Livestock and the Epidemiology of Sleeping Sickness: Mechanisms and Implications

Julianne Meisner

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Peter Rabinowitz, Chair

Jonathan Wakefield

Ali Rowhani-Rahbar

Program Authorized to Offer Degree:
Epidemiology

University of Washington

Abstract

Livestock and the Epidemiology of Sleeping Sickness: Mechanisms and Implications

Julianne Meisner

Chair of the Supervisory Committee:
Professor Peter Rabinowitz
Departments of Environmental and Occupational Health Sciences, Epidemiology, and Global Health

In recent decades, remarkable progress in the control of Human African trypanosomiasis (HAT)—a bloodborne protozoal parasite transmitted by the tsetse fly (*Glossina* species)—has led the WHO to set targets for elimination as public health problem (EPHP) by 2020, and elimination of transmission (EOT) by 2030. Global EPHP targets were met in 2018, however most endemic countries are not yet eligible for national EPHP validation, and there are significant challenges to achieving EOT goals.

Two forms of HAT, which are geographically- and epidemiologically-distinct, exist: the chronic form, caused by *Trypanosoma brucei gambiense* (gHAT) and endemic in western and central Africa, and the acute form, caused by *T. b. rhodesiense* (rHAT) and endemic in eastern and southern Africa. Due to the known importance of animal reservoirs for rHAT EOT targets are set for gHAT alone, resulting in significantly lower investment in rHAT surveillance and control compared with gHAT. Combined with rHAT's acute progression this results in significant underreporting, raising concerns that rHAT will emerge as a major public health problem once gHAT EOT is achieved and donor attention moves away from HAT. With regards to gHAT, uncertainty surrounding animal reservoirs—in particular domestic pigs—as well as latent human reservoirs and undercoverage of high-risk groups by active surveillance activities threaten both the probability of EOT, and re-emergence

following EOT.

In this dissertation, we used data from the WHO Atlas of HAT, spatial epidemiologic methods, methods drawn from the potential outcomes framework of causal inference, and a stochastic compartmental model to estimate the effect of pig density on HAT risk, to evaluate the feasibility of rHAT EOT with control of domestic animal reservoirs alone, and to decompose the livestock density - HAT effect into three components: (1) the reservoir effect, whereby domestic cattle and pigs infected with trypanosomes serve as a source of human infection, (2) the zooprophylactic effect, whereby tsetse fly preference for livestock protects humans from bites—and therefore trypanosome infection—when livestock are nearby, and (3) the environmental change effect, whereby livestock keeping results in environmental changes that in turn modify tsetse distribution and HAT risk. We conducted this work in four high-burden countries which do not meet WHO criteria for EPHP validation: Uganda (gHAT and rHAT), South Sudan (gHAT), Malawi (rHAT), and Democratic Republic of Congo (DRC, gHAT).

Our results suggest pigs may indeed be gHAT reservoirs and, if so, EOT will not be achieved without intervention—trypanocide or, preferably, insecticide treatment—on this reservoir. With regards to rHAT, we found control of the domestic cattle and pig reservoir is critical to control of the disease, in particular in Uganda, but does not lead to EOT. We found evidence of a zooprophlyactic effect in Malawi and South Sudan for both cattle and pigs and in gHAT foci in Uganda for cattle, however we did not detect compelling evidence of an environmental change effect.

These results point to the utility of a One Health approach to HAT control, and represent an important contribution to the HAT literature and the efforts of National Sleeping Sickness Control Programs in the study countries. In conjunction with the high-resolution livestock density maps we have produced, our findings will support targeted delivery of expanded and/or enhanced HAT control efforts. Delivering these efforts in a One Health framework,

whereby control of animal African trypanosomiasis is coordinated with that of HAT, will increase the likelihood and sustainability of gHAT elimination and ensure rHAT does not subsequently emerge as a major public health problem, reducing the burden of this highly-fatal and poverty-reinforcing disease.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY OF KEY TERMS AND ACRONYMS

AAT: Animal African trypanosomiasis, infection with a bloodborne protozoal parasite transmitted by the tsetse fly. An important sources of animal morbidity and production losses in sub-Saharan Africa.

CLUSTER: While cluster can take many definitions, in this dissertation "cluster" refers to a pixel on the corresponding livestock map. In Chapter 2, "cluster" is used instead to refer to survey cluster, defined below.

COUNTERFACTUAL: A scenario which is counter to fact. Under the potential outcomes framework for causal inference, counterfactuals are generally used to reflect the outcome that would have occurred if exposure (or mediator) was fixed at a given level.

CONFOUNDER: A variable which is a cause of the outcome and is associated with exposure but is not on the causal pathway (i.e., is not a result of exposure).

DAG: Directed acyclic graph, a directed graph with no directed cycles. In a causal inference framework, nodes represent variables, and vertices represent causal relationships between these variables.

DHS: Demographic and Health Survey.

EOT: Elimination of transmission, targeted for 2030 for gHAT. Defined as (1) an 80% reduction in high and moderate risk areas relative to the 2000 baseline, (2) > 50% of the at-risk population located within 1 hours' travel time to a fixed health facility capable of gHAT diagnosis, and (3) > 95% of the at-risk population located within 5 hours' travel time to a fixed health facility capable of gHAT diagnosis.

EPHP: Elimination as a public health problem. Targeted for 2020 for gHAT and rHAT, global EPHP is defined as (1) less than 2,000 cases reported annually and (2) a 90% (638,000 km^2) reduction over 2016-2020 in the area at risk reporting 1 or more cases/10,000 people/year, compared to the 2020-2004 baseline. National EPHP is defined as fewer than 1 reported case/10,000 inhabitants/year, averaged over a 5-year period in each health district of the country, in conjunction with adequate, functional control and surveillance.

E-VALUE: The minimum strength of association an unmeasured confounder would need to have with exposure and outcome, on the relative risk scale and conditional on the measured covariates, to "explain" the exposure-outcome effect detected.

ECOLOGICAL STUDY: A study in which the unit of analysis is a group of individuals, commonly defined by geographic location.

GHAT: Human African trypanosomiasis caused by *Trypanosoma brucei gambiense*, the chronic form endemic in west and central Africa. Importance of animal reservoirs is uncertain.

GMRF: Gaussian Markov random field. Gaussian latent fields exhibiting this property are have the continuous Markov property, whereby a particular location's spatial process depends only on that of its neighbors, not the entire map, resulting in a "sparse" precision matrix.

GP: Gaussian process. A latent process that is an n dimensional Gaussian distribution for any integer n.

HAT: Human African trypanosomiasis, infection with *Trypanosoma brucei*, a protozoal parasite transmitted by the tsetse fly. There is no vaccine for HAT, and the disease is generally fatal without treatment.

HAT FOCUS: A historically-stable and geographically-distinct location of active HAT transmission.

ICAR: The intrinsic conditional autoregressive model, a spatial model used for discrete space (areal approaches). The ICAR model smooths each area's random effect to that of its neighbors, with more smoothing performed for areas with fewer neighbors.

IDENTIFIABILITY CRITERIA: A set of assumptions which confer identifiability of causal effects from observational data: exchangeability (no unmeasured confounding), positivity (every unit has a non-zero probability of receiving every exposure level under consideration), and the stable unit treatment value assumption, comprised of consistency (well-defined exposure) and no interference (each unit's counterfactual outcome is independent of the all other units' counterfactual outcomes).

ITC: Insecticide treatment of cattle; used here to refer to treatment of domestic cattle and pigs with synthetic pyrethroids.

LIVESTOCK REVOLUTION: Increases in livestock production anticipated in response to population growth and economic development, largely met through intensification of livestock production systems.

LST: Land surface temperature, a remote-sensed indicator of surface temperature.

MEDIATOR: A variable which lies in the middle of the causal pathway from exposure to outcome.

NDVI: Normalized Difference Vegetation Index, a remote-sensed indicator of vegetation coverage.

PARAMETRIC G-FORMULA: A model-based standardization approach to estimating an exposure-outcome effect in the face of longitudinal data with time-varying exposures and confounders and treatment-confounder feedback. If the identifiability criteia are met, the resulting estimand can be treated as causal.

PRIOR: Used in Bayesian inference, a probability distribution that reflects one's beliefs about a parameter or hyperparameter.

RASTER: A representation of spatial data as a matrix of cells arranged in a grid, where each cell's value represents the level of the variable in question at the corresponding location.

RHAT: Human African trypanosomiasis caused by *T. b. rhodesiense*, the acute form endemic in east and southern Africa. Domestic and wild animal reservoirs are of known importance.

SMALL AREA ESTIMATION: An approach to estimating a parameter for areas with sparse data, commonly using Bayesian hierarchical models to perform spatial smoothing and thereby stabilize variance.

SPDE: Stochastic partial differential equation. A Matérn GP can be represented as the solution to an SPDE, which, under certain constraints, is a GMRF. The SPDE can be solved using the finite element method by discretizing the spatial field using a fine triangular mesh and representing the GP as a weighted sum of the resulting piecewise linear basis functions, where the wights are multivariate Gaussian and picked to have a sparse precision matrix. The spatial process can then be projected from the mesh vertices to a regular grid, generating a high-resolution approximation of the spatial field.

SURVEY CLUSTER: The smallest geographical sampling unit of a household survey, typically an enumeration area consisting of approximately 100 households.

TREATMENT-CONFOUNDER FEEDBACK: A relationship between exposure (treatment) and confounders that can arise in longitudinal data, whereby the confounder is a cause of exposure at one point in time, and a result of exposure at a later point in time, rendering it both a confounder and a mediator.

TT: Trypanocide treatment of domestic animals.

VC: Vector control. Includes stationary (traps and targets) and live (insecticide-treated domestic animals) baits, aerial spraying, and sterile male insect releases.

ZOOPROPYLAXIS: The protective effect exerted by the presence of animals, when an insect vector of a disease that infects humans exhibits strong preference for animal hosts. These animal hosts do not need to be reservoirs of the pathogen in question.

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to the World Health Organization's Department of Control of Neglected Tropical Diseases for use of their Atlas of HAT data, and to the HAT research community, in particular Prof. Eric Fevré and Dr. Richard Selby, which has lent a remarkable amount of support to this relative outsider.

The author is also profoundly thankful for efforts of her Doctoral Supervisor Committee, in particular Profs. Peter Rabinowitz (chair) and Jonathan Wakefield. This dissertation owes so much to the breadth and depth of their expertise, and to their patience and support. Their skill as mentors has provided the author with fantastic examples of the mentor she would like to be—and meant she is submitting this dissertation without a single committee-related cautionary tale!

Most importantly, the author would like to acknowledge the efforts of her collaborators in the study countries, in particular Dr. Marshall Lemerani of the Malawi Ministry of Health, Mr. Acaga Ismail Taban of IntraHealth International in South Sudan, Dr. Agapitus Kato of the Ministry of Agriculture, Animal Industry and Fisheries in Uganda, and Drs. Jean-Marie Vianney and Erick Mwamba Miaka of Programme National de Lutte contre la Trypanosomiase Humaine Africaine in the Democratic Republic of Congo. Without the generosity of time and support provided by these individuals, this dissertation would not have been possible.

DEDICATION

This dissertation is dedicated to my family: to my parents, who have been boundlessly supportive of my somewhat circuitous career trajectory. From the designs of a stubborn teenager intent on moving to the UK for university, to a later (and more academic) mission of becoming a veterinarian, and eventual arrival at public health, my appreciation for your ability to roll with the many punches I've thrown over the decades cannot be understated. To my sister, who has lent an enormous amount of support to me as a statistician and friend, while navigating early motherhood and academia alongside me. From our study sessions in the Health Sciences Library in the early days of my MS program and the waning days of your PhD, to Google Chat sessions at all hours of the day and night answering my statistics questions and searching with me for that elusive key to infant (and toddler) sleep, thanks to you I have not had to go any of this alone. To my son, Virgil, and my soon-to-be-born daughter, who together have taught me productivity and time off can coexist, and that sleep is really just a luxury. Despite how much I love my work, the opportunity to be your mother has shown me how to prioritize what will still be there on the days that my work doesn't love me quite as much as I love it. And finally, to my husband, Mike, who has been pulling way more than his weight in terms of housework and parenting tasks even before this pandemic, allowing me time to get my work done and be a more relaxed and present mother than I ever could have been otherwise. Your efforts—including spending four months squeezing your work into naptimes and the post-bedtime hours when we had no other childcare, without any resentment or complaint—are the sole reason I have been able to finish this dissertation on time with a toddler in a pandemic. Quite literally, none of this would be possible without you.

Chapter 1
INTRODUCTION

1.1 Current epidemiology of human African trypanosomiasis

An estimated 61 million people are at risk of human African trypanosomiasis (HAT) worldwide, with 15 million at moderate risk and 1.2 million at high or very high risk [1].Commonly known as sleeping sickness, HAT is a largely-bloodborne (with some evidence of dermal infection) protozoal parasite transmitted by the tsetse fly. There is no vaccine available for HAT and infection is generally fatal without treatment, however treatment requires inpatient care followed by six months of convalescence, imposing often-catastrophic health expenditures on patients and their families [2].

HAT is a highly focal disease, characterized by distinct outbreaks ranging in size from a single village to an entire region. Intensity can vary within larger foci, and the the geographic distribution of foci can change over time due to human migration, environmental changes, and reservoir movement [3]. The two major forms of HAT are also geographically-distinct: the chronic form caused by *Trypanosoma brucei gambiense* (gHAT), endemic in west and central Africa and considered largely anthroponotic (transmitted human-to-human), and the acute form caused by *T. b. rhodesiense* (rHAT), endemic in east and southern Africa and known to have important animal reservoirs. The two forms coexist only in Uganda, where gHAT occupies a focus in the northwest of the country, and rHAT occupies a distinct focus in the southeast [4]. The majority of gHAT cases occur in Democratic Republic of Congo (DRC), Central African Republic (CAR), Chad, and South Sudan, with nearly 40,000 out of the 48,000 total cases reported from 2009-2018 occurring in DRC. Over 2014-2018, the estimated populations at risk of gHAT in these countries were 24.5%, 8.5%, 1.5%, and 8.2% of the total national population, respectively. In the case of rHAT, the majority of case

reports from 2009-2018 came from Uganda (564) and Malawi (263). In these countries, 4.2% and 14.4% of the total national population is estimated to be at risk of rHAT, respectively [5].

Elimination of HAT was nearly achieved in the 1960s-1970s, however a collapse in control activities and political instability following decolonization lead to a resurgence in the 1990s [6]. Due to a lack of surveillance and funding for treatment programs, HAT incidence climbed to alarming levels in the late 20th century. This resulted in a 1997 WHO resolution to improve access to diagnosis and treatment and strengthen surveillance and control. In 2001, a public-private partnership between WHO and sanofi-aventis was launched, which allowed WHO to distribute drugs for HAT treatment free-of-charge, and to reinforce capacity for screening populations at risk. As a result of this partnership, efforts from NGOs and National Sleeping Sickness Control Programmes of endemic countries, and cessation of political unrest in endemic countries (particularly Angola, DRC, and South Sudan), field activities were scaled up, reducing transmission and producing more reliable estimates of HAT distribution. With elimination again within reach, in 2012 the WHO set targets for elimiation as a public health problem (EPHP) by 2020, and elimination of gHAT transmission (EOT) by 2030. Because of its zoonotic nature and limited options for control, rHAT is excluded from EOT goals.

EPHP targets are defined at the global and national levels, with global EPHP defined as (1) less than 2,000 cases reported annually and (2) a 90% (638,000 km^2) reduction over 2016-2020 in the area at risk reporting 1 or more cases/10,000 people/year, compared to the 2020-2004 baseline. Three secondary indicators are also monitored: (a) the geographic distribution of HAT, (b) the size of the at-risk population, and (c) coverage of the at-risk population by control and surveillance activities. National EPHP is defined as fewer than 1 reported case/10,000 inhabitants/year, averaged over a 5-year period in each health district of the country, in conjunction with adequate, functional control and surveillance. The latter is defined by intensity and effectiveness of HAT control and surveillance activities, including medical interventions (active and passive surveillance), vector control, and animal African

trypanosomiasis (AAT) control [5]. EPHP indicators are applied to both gHAT and rHAT.

Conversely, EOT goals exclude rHAT due to its zoonotic nature and perceived lack of effective control measures. EOT is defined as (1) an 80% reduction in high and moderate risk areas relative to the 2000 baseline, (2) > 50% of the at-risk population within 1 hour's travel time of a fixed health facility capable of gHAT diagnosis, and (3) > 95% of the at-risk population within 5 hours' travel time of a fixed health facility capable of gHAT diagnosis.

Over the past decade, remarkable progress has been made in the control of HAT: 977 cases were reported in 2018, down from over 25,000 cases in 2000. This represents a 98% reduction in gHAT cases and a 97% reduction in rHAT cases from 2000 to 2018, and achievement of global EPHP targets ahead of the 2020 goal [5]. With 2.7 million people screened in 2018, medical interventions, including active screening and passive surveillance, are the mainstay of elimination strategies in most HAT-endemic areas. Recent approval of fexinidazole, an oral 10-day treatment approved for both stages of gHAT (excluding severe cases), replaces the need for inpatient care and is expected to improve access to and uptake of treatment, increasing effectiveness of medical intervention-based approaches and feasibility of achieving EOT by 2030 [7]. Other control methods, including application of top-down strategies targeting the tsetse vector or animal reservoirs, are infrequently deployed.

