsymbol{\rho}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}\right) \\
&= \left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\left(\frac{1}{1+\left\{\frac{\hat{\varepsilon}_{11}+1}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)},\frac{\hat{\varepsilon}_{12}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)},...,\frac{\hat{\varepsilon}_{KL}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}\right\}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}\right) \\
&= \left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\left(\frac{1}{1+\frac{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)-\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}+\left\{\frac{\hat{\varepsilon}_{11}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)},\frac{\hat{\varepsilon}_{12}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)},...,\frac{\hat{\varepsilon}_{KL}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}\right\}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\left(\frac{1}{\frac{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)} + \left\{\frac{\hat{\varepsilon}_{11}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}, \frac{\hat{\varepsilon}_{12}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}, ..., \frac{\hat{\varepsilon}_{KL}}{\chi^{1,1}(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1)}\right\}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}\right) \\
&= \left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\left(\frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1) + \{\hat{\varepsilon}_{11},\hat{\varepsilon}_{12},...,\hat{\varepsilon}_{KL}\}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}\right) \\
&= \left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\left(\frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)/\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)}{1+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)}\right) \\
&= \underbrace{\left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\left(\frac{1}{1+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)}\right)}_{\hat{p}_{ij}}\left\{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)/\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\right\} \\
&= \hat{p}_{ij}\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)/\{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}
\end{aligned}
$$

B.3证明 $\frac{\sum_{i=1}^{s} m_i}{\sum_{i=1}^{M} n_i} \xrightarrow{P} \chi^1,\ {}_*{}^1$

$$
\begin{aligned}
\frac{\sum_{i=1}^{s} m_i}{\sum_{i=1}^{M} n_i} &= \frac{\sum_{i=1}^{M}\sum_{j=1}^{n_i} R_{ij}C_i}{\sum_{i=1}^{M} n_i} \\
&= \frac{1}{M}\sum_{i=1}^{M}\left(\frac{M}{\sum_{i=1}^{M} n_i}\right)\sum_{j=1}^{n_i} R_{ij}C_i \\
&\xrightarrow{P} \lim_{M\to\infty}\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}\left[\left(\frac{M}{\sum_{i=1}^{M} n_i}\right)\sum_{j=1}^{n_i} R_{ij}C_i\right] \\
&= \lim_{M\to\infty}\frac{1}{\sum_{i=1}^{M} n_i}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\mathbb{E}\left[R_{ij}C_i\right] \\
&= \lim_{M\to\infty}\frac{1}{\sum_{i=1}^{M} n_i}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\mathbb{E}\left\{\mathbb{E}\left[R_{ij}C_i \mid \boldsymbol{X}_{ij},\boldsymbol{Z}_i,A_i\right]\right\} \\
&= \lim_{M\to\infty}\frac{1}{\sum_{i=1}^{M} n_i}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\mathbb{E}\left\{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)\right\} \\
&= \mathbb{E}\left\{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)\right\} \\
&= \chi_*^{1,1}
\end{aligned}
$$

B.4 证明 $\hat{}w_{ij}^{MR} =$ $1+o_p(1)\ \underline{\left(\sum_{i=1}^{M} n_i\right)\phi_{ij}^1(\boldsymbol{\eta_0})\lambda_i^1(\boldsymbol{\gamma_0})}$

$$
\begin{aligned}
\hat{w}_{ij}^{MR} &= \hat{p}_{ij}\frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)} \\
&= \left\{\left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\frac{1}{1+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)\}}\right\}\frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)} \\
&= \left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}{\phi_{ij}^1(\hat{\boldsymbol{\eta}}^1)\lambda_i^1(\hat{\boldsymbol{\gamma}}^1)+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})} \\
&= \left(\frac{1}{\sum_{i=1}^{s} m_i}\right)\frac{\chi_*^{1,1}+o_p(1)}{\phi_{ij}^1(\boldsymbol{\eta}_0^1)\lambda_i^1(\boldsymbol{\gamma}_0^1)+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sum_{i=1}^{M} n_i} \underbrace{\left(\frac{\sum_{i=1}^{M} n_i}{\sum_{i=1}^{s} m_i} \right)}_{\xrightarrow{P} \left(\chi_*^{1,1}\right)^{-1}} \frac{\chi_*^{1,1} + o_p(1)}{\phi_{ij}^1(\boldsymbol{\eta}_0^1)\lambda_i^1(\boldsymbol{\gamma}_0^1) + \hat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \hat{g}_{ij}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}})} \\
&= \left(\frac{1}{\sum_{i=1}^{M} n_i} \right) \frac{1 + o_p(1)}{\phi_{ij}^1(\boldsymbol{\eta}_0^1)\lambda_i^1(\boldsymbol{\gamma}_0^1) + \hat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \hat{g}_{ij}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}})} \\
&= \left(\frac{1}{\sum_{i=1}^{M} n_i} \right) \frac{1 + o_p(1)}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}
\end{aligned}
$$

B.5证明 $\mathbb{E}\left\{ \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}_0, A_i)^{\mathrm{T}} \boldsymbol{V}_i^{-1} \underline{\mathrm{diag}\left[\frac{R_{ij}}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)} \right]} \underline{(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i))} \right\} = \mathbf{0}$

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{Y}_i, \boldsymbol{R}_i, \boldsymbol{C}_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i} \left\{ \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \mathrm{diag}\left[\frac{R_{ij} C_i}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)} \right] (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \right\} \\
&= \mathbb{E}_{\boldsymbol{Y}_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i} \left\{ \mathbb{E}_{\boldsymbol{R}_i, \boldsymbol{C}_i | \boldsymbol{Y}_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i} \left\{ \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \mathrm{diag}\left[\frac{R_{ij} C_i}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)} \right] (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \right\} \right\} \\
&= \mathbb{E}_{\boldsymbol{Y}_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i} \left\{ \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \underbrace{\mathbb{E}_{\boldsymbol{R}_i, \boldsymbol{C}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, A_i} \left\{ \mathrm{diag}\left[\frac{R_{ij} C_i}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)} \right] \right\}}_{=I} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \right\} \\
&= \mathbb{E}_{\boldsymbol{Y}_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i} \left\{ \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \right\} \\
&= \mathbb{E}_{\boldsymbol{Y}_i, A_i} \left\{ \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \right\} \\
&= \mathbb{E}_{A_i} \left\{ \mathbb{E}_{\boldsymbol{Y}_i | A_i} \left\{ \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)) \right\} \right\} \\
&= \mathbb{E}_{A_i} \left\{ \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i)}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \underbrace{\mathbb{E}_{\boldsymbol{Y}_i | A_i} \{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0, A_i) \}}_{=\mathbf{0}} \right\} \\
&= \mathbf{0}
\end{aligned}
$$

ˆ 网络附录C $\boldsymbol{\beta}_{MR}$ 的渐近分布

ˆ 本节证明 $\boldsymbol{\beta}_{MR}$ 的渐近正态性，定理 2 在网络附录C.2 中对此进行了总结。在证明定理 2 之前，我们首先在网络附录C.1 中建立引理1。与网络附录B 中的假设类似，我们考虑包含 $n_i = n$和 $M \to \infty$ 的设定。此外，我们用下标星号表示概率极限，用 $(\boldsymbol{\eta}_0, \boldsymbol{\gamma}_0)$ 表示 PS模型的真实参数，并用 $\boldsymbol{\beta}_0$ 表示边际模型的真实参数。

C.1 引理1的证明

引理1。当$\phi_{ij}^1(\boldsymbol{\eta}^1)$和$\lambda_i^1(\boldsymbol{\gamma}^1)$分别是 $\phi_{ij}(\boldsymbol{\eta})$ 和 $\lambda_i(\boldsymbol{\gamma})$ 的正确设定的模型时，我们有：

$$
\begin{aligned}
M^{1/2}\hat{\boldsymbol{\varepsilon}} =& \boldsymbol{G}^{-1}\left(M^{-1/2}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}} diag\left\{\frac{R_{ij}C_i-\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)\right\}\mathbb{I}\right.\\
&-\mathbb{I}^{\mathrm{T}}\mathbb{E}\left[diag\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\right]\mathbb{I}M^{1/2}\left(\hat{\boldsymbol{\eta}}^1-\boldsymbol{\eta}_0\right)\\
&\left.-\mathbb{I}^{\mathrm{T}}\mathbb{E}\left[diag\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right\}\right]\mathbb{I}M^{1/2}\left(\hat{\boldsymbol{\gamma}}^1-\boldsymbol{\gamma}_0\right)\right)+o_p(1),
\end{aligned}
$$

其中 $\boldsymbol{G}=\mathbb{I}^{\mathrm{T}}\mathbb{E}\left[\mathrm{diag}\left\{\frac{\boldsymbol{g}_{ij}(\boldsymbol{\eta}_*\ \boldsymbol{\gamma}_*\ ^{\otimes 2},)()()}{\phi_{ij}^1\ \boldsymbol{\eta}_0\ \lambda_i^1\ \boldsymbol{\gamma}_0}\right\}\right]\mathbb{I}$, $\mathbb{I}=(1,1,\ldots,1)^{\mathrm{T}}$，且 $\hat{}\boldsymbol{\varepsilon}^{\mathrm{T}}=(\hat{}\varepsilon_1,\ldots,\hat{}\varepsilon_{KL})$ 是一个 $KL\times 1$向量，满足以下方程：

$$
\sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}{1+\boldsymbol{\varepsilon}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}=\mathbf{0}.
$$

