Assessing the impact of aggregating disease stage data in model predictions of human African trypanosomiasis transmission and control activities in Bandundu province (DRC)

Mar´ıa Soledad Castan˜o1,2\*, Martial L. Ndeffo-Mbah3,4, Kat S. Rock5,6, Cody Palmer7, Edward Knock5,8, Erick Mwamba Miaka9, Joseph M. Ndung’u10, Steve Torr11, Paul Verl´e12, Simon E.F. Spencer5,8, Alison Galvani3, Caitlin Bever7, Matt J. Keeling5,6, Nakul Chitnis1,2

1. Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland
2. University of Basel, Basel, Switzerland
3. School of Public Health, Yale University, New Haven, Connecticut, USA
4. College of Veterinary Medicine and Biosciences, Texas A&M University, College Station, Texas, USA
5. Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Coventry, UK
6. Mathematics Institute, University of Warwick, Coventry, UK
7. Institute of Disease Modeling, Seattle, Washington, USA
8. Department of Statistics, University of Warwick, Coventry, UK
9. Programme National de Lutte contre la Trypanosomiase Humaine Africaine, Kinshasa, the Democratic Republic of the Congo
10. Foundation for Innovative New Diagnostics, Geneva, Switzerland
11. Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK
12. Department of Public Health, Institute of Tropical Medicine, Antwerp, Belgium

[\*soledad.castano@swisstph.c](mailto:soledad.castano@swisstph.ch)h (MSC)

# Abstract

Since the turn of the century, the global community has made great progress towards the elimination of gambiense human African trypanosomiasis (HAT). Elimination programs, primarily relying on screening and treatment campaigns, have also created a rich database of HAT epidemiology. Mathematical models calibrated with these data can help to fill remaining gaps in our understanding of HAT transmission dynamics, including key operational research questions such as whether integrating vector control with current intervention strategies is needed to achieve HAT elimination. Here we explore, via an ensemble of models and simulation studies, how including or not disease stage data, or using more updated data sets affect model predictions of future control strategies.

# Author summary

Human African tryposonomiasis (HAT), also known as sleeping sickness, is a parasitic disease with over 65 million people estimated to be living at risk of infection. Sleeping sickness consists of two stages: the first one is relatively mild but the second stage is usually fatal if untreated. The World Health Organization has targeted HAT for elimination as a public health problem by 2020 and for elimination of transmission by 2030. Regular monitoring updates indicate that 2020 elimination goals are likely to be achieved. This monitoring relies mainly on case report data that is collected through medical-based control activities — the main strategy employed so far in HAT control. This epidemiological data are also used to calibrate mathematical models that can be used to analyse current interventions and provide projections of potential intensified strategies.

We investigated the role of the type and level of aggregation of this HAT case data (staging data and truncated data) on model calibrations and projections. We highlight that the lack of detailed epidemiological information, such as missing stage of disease or truncated time series data, impacts model recommendations for strategy choice: it can misrepresent the underlying HAT epidemiology (for example, the ratio of stage 1 to stage 2 cases) and increase uncertainty in predictions. Consistently including new data

from control activities as well as enriching it through cross-sectional (e.g. demographic or behavioural data) and geo-located data is likely to improve modelling accuracy to support planning, monitoring and adapting HAT interventions.

# Introduction 1

Human African trypanosomiasis (HAT) is a neglected tropical disease that affects 2

people in resource-limited settings in sub-Saharan Africa, with more than 65 million 3

people living at risk [[1].](#_bookmark13) HAT is caused by a protozoan parasite and is transmitted 4

between humans by biting tsetse flies. The gambiense form of the disease, caused by 5

*Trypanosoma brucei gambiense*, is responsible for over 95% of human cases. This chronic 6

disease progresses through two stages. The first stage can last for several years with 7

relatively minor symptoms such as fever and headaches. Second stage patients show 8

neuropsychiatric disorders (including sleep disturbances that led to the common name, 9

sleeping sickness) and this stage is usually fatal without treatment. Currently available 10

treatments are stage-dependent and so assessment of a patient’s stage - by analysing the 11

cerebrospinal fluid for parasites and number of white blood cells - is a prerequisite for 12

appropriate treatment. 13

Since the start of the 21st Century, control activities against gambiense HAT have 14

had a substantial impact on reducing disease transmission and burden in the main 15

endemic regions [[2].](#_bookmark14) These control efforts have raised expectations that elimination of 16

gambiense HAT may be achievable [[1,](#_bookmark13) [3].](#_bookmark15) The World Health Organization (WHO) has 17

therefore set indicators that target elimination of transmission (EOT) by 2030. 18

Although there were only 953 cases reported globally in 2018 [[4],](#_bookmark16) persistent foci of 19

disease transmission remain a potential challenge for achieving the EOT goal. The 20

Democratic Republic of Congo (DRC) has suffered from persistent infection, 21

contributing between 78–91% of all global cases since 2010 [[4].](#_bookmark16) 22

Efforts to control HAT have mainly relied on screening, testing and treating the 23

human population using active and/or passive surveillance. This has been the only 24

intervention applied at large scale, and it seems likely that this is largely responsible for 25

the precipitous decline in global incidence, including a 97% reduction in HAT cases in 26

the former Equateur province of DRC between 2000 and 2012 [[5].](#_bookmark17) However, the screen, 27

diagnose and treat strategy has been unable to effectively control transmission to this 28 level in all endemic foci (e.g. some health zones of Kwilu province, DRC), probably due 29 to insufficient levels of coverage, imperfect diagnostics, or people at high risk of 30

transmission not participating in screening activities. 31

Where epidemiological and/or control campaign data of infectious diseases are 32

available, data-driven models have proved to be a valuable tool for quantitatively 33

assessing epidemiological assumptions about disease transmission dynamics or 34

evaluating the effectiveness of intervention measures [[6](#_bookmark18)–[8].](#_bookmark19) For HAT, data arising from 35

several interventions implemented in recent years have enabled modelling and 36

quantitative analyses of the potential advantages of novel interventions in endemic 37

regions such as Kwilu and former Equateur province in DRC [[9–11],](#_bookmark22) Mandoul in 38

Southern Chad [[12],](#_bookmark23) and Boffa in Guinea [[13].](#_bookmark24) Nonetheless, many epidemiological 39

aspects of HAT remain unclear, and additional data are needed to fill these knowledge 40

gaps. For example, the role of certain subpopulation groups in maintaining transmission 41

in endemic areas, such as those not covered by screening programmes or at unusually 42

high risk due to behavioral or geographical characteristics; or the potential existence of 43

reservoir animal hosts or asymptomatic human carriers is not fully understood [**?**]. 44

With the 2030 EOT goal on the horizon, it is crucial to determine which efforts in 45

which locations could maximise the potential benefits of any intervention against HAT. 46