Despite this progress, in some DRC foci transmission remains persistent even with ongoing control efforts. This is in part due to lack of participation in active surveillance activities by high-risk demographic groups, with both observational studies and mathematical modeling providing evidence that young adult men with high occupational risk of tsetse exposure are less likely to present for active surveillance than those at lower exposure risk [2, 5]. In foci where existing medical-only approaches have been insufficient, addition of door-to-door screening and expanded access to newly-available rapid diagnostic tests may be adequate to achieve EPHP and EOT targets, however multiple models have shown that improving passive surveillance systems and addition of vector control (VC), including insecticide treatment of livestock [8], could lead to EOT by 2030 with higher probability than medical-only strategies even if VC is only 60% effective [9]. Furthermore, the cost-effectiveness of VC and

its superiority in this regard to fexinidazole and rapid diagnostic tests has been repeatedly demonstrated through mathematical models [7]. However, in some foci characterized by civil conflict—in particular in DRC, CAR, and South Sudan—insecurity has hampered control efforts to the extent that active surveillance has never been conducted, and EPHP and EOT remain out of reach [2]. Furthermore, in Guinea, a lapse in control programs during the West Africa Ebola outbreak led to a spike in cases in 2016-2017, and it is possible the COVID-19 pandemic will lead to similar resurgence. Finally, uncertainty surrounding the role of animal reservoirs—with some evidence that pigs may be competent reservoirs—and asymptomatic human cases in gHAT transmission, as well as high influxes of refugees or internally-displaced persons into formerly endemic areas, threaten resurgence after EOT [2, 10].

In contrast with gHAT, active surveillance is not conducted in rHAT foci, leaving passive surveillance the sole means of case detection. While reported cases are few in number and declining, a significant proportion (18%) of 2018 cases occurred in international travelers to endemic areas diagnosed after they returned home, raising concerns reported cases do not accurately reflect the true epidemiologic situation [5]. Consistent with this, underdetection of rHAT is believed to have increased in recent years due to overall weakening in control and surveillance, and to increased uptake of rapid diagnostic tests for malaria diagnosis, reducing incidental diagnosis of rHAT on blood smears and availability and use of microscopes.

Thus, while EPHP may be achievable globally and in a subset of HAT-endemic countries, security challenges and poor active surveillance coverage leave EPHP out of reach for others. Further, EOT will be more challenging than EPHP. First, as case counts continue to decline control will shift from active surveillance to passive surveillance and reactive screening. This requires integration of gHAT surveillance into often-weak rural health systems, and recent mathematical modeling work indicates that only a minority of stage 1 gHAT cases are detected via passive surveillance [11]. The effectiveness of passive surveillance will be further challenged by expected losses in health staff expertise with declines in case counts. Second, poor understanding of asymptomatic human carriers and animal reservoirs for gHAT will threaten both EOT, and, similar to the case of guinea worm eradication, re-emergence fol-

lowing EOT [10]. Finally, due to its zoonotic nature, rHAT EOT is unlikely to be attained with currently-used technologies [1].

The proposed research seeks to address the second two challenges to EOT goals, by describing the role of pigs and cattle in the epidemiology of HAT.

1.2 Control measures for HAT

1.2.1 Trypanocide treatment of cattle

In addition to HAT, tsetse flies also transmit animal African trypanosomiasis (AAT), a high-morbidity disease of livestock that manifests as acute or chronic wasting [12]. AAT is broadly distributed in sub-Saharan Africa, where it is widely considered to be the single greatest constraint to increased livestock production, with knock-on effects on development of sustainable crop agricultural systems due to the absence of livestock manure to fertilize crops and livestock power to till fields [13, 14, 15, 16, 17]. While *T. b. rhodesiense* can cause AAT, clinical signs are typically mild in native breeds, in contrast with AAT caused by non-zoonotic trypanosome species including *T. congolense* and *T. vivax* [18].

Where domestic animals are important reservoirs of rHAT, management of AAT through use of trypanocides and/or insecticides can also contribute to HAT control. Trypanocide treatment of cattle (TT) of cattle is a mainstay of farmer-led AAT management, with these drugs being readily obtained low-cost at local markets [12, 19]. However, most currently licensed veterianry trypanocides were developed in the first half of the 20th centuy, and reports of resistance are increasing, in particular for *T. congolense* [20, 21]. Furthermore, mathematical modeling work has shown that even in settings where cattle and humans are the only tsetse fly host and 30% of rHAT cases are detected and treated, TT coverage must be high to interrupt transmission (i.e., 86% of cattle kept constantly immune to *T. b. rhodesiense*) [22]. This is consistent with coverage levels required to control rHAT transmission in field studies of mass chemoprophylaxis of cattle [23, 24, 25].

In addition to the high coverage required for TT to lead to rHAT control even when tsetse

take a significant proportion of bloodmeals from cattle, and concerns about drug resistance, TT does not contribute to HAT control in foci where cattle are not an important reservoir of HAT. Vector control, including insecticide treatment of cattle (ITC), is an attractive alternative.

1.2.2 Vector control

After World War II, cheap, persistent organochlorine insecticides such as DDT formed the basis for large-scale tsetse fly eradication campaigns. These were largely successful, however due to environmental concerns and the end of colonial rule tsetse control shifted in the 1990s from centrally organized campaigns to decentralized "field-by-field" approaches that fell to individual livestock keepers, often relying on a single control tactic selectively applied to areas with high economic value to that individual. These, in general, resulted in only short-term improvements in tsetse and trypanosomiasis burden [13, 19, 26, 27].

In 2000, the African Heads of State and Government at the AU Summit created the Pan African Tsetse and Trypanosomiasis Campaign (PATTEC), in collaboration with the African Development Bank. The mission of PATTEC is to eradicate the trypanosomiases and tsetse flies [28]. Under this framework, selective and integrated application of more environmentally-friendly approaches on an area-wide basis has resulted in sustainable management of tsetse populations where invasion pressure is low [19].

PATTEC's efforts have largely focused on areas of high AAT burden. In gHAT settings where tsetse eradication is not targeted, the goal of vector control is to reduce tsetse density below the threshold required for transmission, and maintain this for 4-5 years while medical interventions control the human reservoir [29].

Insecticide treatment of cattle ("live bait") Recognized in the early 2000s to achieve control of AAT and HAT vectors at one tenth the cost of traditional insecticide application approaches (spot-ons, pour-ons, entire body spraying, or dipping), restricted application of insecticides (RAP), specifically synthetic pyrethroids, to the preferred feeding sites (belly and

lower legs) and hosts (older, larger cattle) of tsetse flies has since become the focus of ITC approaches. In addition to its cost-effectiveness, RAP reduces environmental contamination and achieves control of tickborne diseases while still allowing the low levels of exposure required to maintain host tolerance [30, 31, 32, 33, 34, 35].

Successful use of ITC to control HAT and AAT vectors depends on cattle distribution and size of the covered area, with large gaps needing to be closed with stationary baits [36] and greater successes documented for areas >2000km^2 than areas <100km^2 [37], as smaller areas have a larger perimeter:area ratio providing more opportunity for re-invasion. However, in areas where these criteria are not met and thus tsetse control is insufficient, ITC may still achieve trypanosomiasis control [38]: ITC combined with TT was successfully used in Uganda in the early 2010s to reduce the geographical range of *T. b. rhodesiense* and prevalence of AAT [39, 40]. Currently, there are plans to implement intensive mass cattle interventions (trypanocides and insecticides) to 2.7 million cattle in at-risk districts of Uganda [25].

Targets and traps ("stationary bait") Insecticide-impregnated tsetse traps and targets have been a cornerstone of tsetse control campaigns in many countries. Targets, large 1m^2 pieces of blue and/or black cloth impregnated with odor attractants, have successfully controlled AAT and rHAT vectors at low cost. However, no effective odor attractants have been identified for gHAT vectors, which hunt via sight rather than smell, leaving traps—large, biconical devices which are far more costly than targets—the only option for stationary baits until tiny targets were developed in the early 2010s. Found to improve cost-effectiveness for gHAT control by six-fold, tiny targets can be carried in a backpack and deployed by a single person [8, 41, 42, 42, 43, 44]. Such approaches are particularly useful in areas of high tsetse challenge and with insufficient penetration of medical interventions [44]. Addition of tiny targets to HAT control in the Mandoul focus in Chad, a focus of persistent transmission despite several years of active screening and passive surveillance with often high coverage, was found to account for 70.5% of the reduction in reported cases over a 2 year period, with the rest due to medical interventions [45].

Other approaches Ground spraying and sequential aerosol treatment have been largely replaced by traps, targets or ITC, due to operational demands and concerns over environmental contamination [46, 47]. The exception is sequential aerosol treatment (SAT), which is still used in select applications for AAT control under the PATTEC framework. SAT is a non-residual technique based on repeated spraying of deltamethrin at low concentrations, timed to kill females which have emerged after the last treatment and before they deposit their first pupae. As SAT uses low (non-residual) doses, tsetse from neighboring areas can invade within a few days [48]. Successful application of SAT for riverine species, the predominant vectors of HAT, is more challenging as temperature inversions required for SAT occur less frequently in gallery forests than savannah (the preferred habitat of AAT vectors), dense vegetation filters the aerosol, and low windspeed due to tree canopy reduces dispersion [48].

Sterile male insect release (SIT) approaches leverage the fact that female tsetse flies typically mate only once during their lifetime, and thus mating with a sterile male will reduce birth rate. However as SIT does not prevent re-invasion it is most effective for ecological islands, and mathematical modeling work indicates that if bait methods can sufficiently increase mortality, there is little place for SIT and its attendant costs and logistical demands [49]. Nevertheless, SIT preceded by ITC and insecticide treated targets to increase tsetse mortality successfully eradicated *G. austeni* Newstead from Unguja Island (Zanzibar) in 1994 [50].

1.2.3 Medical interventions

Early modeling of various control strategies showed medical interventions, defined by active and passive case detection and treatment, to be the optimal strategy for gHAT [51]. This approach was credited with the "successful" elimination of *T. b. gambiense* in the 1960s [26]. Briefly, these approaches rely on access to and engagement with active screening campaigns and health care systems capable of timely diagnosis and successful treatment. For gHAT, development of new diagnostics and chemotherapeutics is of priority, to optimize the efficacy of active screening and promote integration of surveillance into health care systems.

In September 2017, a new rapid diagnostic test (RDT) for HAT was commercially launched, supported by the Bill & Melinda Gates Foundation and the Swiss and UK governments [52]. During an active treatment campaign in DRC, sensitivity of this test (the SD BIOLINE HAT) was found to be significantly higher than the currently-used card agglutination test for trypanosomiasis (CATT; 92% versus 69.1%), with comparable specificity (97.1% versus 98%) [53]. However, as elimination nears, this modest loss in specificity will correspond to sizeable decreases in positive predictive value.

In November 2018, the European Medicines Agency adopted a positive opinion of fexinidazole for first- and second-stage treatment of gHAT in adults and children over 6 [54]. This approval supported marketing authorization application in endemic countries in the following year. Developed by the Drugs for Neglected Diseases initiative, fexinidazole is the first oral and stage-independent treatment for HAT, a development which reduces the need for hospitalization and staging via lumbar puncture. Unlike other HAT drugs, fexinidazole displays little to no non-specific cytotoxicity. Prior to its approval, treatment was comprised of injectable pentamidine in stage 1 and combined nifurtimox-eflornithine (NECT) in stage 2. While mortality for this combination is low, NECT has many side effects, and resistance potential has been demonstrated *in vitro*. As fexinidazole is not yet approved for rHAT, injectable protocols—suramin in stage I, melarsoprol in stage 2—remain the only approved option. In addition to the need for protracted inpatient care and monitoring via lumbar puncture, melarsoprol injections are extremely painful and the drug is highly toxic, producing encephalopathy in 10% of patients and death in half of encephalopathy cases [55].

1.3 Coordinated HAT and AAT control: untapped potential

Beyond the importance of animals (including domestic animals) as potential gHAT reservoirs, a single coordinated AAT and gHAT "trypanosomiasis control" strategy may support gHAT EPHP and EOT targets. Simo and Rayaisse (2015) argued that low prevalence of HAT has led to demotivation to participate in active screening following repeat negative results, demotivation among technicians for the same reason, and donor demotivation following few

to no new cases detected.

In contrast with large-scale control measures implemented for AAT in areas with significant livestock production activity, the epidemiologic status of AAT in gHAT foci, characterized by forested regions where animal keeping is small in scale, is not well-understood. However, in gHAT foci where AAT burden has been characterized prevalence is generally high, sometimes exceeding 70%, and mixed infections with *T. b. gambiense* and non-zoonotic trypanosomes have been detected in tsetse and animals [56]. Coordinated AAT and gHAT control, for instance via joint delivery of active screening with TT and ITC [56], may retain donor attention on vector control in gHAT foci during and following elimination, mitigating the potential for re-emergence due to an uncontrolled animal reservoir or latent human infections. Concerns that rHAT will emerge as a major public health problem as HAT contol efforts wane following gHAT EOT [57] further motivate a coordinated approach to HAT and AAT.

1.4 Goals of this dissertation

There is ample evidence that gHAT EOT goals will only be achieved under current control strategies if animal reservoirs do not exist [6], yet there is evidence to suggest pigs may be a reservoir host [58, 10]. Conversely, while most experts agree EOT is impossible to achieve for rHAT, this assumption has not been rigorously challenged, and it is possible that efforts which focus on domestic reservoirs may indeed be both successful for AAT control and sufficient for rHAT EOT. Furthermore, livestock keeping is known to exert environmental effects on a range of scales and tsetse flies, which require shade to rest and deposit larvae, are known to be sensitive to these effects [16, 59, 60, 61].

This intersection of livestock, humans, and the environment, in the context of the WHO's EPHP and EOT goals, is the focus of this dissertation. Using household surveys and censuses which ask about the number of cattle and pigs owned, we have created maps of these species in four high-burden countries which do not yet meet criteria for EPHP validation: DRC, Malawi, South Sudan, and Uganda, detailed in Chapter 2. We have then used these maps

to estimate the total effect of livestock on HAT risk, separately for each species, country, and HAT form, detailed in Chapter 3, and to estimate the proportion of this effect which is mediated by remote-sensed indicators of environmental suitability for tsetse flies which we believe are downstream of livestock keeping, detailed in Chapter 4. These results will indicate the importance of domestic cattle and pigs as gHAT reservoirs, and whether control of such reservoirs in rHAT foci may be adequate for EOT. Finally, in Chapter 5 we present a stochastic extension of a deterministic compartmental model we have previously published [62] to estimate the impact of coordinating AAT control with existing HAT control measures on time to and probability of EOT in the study countries.

This research is critical to determining the feasibility of EOT, and to identifying strategies that optimize control of both HAT and AAT. Furthermore, our study of the influence of livestock on environmental factors which in turn determine sleeping sickness risk has broad applicability to diseases transmitted by insects, including malaria, leishmania, and others. Our hope is that these findings will guide control efforts in Uganda, South Sudan, DRC, and Malawi, thereby averting human and animal illness and death and contributing to poverty alleviation.

Chapter 2

ESTIMATING THE DISTRIBUTION OF LIVESTOCK IN SPACE AND TIME: A TIME-SERIES CONTRIBUTION TO LIVESTOCK DENSITY MAPPING

2.1 Background

Globally, one billion people living on less than US$2 per day depend on livestock. For these communities, which include 80% of the poor in Africa, livestock are a critical source of household income, transport, draft power for crop agriculture, and nutrition, providing 11% of energy and 26% of dietary protein among the poor in East Africa, and up to 50% of energy for children under five in pastoralist communities [59, 63]. Livestock keeping also plays an important cultural role, serving as a means to maintain family cohesion and social networks, gain political prestige, and strengthen legal claims on pasture land [64, 65]. These critical roles livestock play, however, commonly result in close human-animal contact, driving transmission of zoonotic diseases and resulting in over 2.5 billion cases of human illness and 2.7 million deaths per year [63].

Livestock also have major environmental impacts across a range of scales. In Africa, where an estimated 1 billion or more of the projected increase in the human population will occur, urbanization and an increased demand for animal products are also expected [66], driving the so-called "livestock revolution [67]." Production systems are generally low-input in sub-Saharan Africa, with very little technological change in the past 40 years [66]; compounded by degraded natural resources, this lack of technological change requires increasing production demands to be met by overgrazing and land-use changes, rather than intensification, compromising biodiversity and ecosystem services [68]. Furthermore, where intensification has been achieved, animals and animal wastes are generally concentrated, facilitating disease

transmission and water pollution in the absence of proper controls [59].

On a larger scale, livestock systems have a bi-directional relationship with climate change. Climate change affects the quantity and quality of feeds available to livestock, drives production losses due to heat stress, promotes uneven distribution of water resources, and modifies the distribution of livestock diseases and disease vectors, representing another manifestation of the disproportionate burden climate change places on the world's resource poor. In turn, livestock and livestock systems are substantial users of natural resources, in particular water and land for grazing and feed production, and contribute to greenhouse gas emissions and climate change [60, 66].

Thus, high spatiotemporal resolution maps of the distribution of livestock hold utility for a range of research and policy applications, from epidemiology and public health to economics and climate science. To this end, the Gridded Livestock of the World (GLW) database was developed in 2007 [69], with GLW-3 being the most recent iteration. Published in 2018 but providing results for 2010, GLW-3 produces two sets of results, one using dasymetric modeling to disaggregate census counts, which is based on weights derived from statistical models that use spatial covariates, and a simple aerial weighting approach that produces estimates free from association with the spatial covariates used in the daysmetric results. These results are produced globally at a resolution of 0.083333° for counts of cattle, buffaloes, horses, sheep, goats, pigs, chickens, and ducks [70]. GLW does not use any household survey data, such as the Demographic and Health Survey or the Living Standards Measurement Studies, as these data use complex survey sampling approaches that would not be well-suited to GLW's current modeling approaches. However, GLW-3 authors acknowledge that comparing the results of models derived from these household surveys would strengthen livestock distribution mapping efforts at large. While numerous prior authors have mapped livestock diseases and disease vectors, beyond GLW mapping of livestock itself remain isolated to highly localized efforts such as grazing intensity in Kazakhstan [71] and livestock movements in Sahelian Africa [72].