证明。为简化记号，令 $N=\sum_{i=1}^{M}n_i$。将方程 (b.5) 左边关于 $\left(\mathbf{0}^{\mathrm{T}},\boldsymbol{\eta}_*^{\mathrm{T}},\boldsymbol{\gamma}_*^{\mathrm{T}}\right)$ 作泰勒展开，得到：

$$
\begin{aligned}
\mathbf{0} &= M^{-1/2}\sum_{i=1}^{s}\sum_{j=1}^{m_i}\frac{\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}{1+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}\\
&= M^{-1/2}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\frac{R_{ij}C_i\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}{1+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}\\
&= M^{-1/2}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left\{\frac{R_{ij}C_i\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}{1+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}\right\}\mathbb{I}\\
&= M^{-1/2}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left\{\frac{R_{ij}C_ig_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)/\left\{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)\right\}}{1+\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)/\left\{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)\right\}}\right\}\mathbb{I}\\
&+\left\{\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}-\frac{R_{ij}C_ig_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\}^2\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right]\mathbb{I}\right\}\\
&\times M^{1/2}(\hat{\boldsymbol{\eta}}^1-\boldsymbol{\eta}_0)\\
&+\left\{\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}-\frac{R_{ij}C_ig_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\left\{\lambda_i^1\left(\boldsymbol{\gamma}_0\right)\right\}^2}\left\{\frac{\partial\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right]\mathbb{I}\right\}\\
&\times M^{1/2}(\hat{\boldsymbol{\gamma}}^1-\boldsymbol{\gamma}_0)\\
&+\sum_{k=2}^{K}\sum_{\ell=1}^{L}\left\{\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\partial\boldsymbol{\eta}_*^k}\right\}^{\mathrm{T}}\right]\mathbb{I}\right\}M^{1/2}(\hat{\boldsymbol{\eta}}^k-\boldsymbol{\eta}_*^k)\\
&+\sum_{k=1}^{K}\sum_{\ell=2}^{L}\left\{\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\partial\boldsymbol{\gamma}^\ell}\right\}^{\mathrm{T}}\right]\mathbb{I}\right\}M^{1/2}(\hat{\boldsymbol{\gamma}}^\ell-\boldsymbol{\gamma}_*^\ell)+o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&= M^{-1/2}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\right]\mathbb{I} - \left\{\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)^{\otimes 2}}{\left\{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)\right\}^2}\right]\mathbb{I}\right\} M^{1/2}\hat{\boldsymbol{\varepsilon}} \\
&+ \left\{\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\left(\left\{\frac{\partial\phi_{ij}^1(\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}^1}\lambda_i^1(\boldsymbol{\gamma}_0)\right\}^{\mathrm{T}} - \left\{\frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\frac{\partial\phi_{ij}^1(\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}^1}\lambda_i^1(\boldsymbol{\gamma}_0)\right\}^{\mathrm{T}}, \mathbf{0}\right)\right.\right. \\
&\left.\left. - \frac{R_{ij}C_i g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\left\{\phi_{ij}^1(\boldsymbol{\eta}_0)\right\}^2\lambda_i^1(\boldsymbol{\gamma}_0)}\left\{\frac{\partial\phi_{ij}^1(\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right]\mathbb{I}\right\} M^{1/2}(\hat{\boldsymbol{\eta}}^1 - \boldsymbol{\eta}_0) \\
&+ \left\{\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\left(\left\{\phi_{ij}^1(\boldsymbol{\eta}_0)\frac{\partial\lambda_i^1(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}} - \left\{\frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\phi_{ij}^1(\boldsymbol{\eta}_0)\frac{\partial\lambda_i^1(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}, \mathbf{0}\right)\right.\right. \\
&\left.\left. - \frac{R_{ij}C_i g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\phi_{ij}^1(\boldsymbol{\eta}_0)\left\{\lambda_i^1(\boldsymbol{\gamma}_0)\right\}^2}\left\{\frac{\partial\lambda_i^1(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right]\mathbb{I}\right\} M^{1/2}(\hat{\boldsymbol{\gamma}}^1 - \boldsymbol{\gamma}_0) \\
&+ \sum_{k=2}^{K}\sum_{\ell=1}^{L}\left\{\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\left(\mathbf{0}, \left\{\frac{\partial\phi_{ij}^k(\boldsymbol{\eta}_*^k)}{\partial\boldsymbol{\eta}^k}\lambda_i^\ell\left(\boldsymbol{\gamma}_*^\ell\right)\right\}^{\mathrm{T}}\right.\right.\right. \\
&\left.\left.\left. - \left\{\frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\frac{\partial\phi_{ij}^k(\boldsymbol{\eta}_*^k)}{\partial\boldsymbol{\eta}^k}\lambda_i^\ell\left(\boldsymbol{\gamma}_*^\ell\right)\right\}^{\mathrm{T}}\right)\right]\mathbb{I}\right\} M^{1/2}(\hat{\boldsymbol{\eta}}^k - \boldsymbol{\eta}_*^k) \\
&+ \sum_{k=1}^{K}\sum_{\ell=2}^{L}\left\{\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left[\frac{R_{ij}C_i}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\left(\mathbf{0}, \left\{\phi_{ij}^k\left(\boldsymbol{\eta}_*^k\right)\frac{\partial\lambda_i^\ell(\boldsymbol{\gamma}_*^\ell)}{\partial\boldsymbol{\gamma}^\ell}\right\}^{\mathrm{T}}\right.\right.\right. \\
&\left.\left.\left. - \left\{\frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\phi_{ij}^k\left(\boldsymbol{\eta}_*^k\right)\frac{\partial\lambda_i^\ell(\boldsymbol{\gamma}_*^\ell)}{\partial\boldsymbol{\gamma}^\ell}\right\}^{\mathrm{T}}\right)\right]\mathbb{I}\right\} M^{1/2}(\hat{\boldsymbol{\gamma}}^\ell - \boldsymbol{\gamma}_*^\ell) + o_p(1) \\
&= M^{-1/2}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\mathrm{diag}\left\{\frac{R_{ij}C_i g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\right\}\mathbb{I} - \mathbb{I}^{\mathrm{T}}\mathbb{E}\left[\mathrm{diag}\left\{\frac{g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)^{\otimes 2}}{\phi_{ij}^1(\boldsymbol{\eta}_0)\lambda_i^1(\boldsymbol{\gamma}_0)}\right\}\right]\mathbb{I} M^{1/2}\hat{\boldsymbol{\varepsilon}} \\
&- \mathbb{I}^{\mathrm{T}}\mathbb{E}\left[\mathrm{diag}\left\{\frac{g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\phi_{ij}^1(\boldsymbol{\eta}_0)}\left\{\frac{\partial\phi_{ij}^1(\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\right]\mathbb{I} M^{1/2}\left(\hat{\boldsymbol{\eta}}^1 - \boldsymbol{\eta}_0\right) \\
&- \mathbb{I}^{\mathrm{T}}\mathbb{E}\left[\mathrm{diag}\left\{\frac{g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\lambda_i^1(\boldsymbol{\gamma}_0)}\left\{\frac{\partial\lambda_i^1(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right\}\right]\mathbb{I} M^{1/2}\left(\hat{\boldsymbol{\gamma}}^1 - \boldsymbol{\gamma}_0\right) + o_p(1).
\end{aligned}
$$

重新整理最后一个方程即可得到结果。 □

C.2 定理 2 的证明

定理 2. 当 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 分别包含针对 $\lambda_i(\boldsymbol{\gamma})$ 和 $\phi_{ij}(\boldsymbol{\eta})$ 的正确设定模型时，$\sqrt{M}(\hat{\boldsymbol{\beta}}_{MR} - \boldsymbol{\beta}_0)$ 服从渐近正态分布，其均值为 0，方差为 $\mathrm{var}(\boldsymbol{Z}_i)$，其中

$$
\boldsymbol{Z}_i = \left[E\left\{\frac{\partial\boldsymbol{U}_i(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\right\}\right]^{-1}\left[\boldsymbol{Q}_i - \left\{\mathbb{E}\left(\boldsymbol{Q}_i\boldsymbol{S}_{1,i}^{\mathrm{T}}\right)\right\}\left\{\mathbb{E}\left(\boldsymbol{S}_{1,i}^{\otimes 2}\right)\right\}^{-1}\boldsymbol{S}_{1,i} - \left\{\mathbb{E}\left(\boldsymbol{Q}_i\boldsymbol{S}_{2,i}^{\mathrm{T}}\right)\right\}\left\{\mathbb{E}\left(\boldsymbol{S}_{2,i}^{\otimes 2}\right)\right\}^{-1}\boldsymbol{S}_{2,i}\right].
$$

ˆ 证明。回想一下， $\boldsymbol{\beta}_{MR}$ 是估计方程的解:

$$\mathbf{0}=\sum_{i=1}^{M}\operatorname{diag}\left\{R_{ij}C_i\hat{w}_{ij}^{MR}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_{MR}\right).$$

我们对上述估计方程在 $\left(\mathbf{0}^{\mathrm{T}},\boldsymbol{\eta}_*^{\mathrm{T}},\boldsymbol{\gamma}_*^{\mathrm{T}},\boldsymbol{\beta}_0^{\mathrm{T}}\right)$ 处进行泰勒展开，并应用引理1中 ˆ$\boldsymbol{\varepsilon}$ 的渐近展开，得到:

$$\begin{aligned}
\mathbf{0}&=NM^{-1/2}\sum_{i=1}^{M}\operatorname{diag}\left\{R_{ij}C_i\hat{w}_{ij}^{MR}\right\}\boldsymbol{U}_i\left(\hat{\boldsymbol{\beta}}_{MR}\right)\\
&=NM^{-1/2}\sum_{i=1}^{M}\operatorname{diag}\left\{R_{ij}C_i\hat{p}_{ij}\frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)}{\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}\right\}\boldsymbol{U}_i\left(\hat{\boldsymbol{\beta}}_{MR}\right)\\
&=NM^{-1/2}\sum_{i=1}^{M}\operatorname{diag}\left\{R_{ij}C_i\left(\frac{1}{\sum_{i=1}^{s}m_i}\right)\frac{\chi^{1,1}\left(\hat{\boldsymbol{\eta}}^1,\hat{\boldsymbol{\gamma}}^1\right)/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}{1+\hat{\boldsymbol{\varepsilon}}^T\hat{g}_{ij}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\gamma}})/\left\{\phi_{ij}^1\left(\hat{\boldsymbol{\eta}}^1\right)\lambda_i^1\left(\hat{\boldsymbol{\gamma}}^1\right)\right\}}\right\}\boldsymbol{U}_i\left(\hat{\boldsymbol{\beta}}_{MR}\right)\\
&=M^{-1/2}\left(\frac{N}{\sum_{i=1}^{s}m_i}\right)\left\{\frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)\right\}\sum_{i=1}^{M}\operatorname{diag}\left\{\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\\
&-M^{-1/2}\left(\frac{N}{\sum_{i=1}^{s}m_i}\right)\left\{\frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)\right\}\sum_{i=1}^{M}\operatorname{diag}\left\{\frac{R_{ij}C_i\hat{\boldsymbol{\varepsilon}}g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\left[\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)\right]^2}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\\
&+\left(\frac{N}{\sum_{i=1}^{s}m_i}\right)\left[\sum_{i=1}^{M}\operatorname{diag}\left\{\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\lambda_i^1\left(\boldsymbol{\gamma}_0\right)\right\}^{\mathrm{T}}\right.\right.\\
&\left.\left.-\frac{R_{ij}C_i\chi^{1,1}\left(\boldsymbol{\eta}_0,\boldsymbol{\gamma}_0\right)}{\left\{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\right\}^2\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\right]M^{-1/2}(\hat{\boldsymbol{\eta}}^1-\boldsymbol{\eta}_0)\\
&+\left(\frac{N}{\sum_{i=1}^{s}m_i}\right)\left[\sum_{i=1}^{M}\operatorname{diag}\left\{\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\frac{\partial\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right.\right.\\
&\left.\left.-\frac{R_{ij}C_i\chi^{1,1}\left(\boldsymbol{\eta}_0,\boldsymbol{\gamma}_0\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\left\{\lambda_i^1\left(\boldsymbol{\gamma}_0\right)\right\}^2}\left\{\frac{\partial\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\right]M^{-1/2}(\hat{\boldsymbol{\gamma}}^1-\boldsymbol{\gamma}_0)\\
&+\left(\frac{N}{\sum_{i=1}^{s}m_i}\right)\left\{\frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)\right\}\left[\sum_{i=1}^{M}\operatorname{diag}\left\{\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\right\}\frac{\partial\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)}{\partial\boldsymbol{\beta}}\right]\\
&\times M^{-1/2}(\hat{\boldsymbol{\beta}}_{MR}-\boldsymbol{\beta}_0)+o_p(1)\\
&=M^{-1/2}\sum_{i=1}^{M}\operatorname{diag}\left\{\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)-\underbrace{\mathbb{E}\left[\operatorname{diag}\left\{\frac{g_{ij}(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\right]}_{=\boldsymbol{L}}M^{1/2}\hat{\boldsymbol{\varepsilon}}\\
&-\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\right]M^{1/2}(\hat{\boldsymbol{\eta}}^1-\boldsymbol{\eta}_0)\\
&-\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\right]M^{1/2}(\hat{\boldsymbol{\gamma}}^1-\boldsymbol{\gamma}_0)
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}\left[\frac{\partial \boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}}\right] M^{1/2}(\hat{\boldsymbol{\beta}}_{MR}-\boldsymbol{\beta}_0)+o_p(1)\\
&=M^{-1/2}\sum_{i=1}^{M}\operatorname{diag}\left\{\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\\
&-\mathbf{L}\mathbf{G}^{-1}M^{-1/2}\sum_{i=1}^{M}\mathbb{I}^{\mathrm{T}}\operatorname{diag}\left\{\frac{R_{ij}C_i-\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)\right\}\mathbb{I}\\
&-\mathbf{L}\mathbf{G}^{-1}\mathbb{I}^{\mathrm{T}}\mathbb{E}\left[\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\right]\mathbb{I}M^{1/2}\left(\hat{\boldsymbol{\eta}}^1-\boldsymbol{\eta}_0\right)\\
&-\mathbf{L}\mathbf{G}^{-1}\mathbb{I}^{\mathrm{T}}\mathbb{E}\left[\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\right]\mathbb{I}M^{1/2}\left(\hat{\boldsymbol{\eta}}^1-\boldsymbol{\eta}_0\right)\\
&-\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\right]M^{1/2}(\hat{\boldsymbol{\eta}}^1-\boldsymbol{\eta}_0)\\
&-\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)\right]M^{1/2}(\hat{\boldsymbol{\gamma}}^1-\boldsymbol{\gamma}_0)\\
&+\mathbb{E}\left[\frac{\partial \boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}}\right] M^{1/2}(\hat{\boldsymbol{\beta}}_{MR}-\boldsymbol{\beta}_0)+o_p(1)\\
&=M^{-1/2}\sum_{i=1}^{M}\underbrace{\operatorname{diag}\left\{\frac{R_{ij}C_i}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)-\mathbf{L}\mathbf{G}^{-1}\mathbb{I}^{\mathrm{T}}\operatorname{diag}\left\{\frac{R_{ij}C_i-\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)\right\}\mathbb{I}}_{=\boldsymbol{Q}_i(\boldsymbol{\gamma}_0,\boldsymbol{\eta}_0)}\\
&-\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)+\mathbf{L}\mathbf{G}^{-1}\mathbb{I}^{\mathrm{T}}\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\mathbb{I}\right]\\
&\times M^{1/2}\left(\hat{\boldsymbol{\eta}}^1-\boldsymbol{\eta}_0\right)\\
&-\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)+\mathbf{L}\mathbf{G}^{-1}\mathbb{I}^{\mathrm{T}}\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\mathbb{I}\right]\\
&\times M^{1/2}\left(\hat{\boldsymbol{\gamma}}^1-\boldsymbol{\gamma}_0\right)\\
&+\mathbb{E}\left[\frac{\partial \boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}}\right] M^{1/2}(\hat{\boldsymbol{\beta}}_{MR}-\boldsymbol{\beta}_0)+o_p(1)\\
&=M^{-1/2}\sum_{i=1}^{M}\boldsymbol{Q}_i(\boldsymbol{\gamma}_0,\boldsymbol{\eta}_0)\\
&-\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)+\mathbf{L}\mathbf{G}^{-1}\mathbb{I}^{\mathrm{T}}\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\mathbb{I}\right]\\
&\times M^{-1/2}\left\{E\left(\boldsymbol{S}_1^{\otimes 2}\right)\right\}^{-1}\sum_{i=1}^{M}\boldsymbol{S}_{1,i}\\
&-\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\partial\boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)+\mathbf{L}\mathbf{G}^{-1}\mathbb{I}^{\mathrm{T}}\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial\boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\mathbb{I}\right]
\end{aligned}
$$

$$\times M^{-1/2}\left\{E\left(\boldsymbol{S}_2^{\otimes 2}\right)\right\}^{-1}\sum_{i=1}^{M}\boldsymbol{S}_{2,i}$$
$$+\mathbb{E}\left[\frac{\partial \boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}}\right]M^{1/2}(\hat{\boldsymbol{\beta}}_{MR}-\boldsymbol{\beta}_0)+o_p(1).$$

最后一个等式是分别对得分函数$\boldsymbol{S}_{1,\ i}(\hat{\boldsymbol{\eta}}^1)$和$\boldsymbol{S}_{2,\ i}(\hat{\boldsymbol{\gamma}}^1)$在$\boldsymbol{S}_{1,\ i}(\boldsymbol{\eta}_0)$和$\boldsymbol{S}_{2,\ i}(\boldsymbol{\gamma}_0)$处进行泰勒展开得到的，结果如下：

$$M^{1/2}\left(\hat{\boldsymbol{\eta}}^1-\boldsymbol{\eta}_0\right)=M^{-1/2}\left\{\mathbb{E}\left(\boldsymbol{S}_{1,i}^{\otimes 2}\right)\right\}^{-1}\sum_{i=1}^{M}\boldsymbol{S}_{1,i}+o_p(1),$$
$$M^{1/2}\left(\hat{\boldsymbol{\gamma}}^1-\boldsymbol{\gamma}_0\right)=M^{-1/2}\left\{\mathbb{E}\left(\boldsymbol{S}_{2,i}^{\otimes 2}\right)\right\}^{-1}\sum_{i=1}^{M}\boldsymbol{S}_{2,i}+o_p(1).$$

经过进一步的代数运算并应用广义信息等式（即 Tsiatis（2006）的引理 9.1），首先有：

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial \phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial \boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)+\mathbf{L}\mathbf{G}^{-1}{}_{\mathbb{I}}{}^{\mathrm{T}}\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial \phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial \boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}_{\mathbb{I}}\right]\\
&=-\mathbb{E}\left\{\frac{\partial \boldsymbol{Q}_i\left(\boldsymbol{\gamma}_0,\boldsymbol{\eta}_0\right)}{\partial \boldsymbol{\eta}^1}\right\}\\
&=\mathbb{E}\left(\boldsymbol{Q}_i\boldsymbol{S}_{1,i}^{\mathrm{T}}\right),
\end{aligned}$$

其次，

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{diag}\left\{\frac{1}{\lambda_i^1\left(\boldsymbol{\gamma}_0\right)}\left\{\frac{\partial \lambda_i^1\left(\boldsymbol{\gamma}_0\right)}{\partial \boldsymbol{\gamma}^1}\right\}^{\mathrm{T}}\right\}\boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)+\mathbf{L}\mathbf{G}^{-1}{}_{\mathbb{I}}{}^{\mathrm{T}}\operatorname{diag}\left\{\frac{\boldsymbol{g}_{ij}\left(\boldsymbol{\eta}_*,\boldsymbol{\gamma}_*\right)}{\phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}\left\{\frac{\partial \phi_{ij}^1\left(\boldsymbol{\eta}_0\right)}{\partial \boldsymbol{\eta}^1}\right\}^{\mathrm{T}}\right\}_{\mathbb{I}}\right]\\
&=-\mathbb{E}\left\{\frac{\partial \boldsymbol{Q}_i\left(\boldsymbol{\gamma}_0,\boldsymbol{\eta}_0\right)}{\partial \boldsymbol{\gamma}^1}\right\}\\
&=\mathbb{E}\left(\boldsymbol{Q}_i\boldsymbol{S}_{2,i}^{\mathrm{T}}\right).
\end{aligned}$$

因此，我们有

$$\begin{aligned}
\mathbf{0}=&M^{-1/2}\sum_{i=1}^{M}\left[\boldsymbol{Q}_i-\left\{\mathbb{E}\left(\boldsymbol{Q}_i\boldsymbol{S}_{1,i}^{\mathrm{T}}\right)\right\}\left\{\mathbb{E}\left(\boldsymbol{S}_{1,i}^{\otimes 2}\right)\right\}^{-1}\boldsymbol{S}_{1,i}-\left\{\mathbb{E}\left(\boldsymbol{Q}_i\boldsymbol{S}_{2,i}^{\mathrm{T}}\right)\right\}\left\{\mathbb{E}\left(\boldsymbol{S}_{2,i}^{\otimes 2}\right)\right\}^{-1}\boldsymbol{S}_{2,i}\right]\\
&+\mathbb{E}\left\{\frac{\partial \boldsymbol{U}_i\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}}\right\}M^{1/2}\left(\hat{\boldsymbol{\beta}}_{\mathrm{MR}}-\boldsymbol{\beta}_0\right)+o_p(1)
\end{aligned}$$

重新整理上述等式即可得到所需结果。 □

网络附录D EM 算法的推导与实现

D.1 期望最大化算法的推导

表 S1总结了 C_i^O和 C_i的不同模式。为推动期望最大化算法的推导，我们考虑将观测数据划分为两组：第一组包含具有 $C_i^O=1$ 的簇，第二组包含具有 $C_i^O=0$的簇。当 $C_i^O=1$时，我们必须有 $C_i=1$。但是，当 $C_i^O=0$时，真实的 C_i 可以是 1 或 0。因此，对于集群