Modelling could provide the HAT community with a better understanding of the 47

important factors affecting observed changes in intensity of disease reporting and 48

explain some of the variations in effectiveness of HAT control and surveillance activities 49

across different settings. 50

In this study we analyse a longitudinal human epidemiological data set of HAT from 51

former Bandundu province in the DRC to outline how the type of data and its level of 52

aggregation may affect projections of HAT transmission models. Four independent HAT 53

models, fitted to three different data aggregation sets (unstaged disease data, staged 54

disease data, truncated staged disease data), are used to investigate how these levels of 55

data aggregation impact the projections of HAT incidence and likelihood of achieving 56

the EOT goal for current and intensified intervention strategies. Although the 2030 goal 57

is defined as EOT for the continent, and therefore meeting EOT within Bandundu is not 58

directly equivalent, failure to meet the goal in this high-endemicity region would imply 59

failure to meet the global EOT target. Implications of data resolution on the estimated 60 effectiveness of strategy is analysed in order to suggest potential improvements in data 61 collection and availability that could contribute to robust assessment of control 62

programme effectiveness and reliable estimates of HAT elimination. 63

# Materials and methods 64

**Data description and assumptions** 65

Former Bandundu province in the DRC has the world’s highest HAT burden despite a 66

significant coordinated effort between national and international HAT control 67

programmes [[5].](#_bookmark17) This province covers an area of 296,500 km2 (12.6% of DRC) and 68

accounts for the largest number of cases reported since 2001 in the country 69

(approximately 47.6%). 70

In this study we used publicly available provincial level human case data from 71

Bandundu province [[5]](#_bookmark17) to calibrate models of HAT transmission. The data contains the 72

annual number of positive cases for each stage of the disease detected through active 73

screening and passive detection (the primary HAT control interventions implemented in 74

this area); and the total screened population across the province for the years 2000-2012. 75

Although the geographical scale of this province-level data is large, this data was chosen 76

because - to the authors’ knowledge - this is the only (either publicly or under-request) 77

available data providing details on the stage of reported cases for many consecutive 78

years. 79

Estimates of the population of Bandundu were taken from publicly available census 80

data [[14]](#_bookmark25) for 2000-2012 and a 3% annual growth rate was assumed for projections. 81

Although target populations are usually estimated prior to each active screening round, 82

this data was not publicly available and the target varies from year to year depending 83

on the health zones screened. To determine a consistent estimate over 13 years, each 84

model assumed a constant proportion of the population at risk over the entire period, 85

either fixed or estimated during model calibration (see details in [S2 T](#_bookmark7)ext). 86

**HAT models** 87

Four independent deterministic models of HAT transmission were used (hereafter 88

named as Model I, Model S, Model W and Model Y) to evaluate the effects of different 89

levels of data aggregation on forward projections. 90

All of them were based on models previously used in either simulation or data-driven 91

studies [[9,](#_bookmark20) [10,](#_bookmark21) [15–17]](#_bookmark28) and include modifications, independently implemented by each 92

group, to improve calibration to the data analysed here. Differences in structural 93

assumptions (e.g. disease progression, heterogeneity in risk to infection) and 94

parameterisation reflect the variety of complexities and biological uncertainties typically 95

found in epidemiological models. Furthermore, a range of different fitting methodologies 96

were employed which also have implications on results. An overview of key aspects of 97

model structure, interventions and fitting procedure is given in Table [1](#_bookmark0) and more details 98

of each of the models can be found in [S2 Text.](#_bookmark7) 99

## Table 1. Models overview

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model I | | Model S | Model W | Model Y |
| Partitions population into N | | Y | Y | N |
| Asymptomatic infection N | | N | N | Y |
| Infectious stages 1 and 2 Y | | Y | Y | Y |
| Population at risk All at risk | | Assumed constant | All at some risk | Fixed population at risk |
|  |  | (fixed at 70%) | (high or low) | estimated during fitting |
| Pulsed AS | N | Y (1*st* month each year) | (1*st* month each year) | (1*st* month each year) |
| AS in all population | Y | N (only low-risk) | N (only low-risk) | N (only at risk population) |

Transmission

model structure

high/low risk

Interventions

PD: stage-specific detection rate

Y Y (one fitted) Y (both fitted) Y (both fitted)

PD: time-dependent detection rate

N Y (fit to staged and subset staged data )

Y (fit to staged data ) N

PD: underreporting N N Y (stage 2 only) Y (stage 2 only)

EPD: improvement in detection in both stages

Y Y Y Y

Nb. of parameters fixed and fitted

fixed:10 fitted:7

fixed:24 fitted:6

fixed:19 (staged fit) & 17 (other fits) fitted:9 (staged fit) & 8 (other fits)

fixed:17 fitted:8

Fitting procedure

Initial conditions Fitted Endemic equilibrium with ongoing PD

Endemic equilibrium with ongoing PD

Endemic equilibrium with ongoing PD

Likelihood-based Y N Y N

7/26

Likelihood for AS Poisson - Beta-binomial -

Likelihood for PD Poisson - Beta-binomial -

Description of key aspects of model structure, interventions and fitting procedure. Abbreviations: AS: active screening; EPD: enhanced passive detection; PD: passive detection.

**Model fitting** 100

The reported number of cases detected through active and passive screening and the 101

number of people tested were used to calibrate the models emulating the effects of a 102

typical medical control strategy. The data do not contain information on the timing and 103

duration of active screening, so each modelling group independently managed these 104

aspects (see Table [1).](#_bookmark0) 105

The models were calibrated to three different configurations of the data to reflect the 106

diversity of data resolution usually available, allowing the analysis of the impact of data 107

detail on both uncertainty and reliability of model projections. The three configurations 108

were labelled: “unstaged data”, “staged data” and “subset staged data”. “Unstaged 109

data” informed the models using the number of HAT cases detected each year 110

(2000-2012), separated by active and passive detection. This type of longitudinal data - 111

where the disease stage is not noted - is typical of data available at smaller 112

administrative levels, such as health zones or health areas in DRC. “Staged data” 113

additionally partitioned the number of cases from the “unstaged data” by disease stage 114

(first or second). The “subset staged data” consisted of a temporal subset of the “staged 115

data”, covering only years 2000-2006. By cutting the data at this point, the 116

improvement observed after 2006 in the detection of stage 1 cases is not yet apparent, 117

and so we expected to see some effects of this in model estimations and projections. 118

Each group independently chose a calibration method adapted to their own model. 119

The list of fixed parameters used (either obtained from the literature or assumed) and 120

those estimated during the fitting are detailed in the description of each model in [S2](#_bookmark7) 121

[Text.](#_bookmark7) Fitting procedures included Bayesian inference using Markov Chain Monte Carlo 122

(MCMC) (Models I and W) and approximate Bayesian computation methods (Models S 123

and Y). In all cases, one thousand samples (i.e. parameter sets) were generated during 124

the fitting step for further estimations and projections. In all cases plots display the 125

median and associated 95% credible interval (CI). For further details on models’ 126

structure, assumptions and fitting procedure, see details in [S2 Text.](#_bookmark7) 127