To build on and support GLW efforts, we have used household survey and census data

on livestock ownership and the stochastic partial differential equations (SPDE) approach to generate high-resolution (0.017°) maps of cattle and pig density. We have produced these maps for every year from 2000-2020 for Malawi and Uganda and 2008-2015 for Democratic Republic of Congo (DRC), facilitating their use in longitudinal analyses. We have additionally used small area estimation to produce density maps at the county level for South Sudan in 2008. In all four countries, we define density as the ratio of animals to humans, producing a map in which the human population is flattened. We present here our methods and resulting products, and discuss our results in the context of GLW-3 and extension of these approaches to other countries and species.

2.2 Methods

2.2.1 Data collection and processing

Data sources were identified by searching GHDx and the IHSN Central Data Catalog [73, 74], and microdata were downloaded from publicly-accessible websites or by request to relevant national agencies. Both Demographic and Health Survey (DHS) data and population and housing census data were downloaded from IPUMS [75, 76]; most other surveys were downloaded from links in the relevant IHSN entry. Data were processed by source; data sources by country are detailed in Table 2.1, and sampling methodology by data source is detailed in Appendix A.

Table 2.1: Data sources by country

Country	Year	Source
Malawi	2000	Demographic and Health Survey*
	2003	World Health Survey*
	2004	Second Integrated Household Survey
	2004-2005	Demographic and Health Survey
	2006	Technology Adoption and Risk Initiative Survey

Table 2.1: Data sources by country

Country	Year	Source
	2010	Demographic and Health Survey
	2010-2011	Third Integrated Household Survey
	2010, 2013, 2016	Integrated Household Panel Survey
	2012	Malaria Indicator Survey
	2013-2014	Multiple Indicator Cluster Survey
	2014	Malaria Indicator Survey
	2015-2016	Demographic and Health Survey
	2016-2017	Fourth Integrated Household Survey
	2017	Malaria Indicator Survey
South Sudan	2008	Population and Housing Census
DRC	2007	Demographic and Health Survey*
	2010	Multiple Indicator Cluster Survey
	2013-2014	Demographic and Health Survey
Uganda	2001	Demographic and Health Survey*
	2006	Demographic and Health Survey
	2009	Malaria Indicator Survey
	2009-2010	National Panel Survey
	2010-2011	National Panel Survey
	2011	AIDS Indicator Survey
	2011	Demographic and Health Survey
	2011-2012	National Panel Survey
	2014-2015	Malaria Indicator Survey
	2016	Demographic and Health Survey
	2018	Malaria Indicator Survey

Table 2.1: Data sources by country

Country	Year	Source

2010 wave of the Malawi Integrated Household Panel Survey not used. *Did not collect livestock data in the necessary form, used only for wealth mapping.

For Malawi, Uganda, and DRC, survey cluster was the unit of analysis; for South Sudan, county (administrative level 2) was the unit of analysis. In Malawi, DRC, and Uganda, clusters with missing values for cattle counts did not contribute to the denominator for cattle density, clusters with missing values for pig counts did not contribute to the denominator for pig density, and clusters missing both pig and cattle counts were removed.

Number of livestock was parameterized as number in sampled households, summed over cluster (Malawi, Uganda, DRC) or county (South Sudan). Similarly, number of humans was parameterized as number of individuals (adults and children) in sampled households, summed over households in a given cluster (Malawi, Uganda, DRC) or county (South Sudan). As surveys which used stratified sampling methods defined strata by geography and urban/rural status (Appendix A), urban/rural status was also extracted for later modeling. Year was defined as survey year; for surveys conducted over two years, the first year was used. These attribute data were then merged with location data. Finally, source-specific data were merged to create one spatial points dataset (Malawi, Uganda, DRC) or spatial polygon dataset (South Sudan), and predictor data were extracted (detailed below).

2.2.2 Model fitting

Malawi, Uganda, and DRC

For Malawi, Uganda, and DRC, we used the SPDE approach to Gaussian process (GP) modeling. The goal of this approach is to smooth over spatial point data and generate a complete surface.

SPDE approach A spatial process $S(\mathbf{s})$ is a GP if the joint distribution of $S(\mathbf{s}_1), ..., S(\mathbf{s}_n)$ (over the whole study area) is an n dimensional Gaussian distribution for any integer n and any set of locations $\mathbf{s}_i$ (e.g., defined by latitude and longitude). That is, the spatial process at each location has a Gaussian distribution. Locations which are nearby are more likely to be similar to each other than locations which are far away, termed "spatial dependence," and this should be reflected in the spatial model.

In Bayesian hierarchical spatial modeling, this dependence is generally represented by a precision matrix (inverse of a variance-covariance matrix), with the Matérn family being a popular choice of covariance function [77]. Let $f(\mathbf{s})$ be the prior for the GP we are modeling. It can be shown that Matérn GPs can be represented as the solution to the SPDE:

$$(\kappa^2 - \Delta)^{\alpha/2} \tau f(\mathbf{s}) = W(\mathbf{s}) \tag{2.1}$$

where:

- $\Delta = \sum_{j=1}^{d} \frac{\delta^2}{\delta x_j^2}$ is the Laplacian
- $\alpha = \nu + d/2 > 0$ where $\nu > 0$ is a smoothness parameter and d is the dimension of $\mathbf{s}$
- W($\mathbf{s}$) is white noise
- τ^2 is related to the variance through:

$$\sigma_s^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}$$

And the solution to equation 2.1 is a stationary GP, $f(\mathbf{s})$, with a Matérn covariance function:

$$cov[f(\mathbf{s}_i), f(\mathbf{s}_j)] = \frac{\sigma_s^2}{\Gamma(\nu + 1 \,/\, 2)\sqrt{4\pi}\kappa^{2\nu}2^{\nu-1}} (\kappa||\mathbf{s}_i - \mathbf{s}_j||)^{\nu} K_\nu(\kappa||\mathbf{s}_i - \mathbf{s}_j||) \tag{2.2}$$

where:

- $K_\nu(.)$ is a modified Bessel function of the second kind

- $\sigma_s^2 > 0$ is a variance parameter

- $\kappa > 0$ is a scale parameter

Thus, the solution to equation 2.1 is a GP with the Matérn covariance function given in equation 2.2. For a GP with n clusters, due to the dense precision matrix a n dimensional normal distribution must be modeled, which quickly becomes computationally intractable. However, if α is an integer, then the resulting Gaussian random field has the continuous Markov property; i.e., it is a GMRF (Gaussian Markov random field). This means a particular location's spatial process depends only on that of its neighbors, not the entire map, resulting in a "sparse" precision matrix (many 0 entries) and markedly simplifying computation. For $\kappa = 0.5$, the Matérn correlation function reduces to an exponential function $\rho_s(u) = exp(-u / \kappa)$, where u is the "practical range," or the distance at which correlation is 0.05 (that is, the distance between two points at which their correlation would be 0.05) [78]. In our application of the the SPDE approach, we parameterize the Matérn in terms of three parameters: σ_s^2 (spatial variance), ρ_s (spatial range), and ν (shape, or smoothness).

Now that we have represented the GP using an SPDE, the next step is to solve the SPDE using the finite element method. We want to replace $f(\mathbf{s})$, a prior on the GP, with discrete space, i.e. approximate it via a weighted combination of basis functions. We want an approximation that not only best fits the SPDE, but works when the observations are not on a regular grid, as spatial point data are rarely collected on a regular grid:

$$f(\mathbf{s}) \approx \sum_{k=1}^{m} w_k \phi_k(\mathbf{s}) \tag{2.3}$$

where:

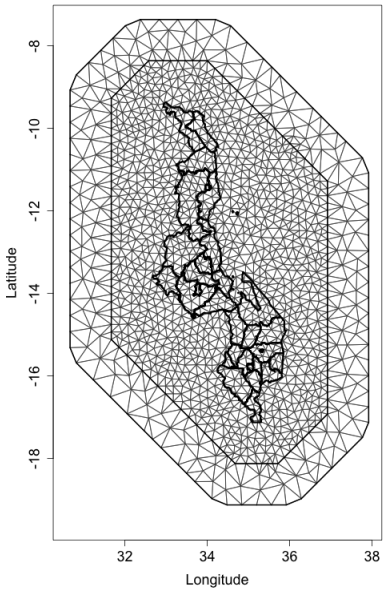
- w_k are multivariate Gaussian weights, which are picked to have a sparse precision matrix

- $\phi_k(\mathbf{s})$ are a set of basis functions

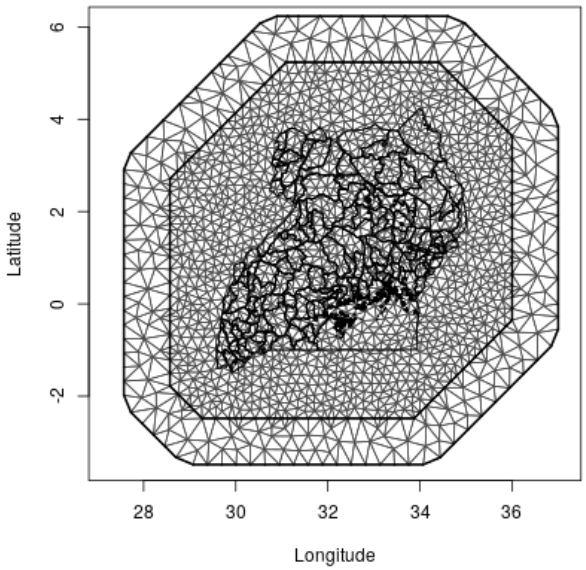
Implemented in `R-INLA`, space is discretized using a set of non-overlapping triangles which together comprise a mesh (Delaunay triangulation) over the study region, with the mesh in equation 2.3 having m vertices. The functions $\phi_k(\mathbf{s})$ are piecewise linear, equal to 1 on vertex k and 0 on all other vertices. If one imagines elevating a single mesh vertex, a three-sided pyramid is formed, and the discrete approximation of the GP is a weighted combination of these pyramids.

A corresponding projection matrix A projects the spatial effect from the mesh vertices to any point in the spatial field. For model fitting, the matrix A projects from the mesh vertices to the n study locations (here, clusters). For locations inside of a mesh triangle, the corresponding entry in A will have three non-zero entries which sum to 1, while locations on a mesh edge (line) will have two non-zero entries in A, and locations on a mesh vertex will have one non-zero entry equal to 1. All other entries in A will be 0, generating the sparse precision matrix.

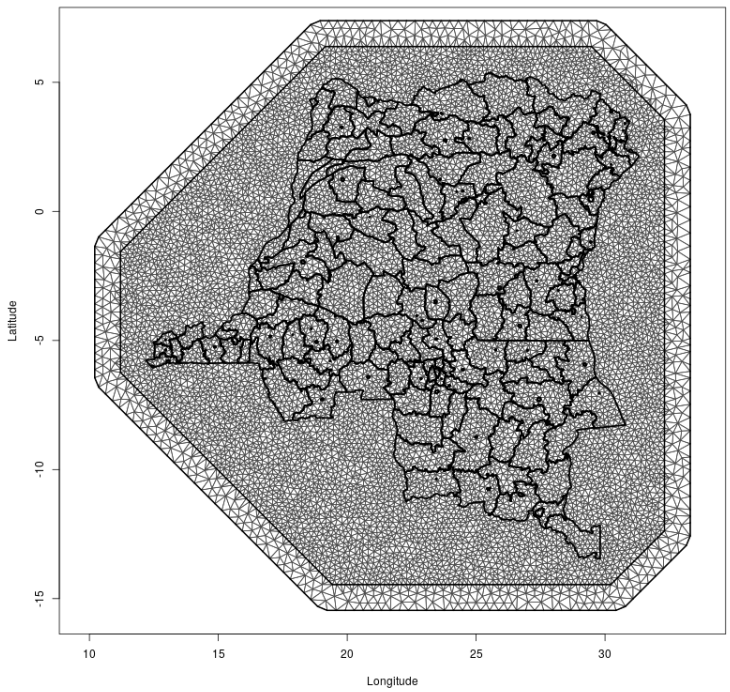
Choice of the mesh dictates the resolution of the spatial effect, with a rule of thumb that features more than two triangles large are resolved well, while features smaller than a triangle will be biased in proportion with the triangle size [77]. Our meshes for Malawi, Uganda, and DRC are presented in Figure 2.1 below. Malawi extends from 9.37°S in the north to 17.12°S in the south (7.75° N-S range), and from 35.92°E in the east to 32.67°E in the west (3.25° E-W range). Uganda extends from 4.23°N in the north to 1.48°S in the south (5.72° N-S range), and from 35°E in the east to 29.57°E in the west (5.43° E-W range). DRC extends from 5.39°N in the north to 13.46°S in the south (18.84° N-S range), and from 12.21°E in the east to 31.31°E in the west (19.10° E-W range). In all three countries, we set our mesh to have an inner and outer mesh, which builds in an extension that moves boundary effects away from the study area (i.e., country borders). We set the triangles to be smaller in the inner mesh (maximum edge length of 0.3°) than the outer mesh (maximum edge length of 0.6°), and the extension to be 1°.



(a) Malawi



(b) Uganda



(c) DRC

Figure 2.1: Meshes used for SPDE

Predictors When the data are complex survey data, the sampling scheme may be accounted for by including design variables in the regression model. As detailed in Appendix A, these included district and urban/rural status. As district is confounded by space, the spatial random effect (detailed above) accounts for this aspect of study design, and district was not included as a predictor in our models.

Predictors included urban/rural status (binary), protected areas (binary), elevation, and bodies of water (binary).

While there are many definitions of urbanicity, as urban/rural status is a design variable, the desired definition for this predictor is the value that a given cluster would take had it been sampled by a household survey in a given year. As the data sources we used for livestock mapping largely used the most recent population and housing census as their sampling frame, for each country we generated a 1km^2 urban/rural surface for each sampling frame year (Appendix A). We accomplished this using logistic regression models, with population density from WorldPop [79] and nighttime lights from the National Oceanographic and Atmospheric Administration's Nighttime Lights Time Series [80], detailed in Appendix B, as the predictors. For validation, we split the data into a 2/3 training and 1/3 test set, and evaluated model performance on the test set. We found these models performed very well for rural clusters (95% correct for Uganda, 96% for Malawi and DRC), however performance varied across countries for urban clusters (43% for Uganda, 86% for Malawi, 78% for DRC). As the vast majority of clusters in all three countries are rural, overall performance remained adequate.

Data on protected areas came from the World Database on Protected Areas [81], which is a joint project between the UN Environment Programme and the International Union for the Conservation of Nature (IUCN). The database recognizes sub-national, national, regional, and international sites which meet either the definition in Article 2 of the Convention on Biological Diversity ("A geographically defined area which is designated or regulated and managed to achieve specific conservation objectives"), or the 2008 IUCN definition ("A clearly defined geographical space, recognized, dedicated and managed, through legal or other

effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values") [82]. The database is updated monthly, however there is no publicly-accessible archive.

Data on bodies of water came from the Global Lakes and Wetlands database [83], which is produced by the World Wildlife Foundation and the Center for Environmental Systems Research at the University of Kassel, Germany. These data are provided on three levels. For our models, we used levels 1 and 2, where level 1 comprises the largest lakes ($\geq$ 50km^2) and reservoirs ($\geq$0.5km^3), and level 2 comprises permanent open water bodies with a surface area of $\geq$0.1km^2.

Elevation data came from GMTED2010 [84], which is a collaboration between the US Geological Survey and the National Geospatial-Intelligence Agency and provides elevation data at 30-, 15- and 7.5-arc-second spatial resolutions. Seven rasters representing different aggregation methods (minimum elevation, maximum elevation, mean elevation, median elevation, standard deviation of elevation, systematic subsample, and breakline emphasis) are available for each resolution. For our analyses we used the 7.5-arc-second data, which has an root mean squared error of 26-30 meters, and median elevation.

Note, these predictors were used to improve model fit. Unlike the approach taken in GLW-3, these predictors were not used to mask unsuitable pixels [70].