表S1：簇大小为 3时 C_i^O和 C_i 的不同模式。 C_i^O是观测簇级缺失指示变量， C_i 是由簇级缺失过程诱导的真实簇级缺失指示变量。每种模式展示了（$\boldsymbol{R}_i$、 C_i^O、 C_i）的不同组合。由于观测数据仅提供了关于（$\boldsymbol{R}_i$、 C_i^O）的信息，因此无法区分模式 1和模式 2。这会导致对 C_i的潜在错误分类，尤其是在簇大小较小时。

模式	R_{ij}	C_i^O	C_i
1	0	0	0
	0	0	0
	0	0	0
2	0	0	1
	0	0	1
	0	0	1
3	1	1	1
	0	1	1
	0	1	1

在第一组中，观测到的 C_i^O始终等于真实的 C_i。然而，对于第二组中的集群， C_i^O可能会被错误分类。我们将 C_i 视为部分观测数据。因此，“完整”数据为（$\boldsymbol{C}$、 $\boldsymbol{R}$、 $\boldsymbol{Y}$、 $\boldsymbol{A}$、 $\boldsymbol{Z}$、 $\boldsymbol{X}$）。完整数据似然为:

$$
\begin{aligned}
& L_c(\boldsymbol{C}, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\gamma}) \\
&= L(\boldsymbol{C}, \boldsymbol{R} \mid \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\gamma}) L(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}) \\
&\propto L(\boldsymbol{C}, \boldsymbol{R} \mid \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\gamma}) \\
&= L(\boldsymbol{C}, \boldsymbol{R} \mid \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\gamma}) \\
&= \prod_{i=1}^{M} L(\boldsymbol{R}_i, C_i \mid A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}, \boldsymbol{\gamma}) \\
&= \prod_{i: C_i^O = 1} L(\boldsymbol{R}_i \mid C_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}, \boldsymbol{\gamma}) L(C_i \mid A_i, \boldsymbol{Z}_i; \boldsymbol{\eta}, \boldsymbol{\gamma}) \\
&\quad \prod_{i: C_i^O = 0} L(\boldsymbol{R}_i \mid C_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}, \boldsymbol{\gamma}) L(C_i \mid A_i, \boldsymbol{Z}_i; \boldsymbol{\eta}, \boldsymbol{\gamma}) \\
&= \prod_{i: C_i^O = 1} P(\boldsymbol{R}_i = \boldsymbol{r}_i \mid C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta})^{I(\boldsymbol{R}_i = \boldsymbol{r}_i)} P(C_i = 1 \mid A_i, \boldsymbol{Z}_i, ; \boldsymbol{\gamma}) \\
&\quad \prod_{i: C_i^O = 0} \underbrace{P(\boldsymbol{R}_i = \mathbf{0} \mid C_i = 0, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta})}_{=1}{}^{(1 - C_i)} P(C_i = 0 \mid A_i, \boldsymbol{Z}_i, ; \boldsymbol{\gamma})^{(1 - C_i)} \\
&\qquad P(\boldsymbol{R}_i = \mathbf{0} \mid C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta})^{C_i} P(C_i = 1 \mid A_i, \boldsymbol{Z}_i, ; \boldsymbol{\gamma})^{C_i} \\
&= \prod_{i: C_i^O = 1} \left\{ \prod_{j=1}^{n_i} P(R_{ij} = 1 \mid C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta})^{R_{ij}} P(R_{ij} = 0 \mid C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta})^{1 - R_{ij}} \right\} \\
&\qquad P(C_i = 1 \mid A_i, \boldsymbol{Z}_i, ; \boldsymbol{\gamma}) \\
&\quad \prod_{i: C_i^O = 0} P(C_i = 0 \mid A_i, \boldsymbol{Z}_i, ; \boldsymbol{\gamma})^{(1 - C_i)} \left\{ \prod_{j=1}^{n_i} P(R_{ij} = 0 \mid C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}) \right\}^{C_i} \\
&\qquad P(C_i = 1 \mid A_i, \boldsymbol{Z}_i, ; \boldsymbol{\gamma})^{C_i}
\end{aligned}
$$

$$
\begin{aligned}
&= \prod_{i=1}^{M} \left\{ \prod_{j=1}^{n_i} P(R_{ij}=1 \mid C_i=1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta})^{R_{ij}} P(R_{ij}=0 \mid C_i=1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta})^{(1-R_{ij})} \right\}^{C_i^O} \\
&\quad P(C_i=1 \mid A_i, \boldsymbol{Z}_i, ; \boldsymbol{\gamma})^{C_i^O} P(C_i=0 \mid A_i, \boldsymbol{Z}_i, ; \boldsymbol{\gamma})^{(1-C_i)(1-C_i^O)} \\
&\quad \left\{ \prod_{j=1}^{n_i} P(R_{ij}=0 \mid C_i=1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}) \right\}^{C_i(1-C_i^O)} \quad P(C_i=1 \mid A_i, \boldsymbol{Z}_i, ; \boldsymbol{\gamma})^{C_i(1-C_i^O)} \\
&= \prod_{i=1}^{M} \left\{ \prod_{j=1}^{n_i} \left[\operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)^{R_{ij}} \left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right)^{(1-R_{ij})} \right] \right\}^{C_i^O} \operatorname{expit}\left(\boldsymbol{Z}_i^* \boldsymbol{\gamma}\right)^{C_i^O} \\
&\quad \left(1 - \operatorname{expit}\left(\boldsymbol{Z}_i^* \boldsymbol{\gamma}\right)\right)^{(1-C_i)(1-C_i^O)} \left\{ \prod_{j=1}^{n_i} \left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right\}^{C_i(1-C_i^O)} \operatorname{expit}\left(\boldsymbol{Z}_i^* \boldsymbol{\gamma}\right)^{C_i(1-C_i^O)}.
\end{aligned}
$$

完整数据对数似然为:

$$
\begin{aligned}
&\ell(\boldsymbol{C}, \boldsymbol{C}^O, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\gamma}) \\
&= \sum_{i=1}^{M} C_i^O \left(\sum_{j=1}^{n_i} R_{ij} \log\left(\operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) + (1-R_{ij}) \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right) + C_i^O \log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \boldsymbol{\gamma}\right)\right) + \\
&\quad (1-C_i)(1-C_i^O) \log\left(1 - \operatorname{expit}\left(\boldsymbol{Z}_i^* \boldsymbol{\gamma}\right)\right) + C_i(1-C_i^O) \left(\sum_{j=1}^{n_i} \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right) \\
&\quad C_i(1-C_i^O) \log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \boldsymbol{\gamma}\right)\right) \\
&= \sum_{i=1}^{M} C_i^O \left\{ \log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \boldsymbol{\gamma}\right)\right) + \sum_{j=1}^{n_i} R_{ij} \log\left(\operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) + (1-R_{ij}) \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right\} + \\
&\quad (1-C_i)(1-C_i^O) \log\left(1 - \operatorname{expit}\left(\boldsymbol{Z}_i^* \boldsymbol{\gamma}\right)\right) + \\
&\quad C_i(1-C_i^O) \left(\log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \boldsymbol{\gamma}\right)\right) + \sum_{j=1}^{n_i} \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right).
\end{aligned}
$$

E 步的条件期望为

$$
\begin{aligned}
&Q\left(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\eta}^{(\nu)}, \boldsymbol{\gamma}^{(\nu)}, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}\right) \\
&= \mathbb{E}_{\boldsymbol{\eta}^{(\nu)}, \boldsymbol{\gamma}^{(\nu)}} \left\{ \ell(\boldsymbol{C}, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\gamma}) \mid \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X} \right\} \\
&= \sum_{i=1}^{M} C_i^O \left\{ \log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \gamma\right)\right) + \sum_{j=1}^{n_i} R_{ij} \log\left(\operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) + (1-R_{ij}) \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right\} + \\
&\quad (1-w_i^{(\nu)})(1-C_i^O) \log\left(1 - \operatorname{expit}\left(\boldsymbol{Z}_i^* \gamma\right)\right) + \\
&\quad w_i^{(\nu)}(1-C_i^O) \left(\log\left(\operatorname{expit}\left(\boldsymbol{Z}_i^* \gamma\right)\right) + \sum_{j=1}^{n_i} \log\left(1 - \operatorname{expit}\left(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}\right)\right) \right),
\end{aligned}
$$

其中

$$
\begin{aligned}
w_i^{(\nu)} &= \mathbb{E}_{\boldsymbol{\eta}^{(\nu)}, \boldsymbol{\gamma}^{(\nu)}} \left[C_i \mid \boldsymbol{R}_i, Y_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i \right] \\
&= P(C_i=1 \mid \boldsymbol{R}_i, Y_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}^{(\nu)}, \boldsymbol{\gamma}^{(\nu)})
\end{aligned}
$$

$$
\begin{aligned}
&= P(C_i = 1 \mid \boldsymbol{R}_i, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}^{(\nu)}, \boldsymbol{\gamma}^{(\nu)}) \\
&= P(C_i = 1 \mid \boldsymbol{R}_i \neq \mathbf{0}, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}^{(\nu)}, \boldsymbol{\gamma}^{(\nu)}) I(\boldsymbol{R}_i \neq \mathbf{0}) \\
&\quad + P(C_i = 1 \mid \boldsymbol{R}_i = \mathbf{0}, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}^{(\nu)}, \boldsymbol{\gamma}^{(\nu)}) I(\boldsymbol{R}_i = \mathbf{0}) \\
&= I(\boldsymbol{R}_i \neq \mathbf{0}) + P(C_i = 1 \mid \boldsymbol{R}_i = \mathbf{0}, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}^{(\nu)}, \boldsymbol{\gamma}^{(\nu)}) I(\boldsymbol{R}_i = \mathbf{0}) \\
&= C_i^O + P(C_i = 1 \mid \boldsymbol{R}_i = \mathbf{0}, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}^{(\nu)}, \boldsymbol{\gamma}^{(\nu)})(1 - C_i^O) \\
&= C_i^O + (1 - C_i^O) \times \\
&\quad \left\{ \frac{P(\boldsymbol{R}_i = \mathbf{0} | C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}^{(\nu)}) P(C_i = 1 \mid A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\gamma}^{(\nu)})}{P(\boldsymbol{R}_i = \mathbf{0}, C_i = 1 \mid A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\gamma}^{(\nu)}, \boldsymbol{\eta}^{(\nu)}) + P(\boldsymbol{R}_i = \mathbf{0}, C_i = 0 \mid A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\gamma}^{(\nu)})} \right\} \\
&= C_i^O + (1 - C_i^O) \left\{ \frac{\prod_{j=1}^{n_i} \left(1 - \operatorname{expit}(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}^{(\nu)})\right) \operatorname{expit}(\boldsymbol{Z}_i^* \boldsymbol{\gamma}^{(\nu)})}{\prod_{j=1}^{n_i} \left(1 - \operatorname{expit}(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}^{(\nu)})\right) \operatorname{expit}(\boldsymbol{Z}_i^* \boldsymbol{\gamma}^{(\nu)}) + 1 - \operatorname{expit}(\boldsymbol{Z}_i^* \boldsymbol{\gamma}^{(\nu)})} \right\} \\
&= C_i^O + (1 - C_i^O) \left\{ \frac{\prod_{j=1}^{n_i} \left(1 - \operatorname{expit}(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}^{(\nu)})\right) \operatorname{expit}(\boldsymbol{Z}_i^* \boldsymbol{\gamma}^{(\nu)})}{1 - \operatorname{expit}(\boldsymbol{Z}_i^* \boldsymbol{\gamma}^{(\nu)}) \left[1 - \prod_{j=1}^{n_i} \left(1 - \operatorname{expit}(\boldsymbol{X}_{ij}^* \boldsymbol{\eta}^{(\nu)})\right)\right]} \right\}
\end{aligned}
$$