**Simulated HAT interventions** 128

Four interventions were considered for simulations. They consisted of three 129

medical-based interventions: “active screening”, “passive detection” and “enhanced 130

passive detection”; and “vector control”. A brief description of these interventions is 131

provided below. 132

* **Active screening (AS).** This is the screening of the population at large in 133

at-risk locations by mobile teams. Once detected, patients travel to medical 134

centres for treatment. In this study, the reported annual number of people 135

screened was used to estimate the mean active screening coverage. Models that 136

included population heterogeneity in exposure to tsetse (Models S and W) assume 137

that only low-risk people are screened actively. 138

* **Passive detection (PD).** This is the diagnosis and treatment of infected people 139

who self-present at medical facilities. HAT models usually assume that passive 140

surveillance detects mainly stage 2 cases, when symptoms are more severe and 141

specific to HAT. The data used in the present work reports a non-negligible 142

proportion of stage 1 cases detected through passive surveillance. For this reason, 143

both stages were assumed to have the potential to be detected in all models. 144

* **Enhanced passive detection (EPD).** This is passive screening where the time 145

to detection of infected people is reduced (i.e. improved detection rate per capita). 146

Such improvement could result from one or a combination of changes in current 147

control activities. For example, increasing the number of health facilities (thus 148

increasing the chances of picking cases), mobilising the population at risk or by 149

reducing the time to detection and treatment through improved HAT diagnostic 150

tools including rapid detection tests (RDTs). In DRC, RDTs have been used in 151

many endemic settings between 2013 and 2016 [[18,](#_bookmark29) [19],](#_bookmark31) although estimates of the 152

improvement on the associated detection rate have not yet been quantified. 153

* **Vector control.** This intervention focuses on increasing the mortality and 154

reducing the density of tsetse flies by, for example, deploying insecticidal baits 155

(e.g., insecticidal targets, insecticide-treated cattle) to attract and kill tsetse. In 156

particular, tiny targets [[20]](#_bookmark32) offer great promise for the large-scale and cost-effective 157

control of the riverine tsetse species which transmit gambiense HAT [[12,](#_bookmark23) [20–22].](#_bookmark34) 158 Tiny targets were first deployed in DRC in 2015, in Yasa Bonga health zone, and 159 they are currently being used in three health zones of former Bandundu province. 160

With these four interventions, three different strategies were investigated that reflect 161

either the current control and surveillance programmes or strengthened strategies to 162

accelerate the elimination of HAT. These are: 163

* Strategy 1: also referred to as “baseline”, this strategy represents the standard 164

control method in Bandundu consisting of continuing active screening and passive 165

detection at present rates. 166

* Strategy 2: consists of vector control in addition to the baseline strategy, as is 167

currently being implemented in Yasa-Bonga, Masi Manimba and Kwamouth 168

health zones of Bandundu. In the models, vector control was assumed to reduce 169

tsetse populations by 60% after one year, which is a conservative estimate from 170

intervention trials conducted in other HAT foci [[12,](#_bookmark23) [20,](#_bookmark32) [21].](#_bookmark33) 171

* Strategy 3: assumes enhanced passive detection, in addition to the annual active 172

screening campaign. For this strategy, Models I and S doubled the passive 173

detection rate of both stages while models explicitly including underreporting 174

(Models W and Y) assumed both a doubled passive detection rate and halving of 175

underreporting. We also assumed that the treatment rate of detected cases 176

remained the same so that increased detection led to a corresponding increase in 177

the treatment rate. 178

The calibrated models were used to simulate the “future” effects of these three 179

strategies (Table [2)](#_bookmark1) in order to compare, for each model, the effects of the different 180

types of data aggregation used for calibration, on projections and associated uncertainty 181

under different control strategies. In all cases the baseline strategy matched the period 182

corresponding to the data, and assumed a continuation of standard passive surveillance 183

and past mean active screening levels informed by the data for projections into the 184

future. 185

Model simulations estimated *(i)* annual stage-specific cases reported from both 186

active and passive screening; *(ii)* new transmissions by year; and *(iii)* year of EOT 187

## Table 2. Different types of future strategies considered in model projections.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Strategy** | **Passive** |  | **Interventions**  **Active** | | | **Vector control** | |
| 1 | Standard | Mean | of | historic | data |  | - |
| 2 | Standard | Mean | of | historic | data | 60% | reduction |
| 3 | Enhanced | Mean | of | historic | data |  | - |

(considering two thresholds: *<*1 new infection per 100,000 and *<*1 new infection per 188

1,000,000 individuals). 189

**Results** 190

**Model fits** 191

**Reported cases** 192

Fig [1](#_bookmark2) shows the data from 2000 to 2012 of the total reported HAT cases in Bandundu 193

and the calibrated simulations of the four models to three different data configurations 194

(median with the 95% credible interval (CI)) under the “baseline” control strategy. All 195

fits of all models consistently reproduced the decreasing trend observed in data. 196

However for most model fits, the 95% CI did not cover all the data points in time series, 197

as is often the case for peaky stage-specific data dominated by a decreasing trend (Fig 198

S1.1 in [S1 T](#_bookmark6)ext). Models provided varying levels of uncertainty, mainly explained by 199

differences in fitting methods as well as model structure and parameterisations. Despite 200

all these differences, the fit to the longer, staged data set generated less uncertainty in 201

all four models, with worse and varying performance for the fits to the other data sets. 202

While for Model W the medians from the fit to staged data gave the lowest 203

estimation compared to the other two fits, for Model S this trend was the opposite for 204

most years. For Models I and Y such a clear trend was not observed among medians. 205