Models We fit a three stage Bayesian hierarchical model as follows for the minimal model, fit separately for cattle and pigs, separately for each country, and for Uganda, separately for gHAT and rHAT:

$$Y_i|\mu_i, p_i \sim \text{ZIP}(\mu_i, p_i)$$

$$\mu_i = exp\Big(\beta_0 + \boldsymbol{\beta}_1\mathbf{z}(\mathbf{x}_i) + S(\mathbf{s}_i) + \text{RW2}(\omega_t) + \epsilon_i + \gamma_k + \eta_t + u(\mathbf{s}_i, t) + log(D_i)\Big)$$

$$S_i|S_j, j \in ne(i) \sim \text{GMRF}(\sigma_s^2, \rho_s)$$

$$\omega_t|\omega_{t-1}, \omega_{t-2}, \omega_{t+1}, \omega_{t+2} \sim N\left(\frac{4}{6}(\omega_{t-1} + \omega_{t+1}) - \frac{1}{6}(\omega_{t-2} + \omega_{t+2}), \frac{\sigma_\tau^2}{6}\right)$$

$$\epsilon_i|\sigma_\epsilon^2 \sim_{iid} N(0, \sigma_\epsilon^2)$$

$$\gamma_k | \sigma_\gamma^2 \sim_{iid} N(0, \sigma_\gamma^2)$$

$$\eta_t | \sigma_\eta^2 \sim_{iid} N(0, \sigma_\eta^2)$$

Where:

- i indexes cluster
- p_i is the hyperparameter that pertains to the model for the zeroes for cluster i
- Y_i is the number of livestock in cluster i
- D_i is the number of humans in cluster i (offset)
- β_0 is the intercept
- $\boldsymbol{\beta}$ is a vector of coefficients for the predictors
- $\mathbf{z}(\mathbf{x}_i)$ is a vector of predictor variables for location i
- $S(\mathbf{s}_i)$ are the spatial random effects (error terms), assumed to follow a Markovian Gaussian random field (GMRF) with variance parameter σ_s^2 and range parameter ρ_s, and $ne(i)$ are the neighbors of cluster i.
- ω_t is a random walk 2 random effect on time (years)
- ϵ_i is an unstructured cluster-level random effect (nugget)
- γ_k is an unstructured random effect on survey
- η_t is an unstructured random effect on time (years)
- $u(\mathbf{s}_i, t)$ is a space-time interaction which is approximated by an SPDE in space combined with an AR(1) process in time

We fit these models as type 1 zero-inflated Poisson (ZIP) models in `R-INLA` [85], which combine a distribution for the proportion of zeros with the Poisson distribution. The type 1 likelihood is given as:

$$\text{Prob}(Y|\mu, p) = p \times 1_{y=0} + (1-p) \times \text{Poisson}(Y|\mu)$$

where the first part is the process that generates the zeros (i.e., clusters with no livestock; $1_{y_i=0}$ is an indicator that cluster i has 0 livestock), and the second part generates the livestock counts in cluster i. This likelihood is a mixture of structural zeros (p_i) and sample zeros ($1-p_i$) [86].

For model selection, we fit a series of five models, with equivalent random effects but varying fixed effects: intercept only (model 1); intercept and urban/rural (model 2); intercept urban/rural, and protected area (model 3); intercept, urban/rural, protected area, and bodies of water (model 4); and intercept, urban/rural, protected area, bodies of water, and elevation (model 5). We also fit a sixth model with an RW1 effect in time, rather than RW2, and all fixed effects included in model 5.

In DRC, due to the limited data availability, we fit models only from 2008-2015 (i.e., extrapolating 2 years in each direction), and we restricted random effects to the structured (SPDE) effect on space, the structured (RW1 or RW2, depending on the model) effect on time, and the unstructured (iid) effect on space (cluster). That is, the space-time interaction and unstructured random effects for survey and time were removed.

Priors As in Wakefield et al. [87], for the spatial random effect we assigned a fixed shape $\nu = 1$, and a "penalized complexity" (PC) prior [88, 89] for the spatial range ρ_s and marginal standard deviation σ_s, such that $Pr(\rho_s < 0.3) = 0.05$, and $Pr(\sigma_s > 1) = 0.05$. For the spatial range parameter, this prior can be interpreted as the 5% quantile corresponds to 0.3°, which is approximately 4% and 9% of the extent of Malawi in the north-south and east-west direction, respectively, approximately 5% of the extent of Uganda in each direction, and approximately 1.5% of the extent of DRC in each direction. This spatial range, as well as the mesh, were

chosen to be adequately fine to allow for construction of a high-resolution raster map, but adequately coarse to account for jittering of cluster locations in the source data (a means to protect privacy) and to make the model computationally feasible. For the marginal standard deviation, given that ZIP models use a log link, this prior corresponds to a posterior 95% credible interval for the residual rate ratio—deviations in livestock density above or below the mean model, attributable to the random effect in question—of (0.37, 2.72). We used `R-INLA`'s default priors for the structured temporal effects (RW1 and RW2).

For the AR(1) process in time, again using the PC prior specification we set $Pr(\rho_t > 0.5) = 0.8$ as the prior. Finally, for the precision (inverse of variance) parameters for each of the independent and identically distributed (iid) random effects ($\sigma_\epsilon^{-2}, \sigma_\gamma^{-2}, \sigma_\eta^{-2}$), we specified the prior as $Pr(\sigma^2 > 0.5) = 0.01$, which results in a posterior 99% credible interval for the residual variance of (0.6, 1.6).

Prediction on a regular grid After fitting our models, we projected livestock density on a 200 (x direction; columns) by 450 (y direction; rows) grid over Malawi (Figure A2), a 330 x 330 grid over Uganda, and an 1170 x 1090 grid over DRC, generating 0.017° x 0.017° grid cells in each country (Figure 2.2). This choice governed the resolution of our final raster maps.

First, we extracted all predictors used in our model fitting for each grid cell. Separately for models 1-6, we took the product of each predictor value and the corresponding fixed effect (model coefficient) for every grid cell. We then used the command `inla.posterior.sample()` to draw 1,000 samples from the posterior distribution of the spatiotemporal random effects, separately for all models and all years; we did not use any of the iid (unstructured) random effects for prediction as these were assumed to reflect measurement error. For each sample we then created a new projection matrix to project the spatial random effect from the mesh vertices to the grid cells. We used the index for each year for the period 2000-2020 in Malawi and Uganda (that is, 2000 would take a value of 1, 2002 would take a value of 2, and so on) and for the period 2008-2015 in DRC to project each temporal random effect. Next, we took

the sum of each of these parts (predictors × fixed effects, random effects × projection) and exponentiated to generate estimates of livestock density at each grid cell. Finally, we took the median and width of the posterior 95% credible interval over these 1,000 estimates.

Because we did not use an offset in our predictions, we essentially "flattened" the distribution of the human population over the map: our final estimate is that of livestock counts in a grid cell with 1 human being, equivalent to the ratio of livestock to humans, which we are calling livestock "density."

Model selection After fitting our models, we performed model selection via leave-one-out-cross validation (LOOCV) for the cattle maps by survey, as we could most readily conceptualize the missing data as hypothetical surveys which were not conducted or samples not collected by a given survey. This was conducted separately for each country. Each fold left out a random 25% sample, distributed evenly among sampling strata defined by district and urban-rural status. We present model performance as mean squared error (MSE), which we calculated as:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_{obs_i} - y_{pred_i}\right)^2$$

where y_{obs_i} is the observed density at location i, and y_{pred_i} is the predicted density at location i.

For external validation, we also regressed livestock density estimated from GLW-3 (dasymetric product), using WorldPop data as the denominator, on our final 2010 estimates of livestock density [90, 91, 79].

South Sudan

Weighted estimates After reading in the data, we used the `svyby()` and `svyratio()` functions in the `survey` package in R [92] to generate weighted estimates of county-level density, using sample weights contained in the IPUMS subset [76] and the ratio estimator

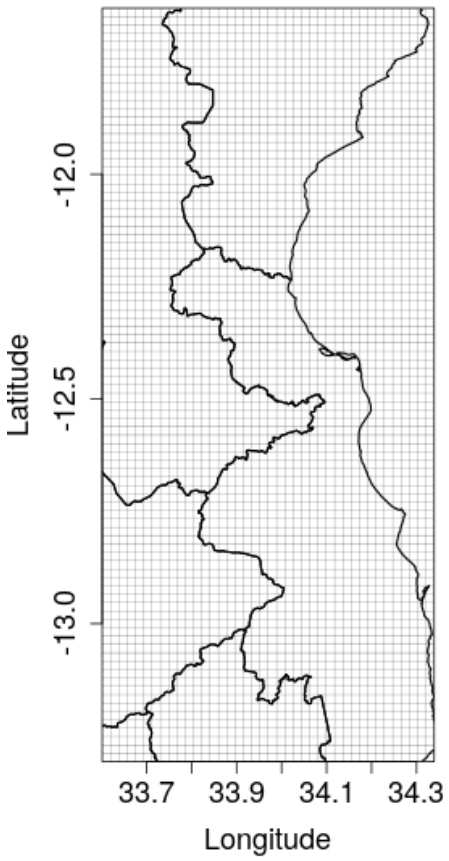


Figure 2.2: Regular grid for prediction, cropped to the northwest of Malawi. Prediction grids were of equivalent resolution in Uganda and DRC

below, where i indexes household and c indexes county:

$$\widehat{Y}_{cHT} = \sum_{i=1}^{n_c} w_{ic} Y_{ic}$$

$$\widehat{D}_{cHT} = \sum_{i=1}^{n_c} w_{ic} D_{ic}$$

$$\hat{\mu}_c = \frac{\widehat{Y}_{cHT}}{\widehat{D}_{cHT}}$$

where

- Y_{ic} is the number of livestock in household i, county c, and $\widehat{Y}_{cHT}$ is the Horwitz-Thompson estimate for the number of livestock in county c
- D_{ic} is the number of residents in household i, county c, and $\widehat{D}_{cHT}$ is the Horwitz-Thompson estimate for the (human) population in county c
- n_c is the total number of households in county c
- w_{ic} is the sampling weight for household i in county c, which is the inverse of sampling probability

We also used the `survey` package to estimate design-based standard errors for livestock density. Adapted from Mercer et al. 2014 [93], we then specified $\theta_c = log(\hat{\mu}_c)$, which, by the delta method, gives us the asymptotic sampling distribution:

$$\theta_c | \mu_c \sim N\left(log(\mu_c), \frac{\widehat{var}(\hat{\mu}_c)}{\hat{\mu}_c^2}\right)$$

Smoothing We performed spatial smoothing to stabilize the variance of the direct estimates by fitting the second and third stages of the three stage Bayesian hierarchical model as:

$$\theta_c = \beta_0 + \epsilon_c + S_c$$

$$\epsilon_c | \sigma_\epsilon^2 \sim_{iid} N(0, \sigma_\epsilon^2)$$

$$S_c|S_k, k \in ne(c) \sim N\left(\overline{S}_k, \frac{\sigma_s^2}{m_c}\right)$$

where

- ϵ_c are county-level iid (unstructured) random effects with variance σ_ϵ^2
- S_c are county-level structured random effects which follow the ICAR model with marginal variance σ_s^2
- $ne(c)$ denotes neighbors (shared boundary) of county c
- m_c is the number of neighbors of county c

and ICAR is the intrinsic conditional autoregressive model, which smooths each county's random effect to that of its neighbors, with more smoothing performed for counties with fewer neighbors.

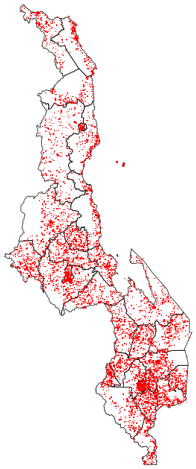
Priors As for the SPDE models, we used PC priors for the smoothing model, with $Pr(\sigma_\epsilon) > 1 = Pr(\sigma_s) > 1 = 0.01$. This yields a posterior 95% credible interval for each random effect's residual rate ratio of (0.36, 2.71) [89, 88].

Model selection As there was only one candidate model for South Sudan, model selection was not performed.

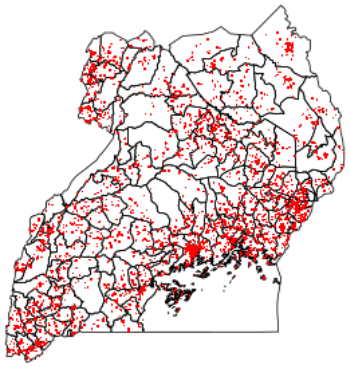
2.3 Results

2.3.1 Input data

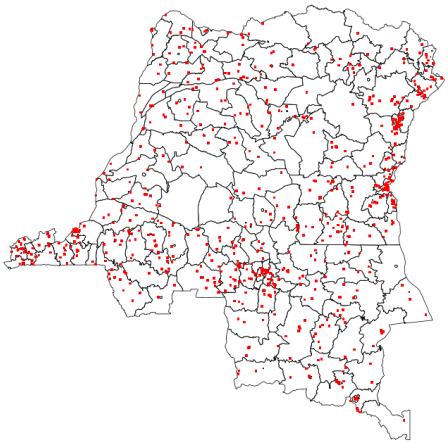
In the final dataset there were 6,330 clusters for cattle and 6,342 clusters for pigs in Malawi; 3,269 clusters for cattle and 3,310 clusters for pigs in Uganda; and 861 clusters for each species in DRC. Cluster distribution in space is presented in Figure 2.3. Data richness (number of survey clusters) by year and country is summarized in Figure 2.4. Cattle were summed over local and exotic breeds, and dairy and non-dairy.



(a) Malawi

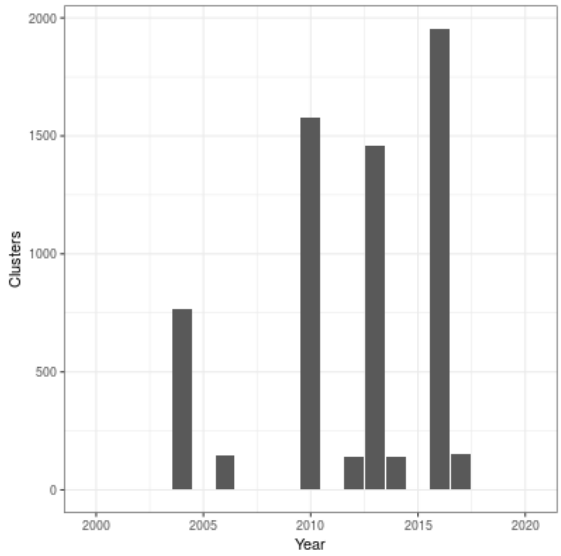


(b) Uganda

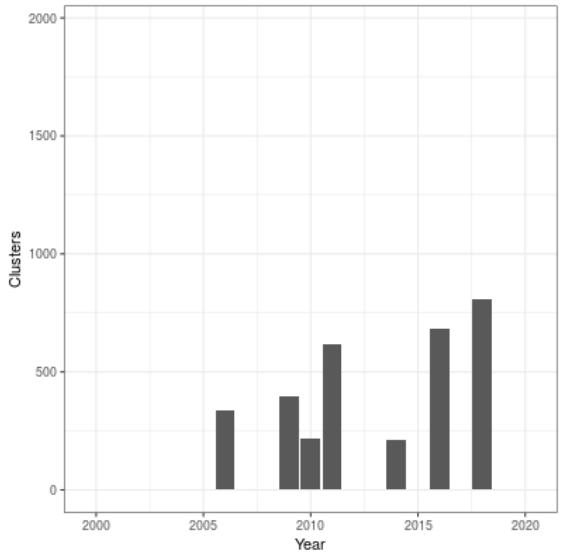


(c) DRC

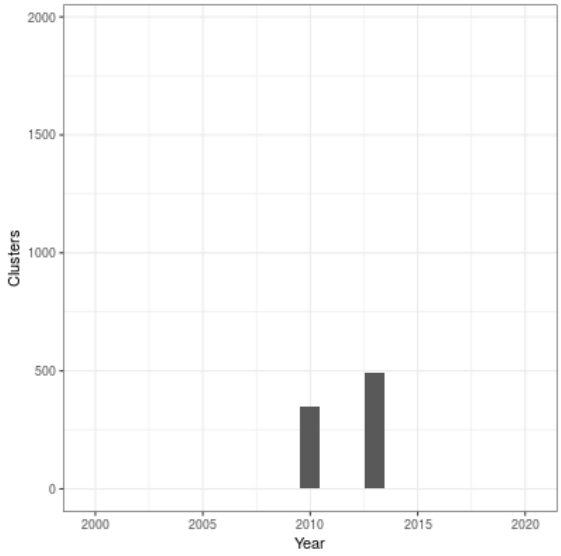
Figure 2.3: Spatial distribution of survey clusters (observed data) across all years. Size of the red dots does not scale with data richness



(a) Malawi



(b) Uganda



(c) DRC

Figure 2.4: Number of survey clusters (enumeration areas) available for cattle density mapping in each year and country. Results were nearly equivalent for pigs

2.3.2 Model selection

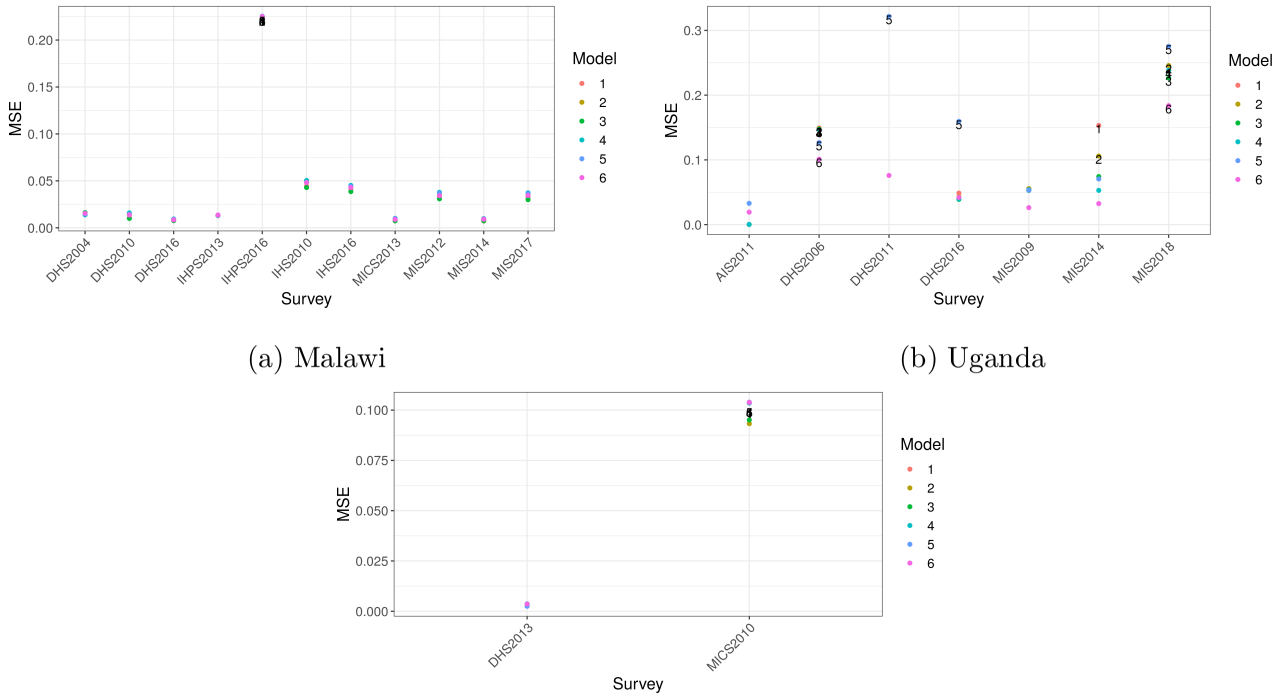
Leave-one-out cross-validation results are presented in Figure 2.5; we have removed models and surveys with very high MSEs for the sake of interpretability of these figures. This is detailed in the accompanying caption, however these surveys and models were not removed when examining overall model performance. Model 6 performed best for Uganda (MSE 0.076), model 3 for Malawi (0.023), and model 2 for DRC (MSE 0.048).

