艾滋病毒感染伴侣关系内与伴侣关系间传播率的推断

Irene Garc´ıa Mu˜noz[1,*]、Ian Hall[1]，和Thomas House[1]

[1]曼彻斯特大学数学系，曼彻斯特，M13 9PL[*]通讯作者：
irene.garciamunoz@manchester.ac.uk

2026年2月5日

arXiv:2602.04638v1 [stat.AP] 4 Feb 2026

摘要

血清状态不一致伴侣中的艾滋病病毒传播仍是一项重大的公共卫生挑战，尤其是在撒哈拉以南非洲。估计此类感染的速率，以及感染自伴侣关系外部引入的速率，是推断组内与组间传播强度这一更普遍的流行病学挑战的一个特殊情况。本研究提出一个随机易感者—感染者（SI）配对模型，用于估计支配伴侣内及伴侣间艾滋病病毒传播的关键流行病学参数，并进一步扩展该模型以考虑感染动态中的性别特异性差异。采用基于似然的推断方法，我们根据观测数据估计传播参数及相关不确定性。这些数值可用于为艾滋病病毒的感染预防策略提供依据，所提出的方法也可推广至其他流行病学情境。

1 引言

在许多传染病系统中，传播既发生在紧密联系的亚群体内部，通过反复接触实现，也发生在这些亚群体之间，通过较弱的接触实现。例如，家庭（Lau 等人 2012；Madewell 等人 2020）、监狱（Beaudry 等人 2020）、医院（L´opez-Garc´ıa 和 Kypraios 2018）以及长期护理机构（Nicolle 2000；Hall 等人 2021）都是此类情形。对于艾滋病病毒而言，稳定性伴侣关系构成了这样一种亚群体结构；研究发现，传播有两种途径——伴侣关系内传播和伴侣关系间传播——这两种途径对传播的相对贡献具有重要的流行病学和公共卫生意义（Morris 和 Kretzschmar 1997；Eames 和 Keeling 2004；Miller 和 Slim 2017）。伴侣关系内传播反映的是已感染伴侣导致继发感染的风险，而伴侣关系间传播则涵盖从更广泛社区中获得感染的情况。Eyawo 等人（2010）的研究详细综合了这一队列中有关艾滋病病毒传播率和风险变量的信息。该荟萃分析整合了多项研究的数据，发现了影响传播概率的重要因素。抗逆转录病毒治疗（ART）或暴露前预防（PrEP）等干预措施往往对其中一种传播途径的作用强于另一种，因此分别量化这两种途径至关重要。

对于撒哈拉以南非洲的艾滋病病毒感染，血清状态不一致伴侣是研究这些动力学的理想人群。他们提供了一个直接的研究场景，可以观察伴侣内部和伴侣之间发生的事件：内部传播会使血清状态不一致的伴侣转为双方均为阳性，而外部传播则会使双方均为阴性的伴侣转为血清状态不一致。追踪这些转变的队列研究（Hugonnet 等人，2002年；Wall 等人，2019年；Brahmbhatt 等人，2014年；Shamu 等人，2021年）因此为根据实证数据估算这两种感染力提供了机会

来自实证数据。

本研究将一种简化的基于配对的建模框架应用于 Wall 等人（2019年）报告的大型稳定异性伴侣回顾性队列。我们的目标是直接根据伴侣状态之间观测到的转变，推断伴侣关系内传播率（τ）和伴侣关系间传播率（λ）。我们分别采用非性别模型和扩展的性别特异性模型进行推断，后者区分男性向女性和女性向男性的内部传播。

就已有研究而言，我们的模型既与 Kretzschmar 和 Heijne（2017年）综述的配对形成模型有关，也与 Ball（1999年）第 3 节中的家庭流行病微分方程模型有关，但经过简化后可以得到闭式解。似然采用了 Kinyanjui 等人（2018年）的方法，但使用精确解而非数值矩阵指数。Ball、Mollison 和 Scalia-Tomba（1997年）使用适用于易感者-感染者-康复者情形的方法，考察了组内和组间传播率的可辨识性；在该情形下，感染者康复后会获得长期免疫力，且数据与最终结果有关。

通过将模型的精确解析解与最大似然推断相结合，我们得到了两个传播率 τ 和 λ的精确估计值及其不确定性区间。这种方法使我们能够量化内部传播和外部传播对维持稳定伴侣关系中艾滋病病毒感染率的相对重要性，并评估考虑性别特异性影响时，这些因素的贡献如何变化。研究结果对干预策略具有直接意义，尤其有助于确定应优先投入资源阻断社区感染，还是减少伴侣间传播。

2 方法

2.1 模型描述

我们构建了一个模型框架，用于描述封闭人群中的疾病传播动力学；在该人群中，每个人在整个研究期间始终处于一对伴侣关系中。由于研究重点是艾滋病病毒，我们采用两个流行病学分舱： S （易感者）和 I （感染者）。根据伴侣双方的状态，我们确定了三种组合： SS （双方均易感或双方均为阴性）、 SI （一方易感、一方感染，即血清不一致）和 II （双方均感染或血清状态一致阳性）；模型的动力学变量为各伴侣类型的预期数量。这些分舱之间的转变由两个传播率驱动：内部传播率 τ，表示伴侣内部的感染；以及外部传播率 λ，表示感染源来自伴侣关系之外。

就对典型性传播感染（STI）配对形成模型（Kretzschmar和Heijne 2017）所作的简化而言，我们所建模的时间尺度并不需要显式的配对动力学，例如伴侣关系持续时间、单配偶制以及在更长时期内具有更重要流行病学意义的配对内再感染。然而，其核心结构——追踪 SS、 SI和 II 配对状态，并使用配对内传播（τ）和外部感染力（λ）——仍忠实于其理论基础。

选择基于配对的建模，而非经典模型（后者假设伴侣关系瞬时形成，并通过性伴更换率刻画传播），有 Ong 等人（2012）的研究作为依据。该研究比较了这两种方法在多种性传播感染中的表现，发现经典模型可能会低估长期性伴关系中的自我维持传播。这验证了我们采用配对模型的合理性，因为该模型能更好地捕捉艾滋病病毒在伴侣关系中的延长传染期，同时避免不必要的模型复杂性。

一组常微分方程组（ODEs）描述了模型的动力学，并确保配对总数随时间守恒。

$$\begin{aligned}\frac{\mathrm{d}P_{SS}}{\mathrm{d}t} &= -2\lambda P_{SS},\\ \frac{\mathrm{d}P_{SI}}{\mathrm{d}t} &= 2\lambda P_{SS} - (\lambda+\tau)P_{SI},\\ \frac{\mathrm{d}P_{II}}{\mathrm{d}t} &= (\lambda+\tau)P_{SI}.\end{aligned} \tag{1}$$

