arXiv:2611.32888v1 [stat.ME] 24 Nov 2026

有限重叠且伴侣记录链接缺失时推广夫妻随机试验

Subir Hait

密歇根州立大学咨询、教育心理学与特殊教育系，美国密歇根州东兰辛
48824 haitsubi@msu.edu ORCID 0009-0004-9871-9677

2026年11月

摘要

我们研究当某些伴侣类型未出现在试验中，且目标调查报告了伴侣特征却未链接配偶记录时，如何推广夫妻随机试验。在伴侣画像项不受限制的加性效应模型下，当妻子与丈夫的边际分布在试验支持图的每个连通分量上赋予相同的概率质量时，目标平均效应恰好可识别。若某一指定交互作用沿至少一个环的交替对比不为零，则该交互作用可识别。在缺少记录链接的情况下，我们通过在符合调查边际分布的配对中进行优化，得到精确界。配偶优势比可收窄二元交互作用的界限。我们推导了索引患者设计的联合界限，量化常见捷径造成的偏差，并使用贝叶斯二元多层回归与后分层估计效应和界限。在一个基于 3,826 对健康与退休研究夫妇及模拟试验结局的实例分析中，索引患者试验与未关联调查所得的伴侣平均效应识别集宽度为 2.93 毫米汞柱。将调查记录链接后，识别集宽度缩小至 2.38 毫米汞柱。加入 300 对无糖尿病夫妇后，未关联调查的识别集宽度缩小至 0.28 毫米汞柱；在拟合模型下，记录链接后则实现点识别。

关键词： 可推广性；可迁移性；二元数据；部分识别；多层回归与后分层；复杂抽样调查；Stan

1 引言

许多健康项目是面向夫妇而非个人实施的（Martire 等，2010；Arden-Close 和 McGrath，2017）。这类项目包括围绕饮食、身体活动、吸烟、减重以及慢性疾病（如2型糖尿病）自我管理的夫妻项目（Arden-Close 和 McGrath，2017；Martire 等，2010；Trief 等，2016）。配偶共享膳食、日常习惯和住所，而且他们在健康状况和健康行为方面也具有相似性（Meyler 等，2007）。例如，配偶患有糖尿病的人，其患糖尿病的风险更高（Leong 等，2014）。研究还表明，当一方配偶改变某种健康行为时，另一方配偶

也经常会发生变化（Falba 和 Sindelar，2008；Jackson 等，2015）。因此，研究人员

认为，伴侣应纳入行为改变项目（Jackson 等，2015；Martire etal.，2010）。

在夫妻随机试验中，随机分配的单位是夫妻双方，且会对双方都进行测量（Trief 等，2016；Arden-Close 和 McGrath，2017）。对一方的影响可能取决于其自身特征，也取决于另一方的特征，这就是二元数据中的行动者-伴侣结构（Kenny 等，2006；Cook 和 Kenny，2005）。夫妻项目也可能使非患者伴侣受益，例如减轻伴侣的糖尿病困扰（Trief 等，2019）。例如，如果妻子的丈夫也患有糖尿病，那么饮食项目可能对她更有效，因为双方都会改变共同的膳食。这种情况是合理的，因为配偶往往会一起改变健康习惯（Falba 和 Sindelar，2008；Jackson 等，2015）。

夫妻试验很少招募具有代表性的夫妻。招募通常通过索引患者进行，例如一位糖尿病控制不佳的人（Trief 等，2016）。由于配偶彼此相似（Meyler 等，2007；Leong 等，2014），伴侣特征也会受到筛选。这样一来，试验平均值可能不同于与政策相关的夫妻总体中的效应（Imai 等，2008；Stuart 等，2011）。要将试验结果推广到该总体，我们需要一项描述目标夫妻的总体调查（Ackerman 等，2021；Degtiar 和 Rose，2023）。健康与退休研究（HRS）就是针对老年伴侣的一项此类调查，因为它会对配偶双方进行访谈（Sonnega 等，2014；Falba 和 Sindelar，2008）。然而，许多其他数据来源只描述个人，而不将配偶联系起来（Ridder 和 Moffitt，2007；D’Orazio 等，2006）。此外，试验本身可能从未招募某些类型的夫妻，这会违反推广方法所需的正值性条件（Petersen 等，2012；Degtiar 和 Rose，2023）。

以上讨论表明，伴侣对试验会同时引发三个问题。每位伴侣的效应都可能取决于另一位伴侣的特征（Cook 和 Kenny，2005；Kenny 等，2006）。试验可能无法招募到所有类型的伴侣对（Trief 等，2016；Petersen 等，2012）。目标调查可能会描述每位伴侣，却不将配偶相互关联（Ridder 和 Moffitt，2007；D’Orazio 等，2006）。这三个问题彼此交织。即使具体配对并不重要，伴侣特征仍可能产生影响；而配对只有通过行动者与伴侣交互作用才会产生影响。这些交互作用恰恰是选择性招募的试验可能无法了解的内容。

2 文献综述

随机试验能够为其纳入的人群提供无偏效应（Rubin，1974），但这些人可能无法代表目标总体（Imai 等，2008；Stuart 等，2011）。既有研究展示了如何通过加权（Cole 和 Stuart，2010；Stuart 等，2011；Westreich 等，2017）、分层分类（Tipton，2013；O’Muircheartaigh 和 Hedges，2014）或结局建模（Kern 等，2016；Dahabreh 等，2020），将试验效应迁移至目标人群。Pearl 和 Bareinboim 给出了此类迁移何时可行的图形条件（Pearl 和 Bareinboim，2014）。Degtiar 和 Rose 回顾了这些方法及其所需的假设（Degtiar 和 Rose，2023）。后续研究将这些方法扩展至由复杂抽样调查描述的目标人群（Ackerman 等，2021），以及将整个集群作为单位进行随机分配的试验（Dahabreh 等，2022；Robertson 等，2024）。所有这些方法都需要正性，即目标人群中的每一种人

都必须有机会进入试验（Petersen 等，2012；Westreich 等，2017）。对于目标人群中未测量的效应修饰因子，也已有相关敏感性分析研究（Nguyen 等，2017；Dahabreh 等，2023；Huang，2024；Asiaee 等，2026）。

多层回归与后分层（MRP）可为目标表提供基于模型的估计（Gelman和Little，1997年；Park等，2004年）。其方法是先对单元格均值拟合多层模型，再根据各单元格在人口中的占比对其进行加权（Gelman，2007年；Si等，2020年）。MRP能够从极不具代表性的样本中获得良好估计，例如自愿参加的选举民调样本（Wang等，2015年）。它还可以估计小亚组之间的深层交互（Ghitza和Gelman，2013年）。在健康研究中，MRP已被用于校正高度选择性样本中的未参与问题（Downes等，2018年）。近期研究使用MRP将试验中的治疗效应推广至总体（Kennedy和Gelman，2021年；Gao等，2021a年）。研究还表明，当样本具有选择性时，MRP可能对其关于方差的先验分布较为敏感（Bruch和Felderer，2022年；Gao等，2021b年）。如今，在Stan中对这些模型进行贝叶斯计算已成为常规做法（Carpenter等，2017年）。

当两个数据集无法链接时，联合分布只能部分识别（Manski，2003；Ridder 和 Moffitt，2007）。统计匹配文献研究如何根据边际分布推知联合表（D'Orazio 等，2006）。Fan、Sherman 和 Shum 推导了当结果和协变量来自不同数据集时治疗效应的精确界（Fan 等，2014）。这类界限通常通过最优传输求得（Galichon，2016）；对于两个二元变量，这些界限可简化为 Fr´echet–Hoeffding 界限（Plackett，1965年；D'Orazio 等，2006）。

二元数据带来了自身的问题。对一方进行治疗可能改变另一方的结局，因此因果效应必须在伴侣对层面定义（Hudgens和Halloran，2008年）。行动者–伴侣相互依存模型将个人自身特征的效应与伴侣特征的效应区分开来（Cook 和 Kenny，2005；Kenny 等，2006）。配偶的健康状况往往一致（Meyler 等，2007；Leong 等，2014），因此，通过一方招募参与者的试验也会选择另一方（Trief 等，2016）。据我们所知，既往研究尚未探讨以下情形下的推广：伴侣特征会改变治疗效应、试验未纳入所有类型的伴侣对，且目标调查未链接配偶数据。

3 目标函数

本研究有四个目的。第一个目的是确定，当伴侣试验未纳入所有类型的伴侣对时，能够识别出什么。我们表明，在伴侣可加效应下，当试验支持图的每个连通分量在妻子和丈夫两侧具有相同的目标权重时，目标效应可识别（定理1）。行动者与伴侣交互作用要求该图中存在一个环（命题4）。第二个目的是确定，当调查未将配偶关联起来时，能够了解什么。我们表明，目标效应的精确界由一个离散最优传输问题给出（命题3）。对于单一二元交互作用，这些界限具有闭式形式，而配偶优势比可作为敏感性参数。当这两个问题同时出现时，联合界限位于两个范围的角点上（推论 2）。第三个目的

是衡量三种捷径所造成的偏差：仅对每个人自身特征进行调整、根据边际分布构建伴侣表，以及当试验按修饰变量进行选择时，标准 MRP 所采用的收缩（命题1、2和5）。第四个目的是使用在 Stan 中拟合的贝叶斯二元多层回归与后分层来估计效应及其界限（Carpenter等，2017年）。