In all three countries, we found extremely strong positive associations between our livestock density estimates and those derived from GLW-3 (coefficients greater than 100), for both cattle and pig density. These remarkably high coefficients can be explained by the coarser resolution of GLW-3 compared with our maps and WorldPop data (0.083˚ versus 0.017˚ and 0.0083˚ pixels, respectively), resulting in markedly higher density estimates for GLW-3 when we calculated density on these data as detailed above. They do, however, indicate general agreement between our estimates and those of GLW-3.

2.3.3 Maps

Density maps for 2005 and 2010 for Malawi and Uganda, 2010 and 2015 for DRC, and for 2008 for South Sudan, are presented in Figures 2.11-2.10. Results (median and width of the posterior 95% credible interval) as .tif files for every year for Malawi, Uganda, and DRC, and as shapefiles for 2008 for South Sudan, are available in a GitHub repository: `https://github.com/JulianneMeisnerUW/LivestockMaps`. Providing estimates of both median and uncertainty allows users to assess the quality of the estimates for each species, country, and year.

In Malawi (Figures 2.11-2.12), cattle density was highest in the northern and southern extents of the country, and pig density generally decreased along a west-east gradient. Across all years and clusters, median cattle density was 0.05, and median pig density was 0.07. Per FAOSTAT total stock estimates [94] and World Bank national population estimates [95], mean national cattle density over the period 2000-2019 was 0.08, and mean national pig



(a) Malawi

(b) Uganda

(c) DRC

Figure 2.5: Leave one out cross validation results. Models with mean squared error (MSE) > 0.1 are labeled. Removed from figures (but not MSE calculations) for interpretablity: (a) Malawi: 2004 IHS, all models; Uganda (b): 2009 NPS, all models, 2011 DHS, models 1-2, 2010 NPS, models 1 and 3-6, and 2011 NPS, models 1-3 and 5

density was 0.15.

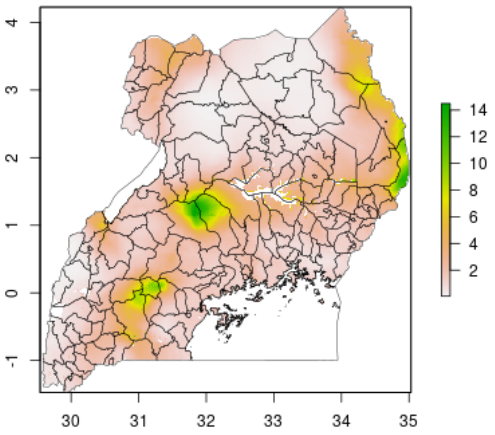
In Uganda (Figures 2.6-2.7), the "cattle corridor" is clearly visible in both years, while pig density appears to be higher in the south of Uganda. Across all years and clusters, median cattle density was 0.23, and median pig density was 0.02. Per FAOSTAT total stock estimates [94] and World Bank national population estimates [95], mean national cattle density over the period 2000-2019 was 0.33, and mean national pig density was 0.07.

In South Sudan (Figure 2.10), cattle density was on average higher than pig density, and there is a clear decreasing southeast-northwest trend in cattle density. For pig density in South Sudan, other than generally low densities in the northwest of the country, there are no clear spatial trends. Across all counties, median density was 1.56 for cattle and 0.007 for pigs in 2008. Per FAOSTAT total stock estimates [94] and World Bank national population estimates [95], mean national cattle density over the period 2000-2019 was 1.33; no FAOSTAT entry is available for pigs in South Sudan.

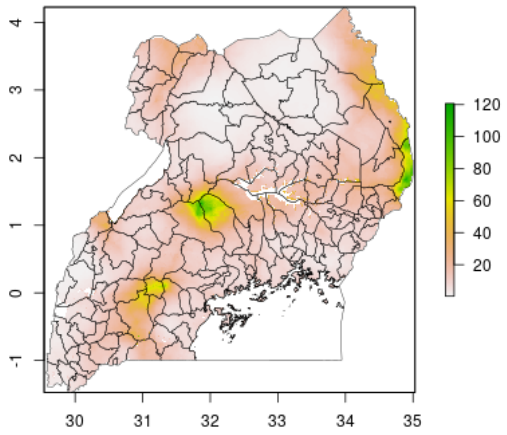
In DRC, there are "hotspots" of cattle density in the east and southwest of the country, and no obvious spatial patterns for pig density (Figures 2.8- 2.9) Per FAOSTAT total stock estimates [94] and World Bank national population estimates [95], mean national cattle density over the period 2008-2015 was 0.013 for cattle and 0.014 for pigs, while across all years and clusters our median estimate was 0.037 for cattle and 0.053 for pigs.

2.3.4 Data and code availability

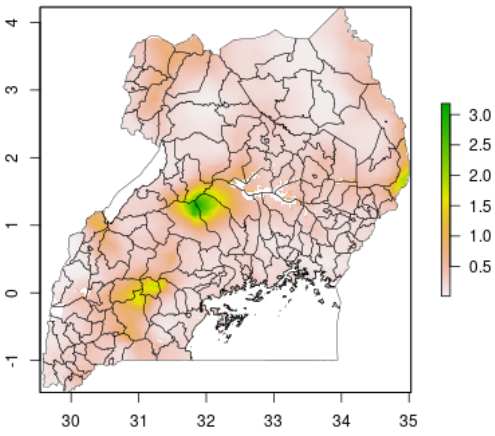
While existing data use agreements do not allow direct sharing of input data, all data were downloaded from publicly-available sources, detailed in Table 2.1, with links included in our Bibliography. All analyses were performed in R, and all code are available in the GitHub repository linked above.



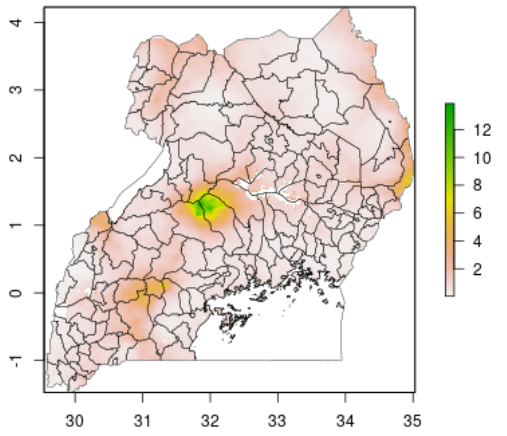
(a) Median, 2005



(b) 95% CI, 2005

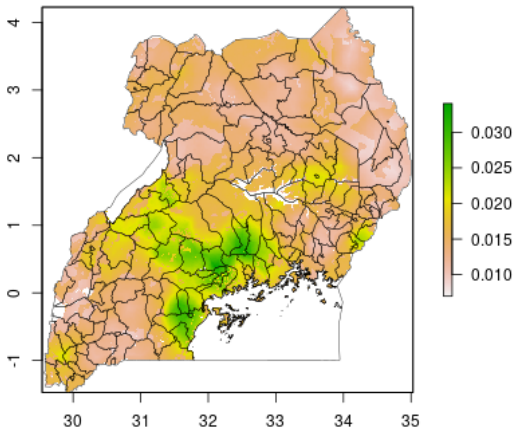


(c) Median, 2010

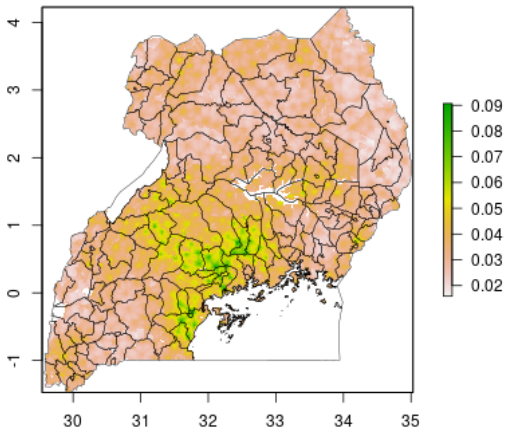


(d) 95% CI, 2010

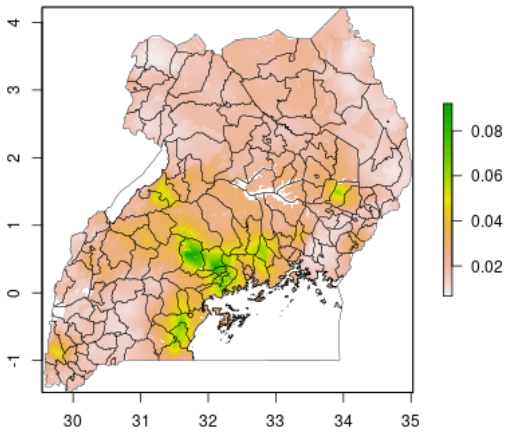
Figure 2.6: Cattle density mapping results, Uganda, 2005 and 2010. Median (a, c) and width of posterior 95% credible interval (b, d)



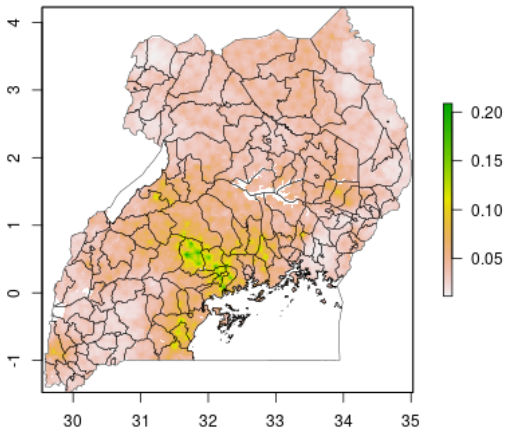
(a) Median, 2005



(b) 95% CI, 2005

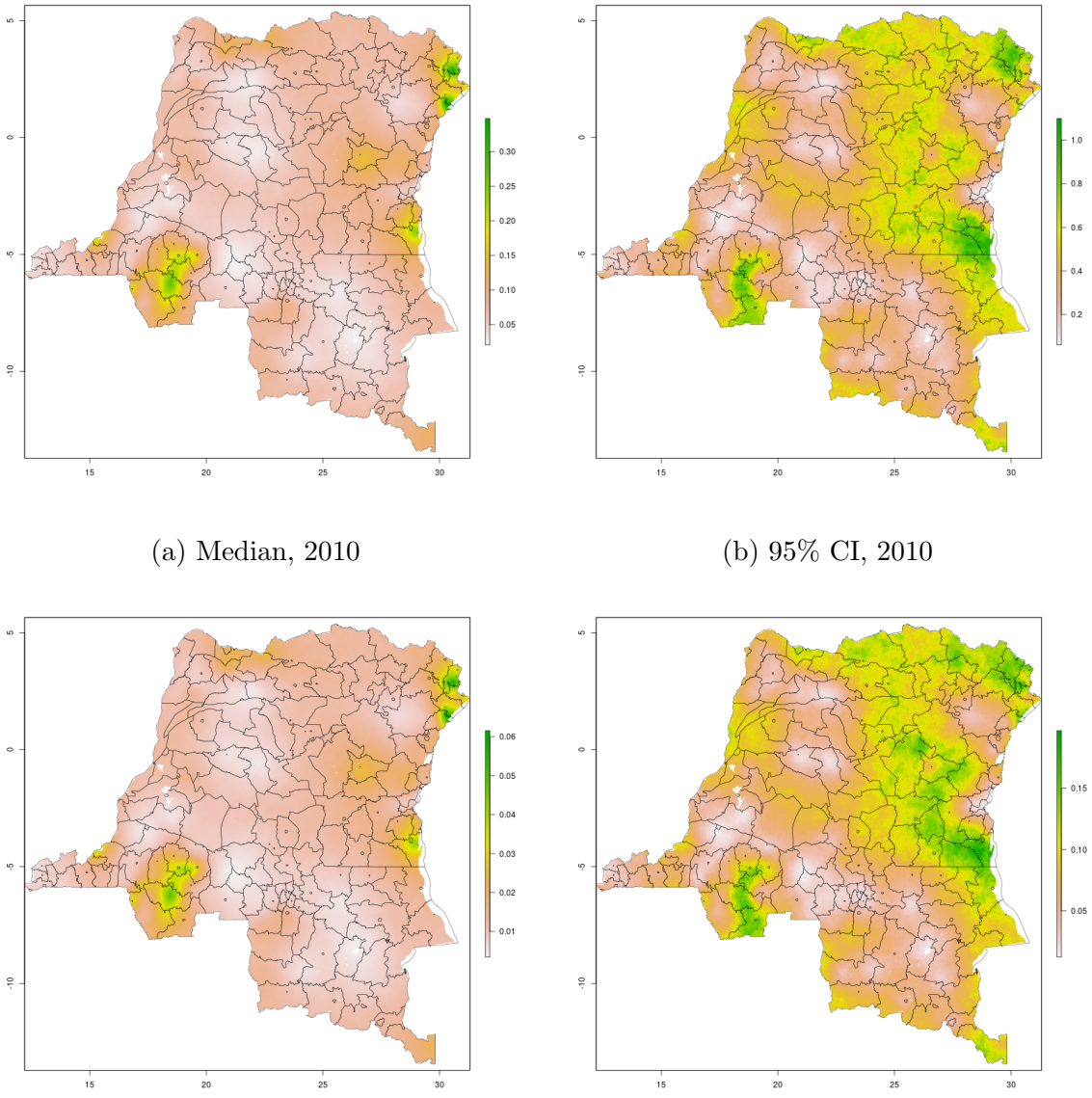


(c) Median, 2010



(d) 95% CI, 2010

Figure 2.7: Pig density mapping results, Uganda, 2005 and 2010. Median (a, c) and width of posterior 95% credible interval (b, d)



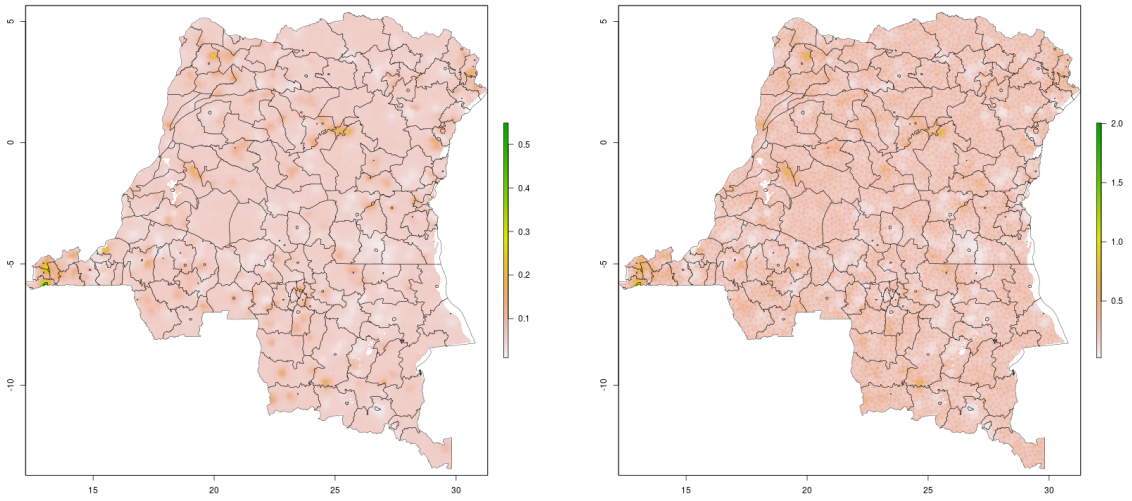
(a) Median, 2010

(b) 95% CI, 2010

(c) Median, 2015

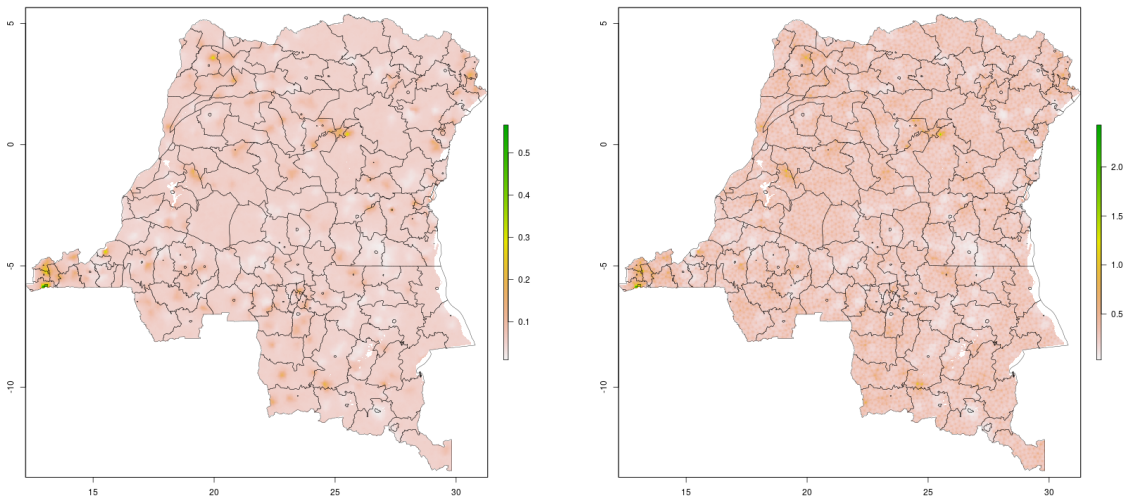
(d) 95% CI, 2015

Figure 2.8: Cattle density mapping results, DRC, 2010 and 2015. Median (a, c) and width of posterior 95% credible interval (b, d)



