Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: De Salazar PM, Cox H, ImhoffH, Alexandre JSF, Buckee CO. The association between gold mining and malaria in Guyana: a statistical inference and time-series analysis. Lancet Planet Health 2021; 5: e731–38.
Supplementary material for “Quantifying the association between gold mining and Plasmodium falciparum malaria in Guyana: a statistical inference analysis”

Dr. med. Pablo M De Salazar 1#, Horace Cox, MD 2#, Helen Imhoff, MD 2, Jean S F Alexandre, PhD 3, Caroline O Buckee, PhD 1

1. Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, US.
2. Vector Control Services, Ministry of Public Health, Georgetown, Guyana.
3. Pan American Health Organization, Georgetown, Guyana.

# These authors contributed equally

Section S1. Adjustment of gold-price time series

Since 1919, the most common benchmark for gold price has been determined twice daily by the London Market Bullion Association up to 2015 (http://www.lbma.org.uk/lbma-gold-price/); since then, it is controlled by a third party, the ICE Benchmark Administration. We adjusted the gold price time series obtained from [1] by extracting the deterministic trend [2], computed using the last 30 years of gold price, to adjust for expected increases in the cost of mining exploration and operationalization.

Figure S1. Adjusted and unadjusted gold price time series

![Figure S1](image)

Figure S1. Grey dots showing monthly average gold price raw, and orange dots after adjusting to the deterministic trend.

Further, we assessed the assumption that a simple deterministic trend could approximate the adjustment to changes in mining operational cost. We obtained data on fuel cost in Guyana (available from World Bank, www.worldbank.org) for the period of analysis, which showed an overall increasing tendency (Figure S2 A) consistent with the deterministic trend adjustment. In addition, we show in panel B that the Consumer Price Index (available from www.statisticsguyana.gov.gy/subjects/price-indices/, an averaged proxy of goods and services costs, also exhibited this similar tendency.
Figure S2. Pump fuel price and Consumer Price Index in Guyana

Figure S2. Panel A showing gasoline (pump fuel) average annual price for Guyana between 2002-2019. Available data is shown as circles and inferred data as rhombus. Panel B showing Consumer Price Index computed for Guyana. Grey areas in both panels represent the years when malaria cases and gold prices exhibited a significant increase trend.
Section S2. Gold-price and deforestation
Available data on annual deforestation due to gold-mining activity 2007-2019, as percentage of forest loss, obtained from [3].

Figure S3. Time series showing weekly average gold price in the left y-axis ($/ounce, orange line) and annual deforestation due to gold mining in the right y-axis (Ha, green line) between 2006-2019. Estimated mean deforestation for 2007-2009 (diamond) was available as a single value for the 3-years period. For facilitating visual inspection, annual deforestation values are plotted by June 1st of each year for 2010-2019 and by January 1st, 2008 for the period 2007-2009.
Section S3. Correlation between gold price and *P. vivax* infections

We performed the similar correlation analysis as in the main methods to infections due to *P. vivax*. In the overall, similar patterns are observed regarding *P. vivax* and gold prices (Figure S7A), while the estimated correlation coefficient was smaller and the shift longer (by 1 month) than when analyzing *P. falciparum* infections. Again, the shift which provides higher correlation is longer in the community vs. miners (3 months vs 2 months) and the correlation coefficient is higher for miners: 0.78 (p<0.001, 0.73-0.81, 95%CI) vs 0.65 (p<0.001, 0.55-0.73, 95% CI), Figure S7C. No correlation was found between gold prices and cases in non-mining regions (Figure S7B).

Figure S4. Stratified correlations between *P. vivax* malaria and gold price time series in Guyana, 2007-2019.

Figure S4 A) *P. vivax* cases in mining regions (Regions 1,7,8 and 10) disaggregated by 1) males 15-50y, representing the mining population (light orange area) and 2) females, children <12y and adults>70y, representing the non-mining population (grey) B) Cases time series in non-mining regions (Regions 2,5,6 and 9) disaggregated by males 15-50y (green) and 2) females, children <12y and elder >70y (grey) C) Scatter plot showing monthly cases among males 15-50y (light orange dots), and among females, children <12y and adults >70y (grey dots) in mining regions versus monthly adjusted gold price with 2 and 4-months shift respectively.
Section S4. Malaria reported cases and El Niño-Southern Oscillation

Previously, Gagnon et al. [4] provided evidence on the association between years exhibiting La Niña events and increased malaria transmission in Guyana, as well as with a 1-year delay since La Niña. In order to understand if climate forcing might also have contributed to the 2008-2014 malaria increase in Guyana, we computed the correlation between the Southern Oscillation Index (SOI) and malaria case counts among males 15-50y as well as females, children <12y and adults >70y. The SOI, designed to measure the strength and phase of the Southern Oscillation [5], was obtained from http://www.bom.gov.au/climate/enso/soi/. We performed similar analysis by disaggregating by regions with a high number of mining camps and those with none or low numbers, and by computing the cross-correlation function coefficients.