这些结果具有实际意义。这两个缺口需要采取不同的补救措施。缺失的夫妇类型需要改进招募方式。链接缺失则需要更完善的总体数据，或配偶一致性的外部估计。第9节展示了如何判断哪个缺口阻碍了某项决策。

我们通过一项 Plasmode模拟（Franklin等，2014年）研究这些方法；该模拟基于2018年健康与退休研究（HRS）中的夫妇（Sonnega 等，2014；RAND老龄化研究中心，2026年）。协变量、层、整群和权重均来自真实数据。试验结局为模拟生成，因此真实总体效应是已知的。

4 符号与假设

夫妇试验样本被随机分配至治疗组 $A = 1$ 或对照组 $A = 0$。每位伴侣 $p \in \{w, h\}$ （妻子、丈夫）均测量结局 Y_p。夫妇目标调查采用包含层、整群和权重的复杂抽样设计进行抽取。两个数据源均记录离散特征 $X = (R,\ X_w,\ X_h)$：夫妇的种族与族裔 R、妻子的个人特征 X_w以及丈夫的个人特征 X_h。在我们的应用中，每项个人特征由年龄组（65岁以下、65–74岁、75岁及以上）、体重指数分组（25以下、25–30、30及以上）和确诊糖尿病构成。因此，共有 $4 \times 18 \times 18 = 1,$ 296个夫妇单元 j。我们将 a 记作妻子的个人特征，将 b 记作丈夫的个人特征，并将 D_w、 D_h 记作两个糖尿病指标。

治疗分配给伴侣对，因此每位伴侣的潜在结局 $Y_p(a)$ 已经包含了另一位伴侣产生的任何溢出效应。我们假设夫妇间不存在干扰（Hudgens和Halloran，2008年）。令 $\tau_p(j) = E\{Y_p(1) - Y_p(0) \mid X = j\}$，并令 $\pi_T(j)$ 表示目标夫妻在单元格 j 中所占的比例。估计目标为

$$\mathrm{PATE}_p = \sum_j \pi_T(j)\,\tau_p(j), \qquad \mathrm{PATE}_c = \tfrac{1}{2}(\mathrm{PATE}_w + \mathrm{PATE}_h). \tag{1}$$

我们作出三个假设。（A1）治疗在试验内随机分配。（A2）二元组可迁移性成立： $E\{Y_p(1) - Y_p(0) \mid X = j,\ \text{试验}\} = E\{Y_p(1) - Y_p(0) \mid X = j,\ \text{目标人群}\}$，且两位伴侣的特征均位于 X 中。（A3）调查设计对于 π_T 是有效的。第5.3节以一个较弱的条件替代通常的正值性假设，第7节研究（A2）不成立时的情况。

5 识别

本节首先给出两种捷径所导致的偏差。随后，我们给出链接缺失时的界限，以及有限重叠时的识别结果。

5.1 两种捷径

命题1（仅行动者调整）。 记 $j = (a, b)$ 为 a 行为者单元（R, X_p）与 b 伴侣特征。若分析仅对行为者单元进行调整，并按目标总体行为者边际分布进行后分层，则其收敛于 $\sum_a \pi_T(a) \sum_b \pi_S(b \mid a)\, \tau_p(a, b)$，其中 π_S 为试验分布。其偏差为

$$\sum_a \pi_T(a) \sum_b \tau_p(a,b)\{\pi_S(b \mid a) - \pi_T(b \mid a)\}. \tag{2}$$

如果伴侣特征不会改变效应，或者试验与目标人群在每个行为者单元内的伴侣构成相同，则偏差为零。由于配偶之间具有一致性，索引患者招募会破坏第二个条件。

命题2（独立配对）。 设 $\pi_{\mathrm{ind}}(r, a, b) = \pi_T(r)\pi_T(a \mid r)\pi_T(b \mid r)$。若 $\tau_p(r, a, b) = f_p(r, a) + g_p(r, b) + \theta_p D_w D_h$，则

$$\sum_j \pi_{\mathrm{ind}}(j)\tau_p(j) - \mathrm{PATE}_p = -\theta_p \sum_r \pi_T(r)\, \mathrm{Cov}_T(D_w, D_h \mid r). \tag{3}$$

独立联合表保留了每位伴侣的所有边际分布，因此只有交互作用会导致偏差。在 HRS 夫妇中，两个糖尿病指标的加权种族组内协方差平均为 0。0127。因此，即使 $\theta_p = -2$ 毫米汞柱也只会使估计值移动约 0。03 毫米汞柱。

5.2 链接缺失：精确界

假设调查提供了种族比例 $\pi_T(r)$ 以及妻子和丈夫的特征边际分布 $p_w(a \mid r)$ 和 $p_h(b \mid r)$，但未提供配对信息。设 Γ_r 为具有这些边际分布的联合分布表 $\gamma_r(a, b)$ 的集合。

命题3（无配对关联时的精确界限）。 在满足（*A1*）–（*A3*）且单元效应已知的条件下，总体平均处理效应$_p$ 的识别集为区间 $[L, U]$，其中

$$L = \sum_r \pi_T(r) \min_{\gamma_r \in \Gamma_r} \sum_{a,b} \gamma_r(a,b)\tau_p(r,a,b), \qquad U = \sum_r \pi_T(r) \max_{\gamma_r \in \Gamma_r} \sum_{a,b} \gamma_r(a,b)\tau_p(r,a,b). \tag{4}$$

每个内层问题都是一个离散最优传输问题，且这些界限是精确的。

推论*1*（单一二元交互作用）。如果 $\tau_p(r, a, b) = f_p(r, a) + g_p(r, b) + \theta_p D_w D_h$，则总体平均处理效应$_p = \Psi_{\mathrm{add}} + \theta_p \sum_r \pi_T(r) P_{11}(r)$，其中 Ψ_{add} 仅取决于边际分布，而 $P_{11}(r) = P(D_w = 1, D_h = 1 \mid r)$在 *Fr´echet–Hoeffding* 区间 $[\max(0, q_w + q_h - 1),\ \min(q_w, q_h)]$ 内变化，其中 q_w、q_h 是种族组 r中的糖尿病比例。如果配偶优势比 ψ 已知，则 P_{11} 是 *Plackett* 方程（*Plackett*，*1965*年）的根，

$$P_{11}(\psi) = \frac{S - \sqrt{S^2 - 4\psi(\psi-1)q_w q_h}}{2(\psi-1)}, \qquad S = 1 + (\psi-1)(q_w + q_h), \tag{5}$$

并且 $\psi = 1$ 满足独立配对。

5.3 有限重叠：试验可识别什么

我们在一个种族组内进行分析。令 S 为试验在两组中均支持的伴侣类型集合（a，b），令 G 为由妻子特征档案和丈夫特征档案构成的二分图，其中每个 $(a,\ b) \in S$对应一条边。在（A1）–（A2）条件下，试验可识别 $(a,\ b) \in S$的 $\tau(a,\ b)$。

定理1 (重叠平衡**)**。 假设 $\tau(a,b) = \alpha_a + \beta_b$对所有 a,b均成立。那么

$$\Psi = \sum_{a,b} \pi_T(a,b)\tau(a,b) = \sum_a p_w(a)\alpha_a + \sum_b p_h(b)\beta_b$$

当且仅当每个连通分量 K （孤立特征档案各自构成一个连通分量）都满足 G $\sum_{a\in K} p_w(a) = \sum_{b\in K} p_h(b)$ 时，才可识别。

只要图是连通图就足够，即使许多伴侣类型未纳入试验。该条件仅使用两个边际分布，因此，即使调查未关联配偶信息，加性效应仍可迁移至该调查。例如，仅考虑糖尿病时，涵盖 00、01 和 11 类型的试验可得 $\tau_{10} = \tau_{00} + \tau_{11} - \tau_{01}$。

命题 4 (交互作用需要环**)**。 设 $\tau(a,b) = \alpha_a + \beta_b + \eta\, h(a,b)$，并假设满足重叠平衡。那么，当且仅当 G 的某个环上 h的交替和非零时，η 才可识别。否则，存在 f, g，使得在 S上有 $h(a,b) = f(a) + g(b)$，且识别集是一条直线

$$\Psi(\eta) = \Psi(0) + \eta\,\Delta, \qquad \Delta = \sum_{a,b} \pi_T(a,b)\{h(a,b) - f(a) - g(b)\}, \tag{6}$$

其中，$\Psi(0) = \sum_{a,\ b} \pi_T$（$a$，$b$）{$\alpha_a + \beta_b + \eta f$（$a$）$+ \eta g$（$b$）} 是对受支持效应的加性外推。$\Psi(0)$ 由 S 上的效应决定，且不依赖于 η。

现在考虑一项索引患者试验，该试验仅招募至少一位伴侣患有糖尿病的夫妇。在 S 上，我们有 $D_w D_h = D_w + D_h - 1$，因此 $\Delta = \pi_T$(00)表示双方均无糖尿病的夫妇在目标人群中的占比。于是，$\tau_{00} = \tau_{10} + \tau_{01} - \tau_{11} + \eta$ 且 PATE(η) = PATE(0) $+ \eta\, \pi_T$(00)，其中

$$\mathrm{PATE}(0) = \sum_j \pi_T(j)\{\tau(j) - \eta(1 - D_w)(1 - D_h)\} = \mathrm{PATE}(\eta) - \eta\, \pi_T(00). \tag{7}$$