D.2 期望最大化算法的实现

下面的算法概述了期望最大化算法的实现。

算法 1 期望最大化算法的实现 1：初始化 $\nu = 0$ 和 $\varepsilon = 10^{-6}$2：根据 C_i^O拟合集群层面和个体层面的 PS模型，以初始化 $\boldsymbol{\gamma}^{(\nu)}$ 和 $\boldsymbol{\eta}^{(\nu)}$ 3：当 TRUE 时循环 4：根据 $(\boldsymbol{\gamma}^{(\nu)}, \boldsymbol{\eta}^{(\nu)})$计算 $w_i^{(\nu)}$ 5：$\boldsymbol{\gamma}^{(\nu+1)} = \text{argmax}\ Q\left(\boldsymbol{\gamma}, \boldsymbol{\eta}^{(\nu)}, \boldsymbol{C}^O, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}\right)$n6：$\boldsymbol{\eta}^{(\nu+1)} = \text{argmax}\ Q\left(\boldsymbol{\gamma}^{(\nu+1)}, \boldsymbol{\eta}, \boldsymbol{C}^O, \boldsymbol{R}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{X}\right)$n7：如果 $\|\ell_c\left(\boldsymbol{\gamma}^{(\nu)}, \boldsymbol{\eta}^{(\nu)}\right) - \ell_c\left(\boldsymbol{\gamma}^{(\nu+1)}, \boldsymbol{\eta}^{(\nu+1)}\right)\|^2 \leq \varepsilon$ 则 8：跳出 9：结束 if 10：$\nu = \nu + 1$11：结束循环 12：输出：$\hat{}\boldsymbol{\eta}^{EM} = \boldsymbol{\eta}^{(\nu+1)}$, $\hat{}\boldsymbol{\gamma}^{EM} = \boldsymbol{\gamma}^{(\nu+1)}$

Web AppendixE Details of Simulations Setup and Result s

E.1 加纳疟疾流行地区贫血治疗研究的结局和基线协变量描述性汇总

表 S2：结局（Y_{ij}）、集群层面协变量（$\boldsymbol{Z}_i$）和个体层面协变量（$\boldsymbol{X}_{ij}$）的描述性汇总统计量

	铁组	无铁组
	780 簇，967 名儿童	772 个簇，991 名儿童
	结局的描述性统计	
疟疾，每100儿童年的发病率 (Y_{ij})	60.9	65.1
缺失集群数（%）	12（1.5）	19 (2.5)
缺失参与者人数（%）	25 (2.7)	29 (2.9)
	集群层面协变量的描述性统计	
集群规模中位数（范围）	1（1 至 4）	1（1 至 5）
家庭教育，是	66.5%	67.6%
财富，高	51.2%	50.9 %
辅食 ≤ 6 个月	87.8 %	87.2 %
家庭规模，均值（标准差）	6.1（2.8）	6.0 (2.8)
	个体层面协变量的描述性统计	
Sex, male	51.3 %	50.7 %
至少食用过 1 种铁强化产品		
入组前 7 天内	1.9 %	2.8 %
月龄，均值（标准差） [范围]	19.5 (8.6) [6至 46]	19.4 (8.6) [6至 36]
消瘦 z 分数，均值（标准差）	-0.64 (0.99)	-0.62 (0.97)
发育迟缓 z 分数，均值（标准差）	-0.90 (1.18)	-0.88 (1.23)
体重不足 z 分数，均值（标准差）	-0.93 (0.99)	-0.91（0.97）
血红蛋白水平，g/dL，均值（标准差）	10.3（1.3）	10.3 (1.3)

E.2 数据生成过程

E.2.1 $\beta_A = 0$

我们生成协变量，使其与表 S2 中的汇总统计量相匹配。连续协变量——消瘦 z 分数、发育迟缓 z 分数和体重不足 z 分数——根据相应的观测样本均值、方差以及假设的 0.1 共同相关系数，从多元正态分布中模拟得出。年龄根据观测样本均值、方差和界限，从截断正态分布中模拟得出。血红蛋白水平根据观测样本均值和方差，从正态分布中模拟得出。对于二元协变量，家庭教育程度、财富、辅食 ≤ 6 个月和性别根据观测比例，从伯努利分布中模拟得出。至少食用过 1 种铁强化产品的数据，则根据逻辑斯蒂-正态模型（Donner 和 Klar，2000），在假设簇间标准差为 0.05 的条件下生成。对于计数数据，簇大小和家庭规模根据相应范围，从离散均匀分布中模拟得出。

治疗分配（A_i）是从概率为 $p = 0.5$ 的伯努利分布中模拟生成的。我们将主要结局 Y_{ij}（每 100 儿童年的疟疾发病率）视为连续变量，并将其生成为

$$Y_{ij} = \beta_I^* + \beta_A^* A_i + \boldsymbol{Z}_i^O \boldsymbol{\beta}_Z + \boldsymbol{X}_{ij}^O \boldsymbol{\beta}_X + A_i \boldsymbol{Z}_i^O \boldsymbol{\beta}_{AZ} + A_i \boldsymbol{X}_{ij}^O \boldsymbol{\beta}_{AX} + \delta_i + \varepsilon_i, \qquad \text{(E.1)}$$

其中$\boldsymbol{Z}_i^O$包含家庭规模、家庭教育和财富，而 $\boldsymbol{X}_{ij}^O$ 包含年龄、消瘦z分数和发育迟缓 z分数。$\delta_i \sim N(0, \sigma_\delta^2)$是簇随机截距，$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$是残差标准误。

通过对($\boldsymbol{Z}_i^O$、 $\boldsymbol{X}_{ij}^O$、 $A_i \boldsymbol{Z}_i^O$、 $A_i \boldsymbol{X}_{ij}^O$)、随机截距和残差标准误进行边际化所诱导的真实边际均值模型为 $\mathbb{E}[Y_{ij}|A_i] = \beta_I + \beta_A A_i$。该研究得出结论

每日摄入含铁 MNP 并未增加干预后期间的疟疾风险，因此选择了 $(\beta_I^*, \beta_A^*, \boldsymbol{\beta}_Z^{\mathrm{T}}, \boldsymbol{\beta}_X^{\mathrm{T}}, \boldsymbol{\beta}_{AZ}^{\mathrm{T}}, \boldsymbol{\beta}_{AX}^{\mathrm{T}})$，使截距 (β_I) 为 63.5，且边际治疗效应 (β_A) 为 0。参数值为 $\beta_I^* = 43.5$, $\beta_A^* = 0.5$,$\boldsymbol{\beta}_Z^{\mathrm{T}} = (0.5, 0.8, 0.7)$, $\boldsymbol{\beta}_X^{\mathrm{T}} = (0.8, 1.5, -0.5)$, $\boldsymbol{\beta}_{AZ}^{\mathrm{T}} = (0.3, -1.5, 0.2)$, $\boldsymbol{\beta}_{AX}^{\mathrm{T}} = (-0.1, 0.5, -1.1)$，由此得到 $(\beta_I, \beta_A) = (63.5, 0)$。

通过以下逻辑回归模型诱导结局缺失数据过程：

$$\operatorname{logit}\left(\lambda_i\left(A_i, \boldsymbol{Z}_i^C; \boldsymbol{\gamma}\right)\right) = \gamma_I + \gamma_A A_i + \boldsymbol{Z}_i^C \boldsymbol{\gamma}_{\boldsymbol{Z}} + A_i \boldsymbol{Z}_i^C \boldsymbol{\gamma}_{A\boldsymbol{Z}}, \tag{E.2}$$

$$\operatorname{logit}\left(\phi_{ij}\left(A_i, Z_i^I, \boldsymbol{X}_{ij}^I \mid C_i = 1; \boldsymbol{\eta}\right)\right) = \eta_I + \eta_A A_i + \eta_Z Z_i^I + \boldsymbol{X}_{ij}^I \boldsymbol{\eta}_X + A_i \boldsymbol{X}_{ij}^I \boldsymbol{\eta}_{AX}. \tag{E.3}$$

$\boldsymbol{Z}_i^C$包括家庭规模、家庭教育和财富。 Z_i^I包括辅食$\leq$ *6* 月。 $\boldsymbol{X}_{ij}^I$ 包括消瘦 z 分数和发育迟缓 z 分数。原研究中约有 2% 的簇和 2.8% 的总体参与者数据缺失。为作说明，我们通过选择 $\boldsymbol{\gamma}$ 和 $\boldsymbol{\eta}$ 来扩大缺失比例，使 12% 的簇数据缺失，且总体参与者中有 30% 的结局数据缺失。倾向评分的参数值为 $\{\gamma_I = 1.10, \gamma_A = -0.29, \boldsymbol{\gamma}_{\boldsymbol{Z}}^{\mathrm{T}} = (0.18, -0.29, -0.22)\, \boldsymbol{\gamma}_{A\boldsymbol{Z}}^{\mathrm{T}} = (0.26, -0.51, -0.36)\}$ 和 $\{\eta_I = 1.73, \eta_A = -0.36, \eta_Z^{\mathrm{T}} = 0.10, \boldsymbol{\eta}_{\boldsymbol{X}}^{\mathrm{T}} = (-0.11, 0.18)\, \boldsymbol{\eta}_{A\boldsymbol{X}}^{\mathrm{T}} = (-0.36, 0.41)\}$，得到 $P(C_i = 1 | A_i, \boldsymbol{Z}_i; \boldsymbol{\gamma}) = 0.14$和 $P(R_{ij} = 1 | C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}) = 0.20$。