这些方程确保配对总数守恒，即 $P_{SS}+P_{SI}+P_{II}=N$，其中 N是配对总数，并涉及 2 个未知参数。

可以求解这个方程组，得到：

$$\begin{aligned}P_{SS}(t) &= P_{SS}(0)\mathrm{e}^{-2\lambda t};\\ P_{SI}(t) &= \left(P_{SI}(0)\mathrm{e}^{-(\tau-\lambda)t} + P_{SS}(0)\frac{2\lambda}{(\tau-\lambda)}(1-\mathrm{e}^{-(\tau-\lambda)t})\right)\mathrm{e}^{-2\lambda t}.\end{aligned}$$

这些结果及其他数学结果的完整推导见补充材料。请注意，全文中我们将用 P 表示给定配对状态的模型预期数量，该数量为连续值；用 N 表示观测到的配对整数数量。由于总和 $P_{SS}+P_{SI}+P_{II}=N$ 由方程 (1) 守恒，因此我们有

$$P_{II}(t) = N - \left(P_{SS}(0) + P_{SI}(0)\mathrm{e}^{-(\tau-\lambda)t} + P_{SS}(0)\frac{2\lambda}{\tau-\lambda}(1-\mathrm{e}^{-(\tau-\lambda)t})\right)\mathrm{e}^{-2\lambda t}.$$

值得注意的是，在上述方程组中，τ 与 λ 组合出现，因此可能值得定义 $\tau=\lambda+\theta$，其中 θ 表示与外部感染力相比，内部接触带来的额外传播潜力。

我们可以像 Brahmbhatt 等人（2014）那样，更明确地追踪男性和女性伴侣。我们区分男性或女性为感染者的血清不一致配对。我们将人群分为两类：m （男性）和 f （女性），使得

$$\begin{aligned}\frac{\mathrm{d}P_{S_mS_f}}{\mathrm{d}t} &= -(\lambda_m+\lambda_f)P_{S_mS_f},\\ \frac{\mathrm{d}P_{I_mS_f}}{\mathrm{d}t} &= \lambda_m P_{S_mS_f} - (\tau_{m\to f}+\lambda_f)P_{I_mS_f},\\ \frac{\mathrm{d}P_{S_mI_f}}{\mathrm{d}t} &= \lambda_f P_{S_mS_f} - (\lambda_m+\tau_{f\to m})P_{S_mI_f},\\ \frac{\mathrm{d}P_{I_mI_f}}{\mathrm{d}t} &= (\tau_{m\to f}+\lambda_f)P_{I_mS_f} + (\lambda_m+\tau_{f\to m})P_{S_mI_f}.\end{aligned}$$

此模组 el现在有四个未知参数。同样，我们可以精确求解该系统，an d

得到

$$
\begin{aligned}
P_{S_mS_f}(t) &= P_{S_mS_f}(0)\mathrm{e}^{-(\lambda_m+\lambda_f)t},\\
P_{I_mS_f}(t) &= \left(P_{I_mS_f}(0)\mathrm{e}^{-(\tau_{m\to f}-\lambda_m)t} + \frac{\lambda_m P_{S_mS_f}(0)}{\tau_{m\to f}-\lambda_m}(1-\mathrm{e}^{-(\tau_{m\to f}-\lambda_m)t})\right)\mathrm{e}^{-(\lambda_m+\lambda_f)t},\\
P_{S_mI_f}(t) &= \left(P_{S_mI_f}(0)\mathrm{e}^{-(\tau_{f\to m}-\lambda_f)t} + \frac{\lambda_f P_{S_mS_f}(0)}{\tau_{f\to m}-\lambda_f}(1-\mathrm{e}^{-(\tau_{f\to m}-\lambda_f)t})\right)\mathrm{e}^{-(\lambda_m+\lambda_f)t},\\
P_{I_mI_f}(t) &= N - \mathrm{e}^{-(\lambda_m+\lambda_f)t}\Bigg(P_{S_mS_f}(0) + P_{I_mS_f}(0)\mathrm{e}^{-(\tau_{m\to f}-\lambda_m)t}\\
&\qquad + P_{S_mI_f}(0)\mathrm{e}^{-(\tau_{f\to m}-\lambda_f)t} + \frac{\lambda_m P_{S_mS_f}(0)}{\tau_{m\to f}-\lambda_m}\big(1-\mathrm{e}^{-(\tau_{m\to f}-\lambda_m)t}\big)\\
&\qquad + \frac{\lambda_f P_{S_mS_f}(0)}{\tau_{f\to m}-\lambda_f}\big(1-\mathrm{e}^{-(\tau_{f\to m}-\lambda_f)t}\big)\Bigg).
\end{aligned}
$$

2.2 数据

我们的初始条件和参数估计基于坦桑尼亚姆万扎开展的一项回顾性队列研究所报告的数据。该研究对 $N = 1$、802 对稳定异性伴侣进行了为期两年的随访（Hugonnet 等人，2002年）。研究详细记录了两个时间点上双方的 HIV 状态，使我们能够追踪不同流行病学状态之间的转变。

研究开始时，$N^0_{SS} = 1$，742 对伴侣被确定为双方均为阴性，即双方均为 HIV 阴性。$N^0_{SI} = 43$对伴侣为血清不一致，其中 $N^0_{I_mS_f} = 22$ 对为 HIV 阳性男性与易感女性伴侣，另有 $N^0_{S_mI_f} = 21$ 对为 HIV 阳性女性与易感男性。$N^0_{II} = 17$ 对伴侣为双方均为阳性，即双方均已感染。这些数值用于设定模型的初始条件，因此 $P_{SS}(0) = P_{S_mS_f}(0) = N^0_{SS}$等。

两年后，在时间 T，对该队列进行了重新评估。血清状态一致阴性的伴侣数量减少至 $N^T_{SS} = 1,721$ 对。与此同时，血清状态不一致伴侣数量增加至 $N^T_{SI} = 58$ 对，其中男性为感染者的有 $N^T_{I_mS_f} = 33$ 对，女性为感染者的有 $N^T_{S_mI_f} = 25$ 对。血清状态一致阳性的伴侣数量增至 N^T_{II} =23 对。这些变化反映了伴侣内部传播和潜在外部感染 HIV 的共同作用，并为估计伴侣关系内部（内部）和社区层面（外部）的传播参数提供了依据。