这个基线保留了三种已纳入糖尿病类型中的效应，并将无糖尿病夫妇中的效应替换为其加性外推 $\tau_{10} + \tau_{01} - \tau_{11}$。从每个单元格中直接删去 $\eta D_w D_h$这一项，结果并不相同。那样会改变双方均患有糖尿病的已纳入夫妇中的效应；与式（7）相比，差值为 $\eta\{\pi_T(00) - \pi_T(11)\}$，在采用 HRS 边际分布时约为 1.05 毫米汞柱。

备注 1（协变量模型）。 定理1和命题 4允许对每种妻子和丈夫特征组合设定不受限制的效应。相反，我们的 Stan 模型使用较少的协变量项，$\tau(j) = z(j)^\top \theta$。在这种情况下，目标泛函 $c^\top \theta$ 可识别，当且仅当 c 位于试验设计的行空间中，这也是线性可估性的通常条件。因此，当特征组合层面的图无法识别目标时，协变量模型仍可能识别目标，但这一结果依赖于模型的函数形式。第9节报告了拟合模型所能识别的泛函。

推论 2（联合界限）。 对于没有记录链接的指标-患者情形， $\pi_T(00) = \sum_r \pi_T(r)\{1 - q_w - q_h + P_{11}(r)\}$ 且 $P_{11}(r)$ 处于其 *Fr´échet* 范围内时，$\mathrm{PATE}(0)$ 仅取决于边际分布，而 η 位于给定范围 $[\eta_-, \eta_+]$内。函数（η， $\pi_T(00)$） $\mapsto \mathrm{PATE}(0) + \eta\,\pi_T(00)$ 是双线性的，因此精确联合界限在两个范围的四个角点处取得。

简而言之，招募决定能否学习交互作用，而记录链接决定它在目标人群中适用的频率。招募缺失的伴侣类型可以闭合一个环并识别 η。已链接记录或一致性估计则可以确定 P_{11}。

6 估计：贝叶斯二元多层回归与后分层

6.1 结局模型

对于单元格 i 中的伴侣对 j 和组别 a，我们将两个结局建模为二元正态分布，其均值为$\mu_p(j) + a\,\tau_p(j)$，协方差为 Σ。协方差体现伴侣对内的残差相关性。每位伴侣的效应为 $\tau_p(j) = \tau_{0p} + z_p(j)^\top \theta_p$，其中 $z_p(j)$包含 14 个参照编码列：自身年龄（2）、自身体质指数（2）、自身糖尿病、伴侣年龄（2）、伴侣体质指数（2）、伴侣糖尿病、自身与伴侣糖尿病的交互作用，以及种族（3）。预后部分包含自身年龄、自身体质指数、自身糖尿病、伴侣糖尿病和种族等项。默认情况下，我们对妻子和丈夫的系数进行部分合并， $\theta_{pk} = \bar\theta_k + \sigma_{b(k)}\varepsilon_{pk}$。先验分布为 $\bar\theta_k \sim N(0, 5^2)$、每个八个批次对应的 $\sigma_b \sim N^+(0，1。52)$、 $N(0, 10^2)$（用于τ_{0p}和预后系数）、 $N^+(0, 20^2)$（用于残差尺度），以及相关性的 LKJ(2) 先验。由于协变量是离散的，似然仅取决于单元格×组别均值和合并散布矩阵。因此，拟合时间不会随试验规模增长。

6.2 构成模型

设 c_j 为单元 j 中的加权调查计数，并将其重新缩放，使其总和等于有效样本量 $n_{\mathrm{eff}} = n_T/\mathrm{deff}$。这里，deff 是所有因子边际分布的设计效应中位数（Savitsky 和 Toth，2016；Williams 和 Savitsky，2021）。我们使用 $\pi_T \mid c \sim$ 狄利克雷分布$(n_{\mathrm{eff}}[(1-\omega)\hat\pi_{\mathrm{HT}} + \omega m])$，其中 $\hat\pi_{\mathrm{HT}}$ 表示加权份额， m 是包含所有主效应和二阶交互作用的对数线性基准测度，而 $\omega = a_0/(a_0 + n_{\mathrm{eff}})$ 中的 a_0 由经验贝叶斯估计。将总浓度设为 n_{eff} ，可避免后验分布声称拥有超出调查数据实际所含信息的更多信息。该后验分布是共轭的，我们在 Stan 的生成量中对其进行抽样。

备注 2（调查自身的不确定性很重要时）。 将单元效应视为固定值，

$$\mathrm{Var}\Big(\sum_j \hat\pi_j \tau(j)\Big) \approx \mathrm{Var}_{\pi_T}\{\tau(X)\}/n_{\mathrm{eff}}.$$

除非各单元格间的效应差异很大，且调查的样本量相对于试验较小，否则调查在后验方差中所占的比例很小。

6.3 为什么标准MRP收缩在选择下存在偏差

命题5 (收缩偏差)。 设 $\tilde{\tau}_j = \lambda_j \hat{\tau}_j + (1-\lambda_j)\hat{\mu}$，其中 $\hat{\tau}_j$ 是τ_j的无偏估计， $\lambda_j \in [0,1]$是固定的（对于没有数据的单元格，$\lambda_j = 0$），且 $\hat{\mu} = \sum_j w_j \hat{\tau}_j / \sum_j w_j$ 采用精度权重 w_j。那么

$$E\Big[\sum_j \pi_T(j)\tilde{\tau}_j\Big] - \text{PATE} = \sum_j \pi_T(j)(1-\lambda_j)(\bar{\mu} - \tau_j), \qquad \bar{\mu} = \frac{\sum_j w_j \tau_j}{\sum_j w_j}. \tag{8}$$

权重w_j 会随着单元格 j中的试验配偶数量增加而增大。因此， $\bar{\mu}$接近试验平均效应，而在试验抽样不足的单元格中， λ_j 最小。如果每个 λ_j 都接近零，MRP就会返回试验平均值。对每个修饰变量采用零均值随机效应的标准MRP，在每个批次上也会表现出相同的模式。补救方法是避免使用以试验为锚点的中心值，因此我们的模型为每个修饰变量指定弱信息先验。先前研究表明，在选择性样本中，MRP对其方差先验十分敏感（Bruch and Felderer, 2022；Gao et al., 2021b）。该命题明确给出了治疗效应迁移中这种偏差的方向和大小。

6.4 界限的后验分布

对于每次单元效应的后验抽样，我们都精确计算第5节中的界限。命题3使用线性规划计算，推论1和推论2则使用闭式形式计算。对于索引患者设计，我们在每次抽样中根据方程（7）将PATE(0)计算为PATE$-\eta\,\pi_T$(00)。沿着此类试验无法学习的那个方向，该量不会变化（第9节）。我们报告两个端点的后验均值、下端点的第2.5百分位数，以及上端点的第97.5百分位数。

7　对未测量的伴侣对层面修饰因素的敏感性　　r

设二元伴侣特征 U 使两位伴侣的效应均改变 δ，并设 $q_S(j)$ 和 $q_T(j)$ 分别表示该特征在试验中的单元格 j 和目标人群中的患病率。将试验的单元效应迁移时忽略 U，会给二元多层回归与后分层带来偏差_　　_$\delta\Delta$，其中 $\Delta = \sum_j \pi_T(j)\{q_S j - q_T j\}$。对于妻子、丈夫和伴侣平均值，该偏差相同。这种乘积形式也出现在早期关于推广的研究中（Nguyen 等，2017；Dahabreh 等，2023；Huang，2024；Asiaee 等，2026）。我们报告使估计值达到某一阈值的临界乘积$\delta\Delta$ 。我们使用一个已观测特征 V对其进行基准比较，采用该特征对伴侣平均效应的估计修饰作用及其原始试验与目标人群间的不平衡。根据定义，已测量特征在单元格内不存在不平衡。因此，其原始不平衡是一个参考尺度，而不是¯对 Δ 的估计；将未测量特征视为在单元格内具有同等不平衡，是一种有意设置的强基准。

8 模拟研究

8.1 方法

在本研究中，我们使用了RAND HRS纵向文件的第14波（2018年）数据（RAND老龄化研究中心，2026年）。目标人群为社区居住、已婚的异性

伴侣对，且双方配偶均接受了访谈并具有正的受访者权重。在剔除104对缺失协变量的夫妇后，剩下3,826对夫妇，分布于80个层中，每层包含2个抽样误差计算单元。2018年家庭权重之和对应2400万对夫妇。我们仅使用该权重构建伪总体，因为它未针对我们的完整案例限制进行调整。不等权重产生的Kish设计效应为1.73。妻子中有21.1%报告患有糖尿病，丈夫中有26.7%报告患有糖尿病；7.3%的夫妇中双方伴侣均患有糖尿病。在59.6%的夫妇中，双方伴侣均未患有糖尿病。

