



Supplementary Information:

Quantifying epidemiological drivers of gambiense human African trypanosomiasis across the Democratic Republic of Congo

Ronald E Crump^{1,2,3}^{*}, Ching-I Huang^{1,2}^o, Ed Knock^{1,4}, Simon E F Spencer^{1,4}, Paul Brown^{1,2}, Erick Mwamba Miaka⁵, Shampa Chancy⁵, Matt J Keeling^{1,2,3}, and Kat S Rock^{1,2}

¹Zeeman Institute for System Biology and Infectious Disease Epidemiology Research, The University of Warwick, Coventry, U.K.


²Mathematics Institute, The University of Warwick, Coventry, U.K.

³The School of Life Sciences, The University of Warwick, Coventry, U.K.

⁴The Department of Statistics, The University of Warwick, Coventry, U.K.

⁵PNLTHA, Kinshasa, D.R.C.

June 23, 2020

 These authors contributed equally to this work.

* Corresponding author: r.e.crump@warwick.ac.uk

S1 Materials and Methods

S1.1 Data

HAT Atlas data The HAT Atlas data for DRC were provided in a spreadsheet format. Records were annually aggregated gHAT case records, aggregated by year, surveillance type and location as defined by multiple fields. There were 117,573 rows in this file; of which 111,408 had an entry in the geolocation (longitude and latitude) fields.

Passive surveillance records with missing or zero case numbers; and active surveillance records with both missing or zero numbers screened and missing or zero case numbers were dropped from the dataset.

This left 111,454 records (105,979 with filled geolocation fields). These records were associated with 23,424 unique combinations of former province, health zone, health area, location and territory identifiers; 20,423 of which had geolocation information and 3,001 did not. We will refer to these 24,424 geographical records as gHAT locations in this document.

Table S1: Number of HAT Atlas records with different combinations of former province, health zone and health area recorded.

Recorded region identifiers:			Number n
Former province	Health zone	Health area	
✓	✓	✓	106823
✓	✓	✓	391
✓	✓		3001
✓		✓	14
✓			1225

DRC Shapefile A recent shapefile for DRC was provided by UCLA (Personal communication). The shapefile contains health zones (an organisational unit with a typical population size around 100,000) across DRC; health areas (nested within health zones, these areas are typically home to around 10,000 people) for the

former province of Bandundu and part of Equateur and Haut Lomami, and post-2015 province identifiers. Former province was added to these records (post-2015 provinces being nested within former province).

Additional geographic information The following geographical information was obtained from the Humanitarian Data Exchange [S9]:

- a health zone shapefile from the United Nations Office for the Coordination of Humanitarian Affairs (OCHA);
- an OCHA file of geolocations of localities; and
- a file of geolocations of health facilities from the Global Healthsite Mapping Project.

These data were used to assist in matching and locating the gHAT data, by providing alternative spellings of names and potentially geolocations for non-geolocated gHAT locations. The locality and health facility lists were concatenated, and this enlarged locality set and the OCHA health zone map were assigned geographical identifiers as per our shapefile of choice.

Matching HAT Atlas records to DRC shapefile Geographical identifiers associated with the HAT Atlas/gHAT location and geographical data were sanitised to assist with matching. This involved removing diacritical marks, conversion to lowercase, collapsing whitespace within identifiers to a single space, removal of leading and trailing whitespace, converting from roman to arabic numerals, removing leading m, n or g from words where they were followed by a consonant, removing leading t from words when followed by an s, and collapsing words into a single string. In addition some specific manual edits were performed during the process as they became apparent.

Matching was then applied sequentially to the gHAT locations; such that once a match had been achieved for any given gHAT location it did not act as input to subsequent steps.

1. gHAT locations with known former province (FP), health zone (HZ), health area (HA) and geolocation were located on the UCLA and OCHA shapefiles. If the FP, HZ and HA matched the values for either the UCLA or OCHA shapefiles at that point, a match was judged to have occurred and the geolocation was accepted. This matched 3,579 of the gHAT locations.
2. gHAT locations with known former province (FP), health zone (HZ) and geolocation were located on the UCLA and OCHA shapefiles. If the FP and HZ matched the values for either the UCLA or OCHA shapefiles at that point, a match was judged to have occurred and the geolocation was accepted. This matched 13,413 of the gHAT locations. *The 16,992 gHAT locations matched in these two steps accounted for 97,520 of the HAT records (87.5%).*
3. Where a gHAT location was associated with a recent active screening event (defined as an active screening record in or after 2012 with a number screened greater than 10), the geolocation was accepted. *This matched a further 454 of the gHAT locations to give a total of 100,747 gHAT records matched (90.4%).*
4. Matching to the locality information:
 - (a) if health zone and location identifier match, the locality's geolocation was assigned to the gHAT location. *Matching 44 gHAT locations, giving a total of 100,811 gHAT records matched (90.5%).*
 - (b) if former province and location identifier match, the locality's geolocation was assigned to the gHAT location. *This matched 331 gHAT locations, giving a total of 101,508 gHAT records matched (91.1%).*
5. Matching to shapefile by geographic identifiers only:
 - (a) Former province, health zone and health area all match. *Matched 523 gHAT locations, giving a total of 102,349 gHAT records matched (91.8%).*
 - (b) Former province and health zone match. *Matched 3636 gHAT locations, giving a total of 110,021 gHAT records matched (98.7%).*
 - (c) Former province and OCHA shapefile health zone name match. *Matched 74 gHAT locations, giving a total of 110,126 gHAT records matched (98.8%).*

(d) Former province and health area match. Matched 60 gHAT locations, giving a total of 110,195 gHAT records matched (98.9%).

(e) Health zone name match. Matched 10 gHAT locations, giving a total of 110,210 gHAT records matched (98.9%).

S1.2 Modelling passive detection and its improvement

There are two sources of passive detection improvements considered in our model: a rapid improvement due to the introduction of the card agglutination test for trypanosomes (CATT) test in all health zones in 1998 and a gradual improvement over time in the former Bandundu and Bas Congo provinces around 2008 and mid-2015 respectively. Prior distributions and percentiles of parameters related to passive detection and its improvement over time are summarised in Table S2.

Improvement in passive surveillance systems over time is considered across the whole of Bandundu and Bas Congo. The province level staging data in Bandundu suggested that passive surveillance systems in Bandundu have improved over time, which was confirmed by PNLTHA and is also supported by previous modelling work [S2]. In Bas Congo, FIND implemented the use of rapid diagnostic tests (RDT) from 2015. Staging information which was available at the province-level from 2000–2012 from the paper of Lumbala *et al* [S4], and in the HAT Atlas data for 2015 and 2016. To inform the health zone-level analyses, a province level fit was carried out to the staged case data of Lumbala *et al*. augmented with the HAT Atlas data aggregated to the former province level for the years 2013–2016. These analyses provided no evidence for improvement in passive surveillance systems (in line with the simple sigmoidal model assumed) for any provinces other than Bandundu and Bas Congo. For Bandundu and Bas Congo, gamma distributions were fitted to the province level posterior samples of $\eta_{H_{amp}}$, $\gamma_{H_{amp}}$ and d_{steep} . The shape (k) and scale (θ) for these fitted distributions were used as the parameters of gamma prior distributions of $\eta_{H_{amp}}$, $\gamma_{H_{amp}}$ and d_{steep} in all health zones of Bandundu and Bas Congo. In Bandundu health zones a scaled and shifted beta distribution was used as the prior for d_{change} . The use of a broader prior for d_{change} than would have resulted from using the province-level posterior distribution resulted from comparing aggregate health zone-level results with province-level observed data. In Bas Congo health zones a fixed value of 2015.5 was used for d_{change} .

Priors for the health zone-level η_H^{post} and γ_H^{post} parameters were also informed by the province-level fits. Gamma prior distributions were used which had the same mode ($mode = (k - 1)\theta$) as Gamma distributions fitted to the province-level posterior distribution of the parameters; and a standard deviation ($s.d. = \sqrt{k}\theta$) of 3×10^{-3} for η_H^{post} and 1×10^{-4} for γ_H^{post} , being arbitrarily selected higher variation than the province-level posterior distributions.

S1.3 Modelling vector control

Fig. S1 shows the dynamics of tsetse populations (under the simple model) where targets are either moderately effective, with a 60% reduction (lower than in Guinea), or highly effective, with a 90% population density reduction in a year (as seen in Uganda).

The function which describes the probability of both hitting a target and dying is time dependent (days) from when the targets were placed:

$$f_T(t) = f_{max} \left(1 - \frac{1}{1 + \exp(-0.068(\text{mod}(t, 182.5) - 127.75))} \right) \quad (S1.1)$$

and f_{max} is chosen such that the tsetse population after one year is at the observed/assumed percentage reduction. For the simplified model this is given by $f_{max} = 0.0305$ for a 60% reduction and $f_{max} = 0.0750$ for a 90% reduction.

S2 Posteriors of fitted parameters

S2.1 Posterior characteristics for example health zones

Table S3 shows summaries of posterior distributions of all fitted parameters in two example health zones: Kwamouth in the former Bandundu province and Tandala in the former Equateur province. Kwamouth is categorised as a high-risk health zone and Tandala is at low-risk. Our fitting results show that Kwamouth has higher R_0 and r but lower active screening specificity, reporting rate and passive detection rate than Tandala.

Table S2: **Parameterisation of passive detection improvement.** Notation and brief description of fitted parameters related to passive detection improvement plus their within former province prior distributions and [2.5th, 50th & 97.5th] percentile.

Parameter		
Province	Prior distribution	Percentiles of prior distribution
η_H^{post} – Treatment rate from stage 1, 1998 onwards		
Bandundu	$\Gamma(3.54, 5.32 \times 10^{-5})$	$[4.59, 17.1, 42.9] \times 10^{-5}$
Bas Congo	$\Gamma(12.0, 2.89 \times 10^{-5})$	$[1.78, 3.36, 5.68] \times 10^{-4}$
Equateur	$\Gamma(4.92, 4.51 \times 10^{-5})$	$[7.12, 20.7, 45.7] \times 10^{-5}$
Kasai Occidental	$\Gamma(10.9, 3.03 \times 10^{-5})$	$[1.64, 3.20, 5.53] \times 10^{-4}$
Kasai Oriental	$\Gamma(2.90, 5.87 \times 10^{-5})$	$[3.38, 15.1, 41.5] \times 10^{-5}$
Katanga	$\Gamma(1.29, 8.79 \times 10^{-5})$	$[5.88, 86.2, 376] \times 10^{-6}$
Kinshasa	$\Gamma(1.26, 8.91 \times 10^{-5})$	$[5.44, 84.4, 376] \times 10^{-6}$
Maniema	$\Gamma(4.25, 4.85 \times 10^{-5})$	$[5.90, 19.0, 44.3] \times 10^{-5}$
Orientale	$\Gamma(1.16, 9.27 \times 10^{-5})$	$[4.24, 79.0, 373] \times 10^{-6}$
γ_H^{post} – Treatment rate from stage 2, 1998 onwards		
Bandundu	$\Gamma(2.45, 1.92 \times 10^{-3})$	$[7.59, 40.7, 121] \times 10^{-4}$
Bas Congo	$\Gamma(1.48, 2.47 \times 10^{-3})$	$[2.54, 28.7, 114] \times 10^{-4}$
Equateur	$\Gamma(1.95, 2.15 \times 10^{-3})$	$[4.88, 35.0, 118] \times 10^{-4}$
Kasai Occidental	$\Gamma(1.71, 2.29 \times 10^{-3})$	$[3.65, 31.9, 116] \times 10^{-4}$
Kasai Oriental	$\Gamma(1.49, 2.46 \times 10^{-3})$	$[2.60, 28.8, 114] \times 10^{-4}$
Katanga	$\Gamma(1.53, 2.43 \times 10^{-3})$	$[2.78, 29.4, 115] \times 10^{-4}$
Kinshasa	$\Gamma(1.68, 2.31 \times 10^{-3})$	$[3.50, 31.5, 116] \times 10^{-4}$
Maniema	$\Gamma(2.60, 1.86 \times 10^{-3})$	$[8.45, 42.3, 122] \times 10^{-4}$
Orientale	$\Gamma(2.54, 1.88 \times 10^{-3})$	$[8.11, 41.7, 122] \times 10^{-4}$
$\eta_{H_{\text{amp}}}$ – Relative improvement in passive stage 1 detection rate		
Bandundu	$\Gamma(2.01, 1.05)$	$[0.258, 1.77, 5.87]$
Bas Congo	$\Gamma(5.23, 1.70)$	$[2.98, 8.33, 18.0]$
$\gamma_{H_{\text{amp}}}$ – Relative improvement in passive stage 2 detection rate		
Bandundu	$\Gamma(1.001, 5)$	$[0.127, 3.47, 18.5]$
Bas Congo	$\Gamma(1.46, 1.26)$	$[0.126, 1.45, 5.81]$
d_{steep} – Speed of improvement in passive detection rate		
Bandundu	$\Gamma(39.6, 2.70 \times 10^{-2})$	$[0.761, 1.06, 1.42]$
Bas Congo	$\Gamma(3.21, 1.45)$	$[1.03, 4.18, 10.9]$