所有模型参数和初始条件汇总于表 1。伴侣总数固定为 $N = 1,802$；数值拟合时，根据 Wall 等人（2019）提供的合理估计值，为 λ （外部传播率）和 τ（内部传播率）等参数设置初始值，随后使用 *scipy* （Virtanen 等人，2020年）的优化算法逐步求得最大似然值。这些参数的性别特异性版本 λ_m、 λ_f、 τ_m 和 τ_f，反映了在该队列中观察到的 HIV 传播方向性。

2.3 参数估计

使用多项概率分布将模型输出与观测数据关联起来，以估计参数 λ 和 τ。由于我们旨在提供具有普适性的方法，因此将采用一般情形：数据由每个观测时间点 $t \in \mathcal{T}$对应的观测计数数组 $(N^t_{SS},\ N^t_{SI},\ N^t_{II})$ 组成，其中 $\mathcal{T}$是一个通用的观测

符号说明		数值／备注
N_{SS}^t	一致阴性（SS）配对	t =时为 1742 0, t =时为 1721 2（Hugonnet 等人，2002年）
N_{SI}^T	血清不一致（SI）配对（1名易感者，1名感染者）	$t=0$时为43对，$t=2$ 时为58对（Hugonnet 等人，2002年）
N_{II^T}	一致阳性（II）配对	$t=0$时为17对，$t=2$ 时为23对（Hugonnet 等人，2002年）
$N_{I_mS_f}$	男性已感染、女性易感配对	22位于 $t=0$，33位于 $t=2$（Hugonnet 等人，2002年）
$N_{S_mI_f}$	女性已感染、男性易感者配对	$t=0$ 处为 21，$t=2$ 处为 25（Hugonnet 等人，2002年）
N	配对总数（在整个期间保持恒定时间）	1802对配对（Hugonnet 等人，2002年）
λ	外部传播率	估计值
λ_m	影响男性伴侣的外部传播率男性伴侣	估计的
λ_f	外部传播率影响女性伴侣	估计的
τ	内部传播率（通用）	估计值
$\tau_{m\to f}$	内部传播率，来自感-染男性至女性	估计值
$\tau_{f\to m}$	内部传播率，来自感-染女性至男性	估计值

表 1：模型参数和数据

时间集合，在我们的特定数据集中等于 $\{0,T\}$。于是，似然函数为

$$L(\lambda,\tau)=\prod_{t\in\mathcal{T}}\binom{N^t}{N_{SS}^t,N_{SI}^t,N_{II}^t}P_{SS}(t)^{N_{SS}^t}P_{SI}(t)^{N_{SI}^t}P_{II}(t)^{N_{II}^t},$$

其中，$P_{SS}(t)$、$P_{SI}(t)$ 和 $P_{II}(t)$ 是方程 (1) 所给出的模型在时间 t 预测的比例，N 是该时间点观测到的配对总数。因此，该似然函数的对数可简化为：

$$\mathcal{L}(\lambda,\tau)=\log(L(\lambda,\tau))\propto\sum_{t\in\mathcal{T}}\big(N_{SS}^t\log(P_{SS}(t))+N_{SI}^t\log(P_{SI}(t))+N_{II}^t\log(P_{II}(t))\big),$$

这是参数估计的基础，因为构成比例常数的多项式系数与待估参数无关。根据观测数据，λ和 τ 的最可能取值可通过最大化该对数似然函数得到。

为衡量参数估计值的不确定性程度，我们考察对数似然曲线在其最大值附近的曲率。具体做法是计算负对数似然的二阶偏导数矩阵，即海森矩阵：

$$\mathbf{H}(\lambda^*,\tau^*)=\left[\begin{matrix}\frac{\partial^2\mathcal{L}}{\partial\lambda^2} & \frac{\partial^2\mathcal{L}}{\partial\lambda\partial\tau}\\ \frac{\partial^2\mathcal{L}}{\partial\tau\partial\lambda} & \frac{\partial^2\mathcal{L}}{\partial\tau^2}\end{matrix}\right]\Bigg|_{(\lambda^*,\tau^*)}.$$

通常，如果我们得到了参数的最大似然估计，

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} = (\hat{\lambda}_{\mathrm{MLE}}, \hat{\tau}_{\mathrm{MLE}}) = \mathrm{argmax}(\mathcal{L}(\lambda, \tau)),$$

然后，可以根据均值为$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$、协方差矩阵为 $-\mathbf{H}^{-1}(\hat{\lambda}_{\mathrm{MLE}}, \hat{\tau}_{\mathrm{MLE}})$的多元正态概率密度函数的等高线，构造一个渐近置信域。

最后，通过在参数值网格上计算$L(\lambda, \tau)$，探究似然曲面的可辨识性和结构。这直观地展示了参数估计的确定程度，并揭示参数估计中的任何相关性或歧义。

2.3.1 似然分析：非性别模型

如上所述，我们考虑特殊情况 $\mathcal{T} = \{0, T\}$。似然包含六项：时间0时三项，以及两年后时间 T 时三项。只有时间 T 时的三项涉及 λ 或 τ项。在时间0时，我们观察到N^0_{SS}、 N^0_{SI}和N^0_{II}；在时间 T时，我们观察到N^T_{SS}、 N^T_{SI}和N^T_{II}。对数似然包含不影响导数的常数项，其表达式为

$$\begin{aligned}\mathcal{L} = N^0_{SS} \log P_{SS}(T; \lambda) + N^0_{SI} \log P_{SI}(T; \lambda, \tau) \\ + N^0_{II} \log P_{II}(T; \lambda, \tau) + \text{constant}.\end{aligned}$$

对一般参数 θ 求对数似然的导数，得:

$$\frac{\partial L}{\partial \theta} = \left(\frac{N^0_{SS}(N - P_{SI}) - P_{SS}(N - N^0_{SI})}{P_{SS}(T)(N - P_{SI}(T) - P_{SS}(T))} \right) \frac{\partial P_{SS}(T)}{\partial \theta} + \left(\frac{(N - P_{SS}(T))N^0_{SI} - (N - N^0_{SS})P_{SI}(T)}{P_{SI}(T)\,(N - P_{SI}(T) - P_{SS}(T))} \right) \frac{\partial P_{SI}(T)}{\partial \theta}.$$

对 τ 求对数似然的导数（记住 P_{SS} 不涉及 τ），可得

$$P_{SI}(T; \hat{\lambda}, \hat{\tau}) = N^0_{SI} \frac{(N - P_{SS}(T; \hat{\lambda}))}{(N - N^0_{SS})},$$