我们按权重比例复制每对伴侣，构建了一个约有 150,000 对伴侣的伪总体，并将每个真实整群拆分为四个。结局是收缩压的六个月变化。妻子的真实效应为 $-4 - 2D_w - 1.5\,\mathrm{Ob}_w + 1.5\,\mathrm{Old}_w - 2D_h - 2D_wD_h - 1.0$ 黑人–0.5 西班牙裔 $+ 1.0\mathrm{Old}_w\mathrm{Ob}_w + \nu$。丈夫的效应以 -3 为基数，采用相同方式定义。最后一项未纳入拟合模型。总体效应分别为 -5.49 （妻子）、 -4.35（丈夫）和 -4.92 毫米汞柱（夫妻平均值）。每个重复样本均抽取一次调查和一次试验。调查在每层的八个整群中抽取两个，并将黑人和西班牙裔伴侣对的抽样比例提高至其他伴侣对的两倍。试验对糖尿病伴侣对超额招募，试验平均值约为 -7.7毫米汞柱。我们运行了三个情景。S1 包含 1,000 对试验配偶和约 3,800 对调查配偶（150 个重复样本）。S2 包含 3,000 对试验配偶和 800 对调查配偶（150 个重复样本）。S3 与 S1 相同，但加入了一项未测量的伴侣对特征，$\delta = -3$ （100 个重复样本）。

8.2 数据分析

我们比较了四种估计量：试验均值差、调查加权逆优势加权（IPW；Ackerman 等，2021）并结合整群自助法、仅行动者回归并进行后分层，以及二元多层回归与后分层。对于每种估计量，我们计算了偏差、均方根误差（RMSE）、95%区间覆盖率和区间宽度。

8.3 结果

二元多层回归与后分层几乎消除了试验与目标人群之间的差距（图1）。在 S1 中，其偏差为 -0.18 毫米汞柱，覆盖率为 97.3%；而试验均值差的偏差为 -2.85 。其均方根误差比逆概率加权低 17%（0.80 对 0.97），区间也窄 15%（3.28 对 3.85 毫米汞柱）。仅行动者调整仍留下 -0.79 毫米汞柱的差距，约为原差距的四分之一，覆盖率为 82%。在 S2 中，其覆盖率降至 63%。在 S2 中，调查样本量较小，导致逆概率加权的自助法区间非常宽（5.44 对 1.94 毫米汞柱）。独立配对和固定的调查份额使二元多层回归与后分层的估计值变动不到 0.03 毫米汞柱。调查所占后验方差的比例在 S1 中平均为 0.65%，在 S2 中平均为 3.6%。

在 S3 中，偏差公式中的量是目标加权的单元内不平衡， $\Delta = 0.243$。因此，该公式预测，遗漏 U 会造成 $-3 \times 0.243 = -0.73$ 毫米汞柱的偏差。原始差距相近（试验配偶中有 64.6%，而总体中有 40.1%，相差 0.245），只是因为 U 的生成与测量特征相互独立。二元多层回归与后分层的总偏差为-0.64 毫米汞柱（蒙特卡洛标准误为 0.07），其中也包含模型本身的误差。在配对检验中，我们重新拟合了 100 个 S3 重复样本，并从结果中移除了 U 的效应。遗漏 U

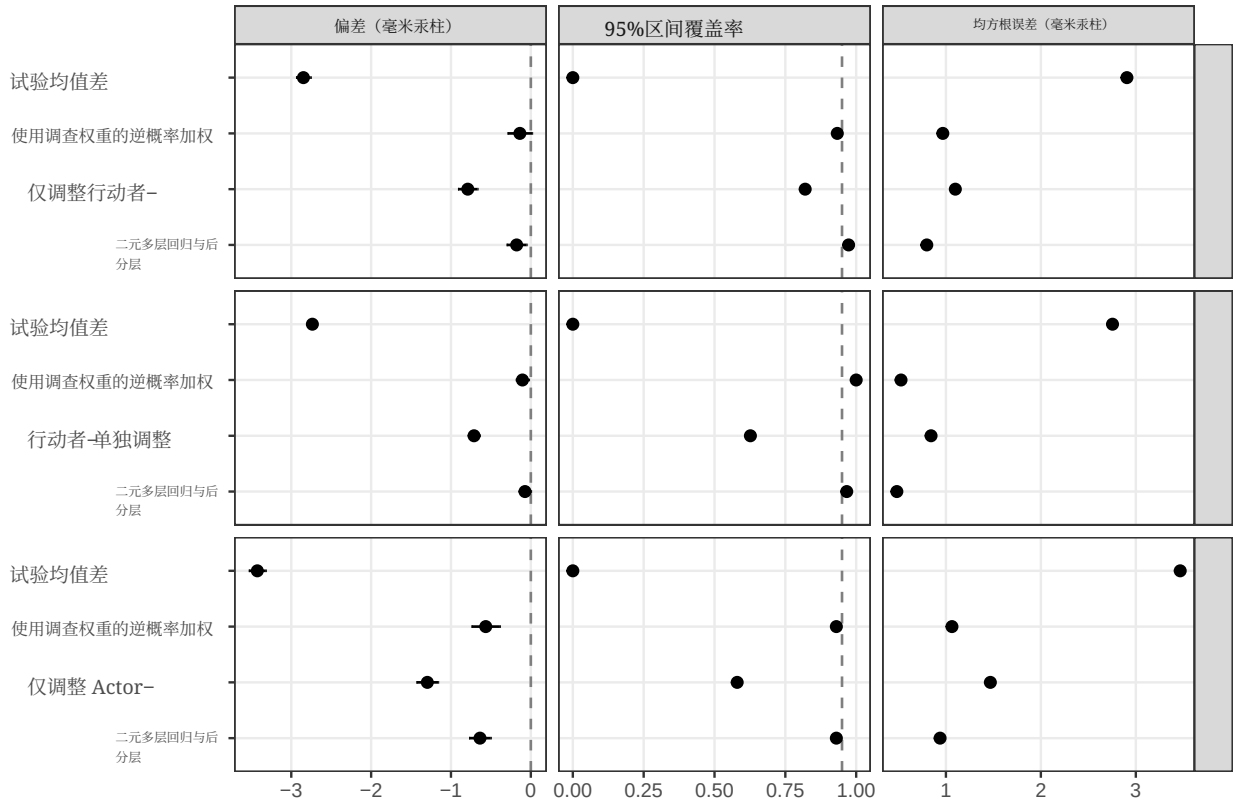


图1：伴侣平均效应的模拟结果。偏差须状线表示95%蒙特卡洛区间；虚线标示零偏差和95%的覆盖率。

造成了 -0.71毫米汞柱的偏差（蒙特卡洛标准误为 0.01）。400 次拟合中有 10 次出现发散转换，ˆ 最大值为 R，即 1.016。

收缩。 在相同的重复样本中，对每个修饰变量水平均设置零均值随机效应的标准 MRP，在 S1 中留下了 -0.72 毫米汞柱（覆盖率为 88.0%），在 S2 的前 100 个重复样本中留下了 -0.39 （覆盖率为 89.0%）。我们的模型留下了 -0.18 和 -0.08，且区间宽度相同。采用经验贝叶斯收缩的饱和单元格层面多层回归与后分层，其估计收缩至试验平均值（-2.82 和 -2.16 毫米汞柱），而命题5预测为 -2.80 和 -2.10。为单独考察合并的影响，我们在 S1 的前 60 个重复样本上重新拟合模型，分别对伴侣进行独立合并、部分合并和完全合并。我们在主要真实情形以及非对称真实情形下进行了分析；在后者中，妻子的反应取决于自身的糖尿病状况，而丈夫的反应取决于妻子的糖尿病状况。伴侣平均效应的偏差介于 -0.23 和 -0.37 毫米汞柱之间，彼此差异均在蒙特卡洛误差范围内。因此，相较于标准 MRP，优势来自不将修饰变量向零收缩，而非跨伴侣合并。

招募。 为了了解需要多少夫妇，我们在一项包含1,000对索引患者伴侣的试验中，加入了0、100、300、600或1,000对随机无糖尿病夫妇。我们对每种组合拟合二元多层回归与后分层（40个重复样本；图2）。我们报告识别集的后验平均宽度。同时，我们还报告从其下端点的第2.5百分位数到上端点的第97.5百分位数之间的宽度。