These factors provide some possible explanation for why gHAT is more persistent in Kwamouth, as well as having more cases reported every year. 109 110

S2.2 Posterior distribution maps 111

The posterior distribution maps (Figure S2–S15) illustrate both the level and variability of the parameter estimates between health zones. We partition the national or province level map into tessellated equilateral hexagons, or partial hexagons at national and health zone borders, and then fill each of these shapes with colour based on the value of randomly sampled values from the posterior distribution of the parameter being plotted from the analysis of the health zone in which that hexagon, or partial hexagon, lies. In this way, the 'overall' impression of the colour for a health zone gives an indication of the most probable value of the parameter, while the variation in colour reflects the uncertainty in the parameter estimates. 112 113 114 115 116 117 118

Table S3: **Posteriors of fitted parameters.** Notation, brief description, and [2.5th, 50th & 97.5th] percentile of posteriors for fitted parameters.

Notation	Description	Kwamouth	Tandala
R_0	Basic reproduction number (NGM approach)	[1.06, 1.09, 1.14]	[1.006, 1.009, 1.014]
r	Relative bites taken on high-risk humans	[3.15, 6.61, 10.75]	[1.30, 2.04, 4.26]
k_1	Proportion of low-risk people	[0.82, 0.90, 0.95]	[0.85, 0.95, 0.99]
η_H^{post}	Treatment rate from stage 1, 1998 onwards (days ⁻¹)	[0.60, 1.24, 2.74] $\times 10^{-4}$	[1.11, 2.74, 4.99] $\times 10^{-4}$
γ_H^{post}	Exit rate from stage 2 (treatment or death), 1998 onwards (days ⁻¹)	[0.46, 1.88, 5.42] $\times 10^{-3}$	[1.72, 3.60, 8.98] $\times 10^{-3}$
$b_{\gamma_H^{\text{pre}}}$	Relative exit rate from stage 2 factor, pre-1998	[0.74, 0.93, 1.00]	[0.54, 0.69, 0.94]
γ_H^{pre}	Exit rate from stage 2 (treatment or death), pre-1998 (days ⁻¹)	[0.38, 1.72, 4.88] $\times 10^{-3}$	[1.03, 2.53, 7.15] $\times 10^{-3}$
Spec	Active screening diagnostic specificity	[0.9987, 0.9991, 0.9997]	[0.9997, 0.9998, 0.9999]
u	Proportion of stage 2 passive cases reported	[0.18, 0.27, 0.40]	[0.29, 0.39, 0.51]
d_{change}	Midpoint year for passive improvement	[2004.4, 2005.8, 2007.3]	–
$\eta_{H_{\text{amp}}}$	Relative improvement in passive stage 1 detection rate	[0.92, 2.52, 5.46]	–
$\gamma_{H_{\text{amp}}}$	Relative improvement in passive stage 2 detection rate	[0.24, 0.51, 0.97]	–
d_{steep}	Speed of improvement in passive detection rate (years ⁻¹)	[0.68, 0.94, 1.29]	–

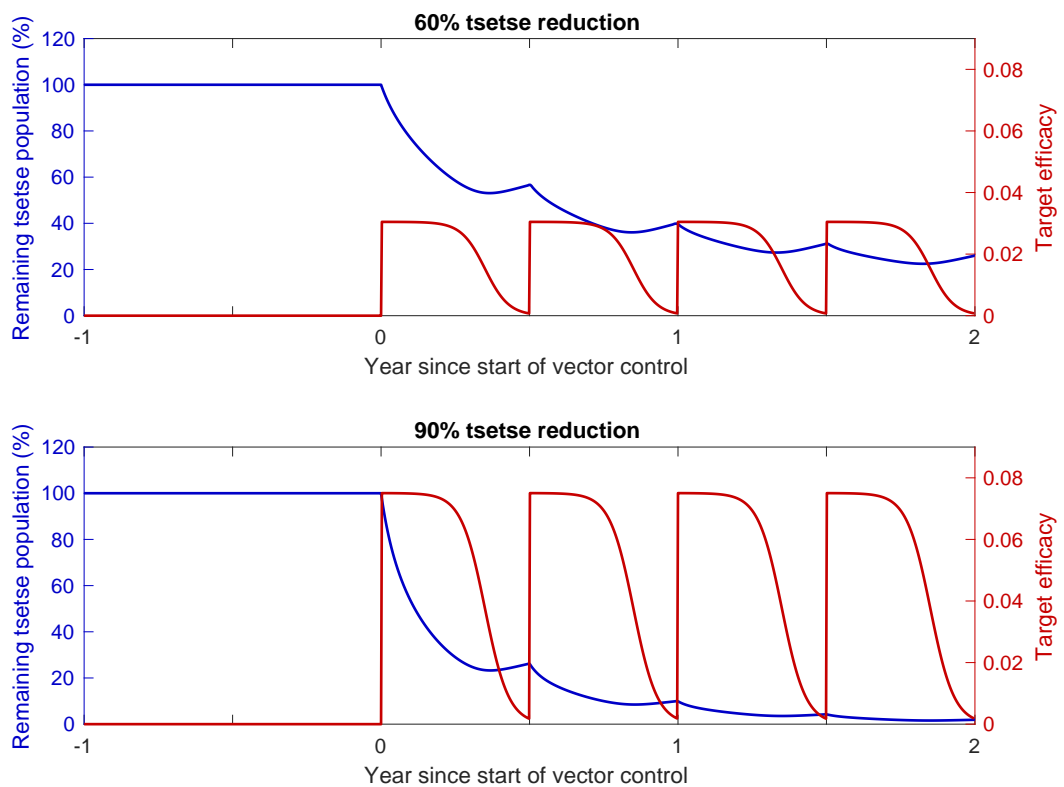


Figure S1: **Impact of tiny targets on tsetse density.** The figures show the how varying target efficacy (red line) impacts tsetse population density (blue line). Target efficacy is measured as the proportion of a host-seeking tsetse which will both hit the tiny target and die as a result. The graphs, reproduced from [S7], show the necessary efficacy of targets needed to reduce density by 60% (top) and 90% (bottom) by the end of the first year.

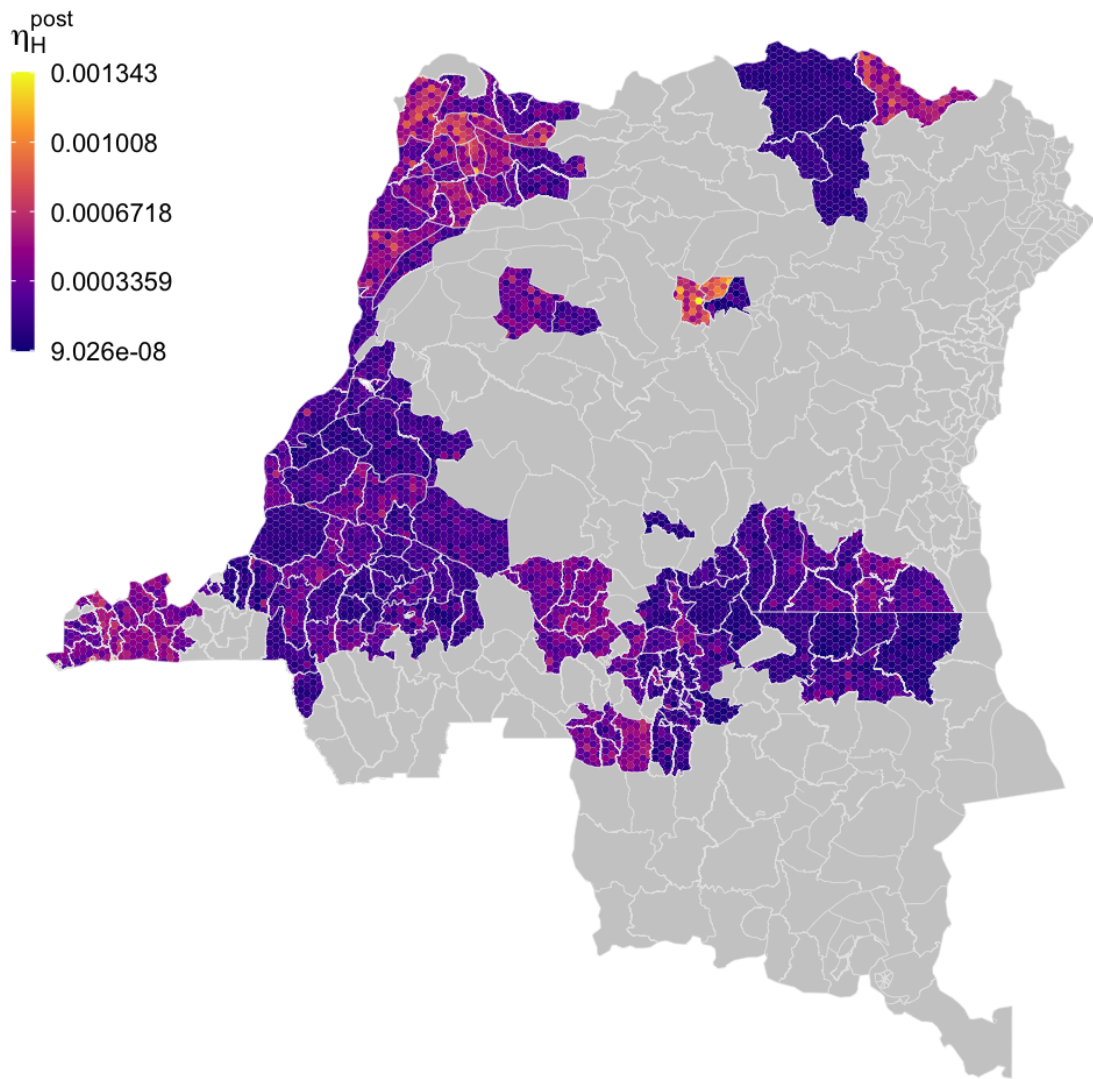


Figure S2: Within health zone posterior distribution of η_H^{post} , treatment rate from stage 1, 1998 onwards. fill colours are a randomly sampled from the posterior distribution for the health zone.