代入偏导数 λ 得 $N^0_{SS}/P_{SS}(T; \hat{\lambda}) = N^0_{SI}/P_{SI}(T; \hat{\lambda}, \hat{\tau})$，由此

$$\hat{\lambda} = \frac{1}{2T} \log \left(\frac{N^0_{SS}}{N^T_{SS}} \right). \tag{2}$$

现在给定 $P_{SI}(T; \hat{\lambda}, \hat{\tau}) = N^0_{SI}$，即可求得 $\hat{\tau}$:

$$\left(N^T_{SI} \mathrm{e}^{-(\hat{\tau} - \hat{\lambda})T} + \frac{N^T_{SS} \log \left(\frac{N^T_{SS}}{N^0_{SS}} \right)}{T} \frac{\left(1 - \mathrm{e}^{-(\tau - \lambda)T}\right)}{(\tau - \lambda)} \right) = \frac{N^0_{SI} N^T_{SS}}{N^0_{SS}}.$$

现在，求解 τ 有多种选择。这里我们重新参数化，使得 $\tau = (2\phi + 1)\lambda$，因此

$$\left(\frac{N^T_{SI}}{N^T_{SS}} \phi - 1 \right) \left(\frac{N^0_{SS}}{N^T_{SS}} \right)^{\phi} = \left(\frac{N^0_{SI} \phi}{N^0_{SS}} - 1 \right).$$

为了得到 τ，我们可以将项 $\binom{N^0_{SS}}{N^0_{SS}}^{\phi}$ 改写为:

$$\left(\frac{N^0_{SS}}{N^0_{SS}} \right)^{\phi} = \left(1 - \frac{N^0_{SS} - N^0_{SS}}{N^0_{SS}} \right)^{\phi} \approx 1 - \phi \frac{N^0_{SS} - N^0_{SS}}{N^0_{SS}} + \dots$$

因此对于sm 对所有 $\frac{N_{SS}^T - N_{SS}^0}{N_{SS}^T}$，我们可以用二项式展开进行近似，方程变为 :

$$\frac{\phi N_{SI}^0}{N_{SS}^0} - 1 = \left(\frac{\phi N_{SI}^T}{N_{SS}^T} - 1\right)\left(1 - \phi\frac{N_{SS}^T - N_{SS}^0}{N_{SS}^T}\right).$$

Then

$$\hat{\phi} \approx \frac{N_{SS}^T}{(N_{SS}^T - N_{SS}^0)}\left(\frac{N_{SI}^0 N_{SS}^T}{N_{SS}^0 N_{SI}^T} - 1 - \frac{(N_{SS}^T - N_{SS}^0)}{N_{SI}^T}\right).$$

最后，我们可以使用重新参数化 $\tau = (2\phi + 1)\lambda$ 来估计 $\hat{\tau}$。我们注意到，一般而言，$\hat{\lambda} = \hat{\lambda}_{\mathrm{MLE}}$，但 $\hat{\tau} = \hat{\tau}_{\mathrm{MLE}} + \mathcal{O}(\phi^2)$。我们将这些结果称为"解析"结果，将通过完整数值优化得到的结果称为"最大似然估计（精炼）"。

2.3.2 似然分析：性别模型

沿用上文的记号，在时间 0，我们观测到 N_{SS}^0、 $N_{I_mS_f}^0$、 $N_{S_mI_f}^0$ 和 N_{II}^0；在时间 T，我们观测到 N_{SS}^T、 $N_{I_mS_f}^T$、 $N_{S_mI_f}^T$ 和 N_{II}^T。性别模型可以按以下方式重新参数化：

$$\begin{aligned}
\lambda_m &= 2\lambda q, \\
\lambda_f &= 2\lambda(1-q), \\
\tau_{m\to f} &= 2\lambda(q + \theta_m), \\
\tau_{f\to m} &= 2\lambda(1 - q + \theta_f).
\end{aligned}$$

Thus

$$\begin{aligned}
P_{S_mS_f}(t) &= N_{SS}^t \mathrm{e}^{-2\lambda t} \\
P_{I_mS_f}(t) &= \left(N_{I_mS_f}^t \mathrm{e}^{-2\lambda\theta_m t} + \frac{qN_{SS}^t}{\theta_m}(1 - \mathrm{e}^{-2\lambda\theta_m t})\right)\mathrm{e}^{-2\lambda t}. \\
P_{S_mI_f}(t) &= \left(N_{S_mI_f}^t \mathrm{e}^{-2\lambda\theta_f t} + \frac{(1-q))N_{SS}^t}{\theta_f}\left(1 - \mathrm{e}^{-2\lambda\theta_f t}\right)\right)\mathrm{e}^{-2\lambda t}.
\end{aligned}$$

The 去除不影响导数的常数项后，对数似然因此简化为 to:

$$\mathcal{L} = N_{SS}^T \log P_{SS}(T) + N_{I_mS_f}^T \log P_{IS}(T) + N_{S_mI_f}^T \log P_{SI}(T) + N_{II}^T \log P_{II}(T).$$

根据与非性别版本类似的论证，我们预期最大似然估计出现在 $P_{SS}(T;\hat{\lambda}) = N_{SS}^0$ 处，因此

$$e^{-2\hat{\lambda}T} = \left(\frac{N_{SS}^0}{N_{SS}^T}\right).$$

随后，我们可以采用与上述相同的方法，推导出在给定 $\hat{q}$ 和 λ 的条件下 θ_m 和 θ_f 的显式形式，得到

$$\begin{aligned}
\frac{\theta_m N_{I_mS_f}^0}{N_{SS}^0} - q &= \left(\frac{\theta_m N_{I_mS_f}^T}{N_{SS}^T} - q\right)\left(\frac{N_{SS}^0}{N_{SS}^T}\right)^{\theta_m} \\
\frac{\theta_f N_{S_mI_f}^0}{N_{SS}^0} - 1 + q &= \left(\frac{\theta_f N_{S_mI_f}^T}{N_{SS}^T} - (1-q)\right)\left(\frac{N_{SS}^0}{N_{SS}^T}\right)^{\theta_f}
\end{aligned}$$

因此

$$\theta_m \approx \left(q + \left(\frac{N_{SS}^T}{N_{SS}^T - N_{SS}^0}\right)\left(\frac{N_{I_mS_f}^T}{N_{SS}^T} - \frac{N_{I_mS_f}^0}{N_{SS}^0}\right)\right)\frac{N_{SS}^T}{N_{I_mS_f}^T} \tag{3}$$

$$\theta_f \approx \left(1 - q + \left(\frac{N_{SS}^T}{N_{SS}^T - N_{SS}^0}\right)\left(\frac{N_{S_mI_f}^T}{N_{SS}^T} - \frac{N_{S_mI_f}^0}{N_{SS}^0}\right)\right)\frac{N_{SS}^T}{N_{S_mI_f}^T}. \tag{4}$$

