Bayesian analysis of coinfection with soil-transmitted helminths

Simon Spencer,
Giorgia Lepschy, Matthew Adeoye, Thomas Crellen, Evandro Konzen, Sharmini Gunawardena, Claudio Nunes Alves, Vachel Gay Paller, Joaquin Prada, Déirdre Hollingsworth

9th July 2026

DEPARTMENT OF
STATISTICS

Acknowledgements

Giorgia Lepschy

Matthew Adeoye

Thanks to Julia Dunn, Alice Easton, Jennifer Keiser, Charles King, Stefanie Knopp, Bruno Levecke and Benjamin Speich for contributing data and helpful discussions.

Introduction

Introduction

- Soil-transmitted helminths (STH) are three worms that share similar transmission mechanisms.
  - Ascaris (*Ascaris lumbricoides*)
  - Hookworm (*Ancylostoma duodenale* and *Necator americanus*)
  - Trichuris (*Trichuris trichiura*)

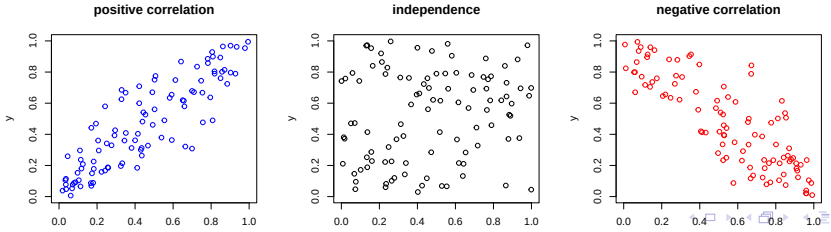
- **Coinfection with multiple species** is common and results in increased morbidity.

Aim: To be able to quantify levels of infection and coinfection from prevalence data on the three species.

The Plackett copula

Copulas

A **copula** $C(u_1, \ldots, u_d)$ is a multivariate CDF on $[0,1]^d$ with $U(0,1)$ marginals.



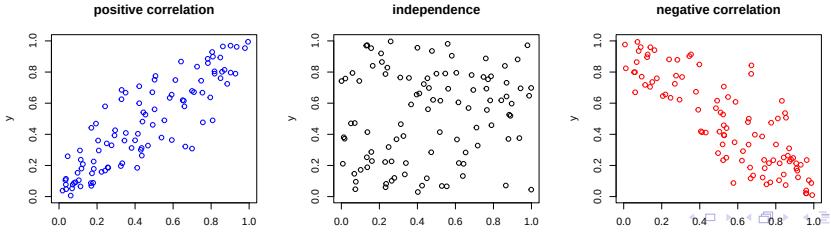
Copulas

A **copula** $C(u_1, \ldots, u_d)$ is a multivariate CDF on $[0,1]^d$ with $U(0,1)$ marginals.

Sklar's Theorem:

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

for some copula $C(u_1, \ldots, u_d)$.



The bivariate Plackett copula

The bivariate **Plackett** copula describes dependence between **binary** variables U and V.

Let $u = \mathrm{P}(U = 1)$, $v = \mathrm{P}(V = 1)$ and $\phi = \dfrac{\mathrm{P}(U = 0, V = 0)\,\mathrm{P}(U = 1, V = 1)}{\mathrm{P}(U = 1, V = 0)\,\mathrm{P}(U = 0, V = 1)}$.

The bivariate Plackett copula

The bivariate **Plackett** copula describes dependence between **binary** variables U and V.

Let $u = \mathrm{P}(U = 1)$, $v = \mathrm{P}(V = 1)$ and $\phi = \dfrac{\mathrm{P}(U = 0, V = 0)\,\mathrm{P}(U = 1, V = 1)}{\mathrm{P}(U = 1, V = 0)\,\mathrm{P}(U = 0, V = 1)}$.

$$
\begin{aligned}
P_{11} &= C(u, v|\phi) \\
P_{10} &= u - C(u, v|\phi) \\
P_{01} &= v - C(u, v|\phi) \\
P_{00} &= 1 - u - v + C(u, v|\phi)
\end{aligned}
$$

The bivariate Plackett copula

The bivariate **Plackett** copula describes dependence between **binary** variables U and V.