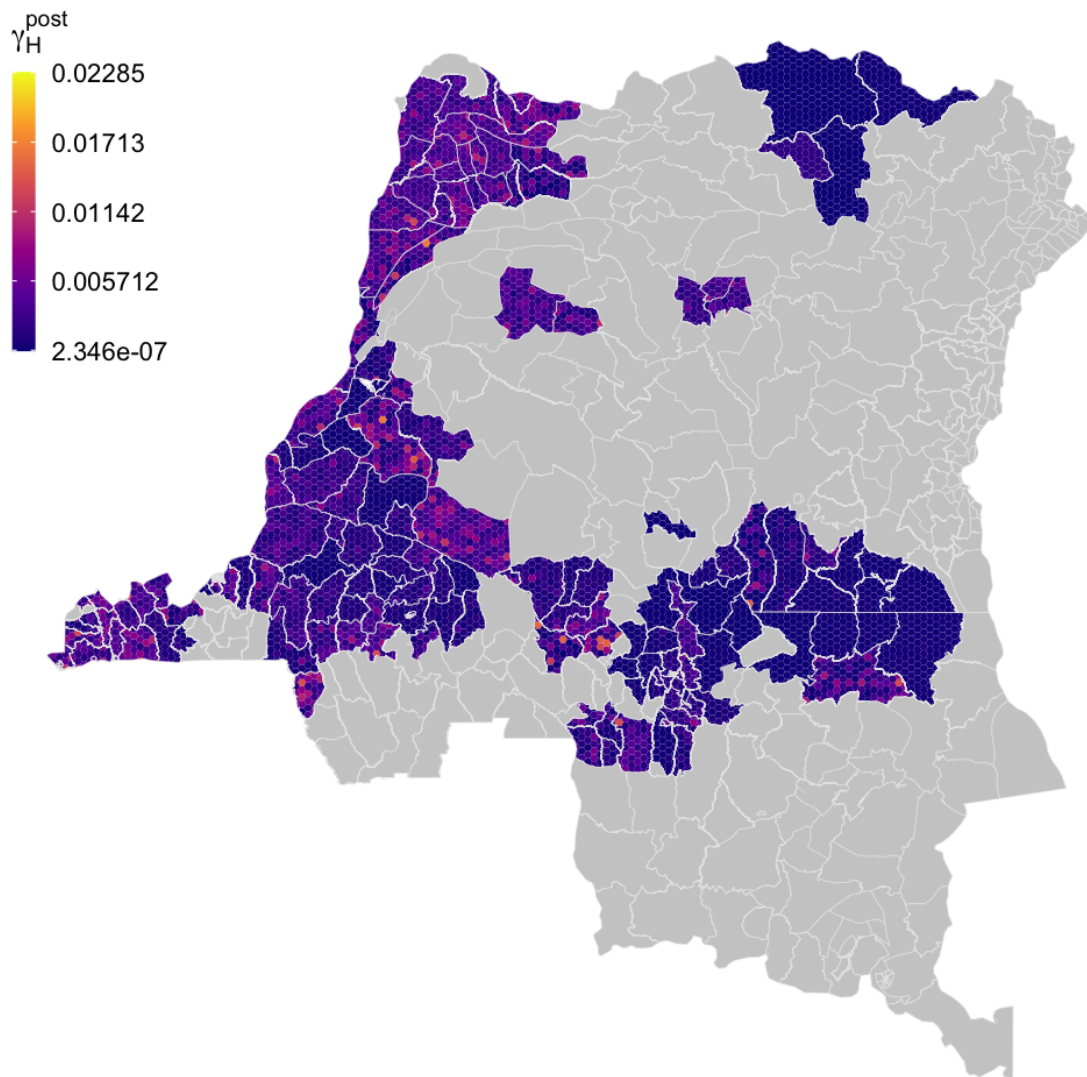


Figure S3: Within health zone posterior distribution of γ_H^{post} , the exit rate from stage 2 (treatment or death), 1998 onwards. The fill colours are a randomly sampled from the posterior distribution for the health zone.

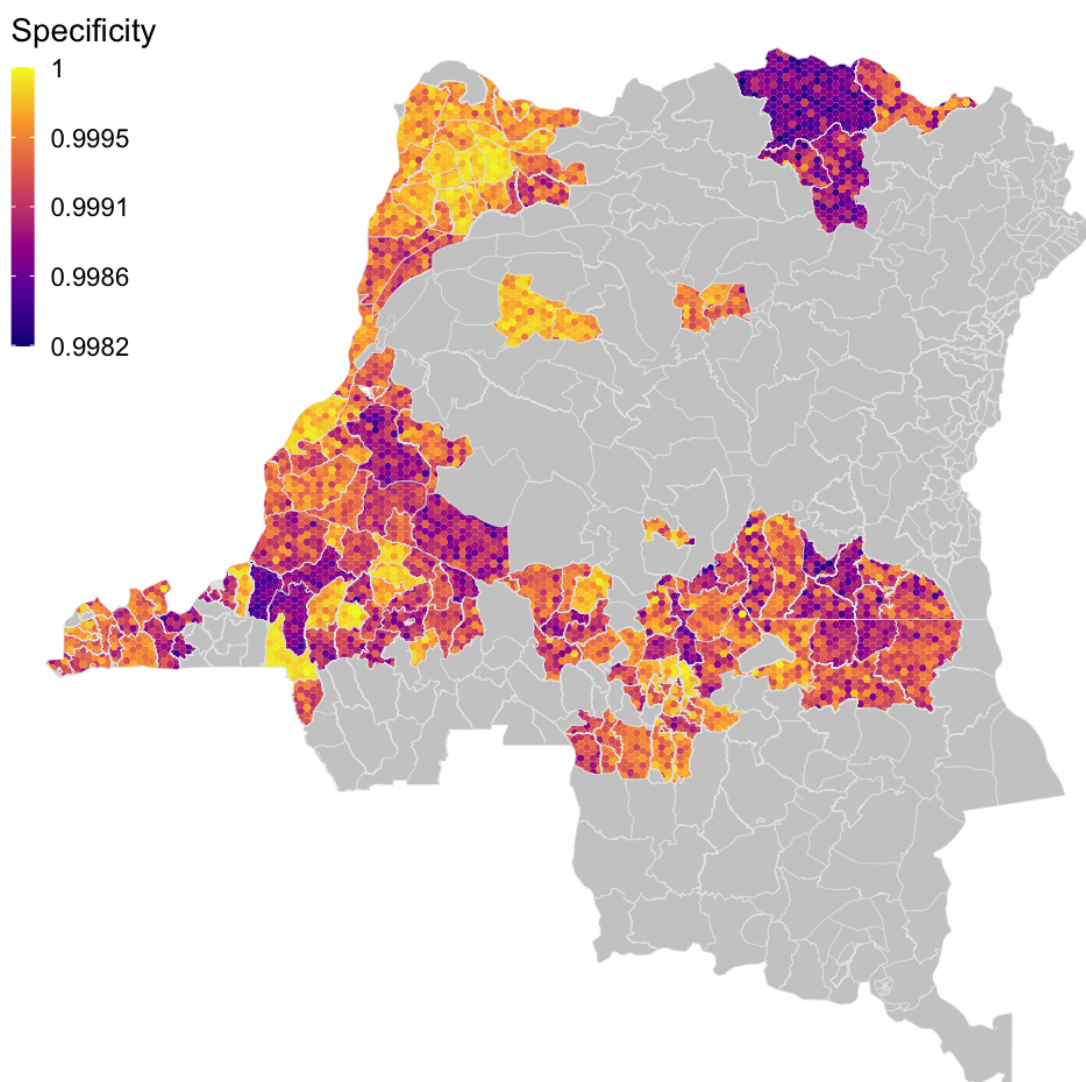


Figure S4: Within health zone posterior distribution of the specificity of the diagnostic algorithm, fill colours are a randomly sampled from the posterior distribution for the health zone.

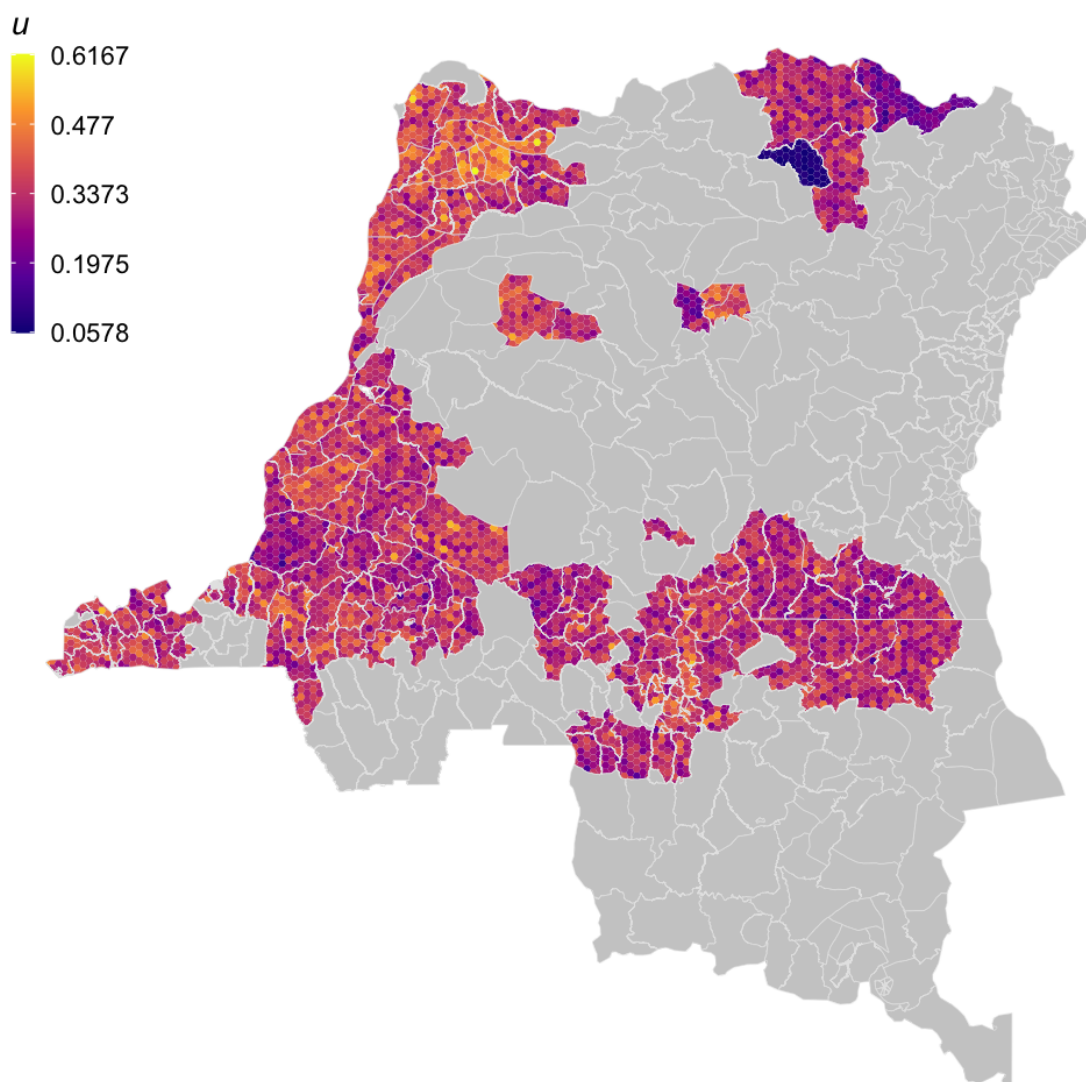


Figure S5: Within health zone posterior distribution of u , the proportion of stage 2 passive cases reported. The fill colours are a randomly sampled from the posterior distribution for the health zone.