通过推导上述“解析”结果，我们可以避免在某些情况下为获得“最大似然估计（精炼）”结果所需的数值优化，并深入了解不同观测如何影响所有模型的估计值和可辨识性。

3 结果

3.1 解析结果

对于非性别版本，解析结果很直接：代入外部感染力率的数值后，我们得到：

$$\lambda = \frac{1}{4}\log\left(\frac{1742}{1721}\right) \approx 0.003.$$

对于内部感染力，我们可以使用重新参数化 $\tau = (\phi + 1)\lambda$ 来估计 $\hat{}\tau$：

$$\hat{\tau} = (16.95 + 1)\hat{\lambda} = 17.95\hat{\lambda} = 0.054$$

对于性别模型，我们得到的 λ 值与非性别模型中的相同。然而，在为性别模型求解 (4) 时，我们发现需要使用泰勒级数中的更高阶项才能得到唯一的极大值，这表明对于这种更复杂的模型，采用数值优化比解析结果更可取。

3.2 数值优化结果

对于配对模型，使用上述参数推断方法计算外部传播率（λ）和内部传播率（τ）的估计值。我们首先根据血清发病率确定初始参数估计值，计算方法与 Wall 等人（2019年）所述相同：

$$\lambda \approx (N_{SI}^{T} - N_{SI}^{0}/(2TN_{SS}^{0}));$$
$$\tau \approx (N_{II}^{T} - N_{II}^{0}/(2TN_{SI}^{0})).$$

我们将这对参数值称为“闭式近似”（CFA）。

$\hat{}$ 最大似然估计为 $\lambda = 0.003$，以及 $\hat{}\tau = 0.056$。在最大似然点处，计算了负对数似然函数的海森矩阵，以评估这些估计值的不确定性。所得协方差矩阵给出的标准误为 $\sigma_\lambda = 0.001$ 和 $\sigma_\tau = 0.046$。

利用标准误，构建了参数的 95% 和 67% 二元置信区域，如图1所示。这些区间表明，观测数据对模型参数的分辨能力良好；与显式似然曲面的比较也显示，只要将相应的高斯分布截断在 λ和 τ 等于 0 处，对海森矩阵采用标准近似就是合理的。

现在，我们来探讨如何将参数解读为所分析人群中的年感染数。根据模型中的初始伴侣数量——1，742 对易感–易感配对（N_{SS}^{0}）和 43 对不一致配对（N_{SI}^{0}）——我们可以估算由内部和外部感染力导致的年新增感染数。采用外部感染率$\lambda = 0.003$，并考虑到共有 3,527 名易感个体（每对 SS 配对中有两名，每对 SI 配对中有一名），我们估计外部来源每年约导致 11 例新感染。

对于内部传播，即感染发生在不一致伴侣对内部的情况，每年新感染人数由 $\tau \times 43$给出，其中 $\tau = 0.056$ 为内部感染力，由此每年约产生2.4例来自伴侣关系内传播的新感染。

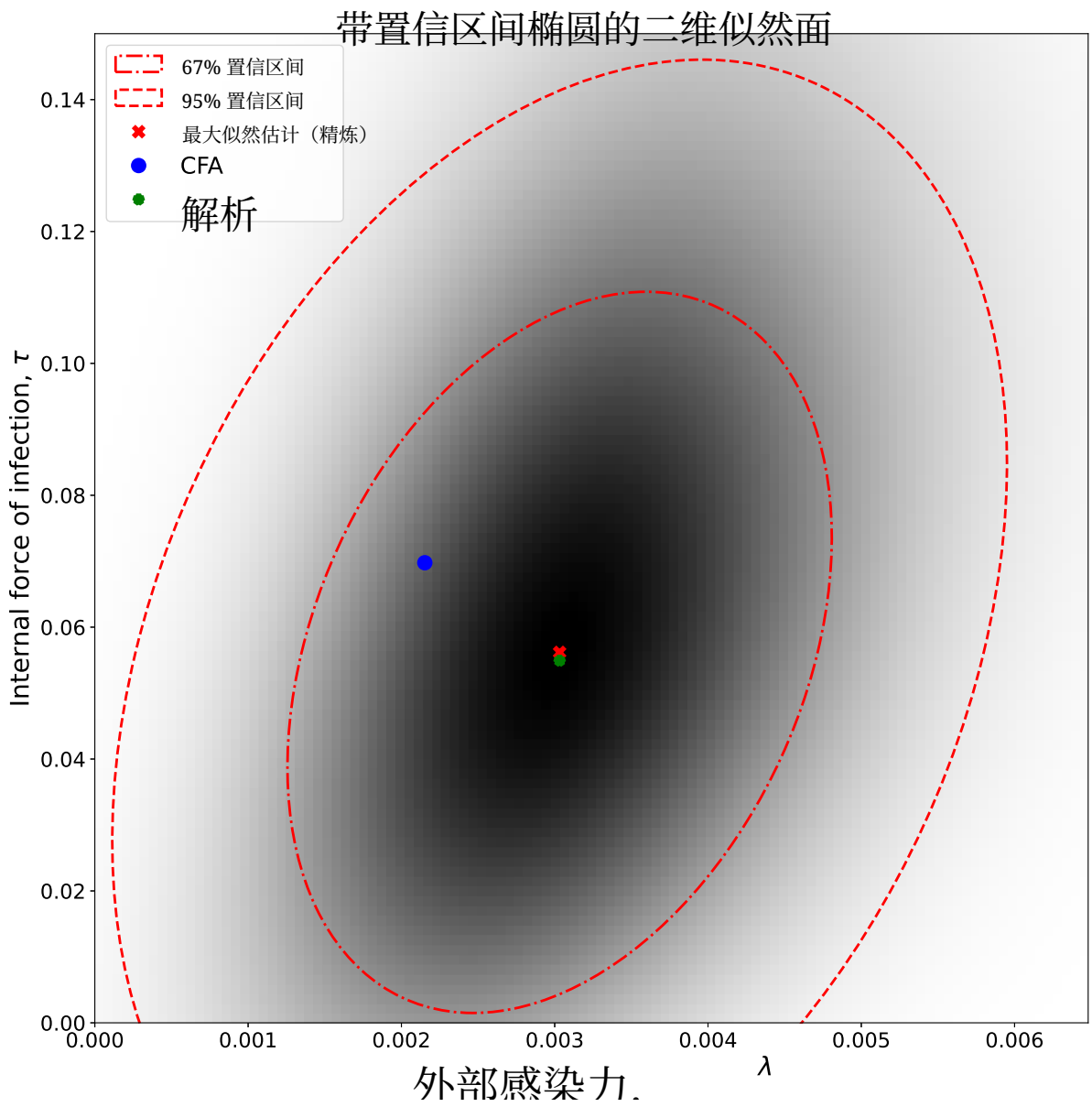

图1：外部（λ）和内部（τ）感染力的二维似然面。图中显示最大似然估计（MLE）、闭式近似（CFA）和解析估计值，以及 67% 和 95% 置信区间（CI）椭圆。CFA 基于底层随机模型的确定性简化得出，可快速提供估计值。