E.2.2 $\beta_A = 1.5$

$\beta_A = 1.5$ 的数据生成过程与 $\beta_A = 0$的数据生成过程相似。治疗分配（A_i）根据概率为 $p = 0.5$ 的伯努利分布进行模拟。对于协变量，每位参与者都有一个个体层面协变量向量 $\boldsymbol{X}_{ij}$ =（X_{ij}^1、X_{ij}^2、X_{ij}^3、X_{ij}^4），其中（X_{ij}^1、X_{ij}^2、X_{ij}^3）来自多元正态分布，其标准误为（1、1.2、0.8），共同相关系数为0.1，且 X_{ij}^4 =（X_{ij}^1）^{2}。每个集群都有一个集群层面协变量向量 $\boldsymbol{Z}_i$ =（Z_i^1、Z_i^2、Z_i^3、Z_i^4）。（Z_i^1、Z_i^2）是（X_{ij}^1、X_{ij}^2）的汇总统计量，即 $Z_i^k = \frac{\sum_{j=1}^{n_i} X_{ij}^k}{n_i}$对于 k = 1，2。Z_i^3根据速率为1.2的泊松分布进行模拟，且$Z_i^4 \sim N$（1.2、1）。我们考虑了以下结局模型：

$$Y_{ij} = \beta_I^* + \beta_A^* A_i + \boldsymbol{Z}_i \boldsymbol{\beta}_Z + \boldsymbol{X}_{ij} \boldsymbol{\beta}_X + A_i \boldsymbol{Z}_i \boldsymbol{\beta}_{AZ} + A_i \boldsymbol{X}_{ij} \boldsymbol{\beta}_{AX} + \delta_i + \varepsilon_i, \tag{E.4}$$

其中， Y_{ij}为连续结局， $\delta_i \sim N(0, \sigma_\delta^2)$为簇随机截距， $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$为残差标准误。对（$\boldsymbol{Z}_i$、$\boldsymbol{X}_{ij}$、$A_i\boldsymbol{Z}_i$、$A_i\boldsymbol{X}_{ij}$）、随机截距和残差标准误进行边际化后，得到的真实边际均值模型为 $\mathbb{E}[Y_{ij}|A_i] = \beta_I + \beta_A A_i$。参数取值为 $\beta_I^* = 0$、 $\beta_A^* = 1.5$、 $\boldsymbol{\beta}_Z^{\mathrm{T}} = (2, -2.5, 1, -1)$、 $\boldsymbol{\beta}_X^{\mathrm{T}} = (1, 1.2, 0.5, -0.5)$、 $\boldsymbol{\beta}_{AZ}^{\mathrm{T}} = (0.8, -0.4, 1, -1)$、 $\boldsymbol{\beta}_{AX}^{\mathrm{T}} = (0.5, 0.3, 1, -1)$，由此得到（$\beta_I$、 β_A） =(0, 1.5)。

通过以下逻辑回归模型诱导结局数据缺失过程：

$$\operatorname{logit}\left(\lambda_i\left(A_i, \boldsymbol{Z}_i^C; \boldsymbol{\gamma}\right)\right) = \gamma_I + \gamma_A A_i + \boldsymbol{Z}_i^C \boldsymbol{\gamma}_{\boldsymbol{Z}} + A_i \boldsymbol{Z}_i^C \boldsymbol{\gamma}_{A\boldsymbol{Z}}, \tag{E.5}$$

$$\operatorname{logit}\left(\phi_{ij}\left(A_i, \boldsymbol{Z}_i^I, \boldsymbol{X}_{ij}^I \mid C_i = 1;\ \boldsymbol{\eta}\right)\right) = \eta_I + \eta_A A_i + \boldsymbol{Z}_i^I \boldsymbol{\eta}_Z + \boldsymbol{X}_{ij}^I \boldsymbol{\eta}_X + A_i \boldsymbol{Z}_i^I \boldsymbol{\eta}_{AZ} + A_i \boldsymbol{X}_{ij}^I \boldsymbol{\eta}_{AX}, \tag{E.6}$$

其中 $\boldsymbol{Z}_i^C = (Z_i^3, Z_i^4)$、 $\boldsymbol{Z}_i^I = Z_i^3$和 $\boldsymbol{X}_{ij}^I = (X_{ij}^1, X_{ij}^2, X_{ij}^3, X_{ij}^4)$。PS 的参数值为 $\{\gamma_I = 2.44,\ \gamma_A = 0.18,\ \gamma_{\boldsymbol{Z}} = (0.12, -0.39)$、 $\gamma_{A\boldsymbol{Z}} = (-0.22, -0.29)\}$ 和 $\{\eta_I = 1.73,\ \eta_A = -0.22,\ \eta_Z = -0.16,\ \boldsymbol{\eta}_{\boldsymbol{X}} = (0.18, 0.26, 0.03, 0.18)$、 $\eta_{AZ} = -0.05$、 $\boldsymbol{\eta}_{A\boldsymbol{X}} = (0.18, 0.26, -0.22, -0.29)\}$，得到 $P(C_i = 1 | A_i, \boldsymbol{Z}_i; \boldsymbol{\gamma}) = 0.14$和 $P(R_{ij} = 1 | C_i = 1, A_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \boldsymbol{\eta}) = 0.23$。

E.2.3 不同簇大小和组内相关系数值

为考察簇大小变化的影响，我们考虑了 $M = 1552$，其 $n_i \sim DU(1, 4)$、$n_i \sim DU(1, 5)$，以及 $n_i = 3$ （对应簇大小较小、簇数较多的情况）；还考虑了 $M = 300$，其 $n_i \sim DU(30, 50)$（对应簇大小较大、簇数中等的情况），其中 DU 为离散均匀分布。铁—疟疾研究报告的组内相关系数（ICC）为 0.0804（Ratovoson 等，2022）。ICC 的计算方式为 $\frac{\sigma_\delta^2}{\sigma_\delta^2+\sigma_\varepsilon^2}$。我们将 σ_ε^2 设为 0.5，得到 $\sigma_\delta^2 \approx 0.043$。此外，通过将 $(\sigma_\epsilon^2, \sigma_\delta^2) = (5, 1.25)$，我们还考察了较大的 ICC 值（0.2）。总而言之，我们的模拟情景包含 $2 \times 3 = 6$种组合，涵盖不同的边际治疗效应（零假设和备择假设）以及不同的簇大小（3 个水平）。

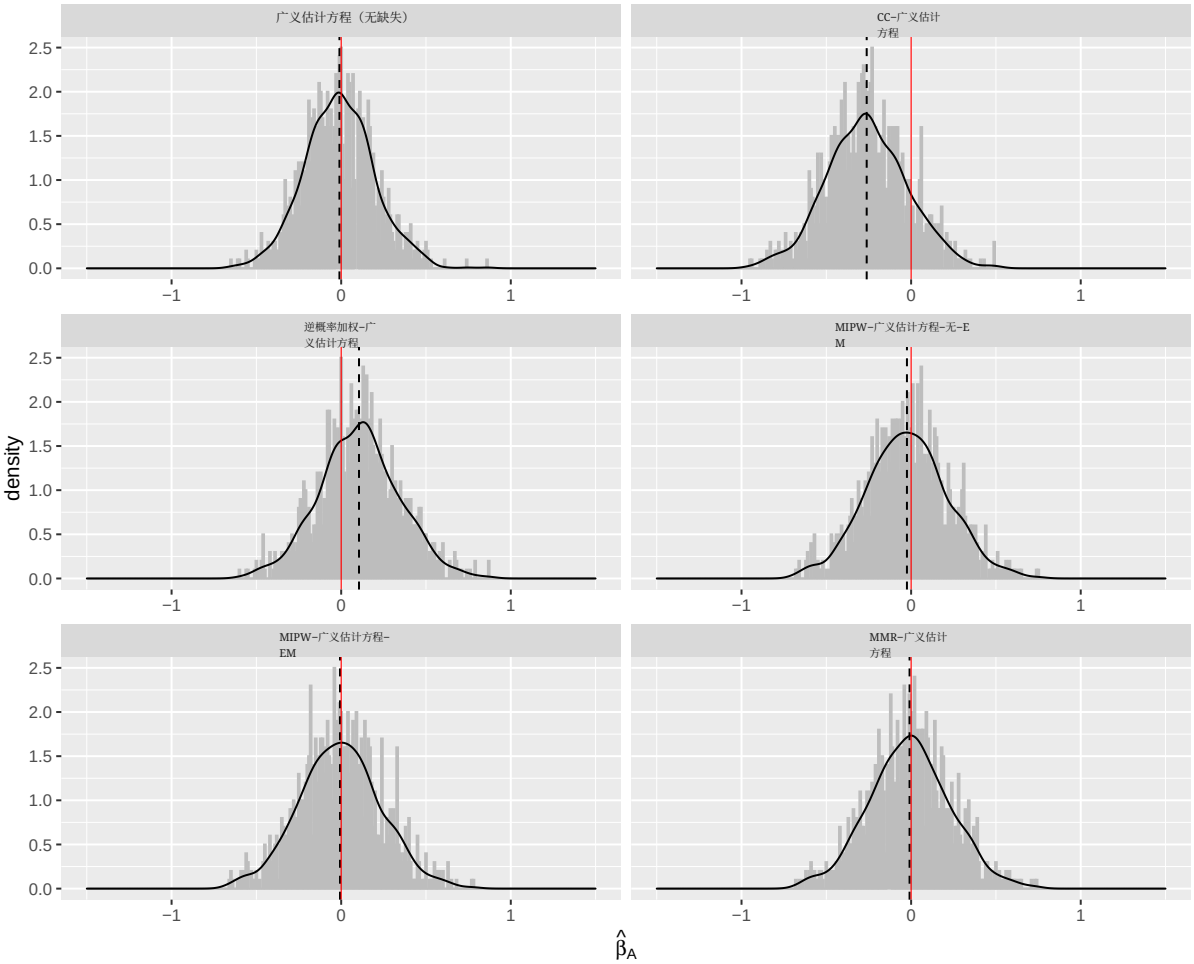
E.3 分析方法

我们比较了以下四种分析方法。首先，我们进行了未加权的 CC-GEE 分析。其次，我们采用了标准逆概率加权广义估计方程方法，其中，倾向评分通过无条件逻辑回归模型进行估计；该模型在 $\beta_A = 0$设定下采用与模型 (E.3) 相同的函数形式，在 $\beta_A = 1.5$ 设定下采用与模型 (E.6) 相同的函数形式，但忽略了整群层面缺失。第三，我们采用了多重逆概率加权广义估计方程方法，其中，集群层面和个体层面的倾向评分模型均设定正确。为了估计倾向评分模型中的参数，我们基于 C_i^O拟合了标准逻辑回归模型（记为多重逆概率加权广义估计方程-无期望最大化），并采用了期望最大化算法（记为 EM多重逆概率加权广义估计方程）。最后，我们通过设定 $\mathcal{P}_1 = \{\lambda_i^k(\boldsymbol{\gamma}^k)\ k = 1、2\}$ 和 $\mathcal{P}_2 = \{\phi_{ij}^\ell(\boldsymbol{\eta}^\ell)\ \ell = 1、2\}$，并使用期望最大化算法估计 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 中的所有参数，实现了我们提出的 MMR-GEE 估计量。 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 均包含一个设定正确的模型和一个设定错误的模型。