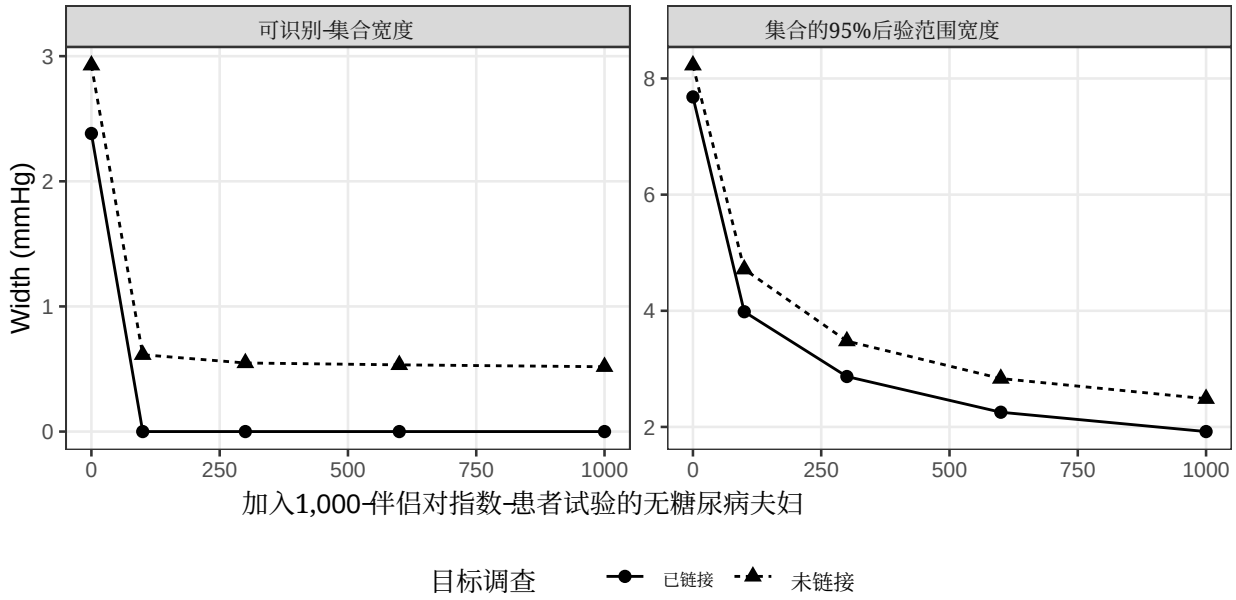


图2：招募模拟（40个重复样本）。左：识别集的后验平均宽度。右：从下端点的第2.5百分位数到上端点的第97.5百分位数的宽度。已链接：调查链接了配偶；未链接：仅知道伴侣边际分布。

在未加入夫妇时，已链接情况下识别集的宽度为2.38毫米汞柱，未链接情况下为2.93毫米汞柱；后验范围分别为7.7和8.2毫米汞柱。加入100对夫妇后，已链接情况下效应变为一个点估计，未链接情况下的识别集缩小至0.61毫米汞柱。后验范围降至4.0和4.7毫米汞柱。增加夫妇数量提高了精确度，但并未带来识别。加入1,000对夫妇后，后验范围分别为1.9和2.5毫米汞柱，而未链接情况下的识别集仍维持在约0.5毫米汞柱。这是因为其宽度取决于未知配对，而不是试验规模。在每种规模下，后验范围在92.5%至100%的重复样本中涵盖真实效应，其 midpoint平均距离真实值不超过0.25毫米汞柱。200次拟合中有10次出现了发散转换。

9 实例分析：基于夫妇的血压项目

在本实例中，我们将真实的2018年HRS夫妇文件作为目标人群，并采用S1招募规则抽取的一个包含1,000对夫妇的模拟试验。在该试验中，样本与目标人群相距甚远。试验中，61.6%的妻子和59.8%的丈夫患有糖尿病，而目标人群中的相应比例为21.1%和26.7%。此外，试验夫妇中有42.3%的夫妇双方均患有糖尿病，而目标人群中这一比例为7.3%。目标权重中只有30.0%落在每个试验组至少包含3对试验夫妇的单元格中，另有17.9%落在不包含任何试验夫妇的单元格中。因此，大部分估计结果都由模型结构所决定。

二元多层回归与后分层将 -8.34 毫米汞柱的试验效应调整为 -6.18 （-7.81 至 -4.58；图3）。$\hat{}$ 真实效应为 -4.92。拟合过程没有发散转换， $R \leq 1.003$，且主体有效样本量超过 1,280。对于双方均未患糖尿病的伴侣对（占目标人群的 59.6%），估计值为 -6.1 毫米汞柱，而真实值为 -3.7。这是基于稀疏支持进行的外推。在

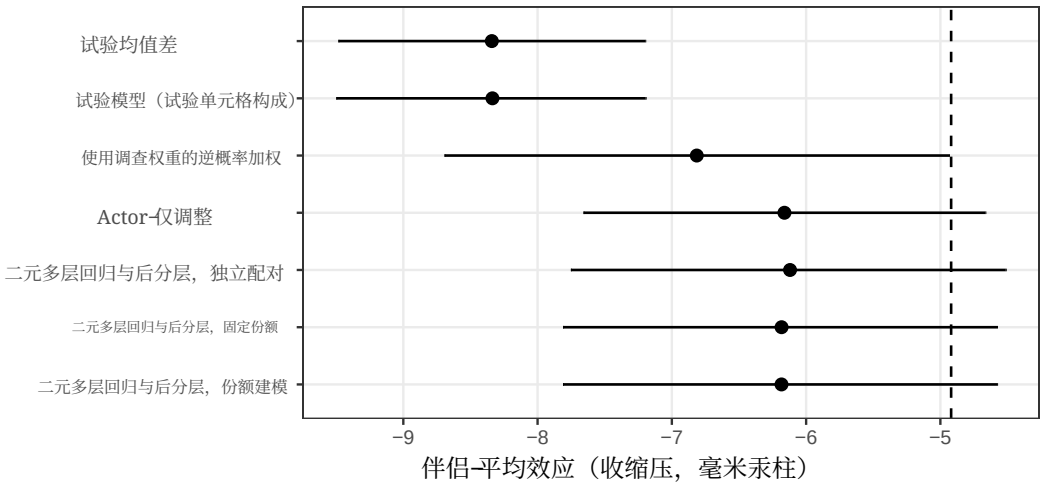


图3：实例分析：按估计量划分的伴侣平均效应及95%区间。虚线表示已知的总体效应（-4.92 毫米汞柱）。

S1 个重复样本中，该亚组的区间有 98% 的时间涵盖真实值。未测量的伴侣对特征需要达到妻子糖尿病效应（$\delta_V = -1.40$ 毫米汞柱，$\Delta_V = 0.405$）的乘积约 5.6 倍，才能将效应缩小至 -3 毫米汞柱。

无记录链接。 随后，我们隐藏了HRS链接，仅保留边际分布。伴侣平均效应位于 -6.79 至 -5.78 毫米汞柱之间，区间宽度为1.01（命题3）。在线性规划的每次抽样中，其结果与推论1的差异均不超过 3×10^{-12}。区间宽度等于 $|\theta| \sum_r \pi_T$（r）min（q_w，q_h）$=0.209|\theta|$，因此由糖尿病-糖尿病系数决定。如果将独立配对估计固定为 -5.5 毫米汞柱，那么只有当该交互作用高于 $+3.36$ 或低于 -8.33 毫米汞柱时，链接缺失才会改变5毫米汞柱的决策（图4）。优势比介于1和3之间时，第一个阈值移至 $+13.15$。

有限重叠与链接缺失并存。 我们又抽取了第二项试验，共纳入 1,000 对夫妇，且仅从至少有一位伴侣患糖尿病的夫妇中抽取。此时，识别依赖于拟合的协变量模型（备注1）。在这项试验中，每位伴侣的回归设计（截距、9 个预后项、治疗和 14 个治疗与修饰变量交互项）的秩为 24（满秩为 25 ）。缺失的方向使交互作用系数和治疗截距增加 c，并使两个糖尿病系数都降低 c。因此，（7）中的 PATE(0) 可估计，但总体平均处理效应不可估计。加入 300 对无糖尿病夫妇后，秩达到 25，两个量都可估计。不使用协变量模型的个体特征层面支持图在每个种族组中有 2 至 31 个连通分量；无论是否加入这些夫妇，图都不平衡。因此，仅凭定理1无法识别任一量。由于拟合效应除 η 外均为加性，使用已链接表和独立配对表得到的 PATE(0) 相同。

索引试验无法提供关于 η 的信息，因此我们将其置于 $[-4, 0]$ 毫米汞柱。份额 π_T(00) 在有记录链接时为 0.596，无记录链接时为 0.523 至 0.732，而当配偶优势比介于 1 和 3 之间时为 0.583 至 0.621。增加 300 对随机分配的无糖尿病夫妇后，拟合模型恢复满秩，因此可识别 η（后验均值 -0.95，标准差 1.38；真值 -2）。图5展示了每种数据

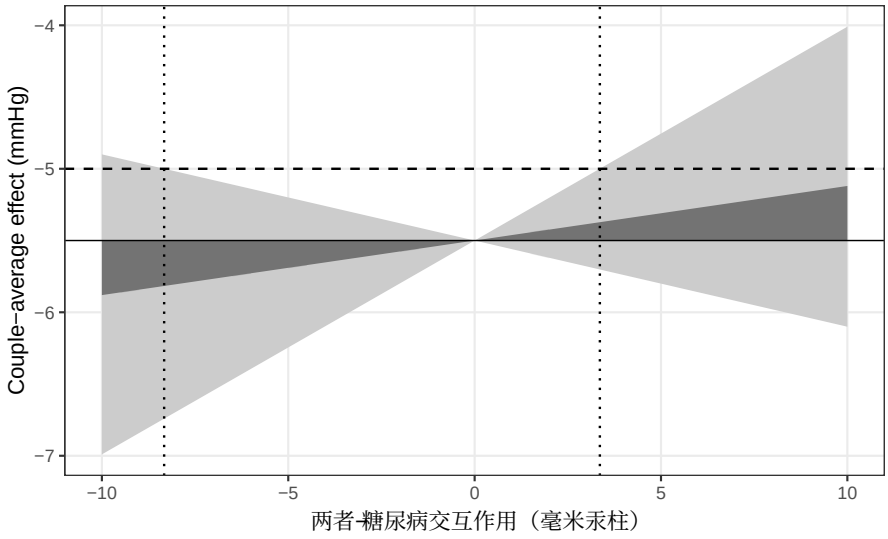


图4：记录链接缺失何时会改变决策。浅色带：无配对关联时的精确界限；深色带：配偶优势比介于1和3之间时的界限。实线：独立配对估计（-5.5）；虚线：5毫米汞柱阈值。