传播类型	**估计年感染数**	**每千人每年**
外部（$\lambda = 0.00303$）	10.6	3.03
内部（$\tau = 0.0561$）	2.48	56.1
总计	**13.1**	**59.1**

表2：使用非性别化传播率计算得出的每年HIV感染估计数。感染按来源分为外部来源（社区传播）或内部来源（伴侣内部传播）。

同样，对于男女配对模型，使用所述参数推断方法计算了外部传播率（λ）和内部传播率（τ）的估计值，方法包括男性-

ˆ女性。使用最大似然估计，$\lambda_m = 0.004$、$\lambda_f = 0.002$、$\hat{\tau}_{m\to f} = 0.047$ 和 $\hat{\tau}_{f\to m} = 0.068$。在最大似然点处，计算负对数似然函数的海森矩阵，以评估这些估计值的不确定性。

参数的95%置信区间根据标准误计算，并在0处截断，得到 λ_m 的区间为 $(0.0006, 0.0073)$，λ_f 的区间为 $(0.00, 0.0051)$，$\tau_{m\to f}$ 的区间为 $(0.00, 0.1819)$，$\tau_{f\to m}$ 的区间为 (0.00,0.2271)。这些区间表明，观测数据为模型参数提供了较为充分的信息。

传播类型	**估计年感染数**	**每千人每年**
外部 ($\lambda_m = 0.004$)	7.05	3.95
外部 ($\lambda_f = 0.002$)	3.52	2.11
内部 ($\tau_{m\to f} = 0.047$)	1.03	46.5
内部 ($\tau_{f\to m} = 0.068$)	1.45	67.9
合计	**13**	**120**

表3：使用性别特异性传播率估计的HIV感染年度数量。外部感染通过社区传播影响易感个体，而内部感染发生在血清状态不一致伴侣中，由男性传至女性，反之亦然。

此外，在 λ和 τ 值的网格上评估了似然曲面，从而生成一幅描绘参数空间结构的热图。该可视化验证了这些参数是可区分的，且似然曲面为单峰，最大似然估计位于最高似然区域的中心。

3.3 参数估计性能

在不同真实参数值下评估了 λ （外部传播率）和 τ （内部传播率）的估计，以检验推断框架的稳健性与准确性。性能分析使用了反映血清状态不一致伴侣真实动力学的模拟数据集。

图 3 展示了估计参数与其各自真实值的对比。参数推断框架的稳健表现凸显了其适用于真实世界数据集的能力。通过准确解析关键流行病学参数，该方法为传播动力学提供了有价值的见解。

4 讨论

本文的主要贡献是通过一个具有精确解的易处理随机配对模型，将血清状态不一致伴侣中艾滋病病毒的内部–外部传播分解加以形式化。针对HIV流行病学中的一个问题，即利用仅在离散时间点观测配对状态的队列数据，将来自社区的感染获得与稳定伴侣关系内的后续传播区分开来。我们的框架同时支持解析方法与数值方法，从而实现透明的似然构建以及闭式近似最大似然估计，既能为本文所考虑的HIV数据提供见解，也能为家庭、护理院、监狱及其他性网络等类似场景提供启示。

非性别化模型的结构简单，但它涵盖了每对伴侣中有多个易感者的情况（ SS中的因子 2）、通过两种途径对 SI 对伴侣持续造成的暴露，以及配对总数的守恒。该模型确实假设伴侣关系不会解体，但考虑到这项研究的观察窗口为两年，这一假设是合理的；这也意味着模型估计的是有效率，而非生物学上的单次行为传播概率。模型的另一项假设是