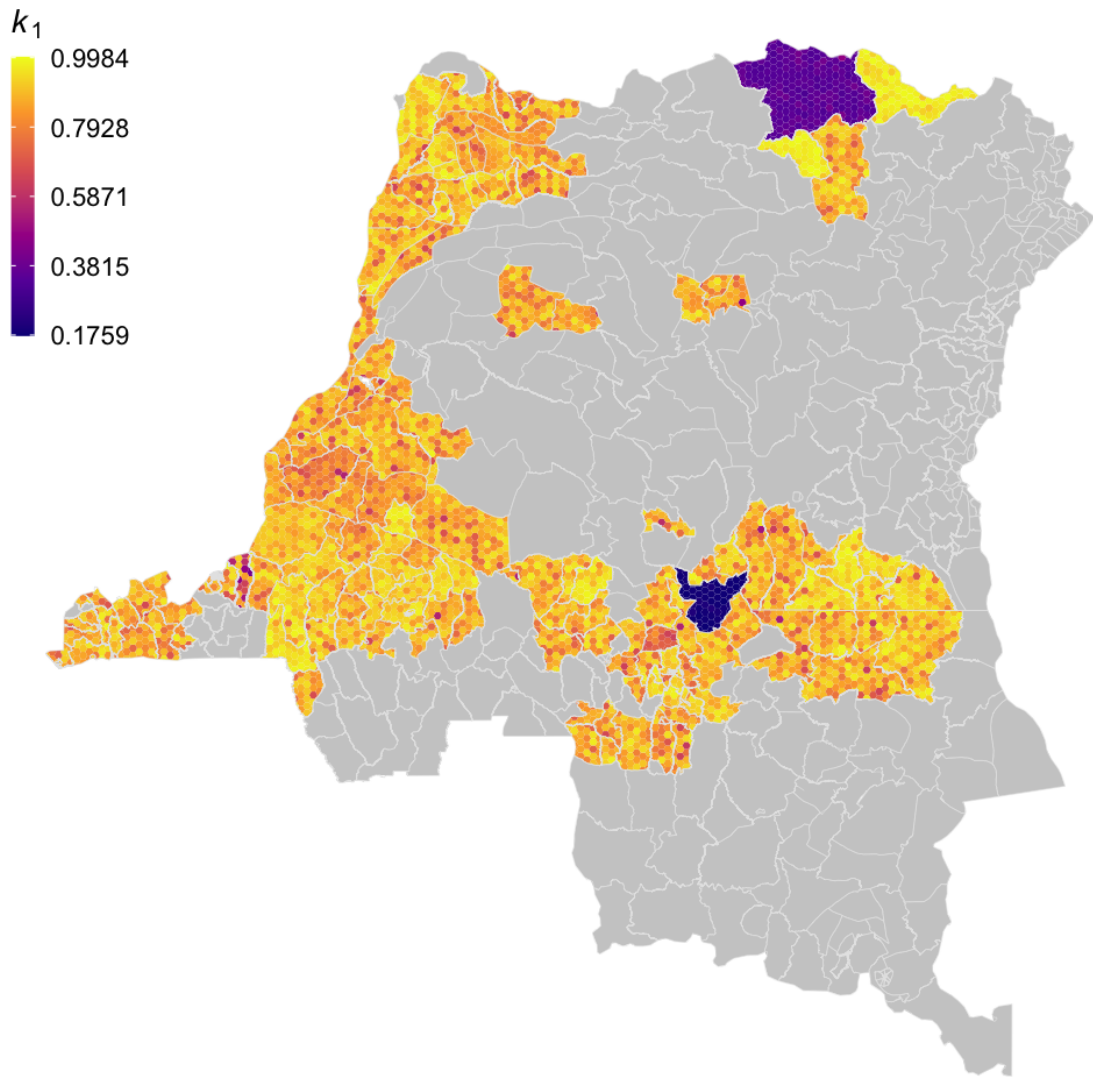


Figure S6: Within health zone posterior distribution of k_1 , the proportion of low-risk people. The fill colours are randomly sampled from the posterior distribution for the health zone.

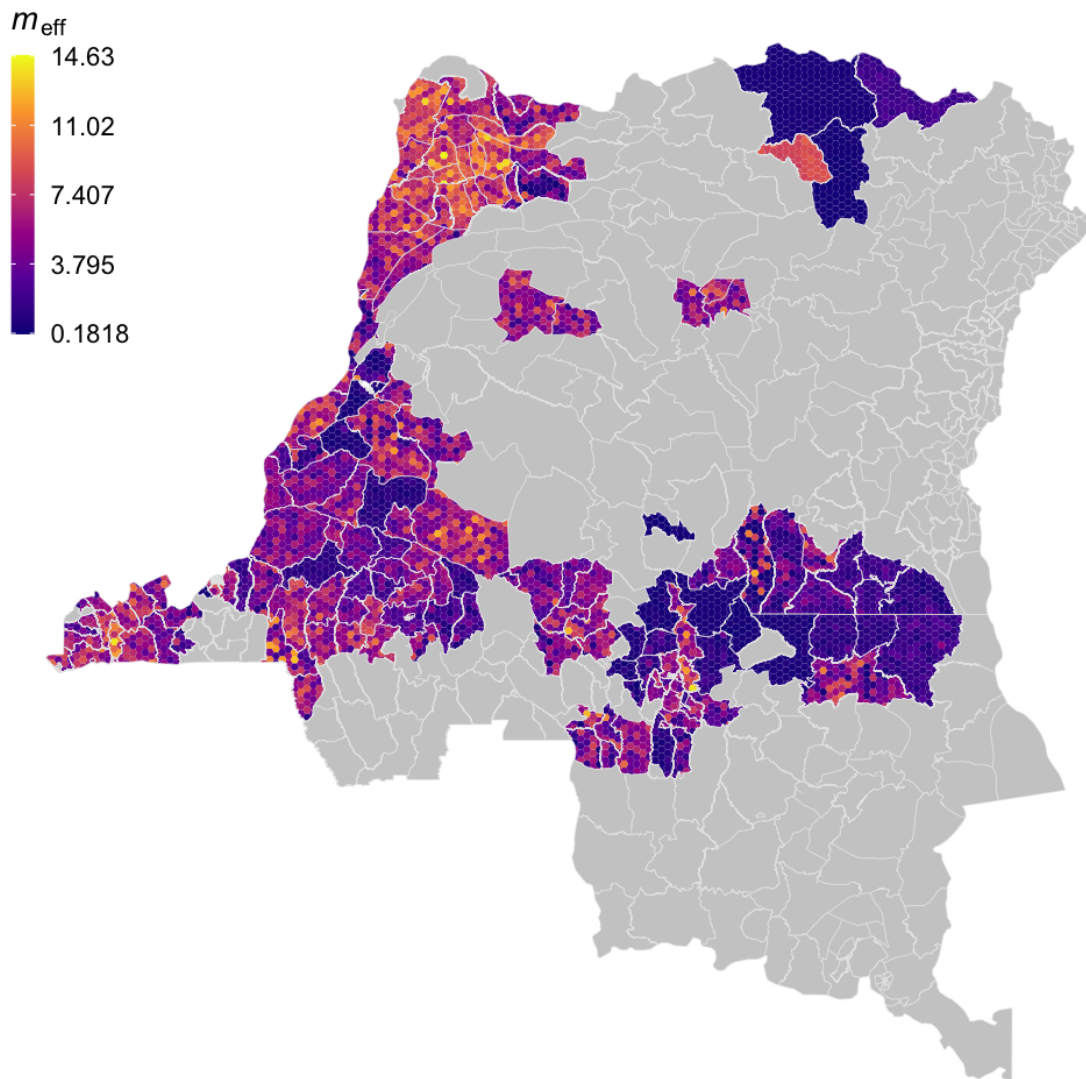


Figure S7: Within health zone posterior distribution of m_{eff} ; the effective, relative density of tsetse to humans. Fill colours are determined by randomly sampled values from the posterior distribution of m_{eff} from the analysis of the health zone.

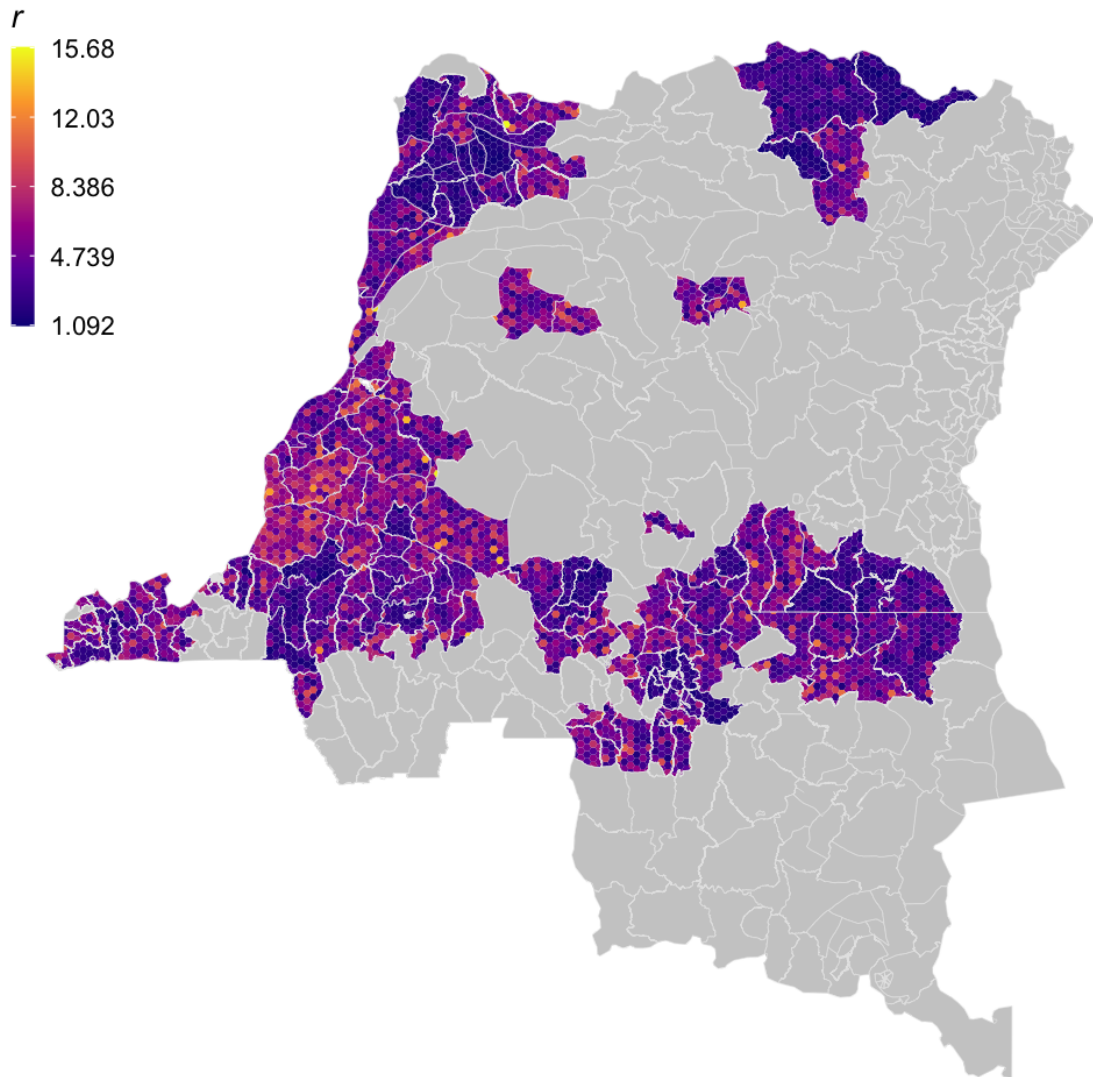


Figure S8: Within health zone posterior distribution of r , the relative bites taken on high-risk humans. Fill colours are determined by randomly sampled values from the posterior distribution of r from the analysis of the health zone.