来源所带来的信息。加入索引试验后，无记录链接时识别集宽 2.93 毫米汞柱，采用完整链接时宽 2.38 毫米汞柱；无论哪种情况，该识别集都跨越 5 毫米汞柱的阈值。加入这些夫妇后，无记录链接时识别集宽 0.28 毫米汞柱，在一致性范围下宽 0.05 毫米汞柱；而在拟合模型下，采用记录链接时则为一个点。

10 讨论

我们已经看到，夫妻试验面临的两种障碍表现各不相同。记录链接缺失的危害往往没有看起来那么大。加性效应只需要边际分布，而配对关系只通过行动者与伴侣交互作用产生影响，配偶一致性会限制这种影响。缺失的伴侣类型往往危害更大。试验无法估计的交互作用会乘以目标人群中缺失类型所占的比例；在我们的示例中，这一比例为60%。在该示例中，增加300对夫妇使识别集缩小了十倍，而仅靠记录链接则使其缩小了五分之一。试验若以夫妇为研究对象，其结果就应推广到夫妇；分析人员首先应判断存在的是上述哪一种缺口。

我们还看到了关于估计的两点启示。第一，当试验根据修饰变量进行选择时，标准的 MRP收缩会将抽样不足的单元格向试验平均值拉近，而命题5给出了这种偏差的大小。弱正则化修饰变量消除了其中的大部分偏差，而在伴侣之间合并修饰变量的影响很小。第二，参数模型承担了大部分识别工作。在示例试验中，个体特征层面的支持图在每个种族组中包含5到34个连通分量，且并不平衡，因此非参数伴侣加性模型无法识别目标效应。分析人员应在报告估计结果的同时，报告试验支持以及拟合设计矩阵的秩。

我们承认，这项研究存在若干局限性。结果是模拟得到的，HRS

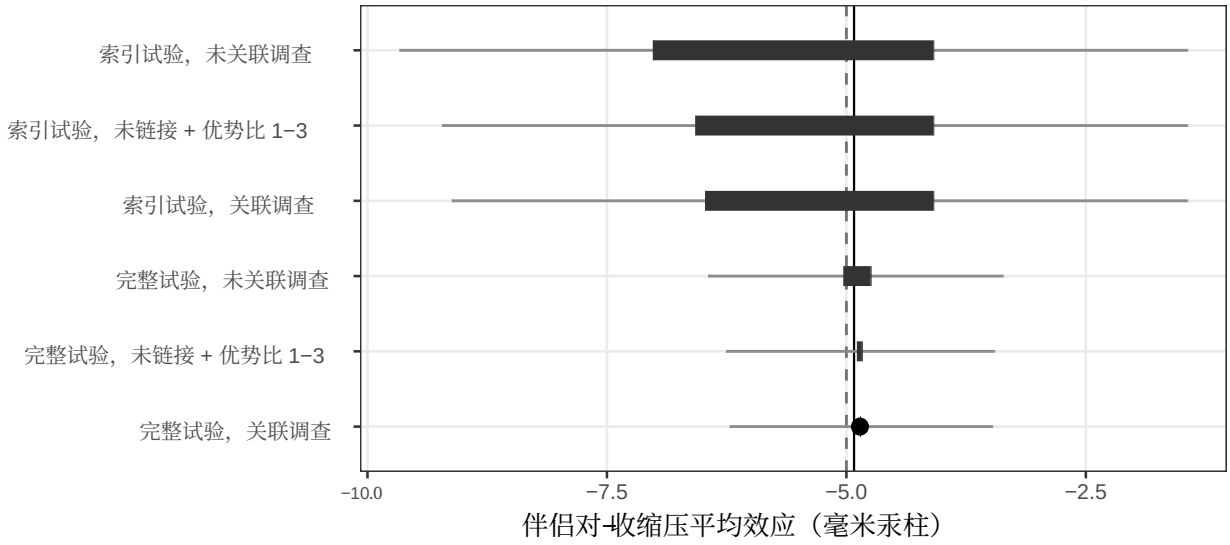


图5：各数据源的作用。粗条：识别集（端点的后验均值）；细线：从下端点的第2.5百分位数到上端点的第97.5百分位数。实线：真实值；虚线：5毫米汞柱阈值。

权重未针对我们的完整案例限制进行调整，且 η 的交互作用范围仅具有示意性。这些界限假设已知每个人的伴侣种族组。我们研究了异性婚姻伴侣，而同性伴侣需要使用可交换模型。命题5将收缩权重视为固定值。

我们希望本研究的发现能为设计和分析基于伴侣的试验的研究人员提供参考。下一步是将该方法应用于一项真实的夫妻随机试验。该方法仅需试验按单元格和试验组汇总的数据，以及一项包含双方伴侣特征的调查；HRS是针对老年伴侣开展试验的自然目标人群。未来研究还应探讨如何以最低成本设计招募方案，从而弥补缺失周期。

声明

数据与代码。注册后，可从健康与退休研究网站获取RAND HRS纵向文件。用于复现所有表格和图表的R和Stan脚本作为附属文件随本预印本一并提供，并将发布在 https://github.com/subirhait。健康与退休研究微观数据不会再分发。**资金与利益冲突**。无专项资助。作者声明不存在利益冲突。

参考文献

Benjamin Ackerman、Catherine R. Lesko、Juned Siddique、Ryoko Susukida 和 Elizabeth A. Stuart。利用复杂调查总体数据将随机试验结果推广至目标人群。《医学统计》，40(5):1101–1120，2021年。

Emily Arden-Close 和 Nuala McGrath。面向夫妇的健康行为改变干预：系统综述。《英国健康心理学杂志》，22(2):215–237，2017年。Amir Asiaee、Samhita Pal 和 Jared D. Huling。推广随机试验的遗漏变量敏感性分析。arXiv:2603.27788，2026年。Christian Bruch 和 Barbara Felderer。高度选择性数据的多层回归与后分层方法中方差参数的先验选择：蒙特卡罗模拟研究。《奥地利统计学杂志》，51(4):76–95，2022年。Bob Carpenter、Andrew Gelman、Matthew D. Hoffman、Daniel Lee、Ben Goodrich、Michael Be-tancourt、Marcus Brubaker、Jiqiang Guo、Peter Li 和 Allen Riddell。Stan：一种概率编程语言。《统计软件杂志》，76(1):1–32，2017年。Stephen R. Cole 和 Elizabeth A. Stuart。将随机临床试验证据推广至目标人群：ACTG 320 试验。《*American Journal of Epidemiology*》，172(1):107–115，2010年。William L. Cook 和 David A. Kenny。行动者–伴侣相互依存模型：发展研究中双向效应的模型。《*International Journal of Behavioral Development*》，29(2):101–109，2005年。Issa J. Dahabreh、Sarah E. Robertson、Jon A. Steingrimsson、Elizabeth A. Stuart 和 Miguel A. Hern´an。将随机试验推论扩展至新的目标人群。《医学统计》，39(14):1999–2014，2020年。Issa J. Dahabreh、Sarah E. Robertson、Jon A. Steingrimsson、Stefan Gravenstein 和 Nina R. Joyce。将整群随机试验的推论扩展至目标人群。arXiv:2203.14761，2022年。Issa J. Dahabreh、James M. Robins、Sebastien J.-P. A. Haneuse、Iman Saeed、Sarah E. Robertson、Elizabeth A. Stuart 和 Miguel A. Hern´an。运用偏倚函数，对将随机试验推论扩展至目标人群的研究进行敏感性分析。《医学统计》，2023年。PMID 36847107。Irina Degtiar 和 Sherri Rose。可推广性与可迁移性综述。《统计学及其应用年度评论》，10:501–524，2023年。Marcello D’Orazio、Marco Di Zio 和 Mauro Scanu。《统计匹配：理论与实践》。Wiley，奇切斯特，2006年。Marnie Downes、Lyle C. Gurrin、Dallas R. English、Jane Pirkis、Dianne Currier、Matthew J. Spit-tal 和 John B. Carlin。多层回归与后分层：一种利用高度选择性调查样本估算总体量的建模方法。《*American Journal of Epidemiology*》，187(8):1780–1790，2018年。