(a) Median, 2010

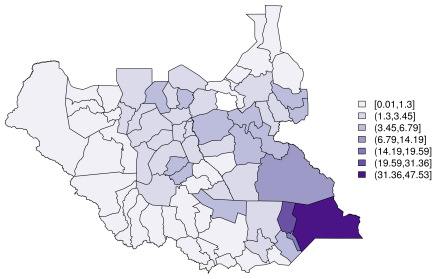
(b) 95% CI, 2010



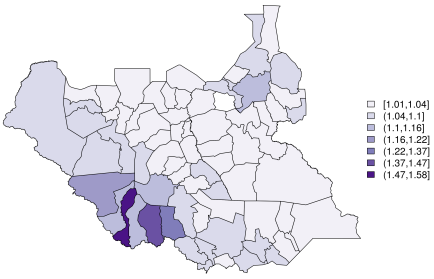
(c) Median, 2015

(d) 95% CI, 2015

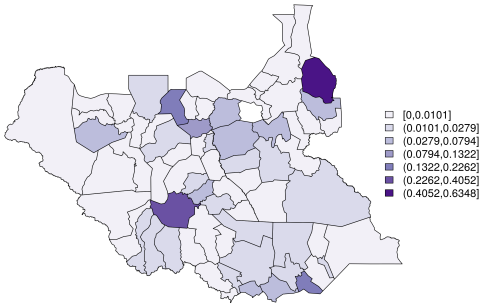
Figure 2.9: Pig density mapping results, DRC, 2010 and 2015. Median (a, c) and width of posterior 95% credible interval (b, d)



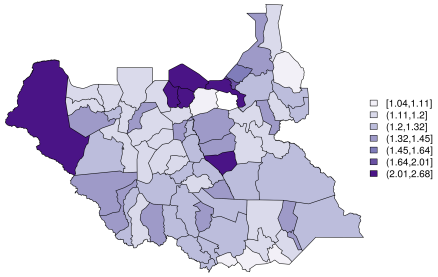
(a) Cattle density, median



(b) Cattle density, sd

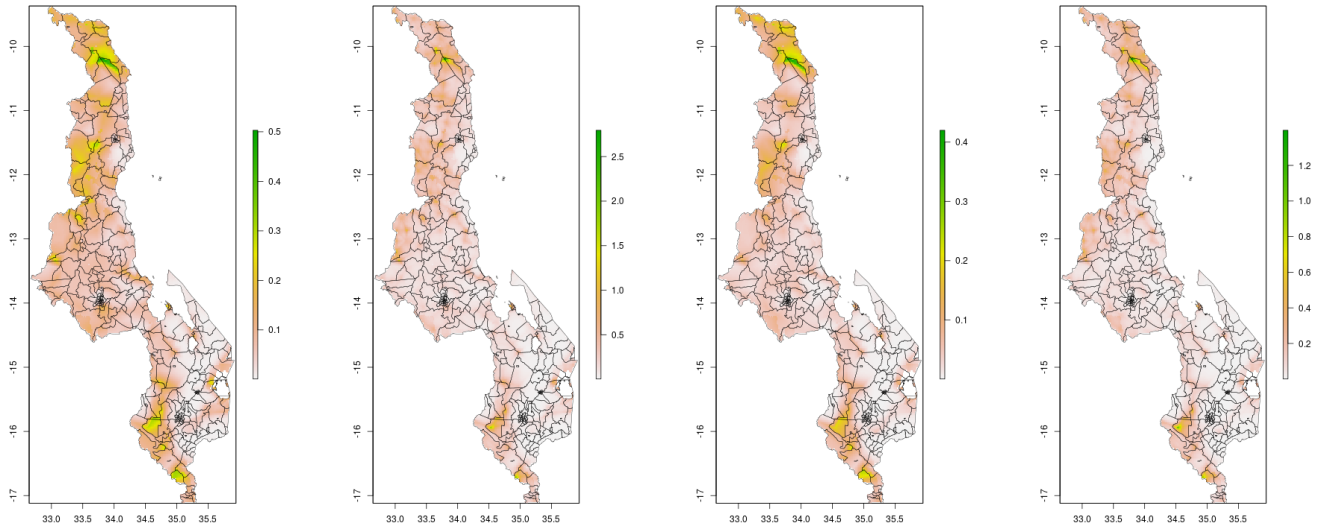


(c) Pig density, median



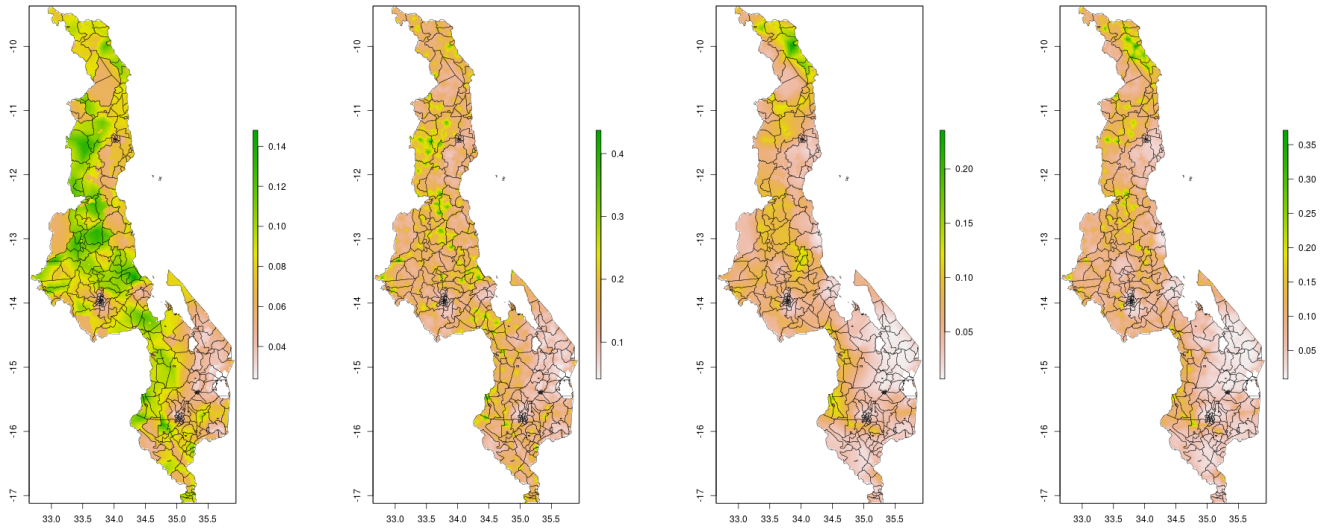
(d) Pig density, sd

Figure 2.10: Cattle (a-b) and pig (c-d) density mapping results, South Sudan, 2008. Median (a, c) and standard error (b,d)



(a) Median, 2005 (b) 95% CI, 2005 (c) Median, 2010 (d) 95% CI, 2010

Figure 2.11: Cattle density mapping results, Malawi, 2005 and 2010. Median (a, c) and width of posterior 95% credible interval (b, d)



(a) Median, 2005 (b) 95% CI, 2005 (c) Median, 2010 (d) 95% CI, 2010

Figure 2.12: Pig density mapping results, Malawi, 2005 and 2010. Median (a, c) and width of posterior 95% credible interval (b, d)

2.4 Conclusions

We have generated a set of cattle and pig density maps for Malawi, DRC, Uganda, and South Sudan, publicly available for use in economic, public health, environmental, agricultural, and other applications. Our results compliment GLW-3 efforts by producing a longitudinal product in Malawi, DRC, and Uganda, parameterizing livestock as the ratio of heads of cattle and pigs to the human population, and utilizing distinct input data and an entirely different modeling approach.

In contrast with GLW-3, we did not mask any pixels as unsuitable for livestock. The result is our findings impose no assumptions with regards to livestock presence in urban and high-altitude pixels, incursion into protected areas, and wet versus dry season changes in the distribution of water bodies, but may result in bias in pixels corresponding to large and permanent bodies of water.

There are several limitations to our results. First, there may be inaccuracies in the input data, for instance over-reporting of livestock ownership due to the social desirability of owning a larger herd, or under-reporting due to concerns about increased taxation. Other inaccuracies may arise due to jittering in the input data, in which clusters in publicly-available datasets are randomly displaced to protect participant privacy [96]. While ground-truthing is both cost-prohibitive and impractical for a longitudinal product, our LOOCV results indicate low MSEs overall, and good agreement between our results and GLW-3 results is a promising indicator of the validity of both products. Furthermore, our findings generally agreed with national density estimates, and we would not expect completely agreement with such estimates as local distribution patterns for humans is unlikely to be equivalent to local distribution patterns for livestock and we are comparing medians to means.

Second, our results treat livestock production systems as equivalent within a given species, however the environmental, socioeconomic, and public health relevance of production system types (e.g., pastoralist vs. intensive) are distinct, potentially reducing the utility of our results for select uses. Furthermore, users will need to take care to ensure the predictors used in our

SPDE models do not introduce circularity in their analyses due to the predictors (spatial covariates) we have selected.

Third, our approach to generating urbanicity surfaces reflects the challenging nature of meaningfully assigning unsampled cluster-years as urban versus rural. Our approach performed well in the test data with the exception of urban clusters in Uganda, for which it performed very poorly. Alternative approaches, such as using Global Human Settlement Layer data [97] and other existing urbanicity surfaces, do not necessarily correspond to the definition of urbanicity needed in our application, nor are they generally available as a longitudinal product.

Fourth, due to limited data in DRC and South Sudan, we could not include a space-time interaction term in DRC—imposing the assumption that spatial trends are constant in time, and temporal trends are constant in space—and could only produce a county-level map for South Sudan, and only for a single year.

Furthermore, in addition to jittering in the input data sources, misclassification of livestock density may arise from the location definition used in our models, whereby livestock are assigned to the centroid of the corresponding enumeration area. Thus our maps should be interpreted as the density of livestock to humans when both are in their place of residence (i.e., at nighttime). While this may be a poor reflection of livestock presence in pastoralist systems, it is difficulty to conceptualize any other definition of livestock density as the daily movements of humans and livestock may be distinct.

Finally, our estimates are accompanied by uncertainty which varies across time, countries, and clusters. Uncertainty was highest for DRC and for early years, rendering our maps less useful in these cluster-years. It is our hope that by producing uncertainty estimates alongside of medians, it will be transparent to users where our results are more or less stable.

We nevertheless present a product which, despite its limitations, represents an important extension to GLW-3 with high demonstrated validity and broad potential utility. While this product is limited to two livestock species and only a few countries, our approach can be readily extended to any species and country with sufficient cluster-level data on household

livestock ownership.

Chapter 3

THE EFFECT OF LIVESTOCK DENSITY ON *TRYPANOSOMA BRUCEI GAMBIENSE* AND *T. B. RHODESIENSE*: A CAUSAL INFERENCE-BASED APPROACH

3.1 Background

Absent from elimination of transmission (EOT) goals for human African trypanosomiasis (HAT) is the acute form of HAT, caused by *Trypanosoma brucei rhodesiense* (rHAT). Both rHAT and its chronic counterpart, caused by *T. b. gambiense* (gHAT) present in two stages. The initial hemolymphatic stage, characterized by non-specific signs, is followed by the meningoencephalitic stage, characterized by central nervous system signs and eventual death in the absence of treatment, occurring within several weeks for rHAT and 18 months or longer for gHAT [6, 98].

Exclusion of rHAT from EOT goals has been justified by its zoonotic nature and the assumption that this makes elimination of rHAT transmission impossible [99]. Field and experimental studies dating back to the early 20th century have demonstrated cattle to be competent reservoirs for rHAT, and it is thought that humans alone cannot maintain rHAT transmission as the duration of stage 1 disease in humans is very short, and the severity of stage 2 disease is assumed to preclude activities that bring humans into contact with tsetse flies. Efforts to re-stock northern districts in Uganda with cattle in the following the cessation of conflict in these areas resulted in the northward spread of *T. b. rhodesiense* towards the *T. b. gambiense* focus in the early 2000s, threatening convergence of the two forms [18, 100, 101]. Coordinated efforts to prevent this convergence via mass chemoprophylaxis of cattle were successful, demonstrating the importance of cattle as an rHAT reservoir in Uganda [23, 25].

There is also some evidence that other domestic animals are competent rHAT hosts, in particular pigs. Pigs have historically been considered less important reservoirs than cattle as they are shorter-lived for economic reasons, however in a survey conducted in Tanzania 4.8% of domestic pigs were found to harbor *T. b. rhodesiense* [57]. Numerous studies have also documented wildlife species to be competent rHAT reservoirs, in particular warthogs and hoofstock such as bushbuck, hartebeest, and waterbuck. Providing further evidence of the importance of these reservoirs for rHAT, proximity to protected areas and occupations and activities that result in frequent entry to these areas are known risk factors for rHAT infection [102].

The omission of rHAT from EOT goals has been largely accepted by virtue of its low incidence, however rHAT is thought to be significantly underreported, due both to its acute presentation—limiting opportunities for diagnosis and utility of active screening activities—and absence of the priority status afforded to gHAT control and surveillance, with modeling work estimating up to 12 deaths occurred for every one reported death in an rHAT focus in Uganda [103]. In rHAT cases that are detected, detection is rarely in stage 1 due both to the disease's rapid progression and absence of active screening activities. With recent approval of a stage-independent oral therapeutic (fexinidazole) limited to gHAT [104], this leaves melarsoprol, a chemotherapeutic agent with an associated fatality of 5% [105], as the only treatment option. Finally, animal African trypanosomiasis (AAT) is an important limiting factor for livestock productivity and poverty alleviation in endemic areas [106], and non-medical control efforts for AAT and HAT coincide. The underreporting of and limited treatment options for rHAT, and the economic importance of AAT and opportunity to coordinate control of both diseases in a One Health framework, suggest that rHAT should not be overlooked.

Furthermore, the Informal Expert Group on Gambiense HAT Reservoirs concluded in a recent review that latent human infections and animal reservoirs threaten EOT goals for gHAT, yet little is known of the role of animal reservoirs in the epidemiology of this form. While gHAT is generally considered as an anthroponotic disease, the absence of animal

reservoirs has never been confirmed. *T. b. gambiense* has been isolated from both wild and domestic animals, however it isn't clear whether these strains are human-infective, nor whether these animals have a role in gHAT transmission or simply represent spillover hosts. Additionally, while trypanosomiasis has historically been considered a bloodborne infection, there is recent evidence of a dermal anatomical reservoir for *T. b. gambiense* in both humans and pigs, suggesting a mechanism by which serological cases with undetectable parasitemia may continue to transmit trypanosomes [107]. This, along with the need to distinguish HAT agents from the morphologically-identical non-zoonotic *T. b. brucei*, renders presence or absence patent infection an imperfect means to identify or rule-out animal reservoirs. Mathematical models have also been limited in their ability to close this gap, largely because the importance of gHAT reservoirs may be focus-specific, and because models depend heavily on accurate parameters. Furthermore, most existing models have been deterministic, while stochastic models are better-suited to capturing the chance events that characterize cryptic reservoirs and near-elimination dynamics.

The maintenance of certain HAT foci at hypo-endemic levels, the finding that human-derived *T. b. gambiense* strains cyclically transmitted between animals for more than a year remained human-infective [10], and the failure of mathematical models to capture low prevalence dynamics without inclusion of an animal reservoir or invasion of infected tsetse, all suggest animal reservoirs for gHAT exist [99, 108, 109, 110]. Due to experimental evidence that *T. b. gambiense* can retain infectivity after repeated passage through pigs and that pigs may remain infected for prolonged durations (up to 18 months), the current consensus is that pigs are the animal reservoir most likely to threaten gHAT elimination [99, 10]. Furthermore, several important gHAT vectors have demonstrated a preference for pigs, who roam in humid and shady areas in the periphery of villages or along small rivers, resulting in high exposure to these riverine tsetse [58].