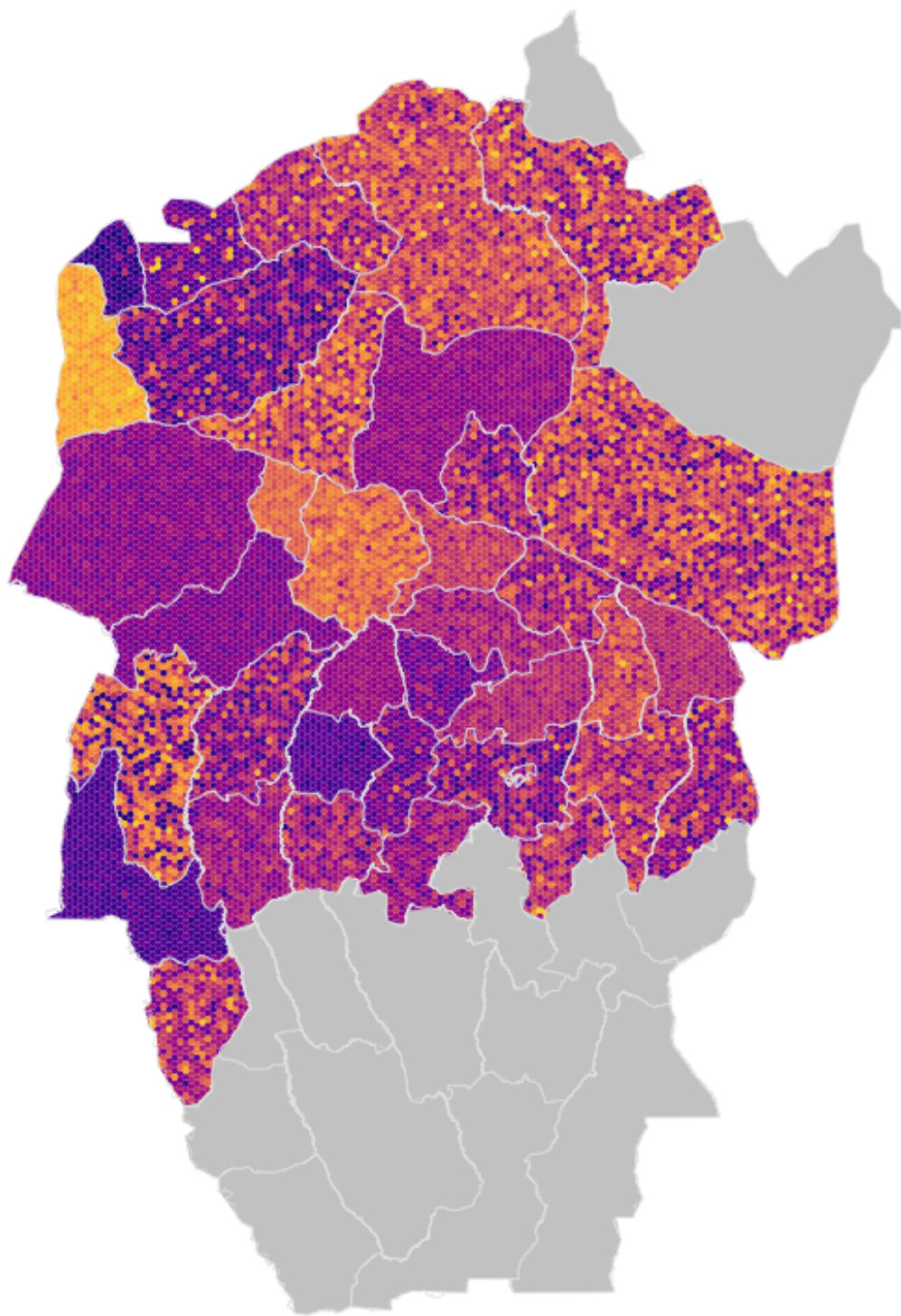


Figure S9: Within health zone posterior distribution of d_{change} , the midpoint year for passive improvement, for the former province of Bandundu. Fill colours are randomly sampled from the posterior distribution for the health zone.

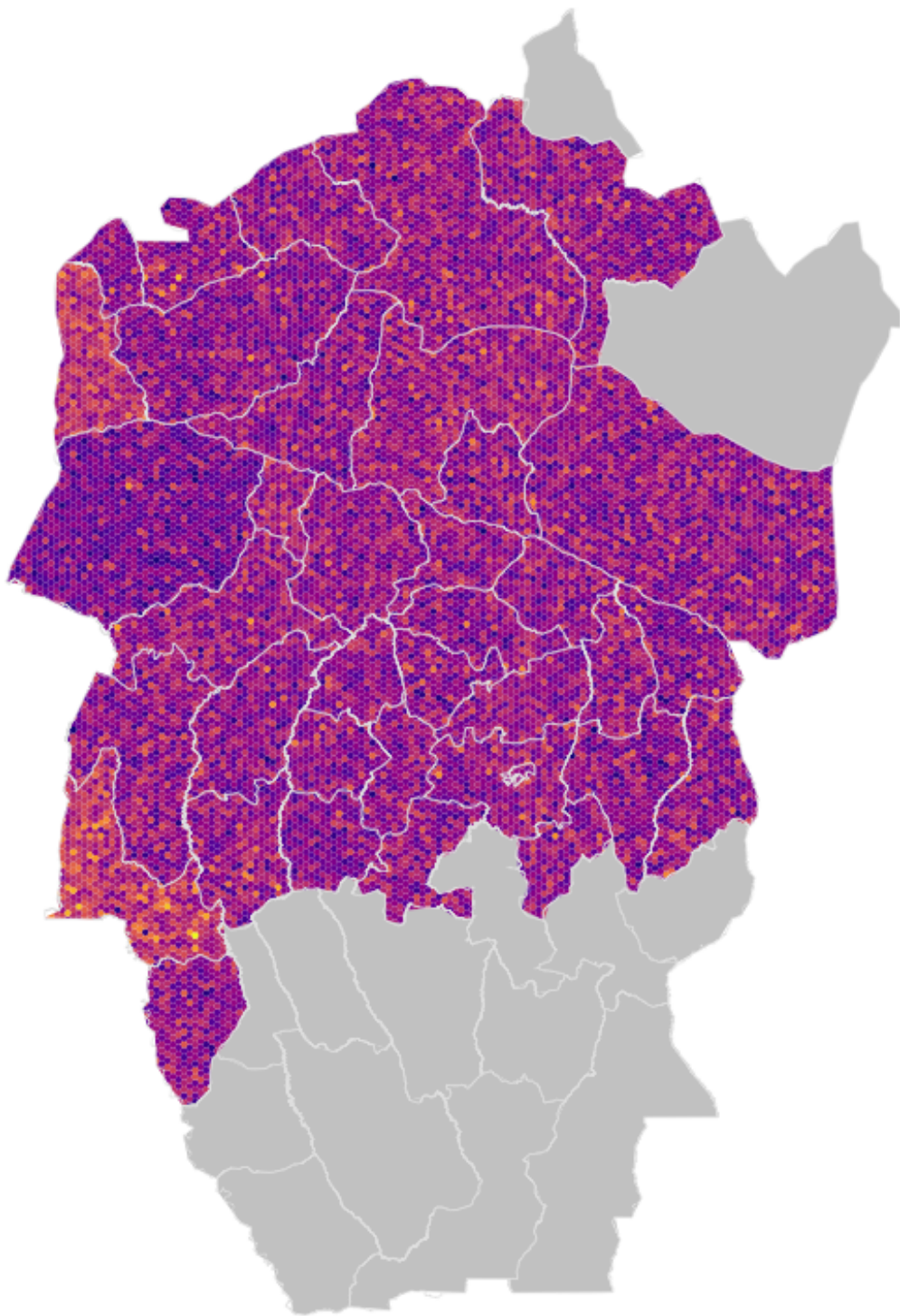


Figure S10: Within health zone posterior distribution of d_{steep} , the speed of improvement in passive detection rate, for the former province of Bandundu, fill colours are randomly sampled from the posterior distribution for the health zone.

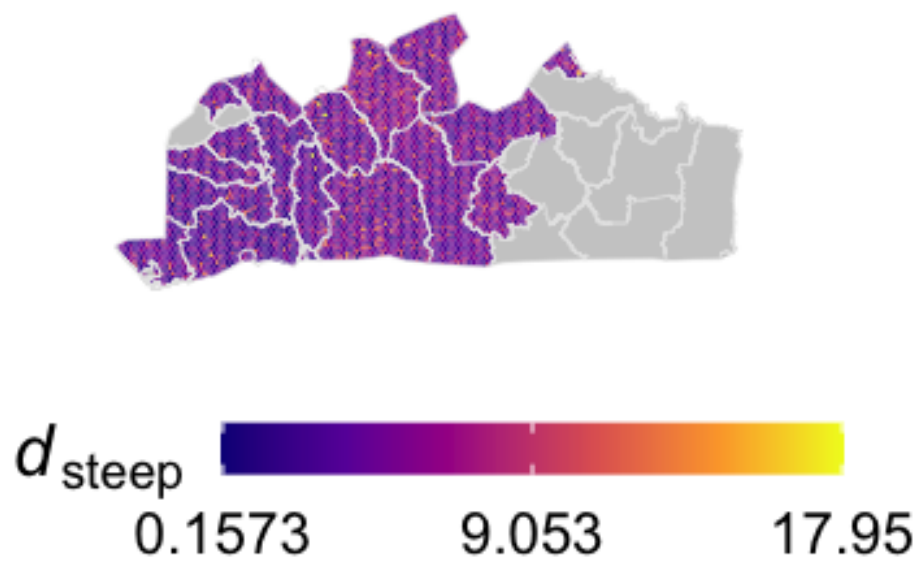


Figure S11: Within health zone posterior distribution of d_{steep} , the speed of improvement in passive detection rate, for the former province of Bas Congo, fill colours are randomly sampled from the posterior distribution for the health zone.