Tracy A. Falba 和 Jody L. Sindelar。健康行为改变中的配偶一致性。*HealthServices Research*, 43(1):96–116，2008年。Yanqin Fan、Robert Sherman 和 Matthew Shum。数据组合条件下治疗效应的识别。*Econometrica*，82(2):811–822，2014年。Jessica M. Franklin、Sebastian Schneeweiss、Jennifer M. Polinski 和 Jeremy A. Rassen。用于评估复杂医疗保健数据库中药物流行病学方法的 Plas- mode 模拟。*Computational Statistics & Data Analysis*，72:219–226，2014年。Alfred Galichon。经济学中的最优传输方法。Princeton University Press，普林斯顿（新泽西州），2016年。Yuxiang Gao、Lauren Kennedy 和 Daniel Simpson。使用多层回归与后分层进行治疗效应估计。arXiv:2102.10003，2021a。Yuxiang Gao、Lauren Kennedy、Daniel Simpson 和 Andrew Gelman。通过结构化先验改进多层回归与后分层。*Bayesian Analysis*，16(3):719–744，2021b。Andrew Gelman。调查加权与回归建模面临的难题。*Statistical Science*，22(2):153–164，2007年。Andrew Gelman 和 Thomas C. Little。使用层次逻辑回归对众多类别进行后分层。*Survey Methodology*，23(2):127–135，1997年。Yair Ghitza 和 Andrew Gelman。MRP 中的深层交互：小型选举子群体的选举投票率与投票模式。*American Journal of Political Science*，57(3):762–776，2013年。Melody Y. Huang。实验结果推广的敏感性分析。*Journal ofthe Royal Statistical Society Series A*，187(4):900–918，2024年。Michael G. Hudgens 和 M. Elizabeth Halloran。迈向存在干扰情形下的因果推断。*Journalof the American Statistical Association*，103(482):832–842，2008年。Kosuke Imai、Gary King 和 Elizabeth A. Stuart。实验研究者与观察性研究者对因果推断的误解。*Journal of the Royal Statistical Society: Series A(Statistics in Society)*，171(2):481–502，2008年。Sarah E. Jackson、Andrew Steptoe 和 Jane Wardle。伴侣行为对健康行为改变的影响：英国老龄化纵向研究。*JAMA Internal Medicine*，175(3):385–392，2015年。Lauren Kennedy 和 Andrew Gelman。了解你的总体，也了解你的模型：使用基于模型的回归与后分层，将研究发现推广至观测样本之外。心理学方法，26(5):547–558，2021年。

David A. Kenny、Deborah A. Kashy 和 William L. Cook。二元数据分析。Guilford Press，纽约，2006年。Holger L. Kern、Elizabeth A. Stuart、Jennifer Hill 和 Donald P. Green。评估将实验影响估计推广至目标人群的方法。*Journal of Research onEducational Effectiveness*，9(1):103–127，2016年。Aaron Leong、Elham Rahme 和 Kaberi Dasgupta。配偶糖尿病作为糖尿病风险因素：系统综述与荟萃分析。*BMC Medicine*，12:12，2014年。Charles F. Manski。概率分布的部分识别。Springer，纽约，2003年。Lynn M. Martire、Richard Schulz、Vicki S. Helgeson、Brent J. Small 和 Ester M. Saghafi。慢性疾病面向伴侣的干预：综述与荟萃分析。*Annals of BehavioralMedicine*，40(3):325–342，2010年。Deanna Meyler、Jim P. Stimpson 和 M. Kristen Peek。伴侣对内部的健康一致性：系统综述。*Social Science & Medicine*，64(11):2297–2310，2007年。Trang Quynh Nguyen、Cyrus Ebnesajjad、Stephen R. Cole 和 Elizabeth A. Stuart。RCT 到目标人群的治疗效应推广中未观测调节变量的敏感性分析。*Annals of Applied Statistics*，11(1):225–247，2017年。Colm O'Muircheartaigh 和 Larry V. Hedges。将非代表性实验的结果推广：一种分层倾向得分方法。*Journal of the Royal Statistical Society: Series C (AppliedStatistics)*，63(2):195–210，2014年。David K. Park、Andrew Gelman 和 Joseph Bafumi。结合后分层的贝叶斯多层估计：基于全国民意调查的州级估计。*Political Analysis*，12(4):375–385，2004年。Judea Pearl 和 Elias Bareinboim。外部效度：从 do 演算到跨人群可迁移性。*Statistical Science*，29(4):579–595，2014年。Maya L. Petersen、Kristin E. Porter、Susan Gruber、Yue Wang 和 Mark J. van der Laan。诊断正值性假设的违反并作出应对。*Statistical Methods in MedicalResearch*，21(1):31–54，2012年。R. L. Plackett。一类二元分布。*Journal of the American Statistical Association*，60(310):516–522，1965年。RAND老龄化研究中心。*RAND HRS*纵向文件*2022*（*V1*）。圣莫尼卡（加利福尼亚州），2026年。由国家老龄研究所和社会保障局资助制作。Geert Ridder 和 Robert Moffitt。数据组合计量经济学。载于 James J. Heckman 和 Edward E. Leamer 编，《计量经济学手册》，第6B卷，第5469–5547页。Elsevier，阿姆斯特丹，2007年。

Sarah E. Robertson、Jon A. Steingrimsson 和 Issa J. Dahabreh。旨在支持可推广推断的整群随机试验。*Evaluation Review*，48(6):1088–1114，2024。Donald B. Rubin。估计随机和非随机研究中治疗的因果效应。*Journal of Educational Psychology*，66(5):688–701，1974。Terrance D. Savitsky 和 Daniell Toth。信息性抽样下的贝叶斯估计。*ElectronicJournal of Statistics*，10(1):1677–1708，2016。Yajuan Si、Rob Trangucci、Jonah Sol Gabry 和 Andrew Gelman。贝叶斯分层加权调整与调查推断。*Survey Methodology*，46(2)，2020。加拿大统计局目录编号 12-001-X。Amanda Sonnega、Jessica D. Faul、Mary Beth Ofstedal、Kenneth M. Langa、John W. R. Phillips 和 David R. Weir。队列概况：健康与退休研究（HRS）。*InternationalJournal of Epidemiology*，43(2):576–585，2014。Elizabeth A. Stuart、Stephen R. Cole、Catherine P. Bradshaw 和 Philip J. Leaf。使用倾向评分评估随机试验结果的可推广性。*Journal of theRoyal Statistical Society Series A*，174(2):369–386，2011。Elizabeth Tipton。利用倾向评分子分类改进实验结果的推广：假设、性质与适用情境。*Journal of Educational and Behavioral Statistics*，38(3):239–266，2013。Paula M. Trief、Lawrence Fisher、Jonathan Sandberg、Donald A. Cibula、Jacqueline Dimmock、Danielle M. Hessler、Patricia Forken 和 Ruth S. Weinstock。对血糖控制不佳的2型糖尿病患者实施电话式伴侣行为改变干预的健康与心理社会结局：一项随机临床试验。*Diabetes Care*，39(12):2165–2173，2016。Paula M. Trief、Lawrence Fisher、Jonathan Sandberg、Danielle M. Hessler、Donald A. Cibula 和 Ruth S. Weinstock。一举两得？伴侣干预对2型糖尿病患者伴侣的影响：一项随机对照试验。*Diabetic Medicine*，36(4):473–481，2019。Wei Wang、David Rothschild、Sharad Goel 和 Andrew Gelman。利用非代表性民调预测选举结果。*International Journal of Forecasting*，31(3):980–991，2015。Daniel Westreich、Jessie K. Edwards、Catherine R. Lesko、Elizabeth A. Stuart 和 Stephen R. Cole。使用抽样逆优势权重迁移试验结果。*American Journalof Epidemiology*，186(8):1010–1014，2017。Matthew R. Williams 和 Terrance D. Savitsky。复杂抽样下伪贝叶斯推断的不确定性估计。*International Statistical Review*，89(1):72–107，2021。

A 证明

命题1。 仅含行动者模型在每个行为者单元中估计 $E_S\{\tau_p \mid a\} = \sum_b \pi_S(b \mid a)\tau_p(a,\ b)$；对 $\pi_T(a)$ 进行后分层并减去总体平均处理效应$_p\ = \sum_a \pi_T(a)\ \sum_b \pi_T(b \mid a)\tau_p(a,\ b)$，即可得到结果。

□

命题2. π_{ind}和 π_T具有相同的边际分布 $\pi_T(r,\ a)$和 $\pi_T(r,\ b)$，因此加性项的平均值相等。在 π_{ind} 下，交互项的平均值为$\theta_p \sum_r \pi_T(r)q_w q_h$；而在 π_T下，其平均值为 $\theta_p \sum_r \pi_T(r)P_{11}(r)$；两者之差为 $-\theta_p \sum_r \pi_T(r)\{P_{11}(r)\ - q_w q_h\}$。 □

命题3. 每个 $\gamma_r \in \Gamma_r$ 都是与观测到的边际分布一致的联合分布，因此，与数据相容的每个目标人群的总体平均处理效应 $\text{PATE}_p = \sum_r \pi_T(r)\ \sum \gamma_r \tau_p$对某个 $\gamma_r \in \Gamma_r$成立，而且每个这样的 γ 都相容。 Γ_r 是一个非空的紧凸多面体，且目标函数是线性的，因此其像是最小值与最大值之间的闭区间，且这两个值均在顶点处取得。各个种族组对应的问题彼此独立。