Figure S5. Stratified correlations between *P. falciparum* malaria and SOI in Guyana.

**Figure S5.** A) Heat map representing cross correlation coefficients computed using gold price time series shifted +/-24 months and 1) monthly cases among males 15-50y in mining regions, 2) females, children <12y and adults >70y in mining regions and 3) all cases in non-mining regions; B) Time series of malaria cases in mining regions among males 15-50y (green), and the SOI with a 13 months shift; C) Time series of malaria cases in mining regions among females, children <12y and adults >70y (brown), and SOI with a 2 months shift; D) Time series of all malaria cases in non-mining regions (grey) disaggregated by males 15-50y (green area) and the Southern Oscillation Index without shift; E) Scatter plot showing monthly cases among males 15-50y (green dots) and the Southern Oscillation Index without shift; F) females, children <12y and adults >70y (brown dots) in mining regions versus monthly SOI values shifted 13 and 2 months respectively.
Section S5. Spectral analysis

We evaluated if the spectrum of the malaria time series changed over time and how this related to the frequency domain of the gold price, under the hypothesis that external forcing of transmission due to gold mining could yield significant changes in the spectrum of the malaria time series. First, we computed the wavelet spectrum (spectrogram) of the monthly malaria time series [6]. Briefly, we used the Morlet wavelet in the \textit{Rwave}-package to identify significant signals [7]. Further, we computed the periodograms of the gold price after adjusting to the stochastic trend [8,9], and the SOI time series and compared them to the one computed from the malaria time series.

The spectral analysis of the malaria cases among males 15-50 in mining regions showed a significant high amplitude between 2009-2013 with a period of around 1.5 years, and a secondary period of around 0.5 years between 2010-2012 (Figure S4A). Consistently, the periodogram of the malaria time series (adjusted for the stochastic trend) for the years 2009-2013 showed the 1.5 and 0.5 period (Figure 4B). Further, the periodogram of the gold price times series (adjusted for the stochastic trend) for the years 2009-2013 showed a significant peak with period around 1.5 years and two secondary peaks with periods 0.75 and 0.5 years (Figure S4C). The SOI time series periodogram for the years 2009-2013 showed a significant peak with period 3 years and a secondary peak of period around 0.75 years (Figure S4D). This analysis suggests that between 2009-2013 a significant periodic oscillatory behavior of period 1.5 years occurred among malaria cases among males 15-50y in mining regions, coincident with the years with the highest gold price. During those years an oscillatory dynamic of the gold price with a similar period of 1.5 years was also observed. Consistently, the normalized residuals of both the malaria and the gold price time series adjusted for the stochastic trend showed very similar trajectories with a significant cross correlation maximized with gold price residuals time series lagged by 2 months. (Figure S4E). The SOI time series showed oscillations of a period around 3 years, consistent with ENSO known periodicity [10].
Figure S6. Spectral analysis performed for *P. falciparum* malaria, gold price and SOI

A) Heat map representing wavelet spectrogram computed using monthly cases among males 15-50y in mining regions 2007-2019; B) Time series periodogram of monthly reported malaria cases among males 15-50y in mining regions during 2009-2013, after adjusting for stochastic trend. C) Time series periodogram of monthly gold price during 2009-2013, after adjusting for stochastic trend D) Time series periodogram of the SOI during 2009-2013 E) Time series of the normalized residuals of 1) monthly reported cases among males 15-50y in mining regions (grey), and 2) monthly gold price (orange), after adjusting for stochastic trend.
Section S6. Details on the forecasting regression model and sensitivity

Regression models have been used widely in epidemiology including communicable diseases; however, features attributable to infectious disease might raise concerns on the adequacy of these models, such as exposure lag times or small observation time units [11]. Nevertheless, the approach aims to underline the usefulness of the model outcome (i.e., predicting malaria burden and trends in the presence of potentially long reporting delays [12] which preclude timely public health interventions) recognizing potential misspecifications. We implemented a generalized linear model, under the assumption monthly malaria case count can be modeled as negative binomial distribution with mean parameter $\mu_t$ and dispersion parameter $\theta_t$ which depend on the natural logarithm of the monthly average gold price $X_m$ and the correlation coefficient $\beta$.

$$Y_t \sim \text{Negative Binomial} \ (\mu_t, \theta_t)$$

$$f(\mu_t) = \beta \ ln(X_t)$$

We used the natural logarithm as the link function of the regression. For each year between 2008-2019, we estimated $\mu_t$ and $\theta$ using Maximum Likelihood Estimates (MLE) from the previous years and computed 95% and 50% prediction intervals (PI) of all falciparum cases nationwide by month for the target year, using a bootstrapping procedure [13]. We test the performance of the model by using an out-of-sample strategy aimed to address the challenges of long reporting delays occurring in Guyana [12]. For each year, and beginning in January, we visually compared the trend of the mean expected number of infections by month ($n=12$) vs. the true observed infections, and assessed if the 12 months observed time series lies within the computed prediction intervals.