In this study, we use a time series of high-resolution cattle and pig density maps, HAT surveillance data collated by the WHO Atlas of HAT, and the parametric g-formula [111] to estimate the effect of livestock density on HAT risk. Defining livestock density as the ratio

of animals to humans, we have conducted this analysis separately for each country, animal species, and gHAT and rHAT, in Malawi, Uganda, Democratic Republic of Congo (DRC), and South Sudan, four high-burden countries that do not yet meet eligibility criteria for elimination as a public health problem (EPHP) validation. While there is ample evidence documenting the reservoir role of cattle in rHAT, the strength of this effect and the corresponding effect for pigs is likely focus-dependent. Country-specific estimates of the livestock effect, in concert with high-resolution data on livestock distribution, will provide both key parameters for rHAT modeling efforts and key inputs for NSSCPs and other stakeholders wishing to identify sub-national foci where transmission may be going undetected, or where ITC or TT may achieve EPHP or EOT. In gHAT foci, estimation of this effect will narrow critical knowledge gaps surrounding the role of domestic animals in gHAT transmission.

3.2 Methods

Our study is a retrospective ecological cohort study, with cluster-year (0.017° pixel) as the unit of analysis in Malawi, Uganda, and DRC, and county as the unit of analysis in South Sudan.

3.2.1 Study population

In Uganda, DRC, and Malawi, our study population included all clusters within five hours' travel time of a fixed health facility capable of HAT diagnosis [1]. Study areas were defined separately for gHAT and rHAT. In South Sudan, we restricted the study area to counties (administrative level 2) which reported one or more HAT cases or conducted active surveillance during the study period.

The study period was defined separately for each country, on the basis of WHO Atlas of HAT data access provided to the authors (2000-2018 for Uganda and DRC, 2000-2014 for Malawi and South Sudan), and then further limited to the bounds of available input data for livestock mapping to mitigate bias due to uncertainty (measurement error) in the exposure data: 2006-2018 for Uganda, 2003-2014 for Malawi, and 2010-2013 for DRC (Table 2.1).

This ensures we are only using data interpolated in time, not extrpolated in time. In South Sudan we further restricted the study period to reflect limited exposure data availability (2008 alone), as detailed below.

3.2.2 DAG formulation

We identified confounders by a priori subject knowledge, which we encoded in a directed acyclic graph (DAG). We used an initial time-stratified DAG (Figure 3.1) and DAGitty.net [112] to identify the minimally sufficient adjustment set.

"Index events" refers to natural and technological disasters and armed conflict. We believe tribal or ethnic group is upstream of religion, and that both variables are upstream of educational attainment and wealth through socioeconomic status and structural inequalities. We also believe wealth causes vegetation coverage (approximated through remote-sensed normalized difference vegetation index, NDVI) and surface temperature (approximated through remote-sensed land surface temperature, LST) as wealth is both a cause and effect of economic development and its associated land use changes. Similarly, we assume index events cause NDVI and LST both directly through landscape changes, and indirectly through economic development and its land use effects. We assume vector control exerts an effect on livestock density through control of AAT, however surveillance (active and passive) intensity and coverage has no such effect on livestock density. For rHAT we added location within a protected area to the our DAG, as this dictates both land use—and thus livestock density—and presence of wildlife reservoirs for rHAT.

Note that all variables are cluster-level, representing proportions for variables which are composites of individual-level categorical variables (ethnicity, educational attainment, and religion), and means for composites of continuous variables (wealth).

The minimally sufficient adjustment set is {elevation, index events, LST, NDVI, wealth, vector control}. The variables marked with a dashed box in Figure 3.1 exhibit treatment-confounder feedback: they are time-varying variables which are causes of the outcome and are both downstream of earlier exposure (mediators) and upstream of later exposure (con-

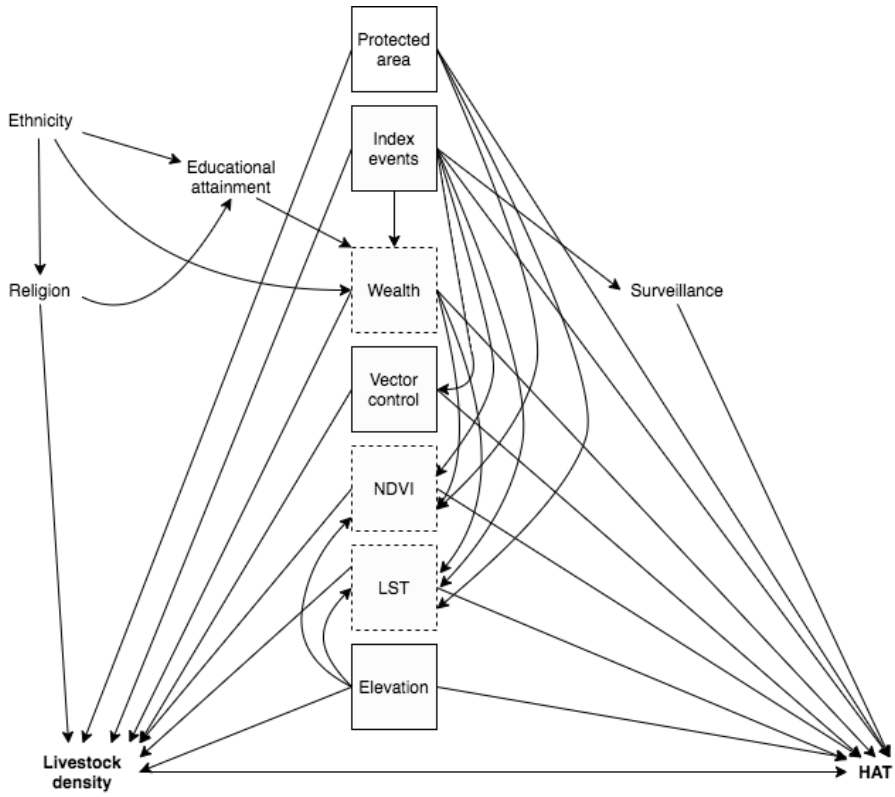


Figure 3.1: Time-stratified directed acyclic graph (DAG). Boxed variables are those in the minimum sufficient adjustment set; those with a dashed line exhibit treatment-confounder feedback. Bolded variables are the exposure and outcome of interest. Protected areas refers to rHAT models only.

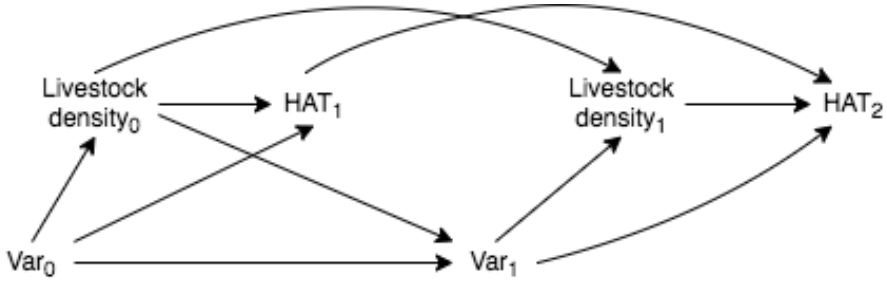


Figure 3.2: Time-varying directed acyclic graph (DAG) simplified to demonstrate the concept of treatment-confounder feedback and the bias that results from adjustment for variables exhibiting this feedback. Var_0 is a vector containing wealth, NDVI, and LST at the first of two hypothetical time points, and Var_1 is a vector containing the same variables at the second time point. Var_1 is a confounder of the livestock density$_1$—HAT_2 pathway, but is a mediator of the livestock density$_0$—HAT_1 and livestock density$_0$—HAT_2 pathways, thus adjustment of Var=$\{\text{Var}_0, \text{Var}_1\}$ will bias the joint (over time) effect of livestock density on HAT

founders) (Figure 3.2). As unbiased estimation of the total exposure-outcome effect requires adjustment of confounders but precludes adjustment of mediators, traditional regression adjustment cannot be utilized. Robin's generalized methods, referred to as "g-methods," allow the identification and estimation of causal effects in the face of time-varying confounding with exposure-confounder feedback. These include marginal structural models, structural nested models, and the g-formula [113]. As marginal structural models are best suited to binary exposures, and structural nested models are best suited to dynamic treatment regimes, we have opted to apply the g-formula in countries with longitudinal exposure data (Malawi, Uganda, and DRC). As we do not have longitudinal exposure data in South Sudan, time-varying confounding is not a concern.

Our final longitudinal DAG, restricted to the minimally sufficient adjustment set and to three time points, is presented in Figure 3.3. We assume index events are upstream of environmental variables (LST and NDVI) and wealth measured at the same time, but

impose lags of one year for the effects of index events, environmental variables, and wealth on both livestock density and HAT risk. While index events, or shocks in environmental variables or wealth, may generate near-immediate changes in livestock density through sales and purchases, we expect most of these effects to be realized instead through livestock births and deaths via slaughter and natural means, resulting in a moderate lag. Note, livestock density is a function of both livestock and human distribution, thus changes to the distribution of the human population even in the absence of changes to the distribution of livestock will affect the distribution of livestock density. However most divergent trends in human vs. livestock distribution (for instance, due to urbanization), occur on a coarser timescale than that of our analyses.

For all time-varying variables, we assume all effects are completely mediated by the same variable's value at the subsequent time point; for instance, the effect of index events at time 0 on HAT risk at time 2 is fully mediated by index events at time 1, with no direct effect. The exception to this is vector control, in which case we assume there is neither a direct effect on HAT at greater than one lag (i.e., no directed edge from vector control_0 to HAT_2), nor an indirect effect mediated by vector control at the subsequent time (i.e., no directed edge from vector control_0 to vector control_1).

3.2.3 Measures

Exposure We first generated maps of livestock density, parameterized as the ratio of livestock to humans, using survey and census data. These maps are detailed in Chapter 2. "Livestock" refers collectively to cattle and pigs, however all analyses are conducted separately for each species.

For Malawi, Uganda, and DRC we generated continuous maps by year for the study period (2000-2020 in Malawi and Uganda, 2008-2015 in DRC due to limited data availability, see Table 2.1). We validated these maps using leave one out cross validation, and by comparing our final 2010 maps against the Gridded Livestock of the World 3 maps [90, 91], using WorldPop data to estimate human population and therefore density [79]. For South Sudan,

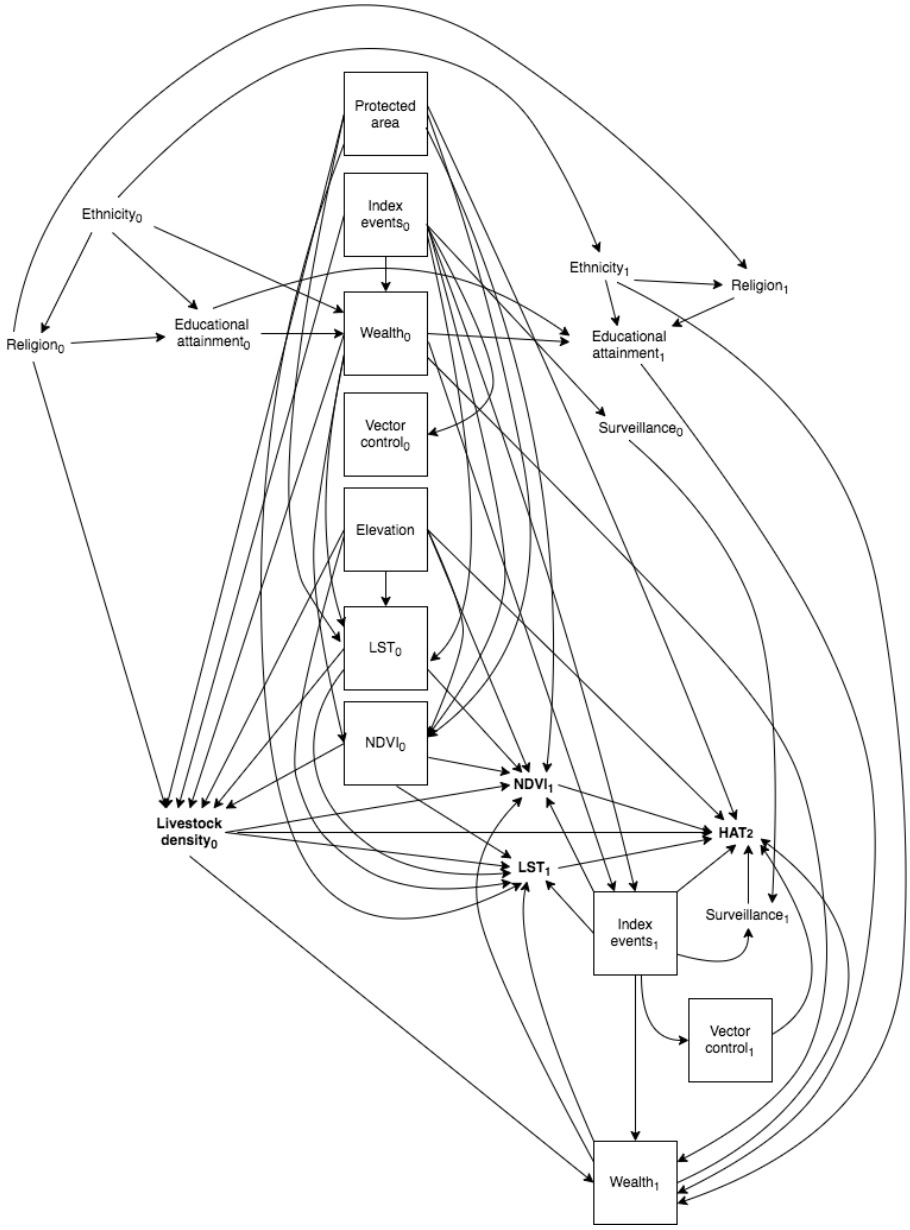


Figure 3.3: Final time-varying directed acyclic graph (DAG). Protected areas pertains to rHAT models only. Exposure and outcome of interest are bolded, and confounders in the minimally-sufficient set are denoted by solid boxes

given limited availability of data that provided both counts (rather than categories of counts) of livestock and location data, it was only possible to produce areal maps, and only for 2008.

Denominator data Denominator data came from the University of Southhampton's WorldPop project, which produces high resolution (1km at the equator) data on human population distributions using a flexible random forest estimation technique [114].

Outcome data Outcome data included all new cases of HAT diagnosed in a given year and given cluster (Malawi, Uganda, DRC) or county (South Sudan) in the WHO Atlas of HAT [4]. For Malawi, Uganda, and DRC, clusters which were not represented in the Atlas and were at least 1km from any clusters with a reported case were assigned 0 cases. In DRC, due to the computational challenges of estimating a distance matrix for such a large spatial object, we first stratified the study region on administrative level 2 areas (territory, n=180) before identifying predicted clusters within 1km of clusters with observed cases, potentially generating misclassification near these borders.

Confounders We estimated wealth using an exploratory factor analysis approach modeled after the DHS Wealth Index [115] but excluding livestock-related measures. We then used spatial modeling to generate continuous annual maps (detailed in Appendix B).

We used the EM-DAT database to identify natural disasters during the study period, which includes all events in which 10 or more people died, 100 or more people were affected (injured, homeless), or the affected country declares a state of emergency or appeals for international assistance are entered [116]. Disaster was parameterized as a binary variable for our models, taking value=1 if an event occurred in a given cluster-year or county. We used the UCDP/PRIO Armed Conflict Database to identify armed conflicts.

Data on NDVI came from the Land Long Term Data Record 5 (LTDR5). In each year, we selected the first product available after January 1st, from among N_19AVH13C1, N_18AVH13C1, and N_16AVH13C1, which are all daily products [117]. Data on LST came

from MODIS/Terra MOD21 for 2000-2002 [118], and from MODIS/Aqua MYD21 for 2003-2014 [119]. For each year, we used the first product available with a maximum of 15% cloud cover.

Elevation data came from GMTED2010 [84]; for our analyses we used the 7.5-arc-second data, which has an root mean squared error of 26-30 meters, and median elevation.

We did not attempt to adjust for vector control as top-down vector control efforts remain limited [120, 121] and it is not practical to parameterize farmer-led efforts. We do, however, present E-values on the relative risk scale to indicate the strength of uncontrolled confounding that would be required to fully explain the detected effects [122, 123].

3.2.4 Parametric g-formula

In Uganda, DRC, and Malawi, we used the parametric g-formula to estimate the total effect of livestock density on HAT risk.

Potential outcomes framework and counterfactual notation Under the potential outcomes framework for causal inference, causal estimands are defined in terms of counterfactual outcomes Y^a, specifically contrasts of these counterfactual outcomes under different exposure values $Y^a - Y^{a*}$, where Y denotes outcome and $A = a$ denotes exposure fixed at level a. While individual-level counterfactual outcomes cannot in general be estimated from observed data, population-average counterfactual outcomes $E[Y^a]$ can be estimated if three conditions are met: (conditional) exchangeability, positivity, and stable unit treatment value assumption (SUTVA).

Identifiability criteria By notational convention, let L be a vector of measured confounders and T be time, and let overbars denote history through time $T = t$. As exchangeability, also known as no exposure-outcome confounding, is rarely satisfied marginally, the weaker assumption of exchangeability conditional on the past and on L is typically adopted. Using counterfactual notation, this assumption is expressed as $Y^{\overline{a}} \perp A_t | \overline{A}_{t-1}, \overline{L}_t$. Positivity,

which refers to a positive probability of being assigned to each of the treatment levels (that is, no structural zeroes) is expressed as $Pr(A = a|L = l) > 0$ for $Pr(L = l) \neq 0$. SUTVA has two components: consistency and no interference. Under consistency, each unit of observation (here, a cluster-year) has one potential outcome for a given treatment level $A = a$, and therefore a cluster's observed outcome equals their counterfactual outcome under their observed exposure: if $A_i = a$, then $Y_i^a = y_i$ for all clusters. Under no interference, each cluster's potential outcome is independent of all other clusters' potential outcomes.