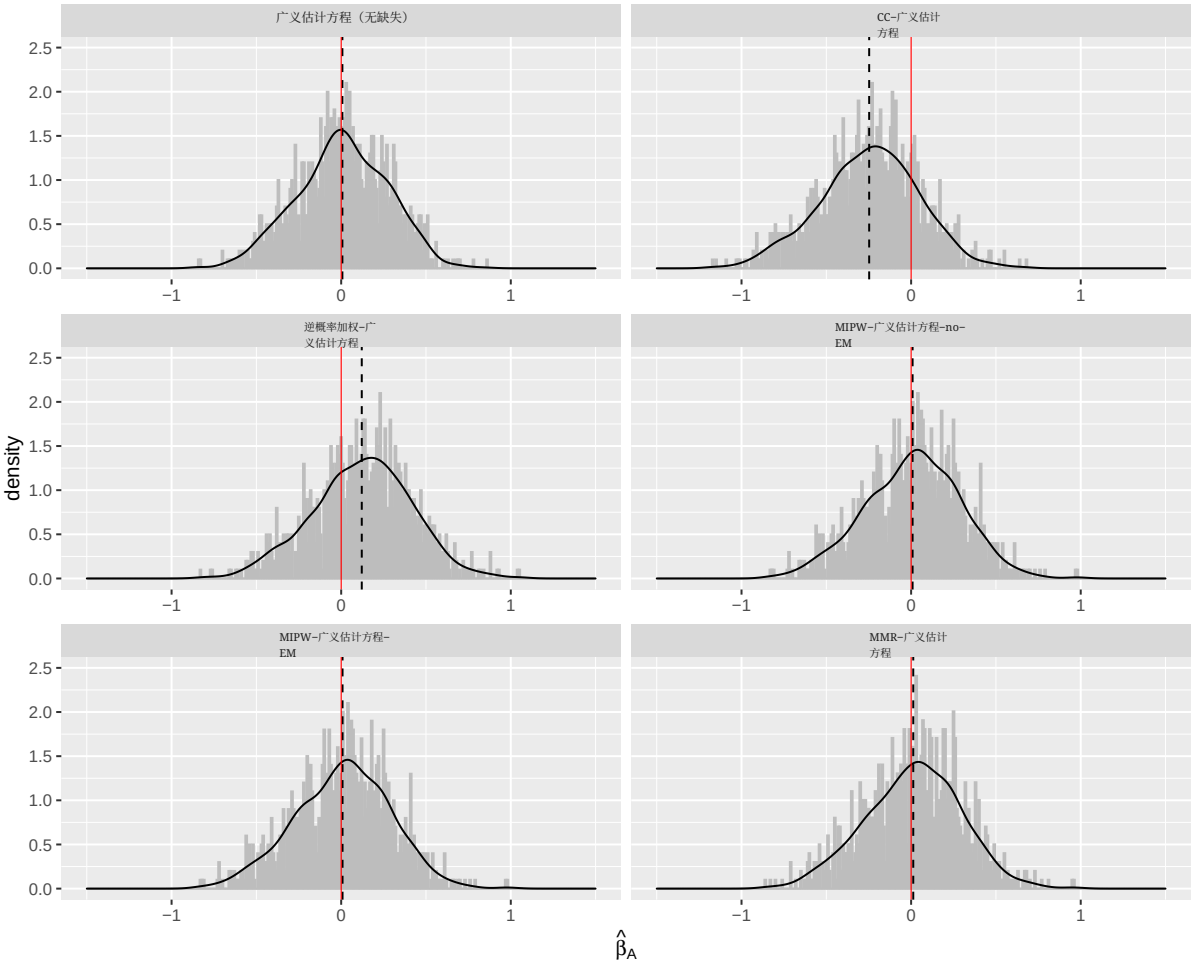
对于 $\beta_A = 0$ 设定，$\lambda_i^1(\boldsymbol{\gamma}^1)$和$\phi_{ij}^1(\boldsymbol{\eta}^1)$是正确设定的模型。$\lambda_i^2(\boldsymbol{\gamma}^2)$未纳入家庭规模、家庭教育、财富及其与处理指示变量的交互项。 $\phi_{ij}^2(\boldsymbol{\eta}^2)$未纳入消瘦 z 分数、发育迟缓 z 分数以及所有处理变量与协变量的交互项。此外， $\lambda_i^2(\boldsymbol{\gamma}^2)$和 $\phi_{ij}^2(\boldsymbol{\eta}^2)$都包含额外协变量，其中 $\lambda_i^2(\boldsymbol{\gamma}^2)$纳入了辅食 ≤ 6 月龄及其与处理变量的交互项，而$\phi_{ij}^2(\boldsymbol{\eta}^2)$纳入了年龄、体重不足 z 分数。对于 $\beta_A = 1.5$ 设定， $\lambda_i^1(\boldsymbol{\gamma}^1)$和 $\phi_{ij}^1(\boldsymbol{\eta}^1)$是正确设定的倾向评分模型。$\lambda_i^2(\boldsymbol{\gamma}^2)$未纳入$(Z_3, Z_4, AZ_3, AZ_4)$，而$\phi_{ij}^2(\boldsymbol{\eta}^2)$未纳入$(Z_3, X_1, X_3, X_4, AZ_3, AX_1, AX_3, AX_4)$。此外， $\lambda_i^2(\boldsymbol{\gamma}^2)$和 $\phi_{ij}^2(\boldsymbol{\eta}^2)$都包含额外协变量（$Z_1$, AZ_1）。

E.4 模拟结果

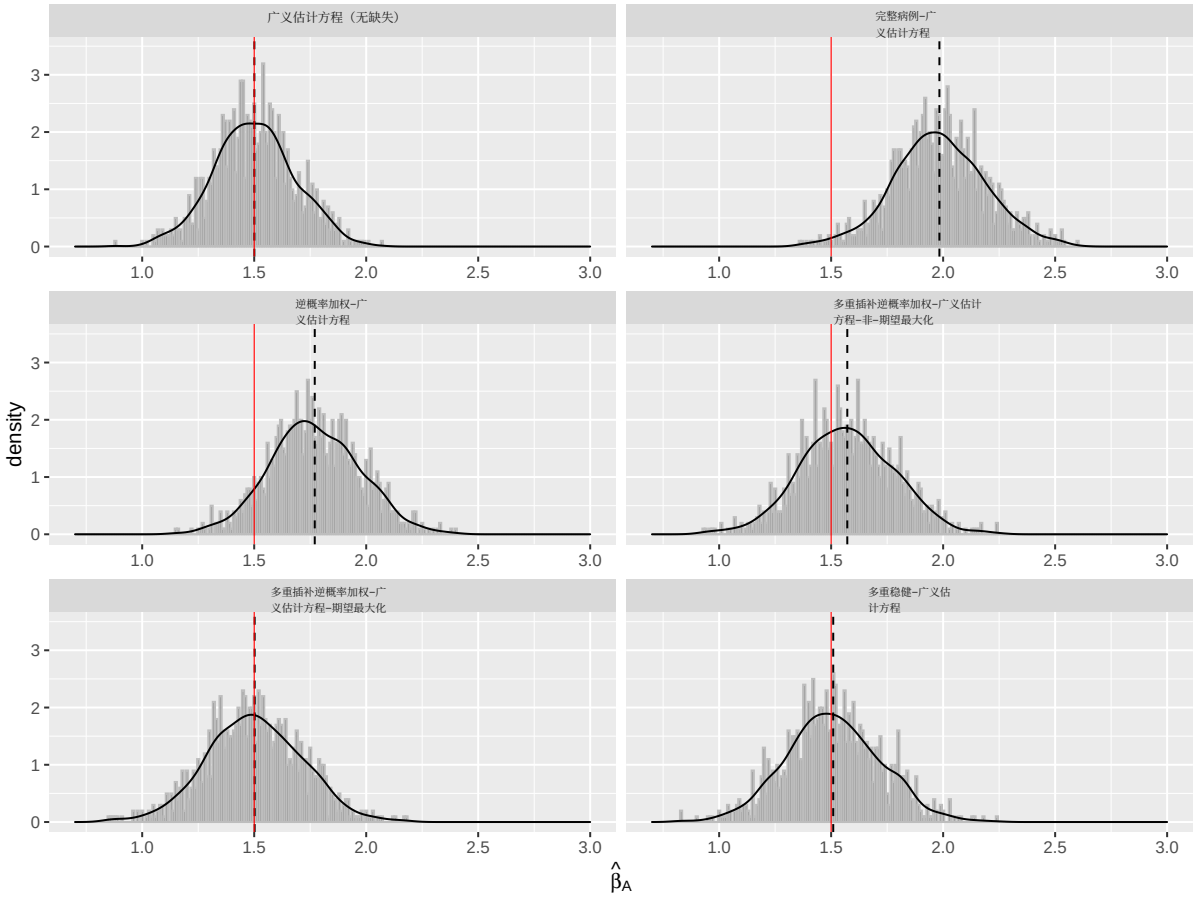
图（1）和（2）分别是在零假设下、 $\{n_i = 3, M = 1552, ICC = 0.0804\}$ 和 $\{n_i \sim DU(30, 50)$ $M = 300, ICC = 0.2\}$条件下效应估计值的经验分布。图（3）、（4）和（5）分别是在备择假设下、 $\{n_i \sim DU(1, 4)$ $M = 1552, ICC = 0.0804\}$, $\{n_i = 3, M = 1552, ICC = 0.0804\}$ 和 $\{n_i \sim DU(30, 50)$ $M = 300, ICC = 0.2\}$条件下效应估计值的经验分布。