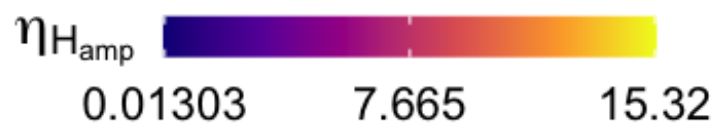
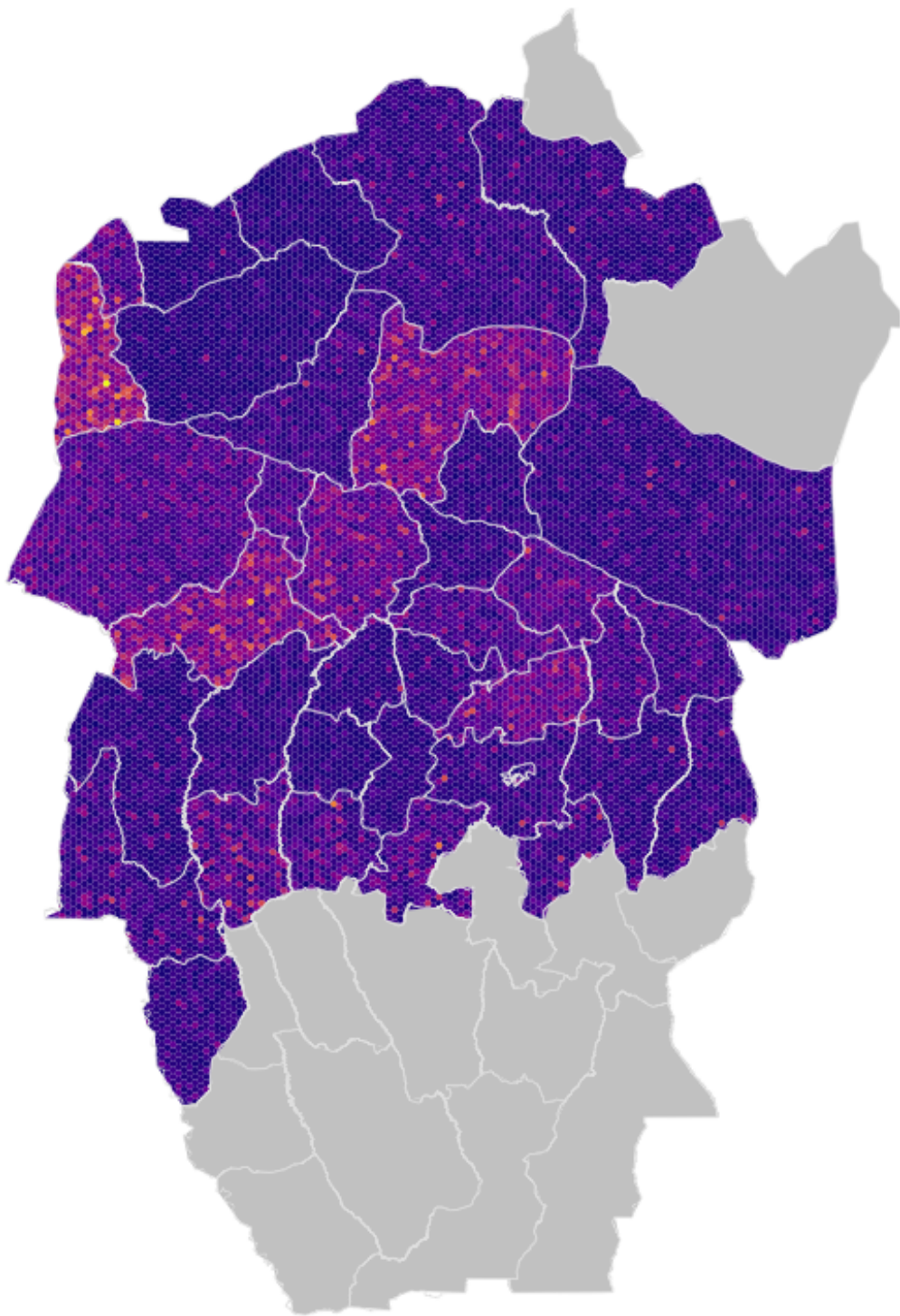


Figure S12: Within health zone posterior distribution of $\eta_{H_{amp}}$, the relative improvement in passive stage 1 detection rate, for the former province of Bandundu. Fill colours are randomly sampled from the posterior distribution for the health zone.

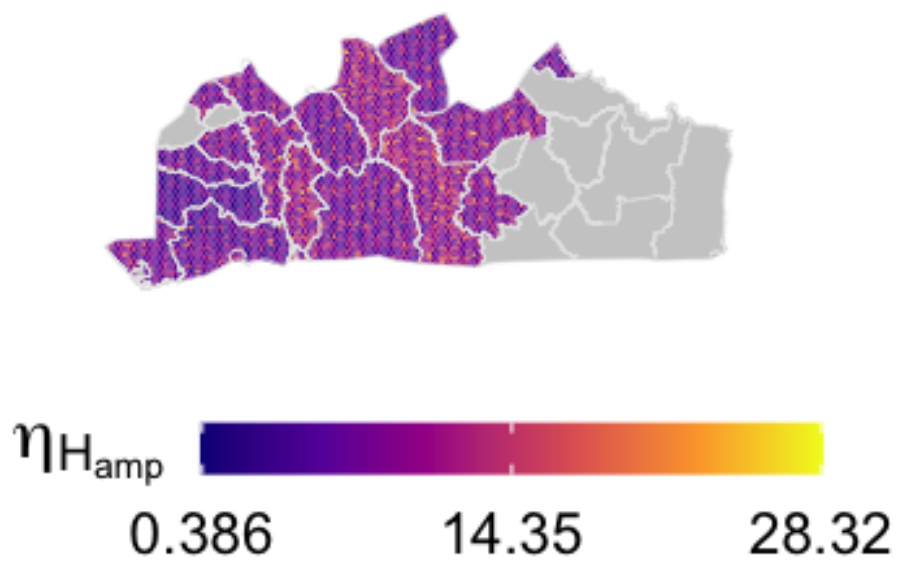


Figure S13: Within health zone posterior distribution of $\eta_{H_{amp}}$, the relative improvement in passive stage 1 detection rate, for the former province of Bas Congo, fill colours are randomly sampled from the posterior distribution for the health zone.

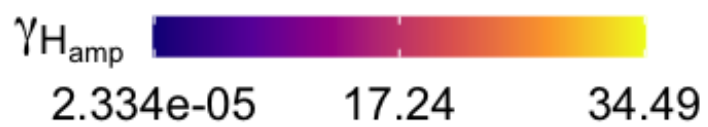
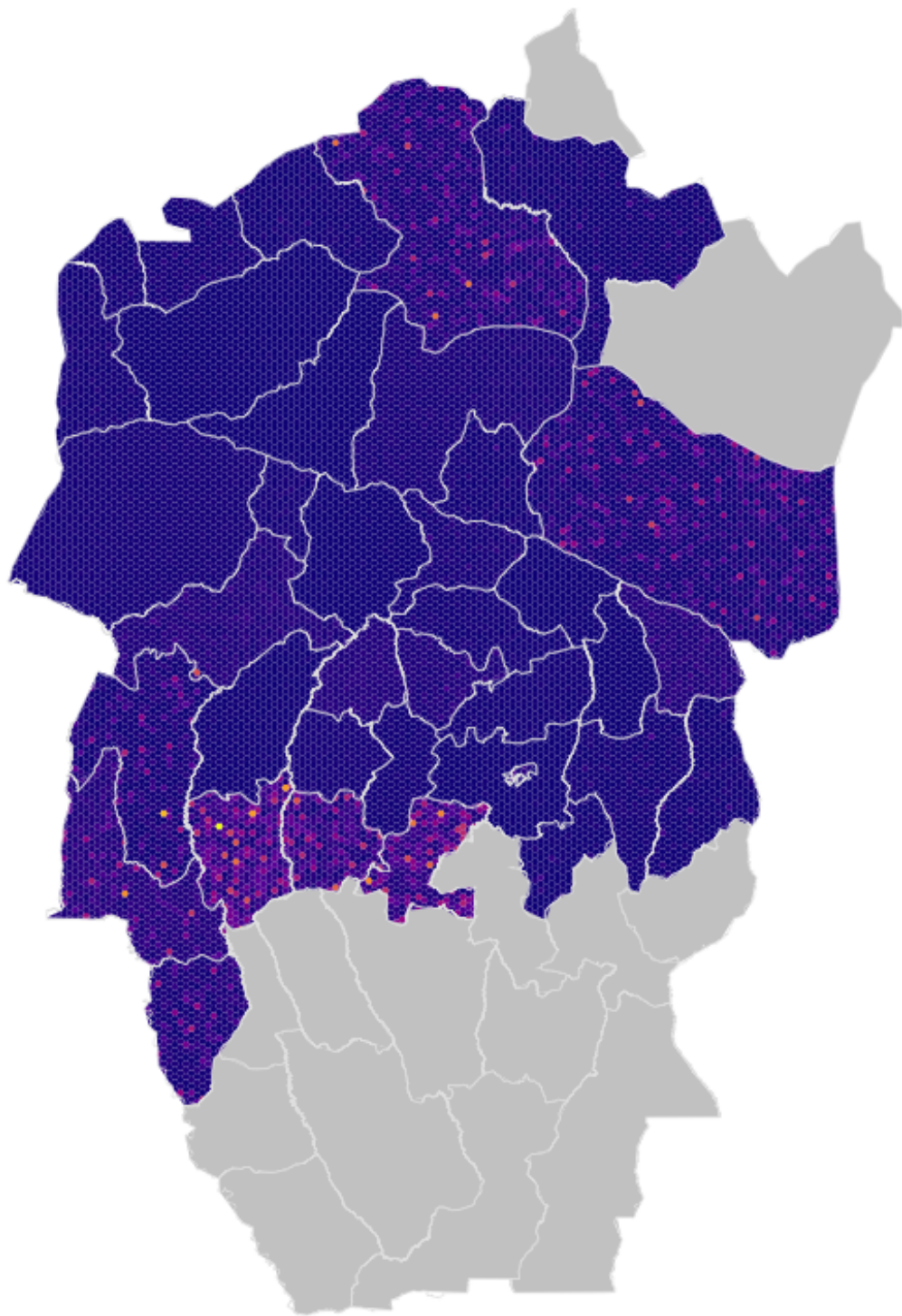


Figure S14: Within health zone posterior distribution of $\gamma_{H_{amp}}$, the relative improvement in passive stage 2 detection rate, for the former province of Bandundu. The fill colours are randomly sampled from the posterior distribution for the health zone.

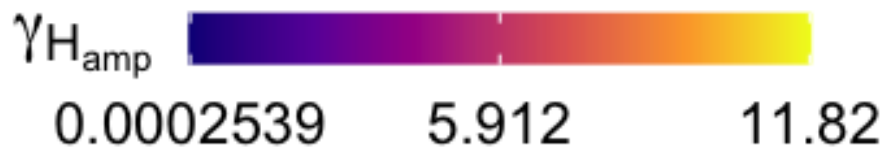
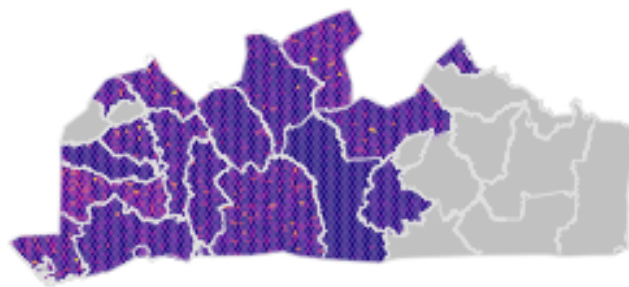


Figure S15: Within health zone posterior distribution of $\gamma_{H_{amp}}$, the relative improvement in passive stage 2 detection rate, for the former province of Bas Congo, fill colours are randomly sampled from the posterior distribution for the health zone.