Let $u = \mathrm{P}(U = 1)$, $v = \mathrm{P}(V = 1)$ and $\phi = \dfrac{\mathrm{P}(U = 0, V = 0)\,\mathrm{P}(U = 1, V = 1)}{\mathrm{P}(U = 1, V = 0)\,\mathrm{P}(U = 0, V = 1)}$.

$$
\begin{aligned}
P_{11} &= C(u, v|\phi) \\
P_{10} &= u - C(u, v|\phi) \\
P_{01} &= v - C(u, v|\phi) \\
P_{00} &= 1 - u - v + C(u, v|\phi)
\end{aligned}
$$

$$
C(u, v|\phi) = \begin{cases} \dfrac{1 + (\phi - 1)(u + v) - \sqrt{[1 + (\phi - 1)(u + v)]^2 - 4\phi(\phi - 1)uv}}{2(\phi - 1)} & \phi \neq 1 \\ uv & \phi = 1. \end{cases}
$$

The trivariate Plackett copula

Introduce a third variable W with $w = \mathrm{P}(W = 1)$.

Let $\psi = \dfrac{P_{111}P_{100}P_{010}P_{001}}{P_{000}P_{011}P_{101}P_{110}}$.

The trivariate Plackett copula

Introduce a third variable W with $w = \mathrm{P}(W = 1)$.

Let $\psi = \dfrac{P_{111}P_{100}P_{010}P_{001}}{P_{000}P_{011}P_{101}P_{110}}$.

$$
\begin{aligned}
P_{111} &= C(u, v, w|\psi) \\
P_{110} &= C(u, v|\phi_{UV}) - C(u, v, w|\psi) \\
P_{101} &= C(u, w|\phi_{UW}) - C(u, v, w|\psi) \\
P_{011} &= C(v, w|\phi_{VW}) - C(u, v, w|\psi) \\
P_{100} &= u - C(u, v|\phi_{UV}) - C(u, w|\phi_{UW}) + C(u, v, w|\psi) \\
P_{010} &= v - C(u, v|\phi_{UV}) - C(u, w|\phi_{UW}) + C(u, v, w|\psi) \\
P_{001} &= w - C(u, w|\phi_{UW}) - C(u, w|\phi_{UW}) + C(u, v, w|\psi) \\
P_{000} &= 1 - u - v - w + C(u, v|\phi_{UV}) + C(u, w|\phi_{UW}) + C(u, w|\phi_{UW}) - C(u, v, w|\psi)
\end{aligned}
$$

Can find $C(u, v, w|\psi)$ numerically by solving the equation for ψ.

A Bayesian model for coinfection

Survey s has count data $n_{000}^{(s)}, \ldots, n_{111}^{(s)}$ on the number of individuals in each state.

Let $\boldsymbol{\theta}_s = (u_s, v_s, w_s, \phi_{UV}, \phi_{UW}, \phi_{VW}, \psi)$.

A Bayesian model for coinfection

Survey s has count data $n_{000}^{(s)}, \ldots, n_{111}^{(s)}$ on the number of individuals in each state.

Let $\boldsymbol{\theta}_s = (u_s, v_s, w_s, \phi_{UV}, \phi_{UW}, \phi_{VW}, \psi)$.

Multinomial likelihood: $L(\boldsymbol{\theta}) = \prod_s P_{000}(\boldsymbol{\theta}_s)^{n_{000}^{(s)}} \ldots P_{111}(\boldsymbol{\theta}_s)^{n_{111}^{(s)}}$.

A Bayesian model for coinfection

Survey s has count data $n_{000}^{(s)}, \ldots, n_{111}^{(s)}$ on the number of individuals in each state.

Let $\boldsymbol{\theta}_s = (u_s, v_s, w_s, \phi_{UV}, \phi_{UW}, \phi_{VW}, \psi)$.

Multinomial likelihood: $L(\boldsymbol{\theta}) = \prod_s P_{000}(\boldsymbol{\theta}_s)^{n_{000}^{(s)}} \ldots P_{111}(\boldsymbol{\theta}_s)^{n_{111}^{(s)}}$.

Priors:

$$u_s, v_s, w_s \sim \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$$

$$\phi_{UV}, \phi_{UW}, \phi_{VW}, \psi \sim F_{3,3}$$

A Bayesian model for coinfection

Survey s has count data $n_{000}^{(s)}, \ldots, n_{111}^{(s)}$ on the number of individuals in each state.