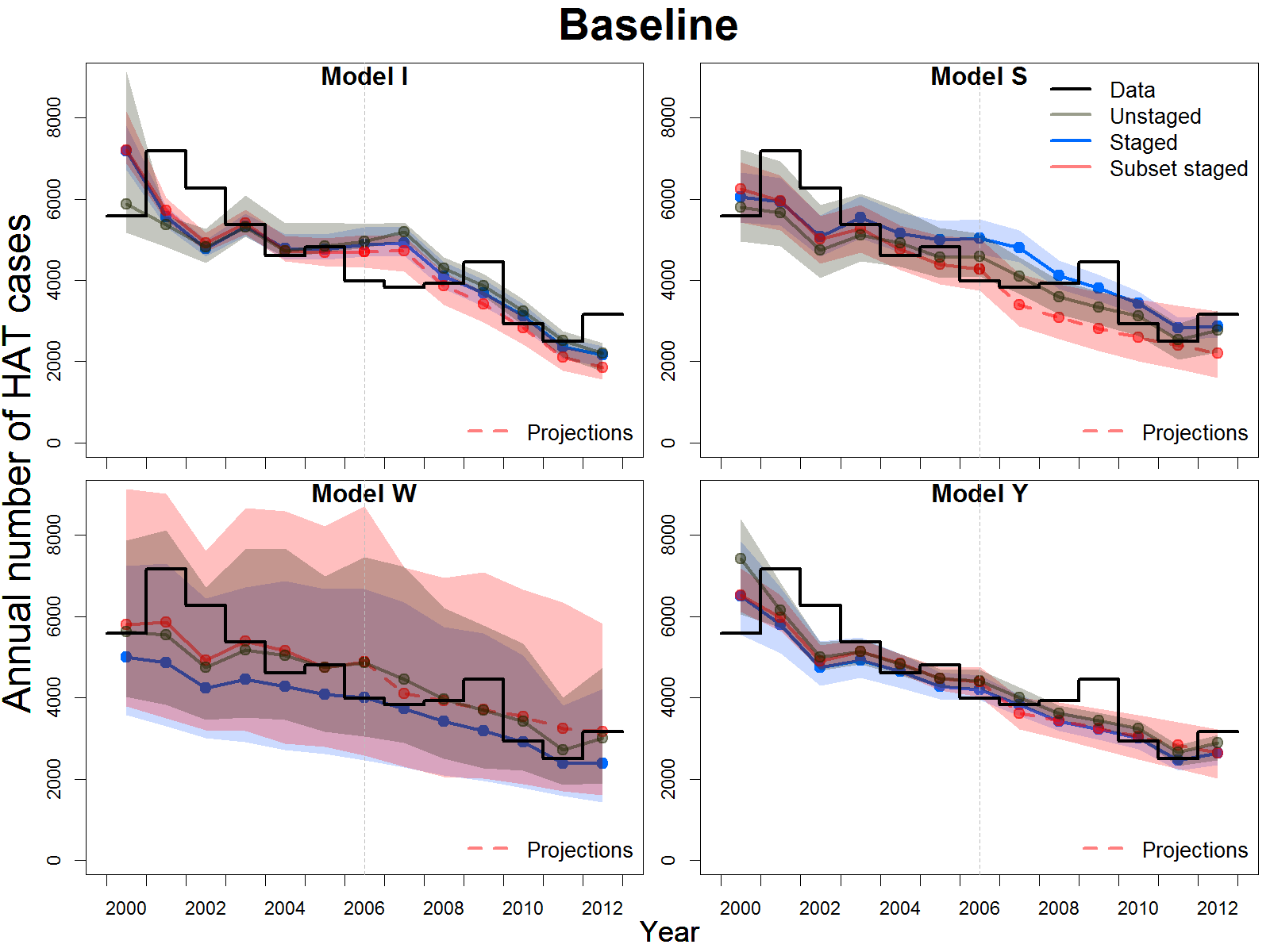
## Proportion of stage 1 cases 206

The increasing trend in the proportion of stage 1 cases out of total reported cases across 207

years (Fig [2)](#_bookmark3) indicates improved screening in Bandundu; this is observed in both active 208

and passive case data [(S1 Fig](#_bookmark11) and [S2](#_bookmark12) Fig). Model fits not informed with staging ratios 209

produced the worst estimates of this proportion and the highest uncertainties (Fig [2),](#_bookmark3) 210



**Fig 1. Former Bandundu province reported data and estimated reported cases.** Estimated reported cases from model calibrations to three different configurations of the data for a baseline strategy composed of annual, pulsed active screening and continuous passive detection. The median (as a point) and the corresponding 95% CI (shaded region of the same color) are shown in each case. Dashed lines indicate projections from the fit to the subset staged data.

reflecting a wide range of possible configurations of the proportion of stage 1 infections 211

compatible with such unstaged data either in active screening [(S1](#_bookmark11) Fig), passive 212

detection [(S2](#_bookmark12) Fig), or both. 213

The variety of assumptions in the models about intervention implementation, 214

including how annual active screening was applied (continuous vs. pulsed, one vs. 215

several per year) or which proportion of Bandundu province population was assumed to 216

be at risk of infection (Table [1),](#_bookmark0) explain in part the variety of results in the proportion 217

of stage 1 cases for different fits. Model W fitted to the full staged data was the only 218

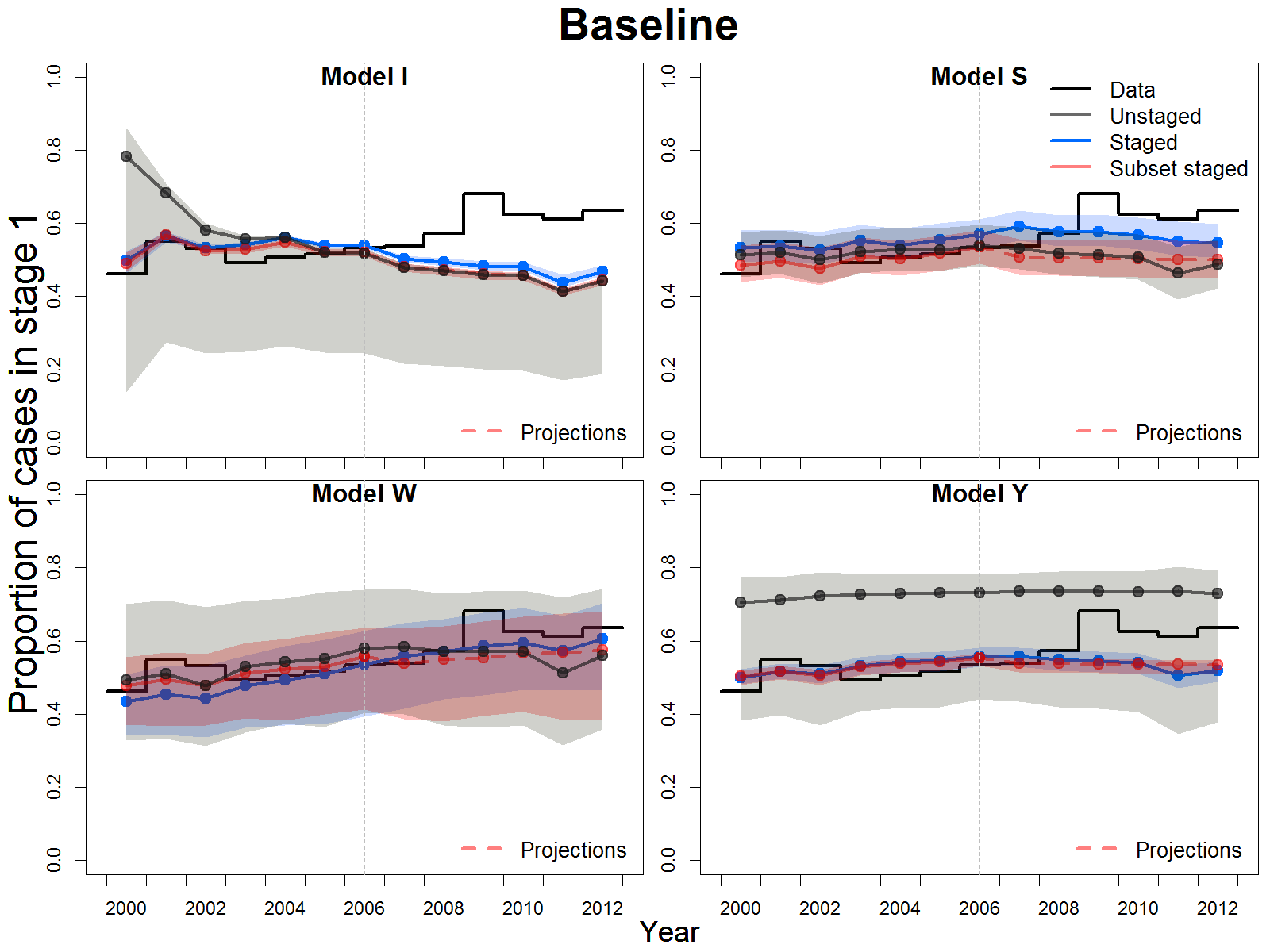
model that reproduced the increasing trend in active screening [(S1](#_bookmark11) Fig); and only 219