S2.3 Joint posteriors for example health zones

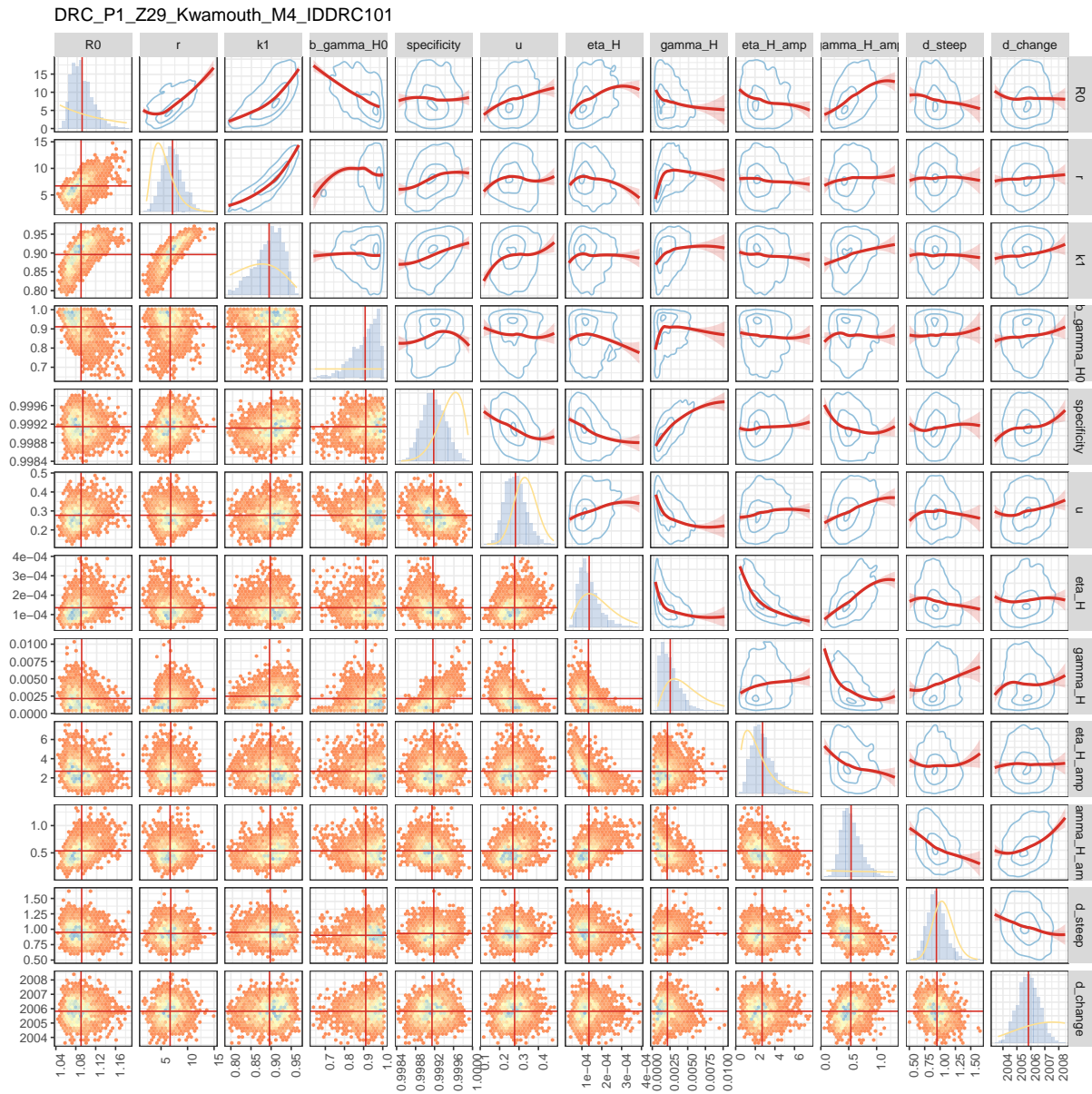


Figure S16: Joint posterior distributions for Kwamouth health zone in the former Bandundu province.

DRC_P3_Z59_Tandala_M4_IDDRRC101

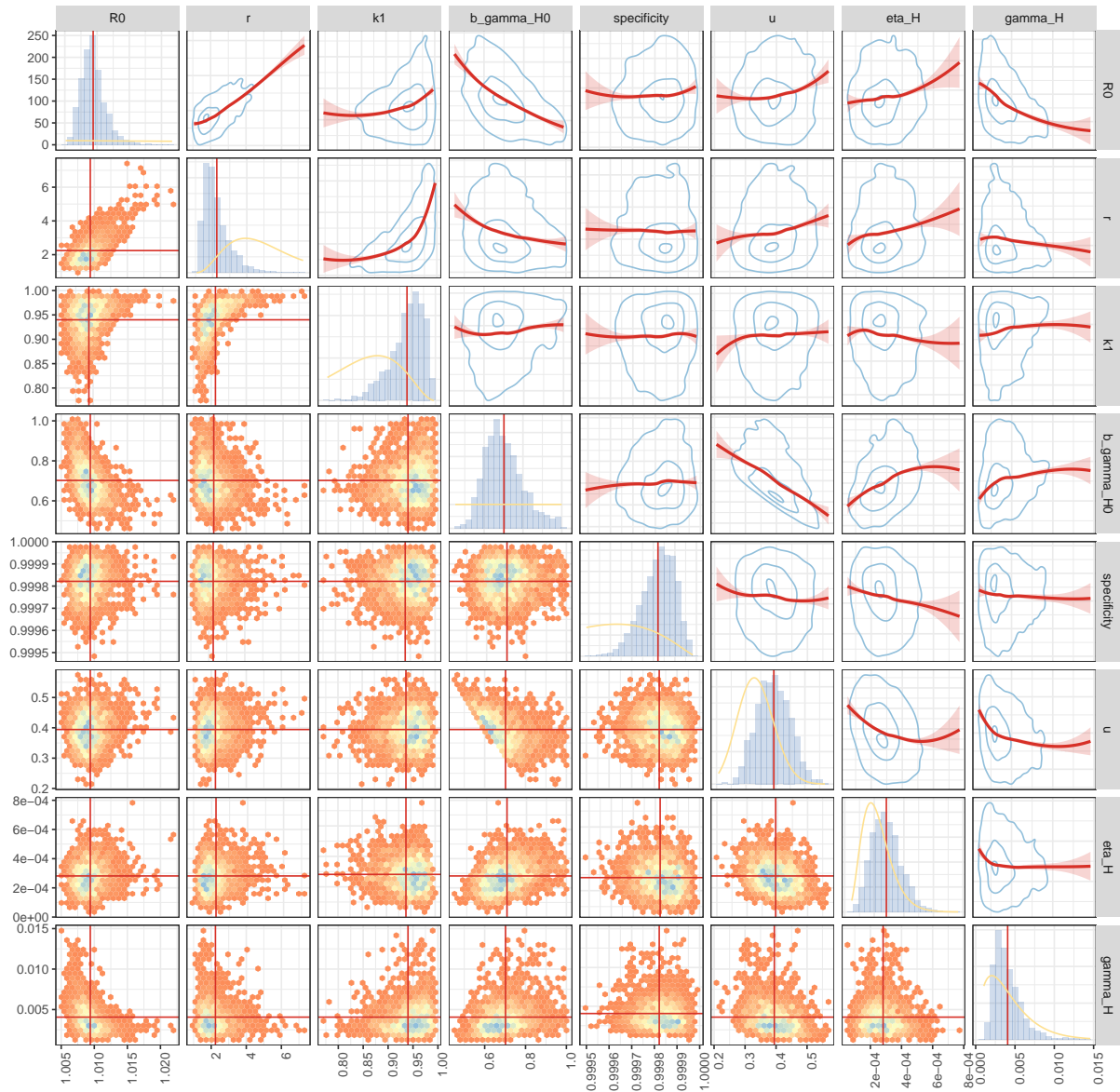


Figure S17: Joint posterior distributions for Tandala health zone in the former Equateur province.

S3 Using posteriors to infer time infected and reporting

120

S3.1 Time spent infected

121

In order to provide a straightforward metric to assess the improvements to passive detection over time we compute the average time spent infected by people not picked up by active screening (for example by people in the high-risk group) using the following equation:

122
123
124

$$\begin{aligned}
 T_{\text{infected}}(Y) &= \mathbb{P}(\text{Passively detected in S1}) \\
 &\quad \times \text{Time spend infected in S1} \\
 &+ \mathbb{P}(\text{Passively detected in S2 or unreported}) \\
 &\quad \times \text{Time spend infected in S1 and S2} \\
 &= \left[\frac{\eta_H(Y)}{\eta_H(Y) + \varphi_H} \times \frac{1}{\eta_H(Y) + \varphi_H} \right] \\
 &\quad + \left[\frac{\varphi_H}{\eta_H(Y) + \varphi_H} \times \left(\frac{1}{\gamma_H(Y)} + \frac{1}{\eta_H(Y) + \varphi_H} \right) \right]
 \end{aligned}
 \tag{S3.1}$$

The inferred change in average time spent infected in Kwamouth and Tandala is shown in Figure S18; in Kwamouth health zone the time has decreased due to improvements in passive surveillance. In Kwamouth the average time changes from 1153 (95% CI: 676–2387) days to 784 (95% CI: 546–1265) days. In Tandala no improvement in passive detection over time was modelled, and hence the amount of time spent infected remains at 716 days on average (95% CI: 548–974). Active screening will have further brought these durations down for those in the population who participate in active screening (low-risk individuals). It is interesting to note that our estimates for mean time spent infected are a little larger than those estimated by Checchi *et al* [S3], even though their estimate for stage 2 duration informed our prior on γ_H^{post} ; their combined estimate of expected duration (S1 and S2 with no treatment) is 778 days (95% CI: 525–1232).

125
126
127
128
129
130
131
132
133

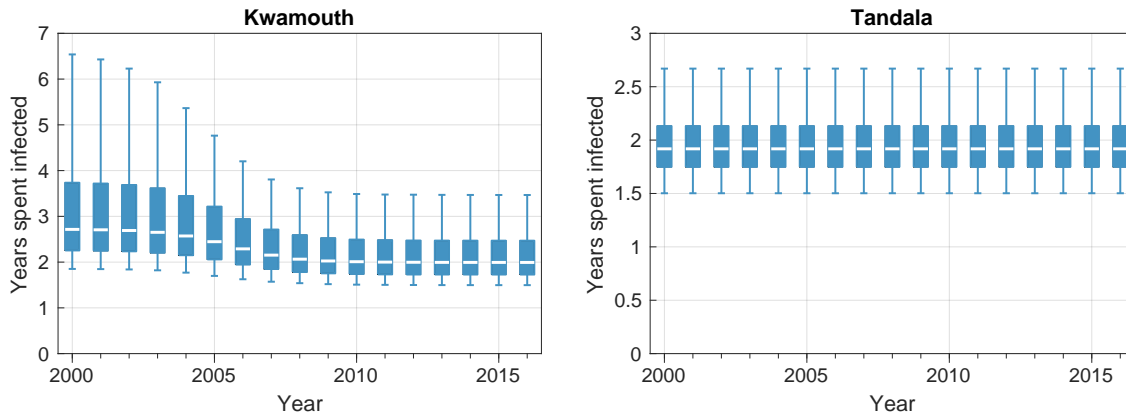


Figure S18: Change in the average time spend infected if not picked up by active screening. Kwamouth is shown on the left, and Tandala on the right.

S3.2 Proportion of successful treatments or deaths reported

134

In order to estimate the change in the proportion of infections reported over time we use sampled model outputs of deaths and case reporting (active and passive for both stages):

135
136

$$\text{Proportion reported} = 1 - \frac{\text{Deaths}}{\text{Active1} + \text{Active2} + \text{Passive1} + \text{Passive2} + \text{Deaths}}
 \tag{S3.2}$$

The estimated change in proportion reported over time in Kwamouth and Tandala is shown in Figure S19; in both locations the proportion of reporting is inferred to vary in time, although the overall decrease in infection and case reporting results in larger credible intervals for 2016, especially in Tandala. In Kwamouth the model estimates the proportion of cases and deaths reported changed from 0.67 (95%: 0.53–0.82) in 2000 to 0.81

137
138
139
140

(95%: 0.62–0.93) in 2016; the consistent high level active screening in conjunction with improving passive surveillance results in an improving trend through time. In Tandala the proportion reported fluctuated over time from 0.65 (95%: 0.53–0.75) in 2000 and finishing back at 0.65 (95%: 0.0–1.0) in 2016 with the median varying between 0.40 (in 2010) and 0.76 (in 2003); the amount of active screening was substantially lower in 2008–2015 than in the 2000–2007 period which is why the average proportion reported is lower in that time period.