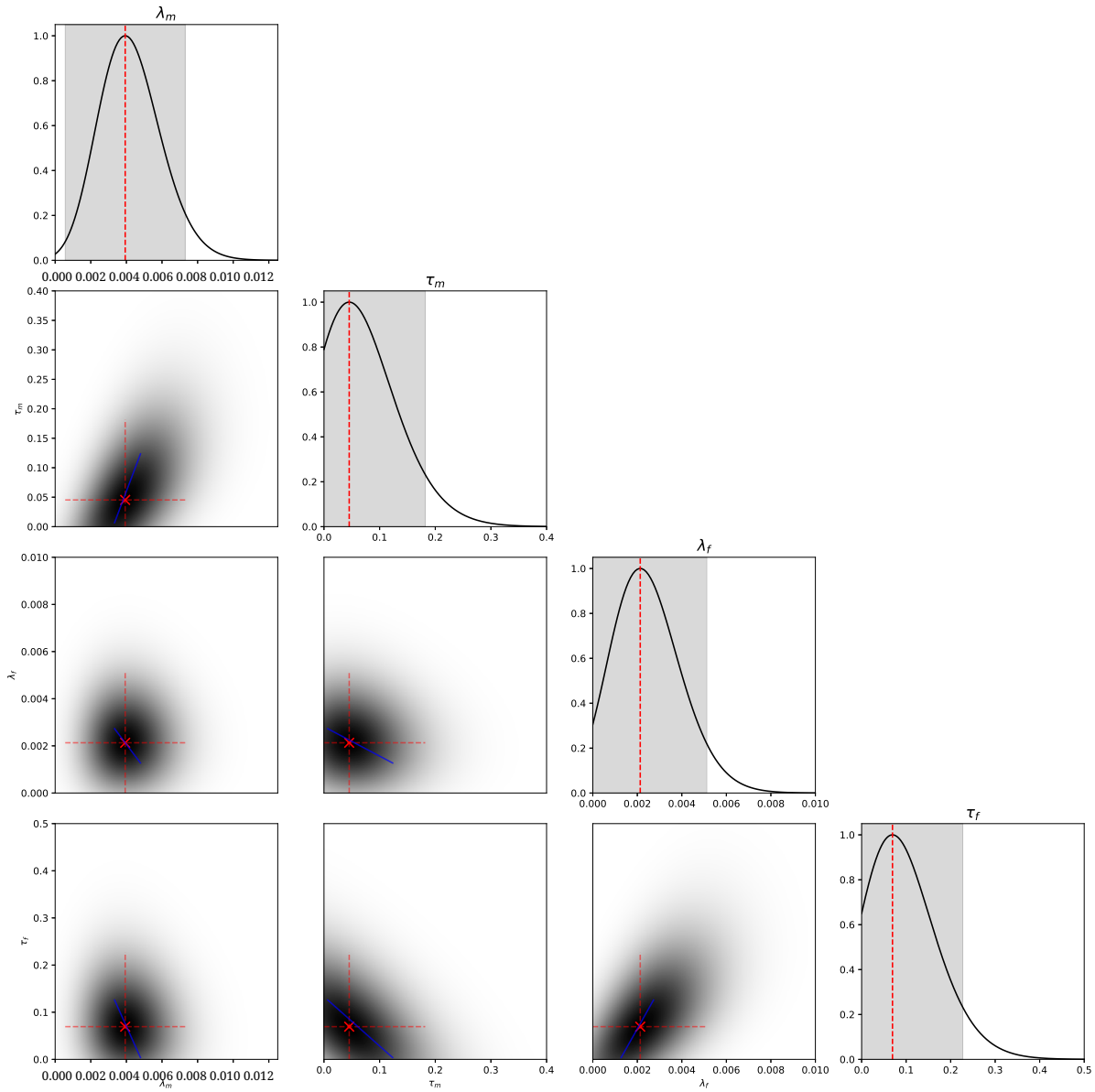

图2：每个对角线面板显示一个参数的剖面似然曲线，该曲线通过改变该参数并将其他参数固定在其最大似然估计值（MLE）处计算得出。红色虚线表示MLE，灰色阴影区域对应95%置信区间。下三角面板显示每对参数的联合似然（原始似然尺度）热图，其余参数同样固定在各自的MLE处。这些面板还显示渐近解的蓝线。

外部传播具有同质性，即所有易感个体面临相同的外部感染力（λ）。这意味着不存在行为异质性，且感染力不取决于伴侣状态。因此，应将 λ 解读为取决于队列关系的背景风险，而非人群层面发病率。

鉴于状态计数涵盖所有情况，即在每个观测时间，研究人群中的每一对都被观测为恰好处于模型的某一个状态；且抽样变异占主导，而非人口随机性，因此采用以配对总数为条件的多项似然是恰当的。从非性别常微分方程来看，它们表明似然由 P_{SS}驱动， λ 、 P_{SI} 和 P_{II}则用于识别在……条件下的 τ

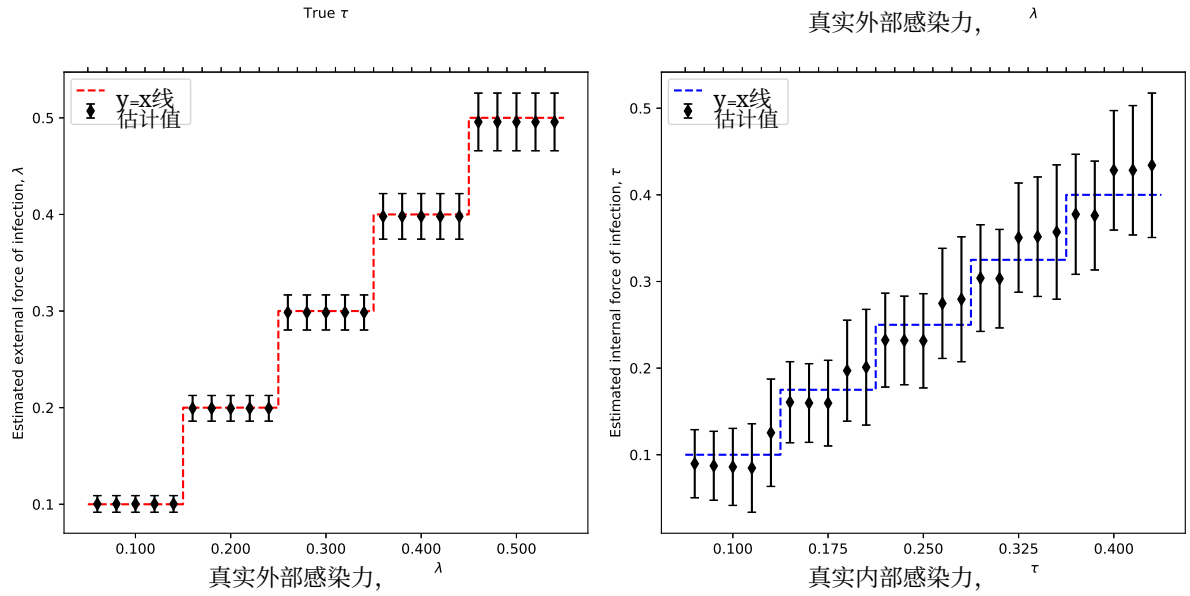

图 3：参数估计性能验证。将估计的外部传播率（λ）和内部传播率（τ）与其真实值进行对比。每个点代表一个模拟数据集，实线对角线对应恒等线（$y = x$），表示完美估计。各点与对角线紧密对齐，表明推断框架在一系列参数值范围内具有准确性和可靠性。

λ。方程 2 的结果表明，λ 仅能根据易感者耗竭来识别，因此与 τ 以及 SI 配对和 II 配对的计数无关。相比之下，τ 是根据小数值之间的差异推断得出的，因此置信区间更宽，且似然曲线呈现明显的不对称弯曲。通过重新参数化和二项式展开进行估计在数学上是正确的；然而，这也揭示了一个更深层的问题：除非 SI 计数呈现显著变化，否则仅凭两个时间点只能弱识别 τ。

可以观察到，内部感染力 τ仅以诸如（$\tau-\lambda$）这样的组合形式出现，因此采用重新参数化 $\tau = \lambda + \theta$。这样，我们便可以将外部感染力 λ解释为基线获得风险，并将 θ 解释为因拥有感染伴侣而产生的超额风险。在性别模型中，采用类似的分解 $\tau_{m\to f} = \lambda_f + \theta_m$ 和 $\tau_{f\to m} = \lambda_m + \theta_f$ 可进一步提升解释的清晰度，并可能降低后验相关性。