Further, in order to test some of the key assumptions of the model, we evaluate the approach by A) predicting only cases in non-mining region B) using 2-months and 13-months lagged SOI, (following findings in section S4) as a predictor in the regression model together with gold price, and C) assuming truncated normal distribution as an alternative distribution of the expected malaria case counts [14].

The model was unable to predict trends in malaria cases in non-mining regions (Figure S5A). Case counts remained low in non-mining regions for the period of analysis, which likely explains why the model can predict total cases in Guyana independent of region of infection. Adding SOI as a predictor did not seem to improve forecasting, while adding extra uncertainty in particular when little information is available for the regression model (see predictions for 2008, Figure S5B). Finally, the truncated normal distribution was capable of predicting observed cases, with the advantage of reducing the right dispersion of uncertainty (see 2008-2014 upper bound predictions in Figure S5C).

While the left truncation of the assumed distribution allows for correcting the misspecification of assuming negative values when using a normal distribution, case counts such as those obtained in infectious disease surveillance are misrepresented by a continuous predicted variable model. Nevertheless, similar limitations as those observed in the main analysis were observed with this distribution. Overall, this alternative model might be a good alternative to that used in the main analysis.
Figure S6. Sensitivity for the forecasting model of *P. falciparum* infections in Guyana

Figure S7. Prediction intervals for expected cases estimated by fitting a linear model to 1-months shifted monthly average gold price with 3 different approaches. A) Predictions of *P. falciparum* cases in only non-mining regions B) Predictions using 2-months and 13-months lagged SOI value as predictor value together with gold price C) Assuming a truncated normal distribution of expected cases. Ribbons represent 25-75 and 2.75-97.5 percentiles of predicted cases. Grey lines represent *P. falciparum* reported cases.
Section S7. Bed net distribution

Table S1. Number of bed nets (insecticide-treated nets, ITNs) distributed per year in Guyana.

| Year | 2007  | 2008  | 2009  | 2010  | 2011  | 2012  | 2013  | 2014  | 2015  | 2016  | 2017  | 2018  | 2019  |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Total ITNs distributed | 2784  | 4287  | 1068  | 11430 | 14550 | 16800 | 27921 | 552996 | 24201 | 8320  | 5534  | 43181 | 1759  |
Supplementary references

1. Website. [cited 6 May 2021]. Available: https://www.quandl.com/api/v3/datasets/WGC/GOLD_DAILY_USD.json?api_key=8rzMrxYvNJCxwA2cX2UU

2. Faticchi S, Barbosa SM, Caporali E, Silva ME. Deterministic versus stochastic trends: Detection and challenges. Journal of Geophysical Research. 2009. doi:10.1029/2009jd011960

3. Guyana Forestry Commission. Guyana REDD+ Monitoring Reporting & Verification System (MRVS). Year 7 Summary Report. Available: https://forestry.gov.gy/wp-content/uploads/2018/11/MRVS-Summary-Report-Year-7_November-2018_Final.pdf

4. Gagnon AS, Smoyer-Tomic KE, Bush ABG. The El Niño southern oscillation and malaria epidemics in South America. Int J Biometeorol. 2002;46: 81–89.

5. Hanf M, Adenis A, Nacher M, Carme B. The role of El Niño southern oscillation (ENSO) on variations of monthly Plasmodium falciparum malaria cases at the Cayenne general hospital, 1996–2009, French Guiana. Malaria Journal. 2011. doi:10.1186/1475-2875-10-100

6. Torrence C, Compo GP. A Practical Guide to Wavelet Analysis. Bulletin of the American Meteorological Society. 1998. pp. 61–78. doi:10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2

7. Bjørnstad ON. Epidemics: Models and Data using R. Springer; 2018.

8. Brockwell PJ, Davis RA. Introduction to Time Series and Forecasting. Springer Science & Business Media; 2013.

9. Retrospective Detection and Estimation of Stochastic Trends. Change-Point Analysis in Nonstationary Stochastic Models. 2016. pp. 55–72. doi:10.1201/9781315367989-5

10. Lu B, Jin F-F, Ren H-L. A Coupled Dynamic Index for ENSO Periodicity. Journal of Climate. 2018. pp. 2361–2376. doi:10.1175/jcli-d-17-0466.1

11. Imai C, Hashizume M. A systematic review of methodology: time series regression analysis for environmental factors and infectious diseases. Trop Med Health. 2015;43: 1–9.

12. Menkir T, Cox H, Poirier C, Saul M, Jones-Weekes S, Clementson C, et al. Adjusting for reporting lags in remote regions to improve malaria surveillance: Guyana as a case study. doi:10.1101/2020.09.08.20178756

13. De Salazar PM, Niehus R, Taylor A, Buckee CO, Lipsitch M. Identifying Locations with Possible Undetected Imported Severe Acute Respiratory Syndrome Coronavirus 2 Cases by Using Importation Predictions. Emerg Infect Dis. 2020;26: 1465–1469.

14. Breen R, Fellow of the Royal Irish Academy Professor of Sociology Richard Breen. Regression Models: Censored, Sample Selected, Or Truncated Data. SAGE; 1996.