Models S and W, which assumed an improvement in passive detection rate, reproduced 220

the increasing trend in passive detection, with systematic overestimation in Model S (S[2](#_bookmark12) 221

[Fig).](#_bookmark12) For these two models, it is clear how the fit to the subset staged data, where the 222

improvement in the passive detection is not yet apparent (contrary to the fit to the full 223



**Fig 2. Proportion of stage 1 cases.** The estimates for the four models fitted to three different configurations of the data under the baseline strategy are shown. The posterior median is shown as a point and 95% CIs shaded. Dashed lines indicate projections from the fit to the subset staged data.

staged data), conditions the models to project lower ratios of stage 1 to stage 2 cases 224

from 2007 onwards. 225

**Projections for future case reporting and transmission** 226

Model projections under all fits came to a consensus that continuing the baseline 227

medical strategy would lead to a sustained but slow reduction of the annual incidence; 228

however some simulations of Model S (86 out of 1000) fitted to unstaged data suggested 229

transmission would increase under baseline strategy (Figs A-D in [S3 T](#_bookmark8)ext). The latter is 230

an example of how some parameters sets, although overall can reproduce unstaged data 231

trends, can have an underlying epidemiology promoting increasing transmission despite 232

continued active screening and passive detection levels. Note that these scenarios are 233

not observed when Model S is fitted to the more informative staged data that impose 234

further constraints to the posterior parameter distributions. 235

As expected, the models indicated that improved or complementary interventions 236

would accelerate this path towards reduced incidence [(S3 T](#_bookmark8)ext). Notably the longer 237

staged data set produced the least uncertainty in all models for projections on annual 238

incidence (Figs A-D in [S3 Text)](#_bookmark8) and associated reported cases (Figs A-D in [S4 T](#_bookmark9)ext). 239

Assuming that projections under the staged data are most robust, the unstaged data 240

generated systematic overestimation in transmission and associated report case 241

projections for any strategy considered in three models (Models S, W and Y); for Model 242

I, a slight discrepancy in projections of new cases was observed, although values from 243

both fits were close and overlapped in projections of reported cases. Model I generated 244

the most optimistic scenarios, with a relatively homogeneous range of projections for 245

the different fits and small uncertainties compared to the other models, with and values 246

on the order of ∼100 new detected cases or fewer by 2030 for Bandundu province. 247

Table [3](#_bookmark4) presents the proportion of simulations (i.e. realisations of different 248

parameter sets) for different fits and models where the 2030 zero transmission goal was 249

achieved, and provides an alternative view on how adding or removing relevant data 250

impacts the models’ projections under different control strategies explored. Here 251

“elimination” is defined as *<*1 transmission case per million individuals per year as in 252

previous work using these deterministic models [[17].](#_bookmark28) 253

## Table 3. Probability of different strategies achieving elimination by 2030.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fit** | Baseline | **Strategy**  Vector control | EPD | **Model** |
| Unstaged | 0.167 | 1 | 0.167 |  |
| Staged | 0 | 0.656 | 0 | I |
| Subset staged | 0 | 1 | 0 |  |
| Unstaged | 0 | 0.206 | 0 |  |
| Staged | 0 | 0.551 | 0 | S |
| Subset staged | 0 | 0.836 | 0 |  |
| Unstaged | 0 | 1 | 0 |  |
| Staged | 0 | 1 | 0.984 | W |
| Subset staged | 0 | 1 | 0 |  |
| Unstaged | 0 | 1 | 0 |  |
| Staged | 0 | 1 | 0 | Y |
| Subset staged | 0 | 1 | 0 |  |

EOT is defined in the models as *<*1 new transmission per 1,000,000 people. In each case simulations of 1000 parameter sets were used.

In all but one case (Model I fitted to unstaged data), the models found that it was 254

extremely unlikely that elimination would occur by 2030 using the baseline strategy. All 255

fits for Models W and Y predicted elimination using vector control tools in addition to 256

the baseline strategy. The least optimistic predictions were observed in Model S, in 257

accordance with higher values and slower reduction in transmission projections when 258

compared to other models’ predictions (Fig B in [S3 T](#_bookmark8)ext). For Model I, the fit using the 259

staged data set showed less optimistic predictions, which is consistent with the 260

transmission projections generated by each fit under this model (Fig B in [S3 T](#_bookmark8)ext). 261

Only two models under different fits (Model I fitted to unstaged data and Model W 262

fitted to staged data) showed that elimination was possible for enhanced passive 263

detection (167 and 984 out of 1000 samples, respectively). 264

For a weaker definition (*<*1 transmission case per 100,000 individuals per year), only 265

Model I suggested elimination could be achieved for the baseline strategy, and all 266

model-fit combinations agreed on vector control achieving elimination by 2030. 267

Substantial improvement in elimination probabilities under enhanced passive detection 268

in Models I and W contrasted to results of Models S and Y where no significant changes 269

were found [(S1 Table).](#_bookmark10) The higher disparity among models in predicting elimination 270

probabilities under enhanced passive detection reflects the influence of structural 271

assumptions, in both HAT transmission dynamics but also in modeling control activities 272

that can lead to such different projections. 273

**Discussion** 274

A suite of independent mathematical models of HAT transmission were calibrated to 275

publicly available data from Bandundu province, DRC, to evaluate the effects of 276

different levels of data aggregation (disease stage and time series length) on model 277

performance and projections under current and improved control strategies. 278

**Informing staging data** 279

Distinguishing cases by stage is inherent to HAT epidemiology due to the way treatment 280

is currently administered. The results here showcase the impact that neglecting staging 281

information in data reporting has on subsequent model estimates and predictions. 282