性别模型引入了更多结构，同时仍受到数据可得性的严格约束。存在更多可辨识性约束：两个时间点、四条转移路径以及较小的 SI 亚组解释了更宽的置信区间、对参数 q 选择的依赖性，以及似然曲面上的近脊。该 q 参数是作用于男性的总外部感染力所占比例；从流行病学角度而言，我们说 q 刻画的是社区感染获得风险中的性别不对称性，而非传播效率。它反映的是性网络中患病率背景、伴侣并行关系、年龄混合以及行为暴露方面的差异，而非生物学单次行为传染性（后者包含在$\tau_{m\to f}$ 和$\tau_{f\to m}$中）。参数 q 也存在可辨识性问题，因为 SS动力学不包含关于 q的信息。因此，关于 q的全部信息来自 $P_{I_m S_f}$ 和 $P_{S_m I_m}$的相对增长。尽管如此，这些分舱起始规模较小且变化幅度中等，同时受到 λ、θ_m 和 θ_f的影响。因此，q 与 θ_m 和θ_f 在解释同一数据时相互竞争。除非固定 θ_m 和 θ_f，或引入强先验分布，或增加更多观测时间，否则似然并不支持$\hat{q}$ 的唯一值。需要强调的是，这并非推断失败，而是对数据

局限性的正确反映。尽管如此，点估计值是合理的。较高的女性向男性的内部传播与实证文献一致，而男性较高的外部风险也与已知的患病率模式相符。

当然，我们的方法存在若干关键局限。现实世界动态可能包含本框架未专门涵盖的其他复杂性，例如伴侣关系的变化、外部干预或共存状况，正如 Perez 等人（1998年）和 Stein 等人（2019年）所述。我们未考虑伴侣关系解除或抗逆转录病毒治疗的接受，并假设存在均一的外部感染力。然而，本研究通过表明基于配对的模型在解析上是可处理的，作出了实质性的方法学贡献；我们相信，若这些复杂性能够在未来工作中得到适当刻画，我们关于组内与组间传播可直接从队列数据加以推断的见解对这些复杂性仍是稳健的。

代码可用性

本文的代码可在以下网址获取：https://github.com/igmsb18/Inference-for-Within–and-Between-Partnership-Transmission-Rates-for-HIV-Infection

致谢

IGM 谨致谢曼彻斯特大学和英国健康安全局提供的博士资助。IH 和 TH 得到威康信托基金会（编号：227438/Z/23/Z）和医学研究理事会（编号：UKRI483）的支持。作者感谢 Maria Zambon 和 Joe Hilton 对本工作提出的有益意见。

参考文献

Ball, Frank（1999）。“家庭划分人群中的SIS流行病随机模型与确定性模型”。载于：数学生物科学 156.1，页41–67。Ball, Frank、Denis Mollison 和 Gianpaolo Scalia-Tomba（1997）。“具有两层混合的流行病”。载于：应用概率年刊 7.1，页46–89。Beaudry, Gabrielle 等（2020）。“监狱中高度传染性疾病疫情暴发的管理：系统综述”。载于：《英国医学杂志全球健康》 5.11，e003201。Brahmbhatt, Heena 等（2014）。“乌干达拉凯地区艾滋病病毒状态一致和血清不一致伴侣中现代避孕方法使用相关因素及艾滋病毒照护项目影响的纵向研究”。载于：《计划生育与生殖健康护理杂志》 40.3，页208–216。Eames, Ken T.D. 和 Matt J. Keeling （2004）。“单配偶网络与性传播疾病的传播”。载于：数学生物科学 189.2，页115–130。Eyawo, Oghenowede 等（2010）。“撒哈拉以南非洲血清状态不一致伴侣中的HIV状态：系统综述与荟萃分析”。载于：《柳叶刀·传染病》 10.11，页770–777。Hall, Ian 等（2021）。“若不加以遏制，护理院疫情暴发可能导致沉重的疾病负担”。载于：《英国皇家学会哲学汇刊*B*辑：生物科学》376.1829，页20200269。Hugonnet, St´ephane 等（2002）。“稳定性伴侣关系中的艾滋病毒感染发生率：坦桑尼亚姆万扎区1802对伴侣的回顾性队列研究”。载于：《*JAIDS*：获得性免疫缺陷综合征杂志》 30.1，页73–80。

Kinyanjui, Timothy 等（2018）。“养老护理院中的疥疮：针对联系紧密的人群子群的建模、推断与干预措施”。载于：*PLOS*计算生物学 14.3，e1006046。Kretzschmar, Mirjam 和 Janneke C. M. Heijne（2017）。“性传播感染的伴侣配对模型：入门介绍”。载于：传染病建模 2.3，页368–378。Lau, Lincoln L. H. 等（2012）。“2009年甲型H1N1流感大流行的家庭传播：系统综述与荟萃分析”。载于：流行病学 23.4。L´opez-Garc´ıa, Mart´ın 和 Theodore Kypraios（2018）。“医院感染传播的统一随机建模框架”。载于：英国皇家学会界面期刊 15.143，页20180060。Madewell, Zachary J. 等（2020）。“SARS-CoV-2的家庭传播：系统综述与荟萃分析”。载于：*JAMA Network Open* 3.12，e2031756–e2031756。Miller, Joel C. 和 Anja C. Slim（2017年11月）。“饱和效应与并发假说：来自解析模型的见解”。载于：*PLOS ONE* 12.11，e0187938。Morris, Martina 和 Mirjam Kretzschmar（1997）。“并发伴侣关系与艾滋病病毒的传播”。载于：艾滋病 11.5。Nicolle, Lindsay E.（2000）。“长期护理机构中的感染控制”。载于：《临床传染病》 31.3，页752–756。Ong, Jimmy Boon Som 等（2012）。“不同性传播感染中伴侣配对模型与经典模型形式所得结果的可比性”。载于：*PLOS ONE* 7.6，e39575。Perez, George 等（1998）。“II型单纯疱疹病毒和生殖支原体作为异性传播艾滋病病毒的危险因素：异性传播艾滋病病毒研究报告”。载于：《国际传染病杂志》 3.1，页 5–11。Shamu, Tinei 等（ 2021 ）。“ 津巴布韦哈拉雷一家艾滋病病毒诊所收治的50岁以上艾滋病病毒感染者的治疗结局：队列研究”。载于：*PLOS ONE* 16.6，e0253000。Stein, Catherine M 等（2019）。“坎帕拉（乌干达）高暴露结核病家庭接触者对潜伏性结核分枝杆菌感染抵抗力的长期稳定性”。载于：《临床传染病》 68.10，页1705–1712。Virtanen, Pauli 等（2020）。“SciPy 1.0：Python科学计算的基础算法”。载于：《自然·方法》 17，页 261–272。Wall, Kristin M 等（ 2019 ）。“ 面向非洲经济可行的艾滋病病毒预防：伴侣共同接受艾滋病病毒检测与咨询”。载于：《国际流行病学杂志》 48.1，页217–227。