141
142
143
144
145
146

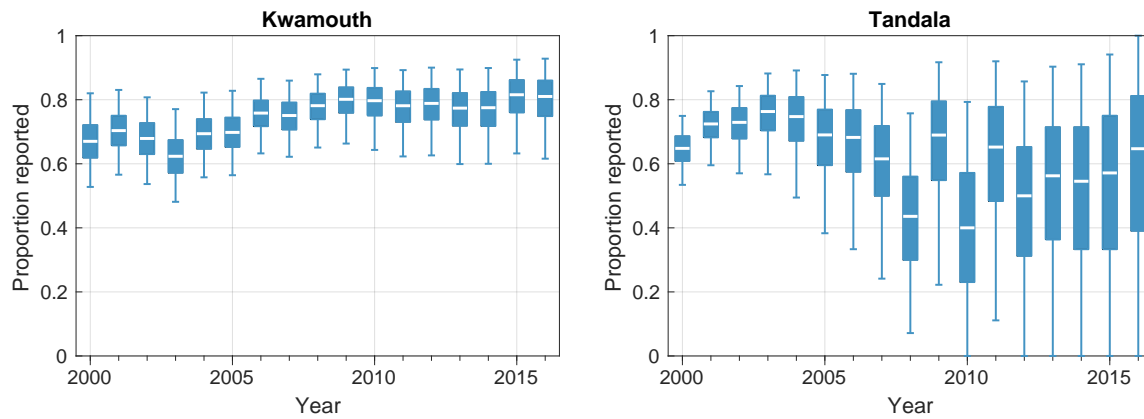


Figure S19: Change in the estimated proportion of infections reported over time. Box plots show estimates for the median (center line), 50% percentiles (boxes) and 95% percentiles (whiskers). Kwamouth is shown on the left, and Tandala on the right.

S4 Online Results

147

Results for each health zone level fit can be viewed at <https://hatmepp.warwick.ac.uk/fitting/v1/>. An example screenshot from this graphical user interface is available on page 25, Fig S20.

148
149

S5 PRIME-NTD criteria

150

It has been recommended that good modelling practises should meet the five key principles relating to communication, quality and relevance of analyses - known as Policy-Relevant Items for Reporting Models in Epidemiology of Neglected Tropical Diseases (PRIME-NTD) [S1]. We demonstrate how these PRIME-NTD criteria have each been addressed in Table S4.

151
152
153
154

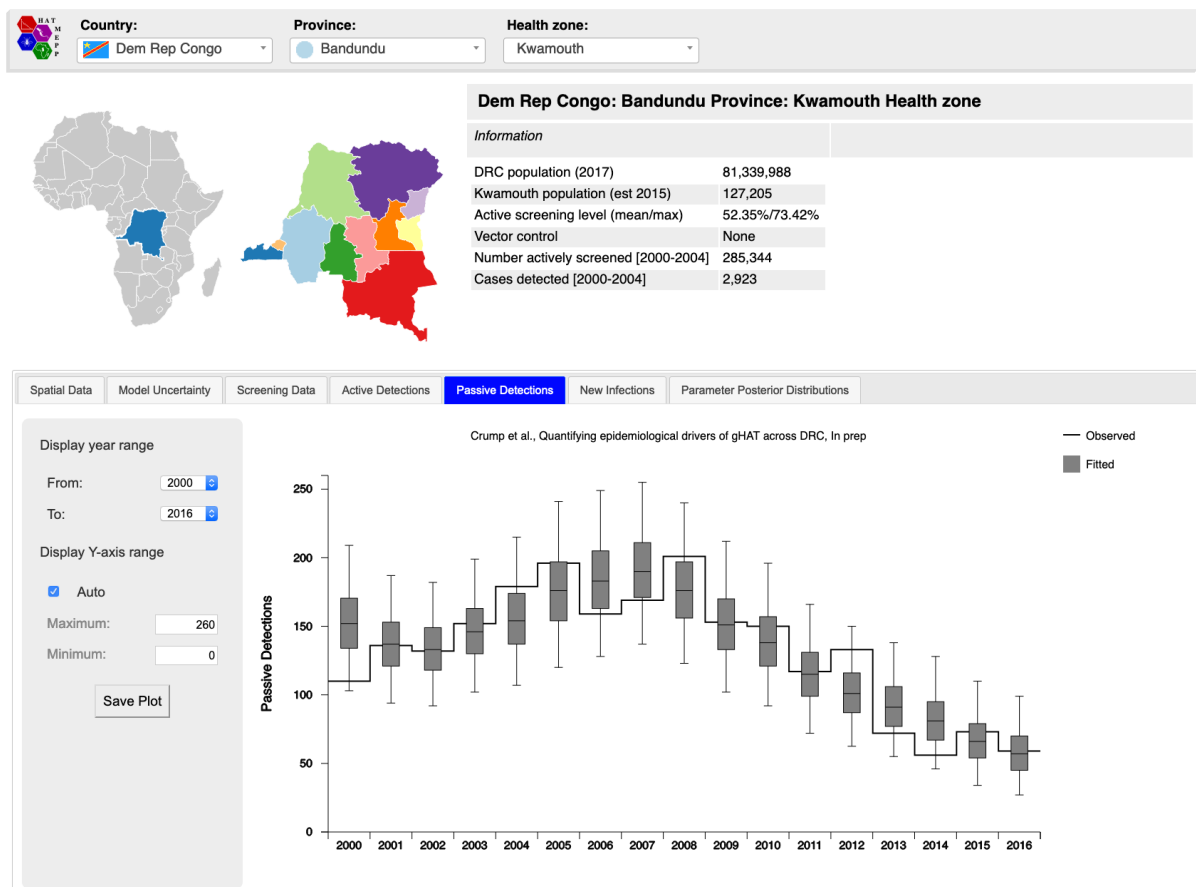


Figure S20: Screenshot of the online results website which accompanies this publication.

Table S4: PRIME-NTD criteria fulfillment. How the NTD Modelling Consortium's "5 key principles of good modelling practice" have been met in the present study.

Principle	What has been done to satisfy the principle?	Where in the manuscript is this described?
1. Stakeholder engagement	<p>This study was lead by modellers and guided by members of the national sleeping sickness control programme in DRC (PNLTHA-DRC) – coauthors E Mwamba Miaka and S Chancy. PNLTHA-DRC have contributed to improved modeller understanding of the epidemiological data and changes to the programme over time and in different geographic regions, both of which impacted model fitting over several rounds of revision (via in-person meetings and email). The GUI (and several variants of it) was designed in conjunction with PNLTHA-DRC to improve communication of the modelling outputs to non-modellers. It has been refined through various in-person meetings with different collaborators with the goal of providing understandable, policy-relevant outputs as well as scientific communication; over 20 non-modellers have had opportunities to interact with and provide feedback on the GUI during development.</p>	<p>Authorship list</p>
2. Complete model documentation	<p>Full model fitting code and documentation is available through OpenScienceFramework (OSF). The model is fully described in the main text and SI.</p>	<p>See Materials and Methods section in the main text, Section S1.3 and at OSF (https://osf.io/ck3tr/?view_only=526344c12324492083db1e49c76136af)</p>
3. Complete description of data used	<p>The original data and how we aggregated the data for fitting were described in detail in the main text and SI. Aggregate data can be viewed next to model fits in our GUI.</p>	<p>See Materials and Methods section, Section S1.1 and the GUI (https://hatmepp.warwick.ac.uk/fitting/v1/)</p>
4. Communicating uncertainty	<p><i>Structural uncertainty:</i> The variant of the model presented here ("Model 4") was chosen as it had good support compared to other plausible model structures when fitting to data sets from Yasa Bonga and Mosango health zones in DRC [S6] and in the Mandoul focus, Chad [S5].</p> <p><i>Parameter uncertainty:</i> We provide estimates for the parameter uncertainty in each health zone within our posterior parameter maps (randomly sampled values from posteriors) and joint posterior distributions of fitted parameters (main text and SI) and also show distributions (histograms) in the GUI.</p>	<p><i>Structural uncertainty:</i> Materials and Methods section in main text.</p> <p><i>Parameter uncertainty:</i> Figures 6–7, Figures S2–S15 and model uncertainty maps in GUI (https://hatmepp.warwick.ac.uk/fitting/v1/)</p>

Continued on next page

Table S4 – continued from previous page

Principle	What has been done to satisfy the principle?	Where in the manuscript is this described?
	<p><i>Prediction uncertainty.</i> We represent uncertainty in our results by: (i) summarising province-level estimates of changes to new infections; and (ii) providing box and whisker plots for fitted dynamics (median, 50% and 95% credible intervals).</p>	<p><i>Prediction uncertainty:</i> (i) Table 3 (ii) Figure 5</p>
<p>5. Testable model outcomes</p>	<p>Previous versions of this model have undergone validation exercises (data censoring) to examine the robustness of the predictive ability of the model [S5, S8]. Whilst this was not performed here, the multiple rounds of model fitting, critical review and refinement in discussion with PNLTHA produced very clear improvements to the fit by altering assumptions about the passive detection rates and changes to diagnostic specificity over time. This was most improved for former Bandundu province where we are now able to match “humped” trends in passive detection.</p>	<p>See main text results for explanation of “humped” trends in passive reporting.</p>

References

- [S1] Matthew R. Behrend, María-Gloria Basáñez, Jonathan I. D. Hamley, Travis C. Porco, Wilma A. Stolk, Martin Walker, Sake J. de Vlas, and for the NTD Modelling Consortium. Modelling for policy: The five principles of the neglected tropical diseases modelling consortium. *PLoS Neglected Tropical Diseases*, 14(4):1–17, 04 2020. 156
157
158
159
160
- [S2] M Soledad Castaño, Martial L Ndeffo-Mbah, Kat S Rock, Cody Palmer, Edward Knock, Erick Mwamba Miaka, Joseph M Ndung'u, Steve Torr, Paul Verlé, Simon E F Spencer, and Others. Assessing the impact of aggregating disease stage data in model predictions of human African trypanosomiasis transmission and control activities in Bandundu province (DRC). *PLoS Neglected Tropical Diseases*, page e0007976, 2020. 161
162
163
164
165
- [S3] F Checchi, S Funk, D Chandramohan, D T Haydon, and F Chappuis. Updated estimate of the duration of the meningo-encephalitic stage in gambiense human African trypanosomiasis. *BMC Research Notes*, 8(1):292, July 2015. 166
167
168
- [S4] Crispin Lumbala, Pere P Simarro, Giuliano Cecchi, Massimo Paone, José R Franco, Victor Kande Betu Ku Mesu, Jacques Makabuza, Abdoulaye Diarra, Shampa Chansy, Gerardo Priotto, et al. Human african trypanosomiasis in the democratic republic of the congo: disease distribution and risk. *International journal of health geographics*, 14(1):20, 2015. 169
170
171
172
- [S5] Mahamat Hissene Mahamat, Mallaye Peka, Jean-baptiste Rayaisse, Kat S Rock, Mahamat Abdelrahim Toko, Justin Darnas, Guihini Mollo Brahim, Ali Bachar Alkatib, Wilfrid Yoni, Inaki Tirados, Fabrice Courtin, Samuel P C Brand, Cyrus Nersy, Oumar Alfaroukh, Steve J Torr, Mike J Lehane, and Philippe Solano. Adding tsetse control to medical activities contributes to decreasing transmission of sleeping sickness in the Mandoul focus (Chad). *PLoS Neglected Tropical Diseases*, 11(7):e0005792, 2017. 173
174
175
176
177
- [S6] Kat S Rock, Steve J Torr, Crispin Lumbala, and Matt J Keeling. Quantitative evaluation of the strategy to eliminate human African trypanosomiasis in the DRC. *Parasites & Vectors*, 8(1):532, 2015. 178
179
- [S7] Kat S Rock, Steve J Torr, Crispin Lumbala, and Matt J Keeling. Predicting the impact of intervention strategies for sleeping sickness in two high-endemicity health zones of the Democratic Republic of Congo. *PLoS Neglected Tropical Diseases*, 11:e0005162, 2017. 180
181
182
- [S8] K.S. Rock, A. Pandey, M.L. Ndeffo-Mbah, K.E. Atkins, C. Lumbala, A. Galvani, and M.J. Keeling. Data-driven models to predict the elimination of sleeping sickness in former Equateur province of DRC. *Epidemics*, 18:101–112, 2017. 183
184
185
- [S9] United Nations Office for the Coordination of Humanitarian Affairs. *The Humanitarian Data Exchange*. 186