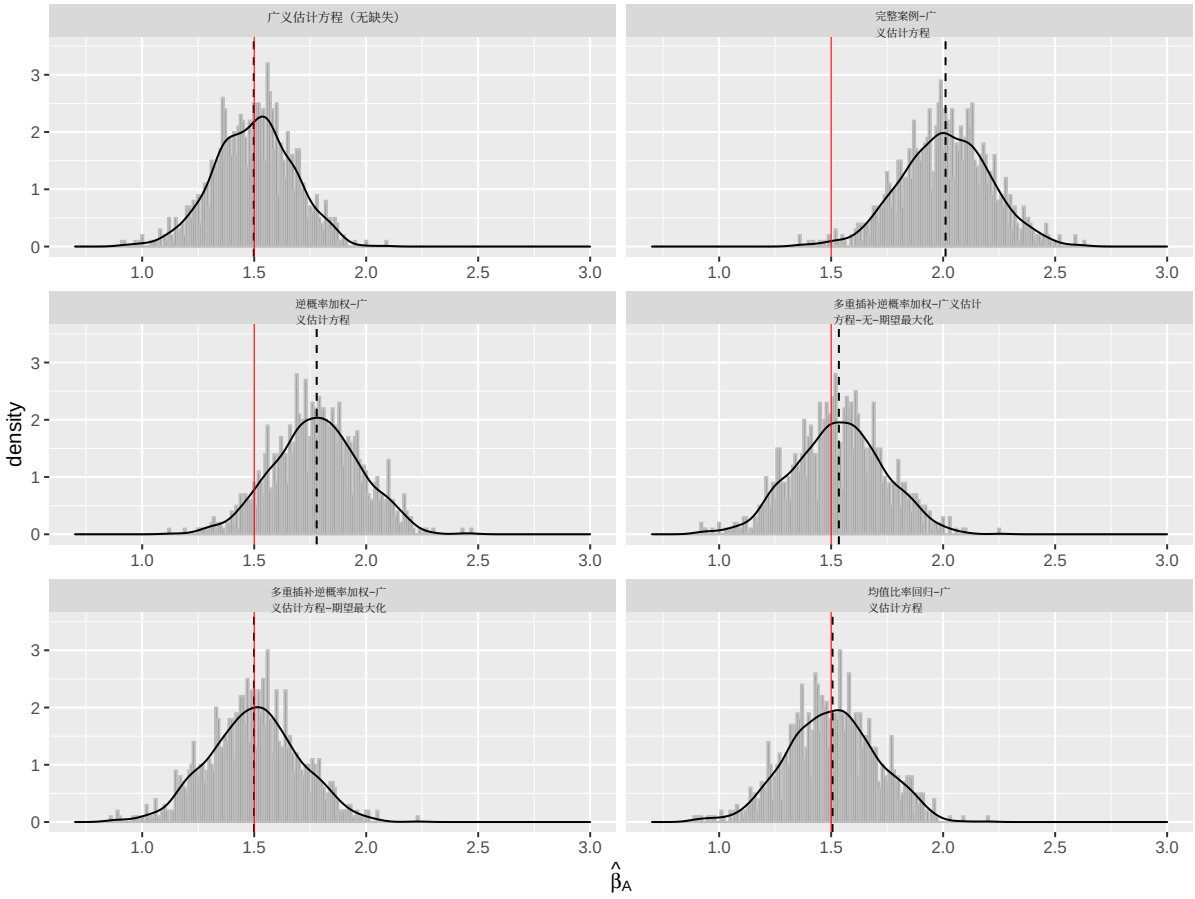
ˆ图1：基于 $n_i = 3$、 $M = 1552$、 $ICC = 0.08$ 设定下1,000次重复的估计的边际处理效应（β_A）的经验分布。红线表示真实边际处理效应（0）。虚线表示估计的边际处理效应的经验均值。



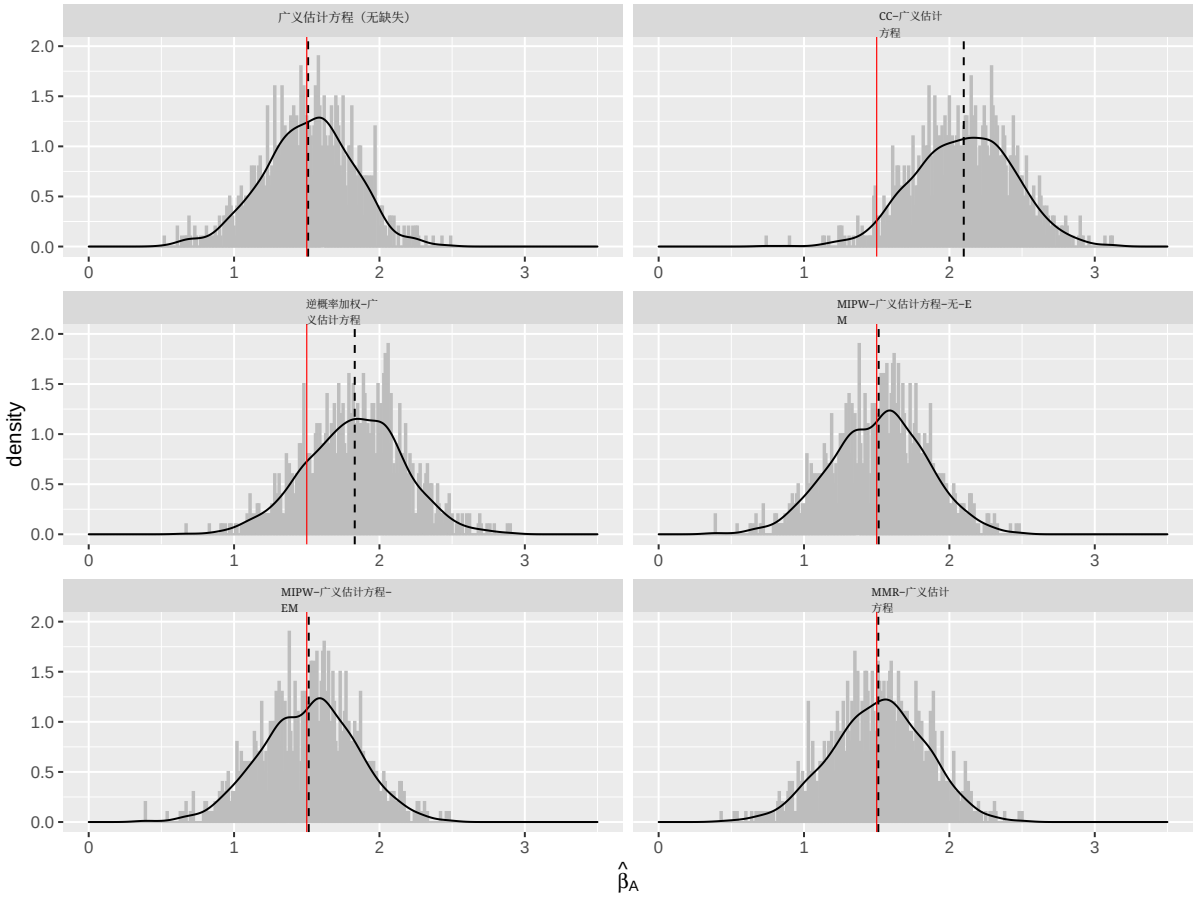
ˆ 图2：在 $ICC = 0.2$ 设定下，基于 $n_i \sim DU(30, 50)$ $M = 300$进行1,000次重复所得的估计的边际处理效应（β_A）的经验分布。红线表示真实边际处理效应（0）。点线表示估计的边际处理效应的经验均值。



ˆ 图3：在 $n_i \sim DU(1,4)$ $M = 1552$、$ICC = 0.08$ 设定下，基于1,000次重复得到的估计的边际治疗效应（β_A）的经验分布。红线表示真实边际治疗效应（1.5）。虚线表示估计的边际治疗效应的经验均值。



ˆ 图4：在 $n_i = 3, M = 1552, ICC = 0.08$ 设定下，基于 1,000次重复得到的估计的边际处理效应 (β_A) 的经验分布。红线表示真实边际治疗效应 (1.5)。虚线表示估计的边际处理效应的经验均值。



ˆ 图 5：基于 1,000 次重复，在 $n_i \sim DU(30, 50)$ $M = 300$、$ICC = 0.2$ 设定下估计的边际处理效应（β_A）的经验分布。红线表示真实边际处理效应（1.5）。点线表示估计的边际处理效应的经验均值。

Pro-CCM 研究网络附录 F 倾向评分模型拟合结果

表 S3：倾向评分模型拟合结果

	与.....相关的 p 值					
	$P(R_{ijk}=1)^a$		$P(C_{ij}^O=1)^b$		$P(R_{ijk}=1\|C_{ij}^O=1)^c$	
	$\theta_X=0$	$\theta_{AX}=0$	$\gamma_X=0$	$\gamma_{AX}=0$	$\eta_X=0$	$\eta_{AX}=0$
家庭层面协变量						
家庭规模	<0.01	<0.01	<0.01	0.89	<0.01	**0.01**
男性比例	<0.01	0.53	0.56	0.92	<0.01	**0.01**
最高教育程度	0.14	0.92	0.22	0.69	**0.02**	0.15
IRS	0.88	<0.01	**0.05**	**0.01**	<0.01	0.18
个体层面协变量						
Sex, male	0.12	0.15			<0.01	0.18
年龄，中位数（标准差）	<0.01	<0.01			<0.01	0.30
小学	0.18	0.43			<0.01	0.99
中学	<0.01	0.20			0.28	<0.01
高级学校	0.71	0.06			0.42	0.18
睡在蚊帐里	0.21	<0.01			<0.01	0.60
睡在院子里	0.22	0.62			0.34	0.30

[a] Wald 检验回归 logit$[P(R_{ijk}=1)]=\theta_0+\theta_A A_i+\theta_X X_{ijk}+\theta_{AX}A_iX_{ijk}$

[b] γ_X γ_{AX} $[P\ C_{ij}^O=1\ ]=\gamma_I+\gamma_A A_i+\gamma_X X_{ij}+\gamma_{AX}A_iX_{ij}$ 中的 θ_X 和 θ_{AX} Wald 检验

[c] 回归 logit ( ) 中的 和 Wald 检验回归 logit$[P(R_{ijk}=1|C_{ij}^O=1)]=\eta_I+\eta_A A_i+\eta_X X_{ijk}+\eta_{AX}A_iX_{ijk}$ 中的 η_X 和 η_{AX}

参考文献

Allan Donner 和 Neil Klar。《健康研究中整群随机试验的设计与分析》。伦敦：Arnold，2000年。

Peisong Han。《缺失数据回归分析中的多重稳健估计》。《美国统计学会会刊》，109(507):1159–1173，2014年。

Rila Ratovoson、Andres Garchitorena、Daouda Kassie、Jemima A Ravelonarivo、Voahangy Andrianaranjaka、Seheno Razanatsiorimalala、Avotra Razafimandimby、Fanjasoa Rako-tomanana、Laurie Ohlstein、Reziky Mangahasimbola等。《主动式社区病例管理降低了马达加斯加农村地区的疟疾患病率：一项整群随机试验的结果》。《BMC Medicine》，20(1):1–15，2022年。

Anastasios A Tsiatis。《半参数理论与缺失数据》。2006年。