Although similar patterns of annual incidence can be obtained from models calibrated 283

to unstaged and staged data, the underlying HAT dynamics for such similar incidence 284

patterns can differ strongly (as indicated by the proportion of stage 1 cases detected), 285

affecting any inference or projection on transmission risk. Contrasting projections 286

between staged and unstaged fits demonstrate how this aspect of HAT epidemiology can 287

impact our optimism about a particular strategy. A key example is that model 288

calibrations using staged data for Bandundu province strongly suggest that passive 289

detection rates have improved over time, whilst this is unobservable in the unstaged 290

data. 291

The data that countries use to determine their elimination policies for HAT are 292

usually limited and come mainly from screening activities. Our results emphasize the 293

need for incorporating staging information in data sets. With current screening 294

protocols, minimal additional effort in data recording is required to systematically 295

include staging, which would help to reduce uncertainties in assessing progress towards 296

elimination goals. 297

In the future, staging information may no longer be collected if new diagnostic tools 298

and treatments are stage-independent. For example, the new drug, fexinidazole [[23],](#_bookmark35) is 299

an all-in-one oral treatment for both stages recently approved by the European 300

Medicines Agency. However, until such tools become part of regularly implemented 301

policy, we emphasise the utility of making routinely collected staging data available. 302

Furthermore, if records of historically collected staging data exist, making these 303

available would substantially improve the reliability and predictive capability of 304

mathematical models. 305

**Time scales and informing on time surveys of active screening** 306

Over half of the total number of stage 1 cases reported between 2000 and 2012 come 307

from active screening. In general, as in this study, data is annually aggregated and so 308

the timing and the duration of active campaigns is unknown. As with current staging 309

data, this information is recorded at lower administrative levels, but is often lost in 310

higher level data sets. Systematically adding temporal data to current routine data 311

collection and collation would enable exploring a variety of case-specific time related 312

epidemiological factors such as the optimal frequency of interventions for achieving 313

specific local goals. 314

**Data delays** 315

There are routinely delays between case detection in the field and the availability of the 316

data for modeling purposes. The extreme example of a six years delay between data 317

collection and availability considered in this study, though unlikely due to improvements 318

in data availability, is chosen to demonstrate how the absence of up-to-date data 319

impacts model predictions. One or two missing years would still provide less accurate 320

results than up-to-date data, especially due to the lack of information on recent active 321

screenings. Nevertheless, we expect that model predictions generated with fewer missing 322

years would generate predictions more similar to predictions using the full data set than 323

those generated with six missing years as investigated in this study. 324

As we approach elimination, including recent data sets is necessary to better assess 325

the actual trends, as our results have suggested. Use of most recent data sets can be 326

sufficient to reproduce current epidemiological trends and the absence of these data sets 327

could affect model projections, especially for short timelines. Improvements in the time 328

between data collection and availability could enable modelling to provide more 329

up-to-date guidance and monitor for early-warning signs of obstacles on the road to 330

elimination. 331

**Province level data vs health zone level data** 332

Aggregated province-level data for endemic HAT regions lose information on the 333

geospatial variation of HAT incidence and screening coverage at lower administration 334

levels that are more compatible with the epidemiological scale of HAT transmission and 335

control. This may explain why although all model fits could capture the decreasing 336

trend in the number of reported cases, they could not reproduce certain peaks observed 337

in stage 1 cases (in 2002 and 2009) from active screening. The models assumed a fixed, 338

spatially homogeneous risk of transmission in Bandundu province, even though large 339

differences between central and southern health zones of Bandundu province had been 340

estimated for this period [[5].](#_bookmark17) Model W uses overdispersion parameters to capture the 341

variation in data between different years, so fitting to finer resolution data would likely 342

explain the source of this variation, and reduce the very large credible intervals from the 343

current results. 344

The peaks observed in the data could arise due to differences in HAT prevalence in 345

the geographical areas in which the active screening occurs between years, due to 346

differences in the quality or coverage of the screening campaigns between years, or 347

reflect true inter-annual variation in HAT epidemiology. Only detailed case data at a 348

finer spatial scale could help models to explore alternative assumptions, capture spatial 349

heterogeneity to better identify geographic reservoirs and improve predictions in global 350

HAT status. Model calibrations at a health zone or finer spatial scale are needed to 351

directly guide practical strategy planning at a local level. The WHO HAT Atlas is one 352

such valuable source of geolocated data in DRC (available upon request from the 353

WHO); and although staging information is typically not available for cases before 2015, 354

recent entries are staged. 355

**Complementary interventions to meet elimination goals** 356

Projections suggest that, at the province level, the continuation of traditional active and 357

passive screening is unlikely to be sufficient to attain EOT by 2030 across most models 358

and fits. The groups therefore simulated other complementary strategies which built 359

upon these baseline interventions to examine if any were sufficient to achieve this goal. 360

**Vector control** 361

Our results agree with previous modelling work indicating that potential strategies that 362

integrate vector control with medical interventions could accelerate progress towards 363

elimination, particularly in high endemicity or persistent hotspots [[10,](#_bookmark21) [11,](#_bookmark22) [13,](#_bookmark24) [16,](#_bookmark27) [17].](#_bookmark28) 364

This is consistent with reductions in HAT transmission reported after implementation of 365

cost-effective vector control methods in highly endemic locations in Guinea [[21]](#_bookmark33) and 366

Chad [[12].](#_bookmark23) 367

Although integrating vector control with current medical interventions at large 368

spatial scales such as Bandundu province (around 296,000 km2) may not be 369

operationally feasible, extending tsetse control interventions to active foci of HAT 370

transmission is feasible and likely to be efficient, particularly as transmission decreases 371

and programmes reduce screening activities. Vector control is currently being 372

implemented in hotspots in Bandundu (totalling approximately 3000 km2) and in the 373

West Nile region of Uganda (covering approximately 5000 km2). Regularly updated 374

epidemiological and entomological data from areas that have added this intervention to 375

HAT screening activities would facilitate the analysis of progress towards elimination 376

objectives, and provide an indication of protection against infection due to vector 377

control. 378

Additionally, secular changes, such as socio-economic development, urbanisation and 379

changes in land use, would likely lead to sustainable reductions in tsetse population 380

densities and consequently in HAT transmission, similarly to what has been reported for 381

other vector-borne diseases [[24].](#_bookmark36) The impact of such secular changes was not addressed 382

in this study but will become more important as transmission reduces further. 383

## Enhanced passive detection 384

This study found that, for passive detection, the increase in the ratio of stage 1 to stage 385

2 cases from 2006 onwards is an indicator of an already improving passive screening 386

system in this part of DRC. Although this is to be expected considering the increased 387

disease control efforts in the region, it is the first time that the improvement in the 388

passive detection rate has been quantified in a mechanistic modelling framework. 389

