

Aberystwyth University

Habitat Hydrology and Geomorphology Control the Distribution of Malaria Vector Larvae in Rural Africa

Hardy, Andrew J.; Gamarra, Javier G. P.; Cross, Dónall E.; Macklin, Mark G.; Smith, Mark W.; Kihonda, Japhet; Killeen, Gerry F.; Ling'ala, George N.; Thomas, Christopher J.

Published in:
PLoS One

DOI:
[10.1371/journal.pone.0081931](https://doi.org/10.1371/journal.pone.0081931)

Publication date:
2013

Citation for published version (APA):

Hardy, A. J., Gamarra, J. G. P., Cross, D. E., Macklin, M. G., Smith, M. W., Kihonda, J., Killeen, G. F., Ling'ala, G. N., & Thomas, C. J. (2013). Habitat Hydrology and Geomorphology Control the Distribution of Malaria Vector Larvae in Rural Africa. *PLoS One*, 8(12), [e81931]. <https://doi.org/10.1371/journal.pone.0081931>

Document License CC BY

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

Supplementary Methods S1

“Habitat hydrology and geomorphology control the distribution of malaria vector larvae in rural Africa”

Andrew J Hardy^{1,2}, Javier G P Gamarra³, Dónall E Cross^{2,3}, Mark G Macklin¹, Mark W Smith⁴, Japhet Kihonda², Gerry F Killeen^{2,5}, George N Ling’ala², Chris J Thomas^{3*}.

1 Institute of Geography & Earth Sciences, Aberystwyth University, Aberystwyth, UK.

2 Biomedical and Environmental Sciences Thematic Group, Ifakara Health Institute, Ifakara, Tanzania.

3 Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK.

4 School of Geography, University of Leeds, Leeds, UK.

5 Vector Biology Department, Liverpool School of Tropical Medicine, Liverpool, UK.

*Corresponding author. Tel: +44 1970 621 987, email: cjt@aber.ac.uk

Statistical analysis of entomological data

Computation of the number of Anopheles arabiensis larvae

Given the highly heterogeneous nature of the data collected, a semiparametric approach that provides robust estimates in the presence of unknown or misspecified covariance structures. In this case we chose Generalized Estimating Equations (GEEs) as opposed to General Linear Mixed Models.

All GEEs were clustered by site, and corrected with sandwich estimators for the standard errors. Quasi-likelihood estimators are produced via iterative algorithms. Two independent GEE models were computed:

1. A GEE modelling the total number of anophelines per dip taken in the field. The family distribution of the larvae was assumed to be Poisson with a log link.
2. A GEE modelling the proportion of PCR samples (a subsample from the total anophelines in the field) that correspond to *An. arabiensis*. The family distribution of the proportion was assumed to be binomial with a logit link.

The product of the two GEE estimators (number of anopheline larvae and fraction of them that corresponds to *An. arabiensis*) is the estimated number of *An. arabiensis* taken in the field. To compute it, a mixture Poisson-binomial distribution was produced with the means and confidence intervals obtained in the previous GEEs. For example, assume two random variables: anopheline larvae L and the fraction of those that are *An. arabiensis* (F) following Poisson (λ) and binomial (p) distributions respectively. The expectation for the number of *An. arabiensis* in the sample is:

$$E_{arab} = E(E(F|L)) = E(pL) = p\lambda$$

Via optimization procedures we calculated the number of events necessary to simulate both the Poisson and binomial distributions with the same confidence intervals as those resulting from the GEE. After stochastic generation, the distributions were mixed.

Next, we bootstrapped the resulting mixture 100 times to calculate the first moment and confidence intervals of the mixture distribution, providing the estimate number of *An. arabiensis* per dip and the corresponding confidence interval.

Contrasts via Method of Variance Estimate Recovery (MOVER)

Once obtained the robust estimators and their confidence intervals, overlaps of the differences between estimators were obtained by pairwise contrasts. The MOVER method [1,2,3] provides a methodology to calculate upper and lower overlaps between confidence intervals and their significance. Assuming 95% two-sided confidence intervals for estimates, the lower and upper limits D_l and D_u for the difference between two log-transformed parameter estimates β_1 and β_2 are:

$$D_l = (\beta_1 + \beta_2) - z_{0.05/2} \sqrt{\text{var}(\beta_1) + \text{var}(\beta_2)}$$

$$D_u = (\beta_1 + \beta_2) + z_{0.05/2} \sqrt{\text{var}(\beta_1) + \text{var}(\beta_2)}$$

where $z_{0.05/2}$ is the upper 95% quantile of a standard normal distribution.

References

1. Zou GY (2008) On the estimation of additive interaction by use of the four-by-two table and beyond. *The American Journal of Epidemiology* 168: 212-224.
2. Zou G, Donner A (2008) Construction of confidence limits about effect measures: a general approach. *Statistics in Medicine* 27: 1693-1702.
3. Kowalik D, Choi Y-H, Zou G (2011) Confidence interval estimation for a difference between two dependent intraclass correlation coefficients with variable class sizes. *Journal of Statistical Theory and Practice* 5: 613-625.