Let $\boldsymbol{\theta}_s = (u_s, v_s, w_s, \phi_{UV}, \phi_{UW}, \phi_{VW}, \psi)$.

Multinomial likelihood: $L(\boldsymbol{\theta}) = \prod_s P_{000}(\boldsymbol{\theta}_s)^{n_{000}^{(s)}} \ldots P_{111}(\boldsymbol{\theta}_s)^{n_{111}^{(s)}}$.

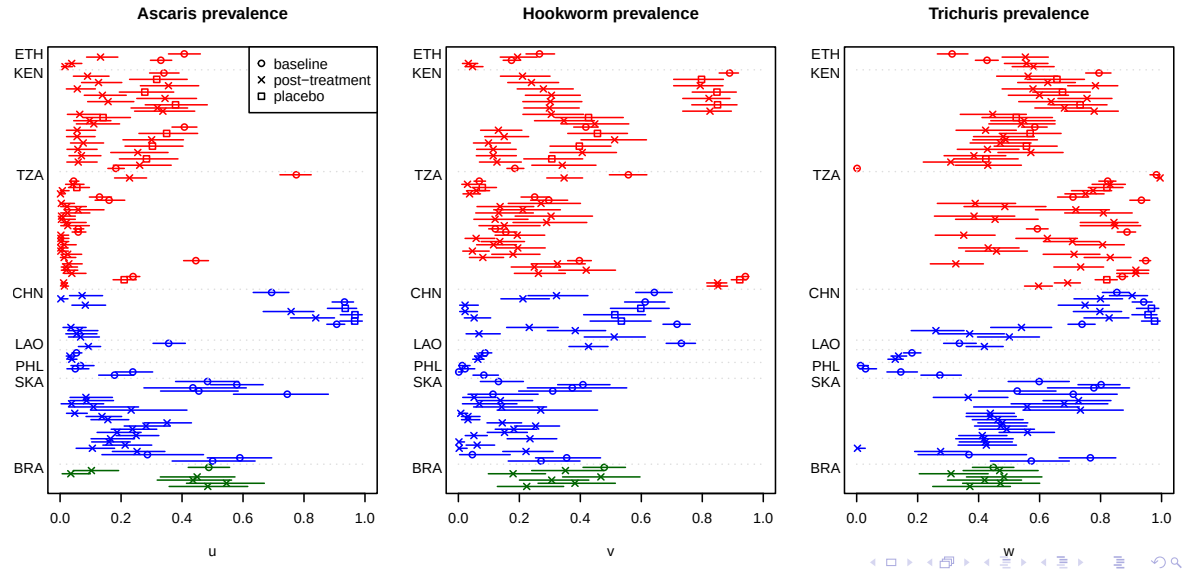
Priors:

$$u_s, v_s, w_s \sim \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right) \quad \rightarrow \quad \text{Independence sampler}$$

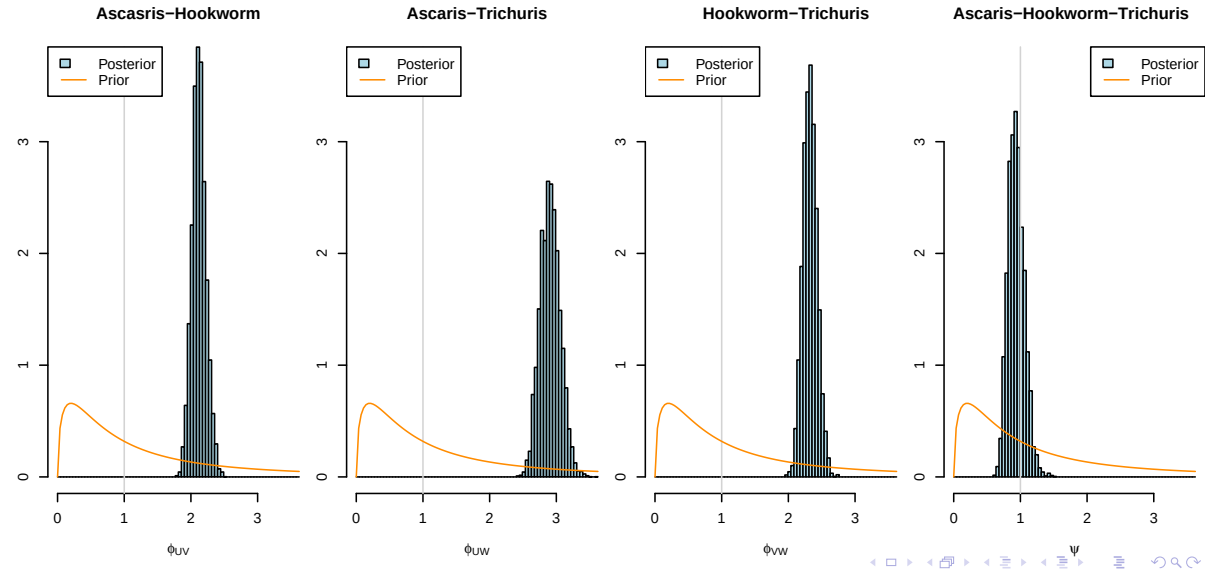
$$\phi_{UV}, \phi_{UW}, \phi_{VW}, \psi \sim F_{3,3} \quad \rightarrow \quad \text{Log-normal random walk}$$