Furthermore, this trend is not observed in other former provinces of DRC for data from 390

the same period [[5].](#_bookmark17) 391

An improvement in time to detection is likely to have been driven by a combination 392

of causes, including improvements in access to care from increased awareness by the 393

population at risk and an increase in the number of health facilities; and improvements 394

in diagnostic tools including the use of digital technologies and RDTs (FIND 395

2016, [[19,](#_bookmark31) [25]).](#_bookmark37) Moreover, new “test-and-treat” strategies combining RDTs with 396

fexinidazole could lead to earlier and more cases treated. 397

Although our results suggest that enhanced passive detection could not be sufficient 398

to achieve short-term reduction goals, its associated sustained effect on reducing 399

transmission, projected by all models, indicates this strategy should be considered for 400

areas in Bandundu where past activities did not reduce HAT transmission as expected. 401

## Reactive screening 402

As the number of reported cases decreases, reactive case detection, i.e., deploying active 403

screening in a given area following detection of a case by passive screening, may be a 404

potential cost-effective strategy. Such a complementary strategy has already been 405

implemented in some regions of Uganda, Chad, Kongo Central and Angola. The 406

inherent spatial aspect of reactive screening implies that modelling elimination would 407

benefit greatly from geolocated and timed case data from different settings. This would 408

allow for an improved assessment of spatially-related measures of HAT transmission risk 409

to inform the appropriate targeting of interventions in space and time to achieve 410

elimination and prevent resurgence. 411

**Cost implications** 412

Naturally each of the different strategies mentioned above will affect the total cost of 413

HAT interventions not only in the Bandundu province but in any affected region, with 414

complementary strategies costing more than the baseline in the short-term due to the 415

extra resources used. Strategies which cost more in the short-term could result in earlier 416

EOT, and therefore may lead to earlier cessation of active screening interventions 417

compared to baseline. This could yield lower long-term costs, but it is non-trivial to 418

assess the costs of the complementary interventions explored in this study without 419

simulating cessation strategies and using a cost model. 420

Cost-effectiveness analyses using dynamic modelling frameworks require assessment 421

of health outcomes (such as years of life lost, and disability adjusted life years due to 422

disease) against a budget or willingness-to-pay threshold which can lead to strategies 423

which are not the least expensive being selected due to the relative gain in health 424

benefits [**?**]. This health-economic work is beyond the scope of the present study, which 425

primarily seeks to address the impact of disease stage data aggregation and truncated 426

data on model fitting and projections. Assessment of cost-effectiveness is clearly an 427

interesting and important objective for future analyses which aim to provide specific, 428

regional recommendations for strategy selection. Such work would ideally provide more 429

local strategy guidance (smaller than the province scale considered here) so that only 430

regions that require complementary interventions include them rather than assuming 431

blanket coverage of additional strategies across large areas. 432

**Extrapolations to other aspects of data** 433

Between 2011 and 2013, a study was performed to analyse the effects of coordinated 434

vector control (using tiny targets) and mass screening in an area of over 300 km2 in the 435

endemic focus of Boffa in Guinea [[21].](#_bookmark33) This study recorded highly detailed 436

pre-intervention geo-referenced data of households and inhabitants (familial clustering 437

via a unique code; name, sex and age of family members); annual screening data; and 438

vector and vector control data (15 targets/km2, estimates of initial tsetse fly densities, 439

trap location, survey duration); as well as subsequent updates including new families 440

and seasonal workers. Although such a comprehensive and rich data set can provide a 441

much deeper understanding of HAT epidemiology and the quantitative impacts of 442

control interventions on transmission, scaling up such studies to cover larger areas is 443

likely to be too costly to be feasible. A potential alternative would be to enrich current 444

standard data collection/collation from screening activities with questionnaires 445

providing additional demographic information on infected individuals (e.g. age, gender, 446

occupation, characteristics of house location) to better assess people at risk, their 447

participation in screening and their impact on transmission. Although this too may be 448

costly in higher transmission areas, it may be feasible close to elimination, where case 449

numbers are low and such enriched data would be particularly useful in identifying 450

potential new cases, as programmes move from untargeted active surveillance to reactive 451

strategies. 452

Table [4](#_bookmark5) summarises different, but not exhaustive, data which, if available, could be 453

used in modelling studies to identify potential beneficial adjustments in future activities 454

and to develop new frameworks for evaluating the path towards elimination and 455

post-elimination scenarios. 456

**Conclusions** 457

We investigated the role of the type and level of aggregation of epidemiological data on 458

recommended control strategy by analysing publicly available HAT case data using four 459

different mathematical models. Our results show that the lack of detailed 460

## Table 4. Summary of relevant data and its potential use in HAT modelling.

Data type

-First-final date of survey (AS)

-Date of presentation at health care centre (PD)

Collected, open access

Collected, available upon request

x

Not routinely collected

Potential use in HAT modelling

Inform time, number and duration of survey

Staging (province level) x x Inform staging ratios

Staging (village or health zone level)

x Inform staging ratios

Explore spatial-related

Geo-referenced x

Age x

Gender x

Occupation x

measures of HAT transmission risk

-Identify at-risk population

-Assess heterogeneity in screening participation

-Identify at-risk population

-Assess heterogeneity in screening participation

-Identify at-risk population

-Assess heterogeneity in screening participation

Socio-economic indicators x Identify at-risk population Better understand feeding

Presence of alternative sources of blood meals (e.g. pigs)

x behaviour of tsetse flies to investigate potential roles of

animal reservoirs

Family clustering x Spatial modeling to better identify foci

The list is not exhaustive. Abbreviations: AS: active screening; PD: passive detection.

epidemiological information, particularly missing staging or truncated time series data, 461

impacts model recommendations for strategy choice: it can increase our prediction 462

intervals and either over or underestimate effectiveness of baseline and intensified 463

interventions. 464

Our study suggests that improved availability of epidemiological data, particularly 465

longer time series which include recent data and information on disease stage, would 466

reduce uncertainties in the prediction of future HAT dynamics. In particular, staging 467

data allow a better estimate of the improvements made in passive detection, and 468

subsequent reduction in HAT transmission. Given the highly focal nature of HAT, we 469

expect that models fitted to recent staged data at smaller spatial scales (e.g. health 470

zone level) will provide valuable information for local planning, monitoring and 471

adapting HAT interventions to reduce transmission and achieve elimination. 472

# Supporting information 473

## S1 Text. Remarks on former Bandundu province case report data. 474

## S2 Text. Model descriptions. 475

**S3 Text. Projections on new infections.** Projections on the annual incidence of 476

new infections for all combinations of models and data sets. 477

**S4 Text. Projections on case reporting.** Projections on the annual HAT cases 478

for all combinations of models and data sets. 479

## S1 Table. Probability of elimination (zero transmission) by 2030 with a 480

## weaker threshold. 481

**S1 Fig. Stage 1 reporting in active screening.** Proportion of stage 1 to total 482

cases reported from active screening, and the corresponding estimation for a baseline 483

strategy under different fitting. The posterior median is shown as a point. Dashed lines 484

indicate projections based on fit to subset staged data. 485

**S2 Fig. Stage 1 reporting in passive detection.** Proportion of stage 1 to total 486

cases reported from passive detection, and the corresponding estimation for a baseline 487

strategy under different fitting. The posterior median is shown as a point. Dashed lines 488

indicate projections based on fit to subset staged data. 489

## S1 Code. Models code. 490

# Acknowledgements 491

The authors thank WHO HAT team for facilitating access to the HAT Atlas data; and 492

to Jos´e R. Franco and Gerardo Priotto for helpful discussion and comments on this 493

manuscript. Calculations for the Model S were performed at the sciCORE 494

[(http://scicore.unibas.ch/)](http://scicore.unibas.ch/)) scientific computing core facility at the University of Basel. 495