□

推论1。 加性项仅取决于边际分布。对于交互作用，Fr´echet区间内的任意 P_{11} 都可由耦合 $\gamma(a,\ b)\ = P(d_a,\ d_b)\ p_w(a \mid d_a)\ p_h(b \mid d_b)$取得，其中 $P(\cdot,\ \cdot)$是具有给定边际分布的 2×2 表，且 P_{11}，$d_a,\ d_b$是档案 $a,\ b$的糖尿病取值；其行和与列和分别为 $p_w(a)$和 $p_h(b)$。不存在任何耦合能给出Fr´echet区间之外的 P_{11} 。对于固定的优势比，具有给定边际分布的 2×2 表是唯一的，且 P_{11} 满足Plackett的二次方程。

□

定理1。 在每个连通分量 K上，边上的取值 $\tau(a,\ b)$可确定（$\alpha,\ \beta$），但存在 $\alpha_a + c_K$、$\beta_b - c_K$的不确定性（固定一个节点，沿生成树传播；生成树之外的边不会增加新的自由度）。因此， Ψ可确定至多相差 $\sum_K c_K(w_K - h_K)$，其中 $w_K = \sum_{a \in K} p_w(a)$且 $h_K = \sum_{b \in K} p_h(b)$。当且仅当对每个 K都有 $w_K = h_K$时，对于所有（c_K）该值为零。如果某个 $w_K \neq h_K$成立，则在 S 上相同的两个参数值会给出不同的 Ψ，因此 Ψ 不可识别。 □

命题 4。限制于 S 的模型关于（α、β、η）是线性的。当且仅当 $h|_S$ 不属于 $\mathcal{A}_S = \{f(a) + g(b)$ 在 $S\}$ 上的空间时，参数 η 才是可识别的。边上的函数 ϕ 属于 $\mathcal{A}_S$，当且仅当它沿 G 的每个环的交替和为零：如果 $\phi = f + g$，则每个节点的值以每种符号各进入环一次；反之，如果所有交替和都为零，则通过传播在生成森林上定义 f、g，而每条非树边都会闭合出一个环，其零和使该边上的 $\phi = f + g$ 得以确定。如果 η 不可识别，则 $h = f + g$ 在 S 上以及参数（$\alpha + \eta f$、$\beta + \eta g$、0）和（α、β、η）在 S 上的取值相同，因此可观测效应确定了所有点对上的 $\tau_0(a$、$b)\ = \alpha_a + \beta_b + \eta\{f(a)\ + g(b)\}$（根据定理1，只差不改变 Ψ 的平移），并且 $\Psi(\eta)\ = \sum \pi_T \tau_0 + \eta \sum \pi_T(h - f - g)$。表示 h 在 S 上的不同（f、g）相差分量平移，这些平移会通过 $\sum_K c_K(w_K - h_K)\ = 0$ 改变 Δ。

□

推论 2。 在指标-患者情形下， $h = D_w D_h$ 以及 $f + g = D_w + D_h - 1$ 在 S上，因此 $h - f - g$ =($1 - D_w$)(1 − D_h)和 $\Delta = \pi_T(00) = \sum_r \pi_T(r)\{1 - q_w - q_h + P_{11}(r)\}$。PATE(0) 对伴侣具有可加性，因此仅取决于边际分布。对于固定的 η，目标效应关于 P_{11}呈线性关系，而对于

固定的 P_{11} ，它关于 η呈线性关系；矩形上的双线性函数在顶点处取得极值，并且根据连续性可以取得每个中间值。 □

命题5。 根据线性性， $E\tilde{\tau}_j = \lambda_j\tau_j + (1-\lambda_j)\bar{\mu}$，因此 $E\sum_j \pi_T(j)\tilde{\tau}_j - \sum_j \pi_T(j)\tau_j = \sum_j \pi_T(j)(1-\lambda_j)(\bar{\mu} - \tau_j)$。

□

B Stan 程序

下方重现了主要模型；补充代码还包含标准的向零收缩比较，以及带有合并开关的版本。

```
// Bayesian dyadic multilevel regression and poststratification (dyadic MRP)
// for transporting a couple-randomized trial to a survey-defined population of couples.//
// Outcome model: bivariate normal model for the wife's and husband's outcomes, fitted to
// cell-by-arm sufficient statistics. Each partner's treatment effect depends on own traits,
// partner traits, an own-by-partner diabetes interaction, and couple race/ethnicity.
// Dyadic partial pooling: the wife's and the husband's coefficient for the same term are
// pooled toward a shared couple-level mean, with one pooling scale per batch of terms.
// Composition model: Dirichlet model for the joint couple-cell distribution,
// pi ~ Dirichlet(n_eff * [(1 - omega) * pi_ht + omega * m]), drawn in generated quantities.

data {
  int<lower=1> C; // number of couple cells
  int<lower=1> KZ; // number of effect-modifier columns
  int<lower=1> KW; // number of prognostic columns
  matrix[C, KZ] Zw; // modifier design, wife (reference coding)
  matrix[C, KZ] Zh; // modifier design, husband
  matrix[C, KW] Ww; // prognostic design, wife
  matrix[C, KW] Wh; // prognostic design, husband
  int<lower=1> B; // number of batches of modifier termsarray[KZ] int<lower=1,
   upper=B> batch; // batch of each modifier column

  // trial: cell-by-arm sufficient statistics
  int<lower=1> G;array[G] int<lower=1, upper=C> g_cell;
  vector<lower=0, upper=1>[G] g_arm;vector<lower=1>[G] g_n;
  matrix[G, 2] g_mean; // (wife, husband) means
  matrix[2, 2] S_tot; // pooled within-group scatter matrix
  int<lower=1> N_trial;



  // target survey
  vector<lower=0>[C] dir_par; // Dirichlet parameters for the composition posterior
  vector<lower=0>[C] pi_ht; // weighted cell shares (plug-in)
  vector<lower=0>[C] pi_indep; // plug-in shares,  independent pairing within race
  vector<lower=0>[C] pi_trial; // trial cell shares

  // subgroups for couple-average CATEs
  int<lower=0> NS;
  matrix[C, NS] sub; // 0/1 cell membership
}

transformed data {
  matrix[G, KZ] gZw = Zw[g_cell];
  matrix[G, KZ] gZh = Zh[g_cell];
```

```
  matrix[G, KW] gWw = Ww[g_cell];
  matrix[G, KW] gWh = Wh[g_cell];
}

parameters {
  vector[2] alpha; // control-arm intercepts
  vector[KW] beta_w; // prognostic coefficients
  vector[KW] beta_h;
  vector[2] tau0; // effect in the reference cell
  vector[KZ] theta_bar; // shared couple-level modifier coefficients
  vector[KZ] z_w; // wife deviations (non-centered)
  vector[KZ] z_h; // husband deviations
  vector<lower=0>[B] sd_dev; // pooling scale per batch
  vector<lower=0>[2] sigma_y;
  cholesky_factor_corr[2] L_y;
}

transformed parameters {
  vector[KZ] theta_w = theta_bar + sd_dev[batch] .* z_w;
  vector[KZ] theta_h = theta_bar + sd_dev[batch] .* z_h;
}

model {
  alpha ~ normal(0, 10);
  beta_w ~ normal(0, 10);
  beta_h ~ normal(0, 10);
  tau0 ~ normal(0, 10);
  theta_bar ~ normal(0, 5);
  z_w ~ std_normal();
  z_h ~ std_normal();
  sd_dev ~ normal(0, 1.5);
  sigma_y ~ normal(0, 20);
  L_y ~ lkj_corr_cholesky(2);
  {
    matrix[G, 2] R;
    R[, 1] = g_mean[, 1] - (alpha[1] + gWw * beta_w) - g_arm .* (tau0[1] + gZw * theta_w);
    R[, 2] = g_mean[, 2] - (alpha[2] + gWh * beta_h) - g_arm .* (tau0[2] + gZh * theta_h);
    matrix[2, 2] L_S = diag_pre_multiply(sigma_y, L_y);
    matrix[2, 2] Sinv = chol2inv(L_S);
    real quad = sum(g_n .* rows_dot_product(R * Sinv, R));
    target += -N_trial * sum(log(diagonal(L_S))) - 0.5 * trace(Sinv * S_tot) - 0.5 * quad;
  }
}

generated quantities {vector[3] pate_full; // wife, husband, couple averagevector[3] pate_fixed;
vector[3] pate_indep;vector[3] sate_model;
vector[NS] tau_sub; // couple-average CATEs in target subgroups
real rho_y = multiply_lower_tri_self_transpose(L_y)[1, 2];{vector[C] tw = tau0[1] + Zw * theta_w;
vector[C] th = tau0[2] + Zh * theta_h;vector[C] tc = 0.5 * (tw + th);
vector[C] pm = dirichlet_rng(dir_par);
pate_full = [dot_product(pm, tw), dot_product(pm, th), dot_product(pm, tc)]';
pate_fixed = [dot_product(pi_ht, tw), dot_product(pi_ht, th), dot_product(pi_ht, tc)]';
pate_indep = [dot_product(pi_indep, tw), dot_product(pi_indep, th), dot_product(pi_indep, tc)]';
```

```
    sate_model = [dot_product(pi_trial, tw), dot_product(pi_trial, th), dot_product(pi_trial, tc)]';
    for (s in 1:NS) tau_sub[s] = dot_product(pm .* col(sub, s), tc) / dot_product(pm, col(sub, s));
  }
}
```