Results

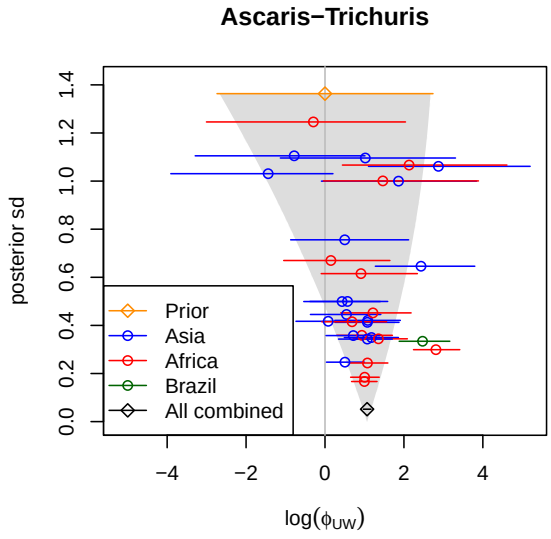
Prevalences

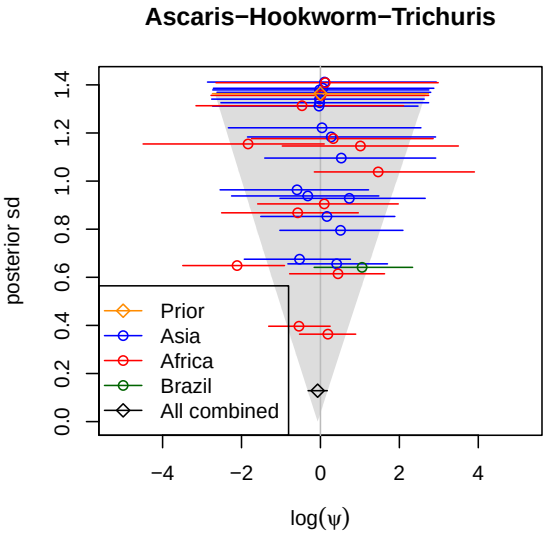


Posteriors of the copula parameters



Bayesian funnel plots





Why is this useful?

STH data is often summarised with just the **marginal prevalences** being retained.

Why is this useful?

STH data is often summarised with just the **marginal prevalences** being retained.

The WHO 2030 goal for STH is described in terms of the proportion of people with **any STH**:

$$\mathrm{P}(\text{any STH}) = 1 - P_{000}.$$

Why is this useful?

STH data is often summarised with just the **marginal prevalences** being retained.

The WHO 2030 goal for STH is described in terms of the proportion of people with **any STH**:

$$\mathrm{P}(\text{any STH}) = 1 - P_{000}.$$

The Plackett copula model allows us to evaluate the **any STH** goal from just **marginal prevalence data**.

Conclusion

Conclusion

- Coinfection with multiple STH species is common and leads to additional morbidity and suffering.
- The Bayesian Plackett copula model describes the levels of coinfection present in survey individuals, including quantifying uncertainty.

Can evaluate progress towards the 2030 goal for STH from the available marginal prevalence data.