# References

1. Simarro PP, Cecchi G, Franco JR, Paone M, Diarra A, Priotto G, et al. Monitoring the progress towards the elimination of gambiense human African trypanosomiasis. PLoS neglected tropical diseases. 2015;9(6):e0003785.
2. Franco JR, Cecchi G, Priotto G, Paone M, Diarra A, Grout L, et al. Monitoring the elimination of human African trypanosomiasis: Update to 2016. PLoS neglected tropical diseases. 2018;12(12):e0006890.
3. Lehane M, Alfaroukh I, Bucheton B, Camara M, Harris A, Kaba D, et al. Tsetse control and the elimination of Gambian sleeping sickness. PLoS neglected tropical diseases. 2016;10(4):e0004437.
4. World Health Organisation. Number of new reported cases of human African trypanosomiasis (T.b. gambiense); 2018. <http://apps.who.int/neglected_diseases/ntddata/hat/hat.html>.
5. Lumbala C, Simarro PP, Cecchi G, Paone M, Franco JR, Mesu VKBK, et al. Human African trypanosomiasis in the Democratic Republic of the Congo: disease distribution and risk. International journal of health geographics. 2015;14(1):20.
6. Basa´n˜ez MG, McCarthy JS, French MD, Yang GJ, Walker M, Gambhir M, et al. A research agenda for helminth diseases of humans: modelling for control and elimination. PLoS neglected tropical diseases. 2012;6(4):e1548.
7. malERA Consultative Group on Modeling, et al. A research agenda for malaria eradication: modeling. PLoS medicine. 2011;8(1):e1000403.
8. malERA Refresh Consultative Panel on Combination Interventions, Modelling. malERA: An updated research agenda for combination interventions and modelling in malaria elimination and eradication. PLoS medicine. 2017;14(11):e1002453.
9. Rock KS, Torr SJ, Lumbala C, Keeling MJ. Quantitative evaluation of the strategy to eliminate human African trypanosomiasis in the Democratic Republic of Congo. Parasites & vectors. 2015;8(1):532.
10. Rock KS, Pandey A, Ndeffo-Mbah M, Atkins K, Lumbala C, Galvani A, et al. Data-driven models to predict the elimination of sleeping sickness in former Equateur province of DRC. Epidemics. 2017;18:101–112.
11. Rock KS, Torr SJ, Lumbala C, Keeling MJ. 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. 2017;11(1):e0005162.
12. Mahamat MH, Peka M, Rayaisse JB, Rock KS, Toko MA, Darnas J, et al. Adding tsetse control to medical activities contributes to decreasing transmission of sleeping sickness in the Mandoul focus (Chad). PLoS neglected tropical diseases. 2017;11(7):e0005792.
13. Pandey A, Atkins KE, Bucheton B, Camara M, Aksoy S, Galvani AP, et al. Evaluating long-term effectiveness of sleeping sickness control measures in Guinea. Parasites & vectors. 2015;8(1):550.
14. Institut National de la Statistique, Ministere du Plan et Revolution de la Modernite de la Republique Democratique du Congo. Annuaire statistique 2014; 2015. [http://www.ins-rdc.org](http://www.ins-rdc.org/).
15. Stone CM, Chitnis N. Implications of heterogeneous biting exposure and animal hosts on Trypanosomiasis brucei gambiense transmission and control. PLoS computational biology. 2015;11(10):e1004514.
16. Sutherland CS, Stone CM, Steinmann P, Tanner M, Tediosi F. Seeing beyond 2020: an economic evaluation of contemporary and emerging strategies for elimination of Trypanosoma brucei gambiense. The Lancet Global Health. 2017;5(1):e69–e79.
17. Rock KS, Ndeffo-Mbah ML, Castan˜o S, Palmer C, Pandey A, Atkins KE, et al. Assessing strategies against Gambiense sleeping sickness through mathematical modeling. Clinical infectious diseases. 2018;66(suppl 4):S286–S292.
18. Lumbala C, Bessell PR, Lutumba P, Baloji S, Bi´eler S, Ndung’u JM. Performance of the SD BIOLINE® HAT rapid test in various diagnostic

algorithms for gambiense human African trypanosomiasis in the Democratic Republic of the Congo. PloS one. 2017;12(7):e0180555.

1. Lumbala C, Bi´eler S, Kayembe S, Makabuza J, Ongarello S, Ndung’u JM. Prospective evaluation of a rapid diagnostic test for Trypanosoma brucei gambiense infection developed using recombinant antigens. PLoS neglected tropical diseases. 2018;12(3):e0006386.
2. Tirados I, Esterhuizen J, Kovacic V, Mangwiro TC, Vale GA, Hastings I, et al. Tsetse control and Gambian sleeping sickness; implications for control strategy. PLoS neglected tropical diseases. 2015;9(8):e0003822.
3. Courtin F, Camara M, Rayaisse JB, Kagbadouno M, Dama E, Camara O, et al. Reducing human-tsetse contact significantly enhances the efficacy of sleeping sickness active screening campaigns: a promising result in the context of elimination. PLoS neglected tropical diseases. 2015;9(8):e0003727.
4. Stanton MC, Esterhuizen J, Tirados I, Betts H, Torr SJ. The development of high resolution maps of tsetse abundance to guide interventions against human African trypanosomiasis in northern Uganda. Parasites & vectors. 2018;11(1):340.
5. Chappuis F. Oral fexinidazole for human African trypanosomiasis. The Lancet. 2018;391(10116):100–102.
6. Tusting LS, Willey B, Lucas H, Thompson J, Kafy HT, Smith R, et al. Socioeconomic development as an intervention against malaria: a systematic review and meta-analysis. The Lancet. 2013;382(9896):963–972.
7. Wamboga C, Matovu E, Bessell PR, Picado A, Bi´eler S, Ndungˆa€™u JM. Enhanced passive screening and diagnosis for gambiense human African trypanosomiasis in north-western Uganda–Moving towards elimination. PloS one. 2017;12(10):e0186429.