These criteria, known as the identifiability criteria as they confer identifiability of causal estimands to observational data, can be thought of as criteria under which an observational study emulates an idealized randomized controlled trial in which sequence allocation is concealed, blinding is maintained, there is no loss to follow-up, and adherence to assigned treatment is complete. Exchangeability is met by virtue of randomization, and positivity is met by by sequence generation and concealment which ensures all participants have a positive probability of assignment to any treatment under consideration. Consistency is met by both a well-defined intervention (i.e., a specific dose and administration route for a medication, or a well-formulated non-pharmacologic intervention) and complete adherence to assigned treatment. Finally, no interference is met through a design which considers whether a given participant's treatment assignment could "contaminate" another participant's treatment and, for instance, replacing an individual-level design with a cluster-level randomized trial if this is the case.

As the no interference assumption is certainly violated in our study, we instead turn our attention to partial interference, in which each unit of observation has a specified interference set $\chi_i = \{i_1, i_2, ...\}$ [124, 125]. Partial interference is satisfied if each unit's potential outcome is independent of the rest conditional on this interference set. Previous authors have defined the interference set in terms of its member's counterfactual outcomes, however applications to correlation in space have been limited to linear space (that is, a river) [125], and extension to (a) two-dimensional, and (b) continuous space has not been previously described.

In our application, interference arises through movement of tsetse flies, livestock, and

people. As grazing radius can vary from less than one to 10 or more kilometers per day, and tsetse flies have a range with a radius of approximately 300 meters per day, we define the interference set as all other clusters within a 5km radius [126, 127]. As for the outcome data, in DRC we stratified on administrative level 2 before defining the interference set, potentially creating a boundary effect for clusters near the border of each administrative area.

Implementation The g-formula applies standardization to estimate counterfactual outcomes, and the parametric g-formula uses parametric models to extend the g-formula to settings with numerous or high-dimensional confounders. Thus, when the parametric g-formula is applied, model misspecification will bias causal estimands even if the identifiability criteria hold [128].

Let $\overline{A} = \overline{a}^*$ be the deterministic regime under which A is set to a^* at all intervals $t \in \{1, ..., T\}$. The parametric g-formula is given as [128]:

$$E[Y_t^{\overline{a}}] = \int_{\overline{l}} E[Y_t | \overline{A} = \overline{a}^*, \overline{L} = \overline{l}] \prod_{t=0}^{T} f\{l_t | \overline{A}_{t-1} = \overline{a}^*_{t-1}, \overline{L}_{t-1} = \overline{l}_{t-1}\}$$

$E[Y_t^{\overline{a}}]$ is estimated by simulating the joint distribution of $\overline{L}$ and Y that would have been observed had all units received exposure $\overline{A} = \overline{a}$ (i.e., model-based standardization). First, we defined our causal estimand $E[Y^a / Y^{a^*}]$, where a^* = empirical mean livestock density, and $a = 1.5 \times$ empirical mean livestock density, thereby generating a relative risk analog corresponding to a 50% increase in exposure.

Next, we examined our DAG (Figure 3.3) to identify the causal sequence of our time-varying variables, in order to write down the model for each time-varying confounder (wealth, NDVI, and LST) and the outcome (HAT cases). For instance, the model for HAT cases at time t included livestock density at specified level a, elevation, index events at $t-1$, wealth at $t-1$, NDVI and LST at $t-1$, location within a protected area (rHAT models only), and elevation. We handled partial interference by adjusting for mean livestock density in the interference set in the outcome (HAT) models.

Finally, we tie each of these models together by invoking the law of total probability ($P(A) = \sum_B P(A|B)P(B)$), allowing us to obtain a marginal probability by averaging over conditional probabilities. Using the notation given above, a simplified version of the outcome model is given as:

$$E[Y] = \sum_A \sum_L E[Y|A, L]P(A|L)P(L) \tag{3.1}$$

As we will implement the g-formula by setting $A = a$, we can remove $P(A|L)$ from equation 3.1. Under the identifiability criteria, we can thus identify the counterfactual parameter $E[Y^a]$:

$$E[Y^a] = \sum_L E[Y|A = a, L]P(L) \tag{3.2}$$

As the first term in equation 3.2 corresponds to our outcome regression model, we need to marginalize over $P(L)$ to get $E[Y^a]$. We do this through empirically modeling the joint distribution of all confounders (time-varying and fixed) by taking 10,000 Monte Carlo samples and using these samples for prediction (detailed below), ensuring all clusters which reported cases were retained. We used non-parametric bootstrapping (100 bootstrapped samples) to obtain confidence intervals.

We implemented all models using the `glm()` function in R. We did not use spatial models due to computational challenges and concerns regarding spatial confounding. We fit Gaussian models for wealth, NDVI, and LST, and Poisson models for HAT cases. Our procedure was as follows, iterating over each t with t_0 defined as 2001 (2000 was not modeled due to the need to implement lags):

1. For t_0, fit all models except the outcome model (HAT cases): wealth, NDVI, and LST. Outcome model not fit due to the need to lag livestock by two time points (years)
2. Set livestock density to $A = a$, and for each model in turn use the `predict()` function to predict the variable in question under $A = a$ on the Monte Carlo sample
3. Repeat for all times $t_1, ..., t_T$, adding in the outcome model for $t > 0$.
4. Iterate steps 1-4 over each bootstrapped sample

The result of this procedure is a vector of counterfactual outcomes for each simulation s, with $Y^{\overline{a}}_{t(s)i}$ defined as the 1-year cumulative incidence of HAT in cluster i, year t, simulation s, and with exposure set to $\overline{A} = \overline{a}$.

$$\mathbf{Y}^{\overline{a}}_{(s)i} = \left\{ Y^{\overline{a}}_{1(s)i}, Y^{\overline{a}}_{2(s)i}, ..., Y^{\overline{a}}_{T(s)i} \right\}$$

As noted above, we take the ratio of the counterfactual outcomes under $\overline{A} = \overline{a}$ versus under $\overline{A} = \overline{a}^*$ as the target of inference. Across clusters, time and simulations, we implement this as follows, separately for each country (and for gHAT and rHAT in Uganda):

$$E_{(s)}\left[E_t\left[E_i[Y^{\overline{a}}_{t(s)i}] \,/\, E_i[Y^{\overline{a}^*}_{t(s)i}] \right] \right]$$

with uncertainty bounds estimated by taking the 2.5% and 97.5% quantiles across simulations.

Validation Validation of the parametric g-formula is conducted through implementation of the "natural course" model, whereby exposure (here, livestock density) is not fixed, but rather modeled as described for time-varying confounders above. The simulated outcome distribution is then compared against the observed to determine how successfully the observed distribution was recovered.

3.2.5 Regression

As we did not have longitudinal data in South Sudan, we used spatial regression models to generate total effect estimates rather than the parametric g-formula. Specifically, we fit Bayesian hierarchical models with ICAR and iid random effects in space (county), as detailed in Chapter 2, also using the same priors detailed in that chapter. We implemented these models using the `R-INLA` package [85].

Using Poisson models, we first fit naive models which did not account for measurement error in livestock density or wealth—that is, did not acknowledge that these variables were estimated. We parameterized exposure (livestock density) such that exponentiated effect

estimates could be interpreted as the rate ratio for a 50% increase in density, allowing direct comparison with g-formula results in the other countries; we accomplished this by log-transforming livestock density and dividing it by the log of 1.5.

Next, we re-fit these models, this time specifying livestock density and wealth as random effects following a classical measurement error (MEC) model (implementing these models within the g-formula was not practical for the other study countries). Under this model, the true value x cannot be observed directly, but is instead measured with error. This model is defined as:

$$w_c = x_c + u_c, c = 1, ..., n$$

where c indexes county, u_c are the measurement error terms distributed as $u_c \sim N(0, \sigma_u^2)$, and x_c and u_c are independent. Since the u terms have mean 0, $E(W|X = x) = x$, that is, W is unbiased for a given unobserved x. Failing to account for measurement error will result in biased effect estimates and inappropriate standard errors [129].

Following the heteroscedastic errors-in-variables approach detailed in Wang et al. (2018), we defined our hierarchical model as follows:

$$Y_c|\lambda = exp\Big(\beta_0 + \beta_1\mu_c + \beta_2\eta_c + \boldsymbol{\alpha}\mathbf{z}(\mathbf{k}_c) + S(\mathbf{s}_c) + \epsilon_c + log(P_c)\Big)$$

$$\mu_c = x_{lc} + u_{lc}, \eta_c = x_{wc} + u_{wc}$$

$$x_{lc} = \lambda_l + \zeta_{lc}, x_{wc} = \lambda_w + \zeta_{wc}$$

$$\epsilon_c|\sigma_\epsilon^2 \sim_{iid} N(0, \sigma_\epsilon^2)$$

$$u_{lc} \sim N(0, d_c\sigma_{ul}^2), u_{wc} \sim N(0, d_c\sigma_{uw}^2)$$

$$\zeta_{lc} \sim N(0, \sigma_{xl}^2), \zeta_{wc} \sim N(0, \sigma_{xw}^2)$$

where

- Y_c is the number of cases in county c
- μ_c is estimated livestock (cattle or pig) density in county c

- η_c is estimated wealth score in county c
- $\boldsymbol{\alpha}$ is a vector of coefficients
- $\mathbf{z}(\mathbf{k}_c)$ is a vector of predictors measured without error
- $S(\mathbf{s}_c)$ are the structured (ICAR) spatial (county) random effects
- ϵ_c are the unstructured (iid) spatial (county) random effects
- P_c is the offset, given as population in county c
- x_{lc} is the true livestock density in county c
- u_{lc} is the measurement error for livestock density in county c
- x_{wc} is the true wealth score in county c
- u_{wc} is the measurement error for wealth score in county c
- λ_l is the mean of the true livestock density
- λ_w is the mean of the true wealth score
- ζ_{lc} is the residual for livestock density in county c
- ζ_{wc} is the residual for wealth score in county c
- d_c is a weight which allows for heteroscedasticity in the error structure

For both livestock density and wealth score, we assume $x_c \sim N(\theta, \sigma_x^2)$, where θ is the mean of x_c and σ_x^2 is the variance.

There are therefore four parameters needed for each of the two resulting measurement error models: β_1 (β_2 for wealth score), $\log(1 \,/\, \sigma_u^2)$, λ, $\log(1 \,/\, \sigma_x^2)$, as well as the scale factor

d_c. For livestock density, we specified the priors and starting values for these parameters as follows for livestock, substituting μ for η and β_1 for β_2 for wealth:

- $\log(1 \,/\, \sigma_u^2) \sim \text{logGamma}(10, 10)$, starting value $\log\big(1 \,/\, var(\sigma_\mu)\big)$
- $\beta_1 \sim \text{Normal}(\hat{\beta}_1, 1 \,/\, \hat{\sigma}^2_{\hat{\beta}_1})$.
- λ fixed at mean $E[\mu]$, with a Gaussian prior
- $\log(1 \,/\, \sigma_x^2) \sim \text{logGamma}(10, 10)$, starting value $\log(1 \,/\, var(\mu))$
- $d_c \sim \text{Unif}(0.5, 1.5)$

where $var(\sigma_\mu)$ is the empirical variance of the posterior standard deviation for livestock density, $\hat{\beta}_1$ is the posterior mean of the coefficient for livestock density from the naive model, $\hat{\sigma}_{\hat{\beta}_1}$ is the posterior standard deviation of the same coefficient from the naive model, and $var(\mu)$ is the empirical variance of the posterior median for livestock density. The logGamma prior specification is equivalent to that used in Wang et al. [129]. The MEC model for wealth was specified equivalently, and in our final models both livestock density and wealth were included in this form.

As a sensitivity analysis, we also repeated these analyses restricted to the active surveillance data, as we expect bias in HAT reporting to follow different mechanisms in the active surveillance versus full dataset.

3.2.6 Data and code availability

HAT outcome data can be requested from the WHO (`https://www.who.int/trypanosomiasis_african/country/foci_AFRO/en/`) and livestock density maps can be downloaded from `https://github.com/JulianneMeisnerUW/LivestockMaps`. Confounder data are available for download from the websites linked in our Bibliography. All analyses were performed in R, and all code are available in the same GitHub repository.

3.3 Results

3.3.1 Descriptive statistics

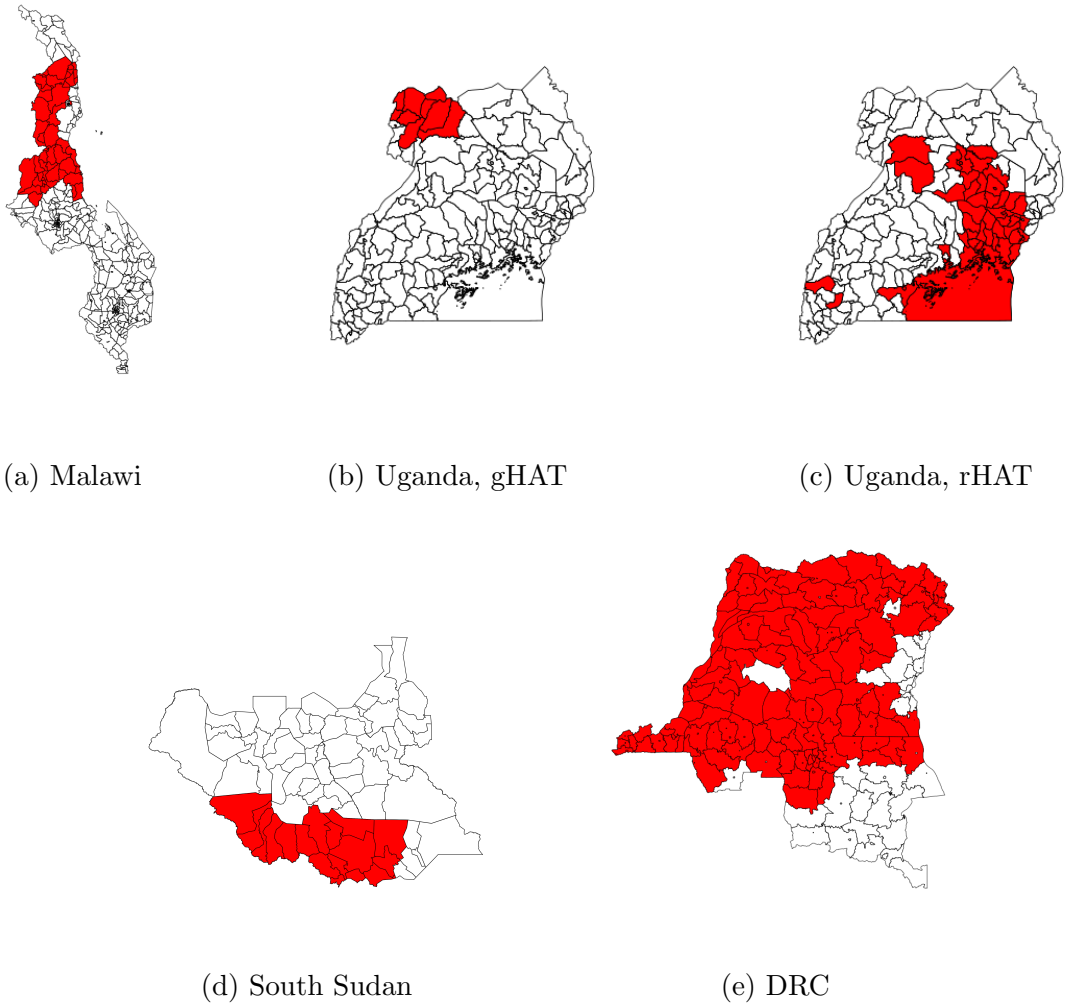
After removing predicted clusters within within 1km of locations in the WHO Atlas of HAT data and more than 5 hours' travel time from fixed health facilities capable of HAT diagnosis, we were left with 3,982 clusters/year in Malawi, 5,746 clusters/year for gHAT models in Uganda, 13,570 clusters/year for rHAT models in Uganda, and 247,205 clusters/year in DRC. In South Sudan, the study area was comprised of 17 counties, with active surveillance being performed in 14 of them in 2008. General study areas are presented in Figure 3.4. Note this figure represents the administrative areas represented in the study data in each country, however in Malawi, Uganda, and DRC, the study area is defined at the cluster-level, not the areal-level (i.e., not all clusters within a given administrative area highlighted in Figure 3.4 appear in the study data).

No armed conflicts in Malawi appeared in the UCDP/PRIO database during the study period, and no natural disasters in South Sudan appeared in the EM-DAT database in 2008. Descriptive statistics are presented in Table 3.1, with additional descriptive statistics figures presented in Appendix C, including temporal trends (aggregated over clusters) and maps for key variables.

3.3.2 Parametric g-formula

In Uganda, for the rHAT models the mean of the squared difference between observed livestock density and that predicted by the "natural course" model across cluster-years was less than 1 x 10^{-4} for both cattle and pigs. For the gHAT models, the pig models performed very badly in 2010 and 2012; removing these years from all subsequent analyses for pigs, this value was < 0.0005 for both cattle and pigs. In Malawi and DRC, this value was $< 1 \times 10^{-4}$ for both species across all years.

In the Malawi models, several confounders needed to be dropped due to the very low number of observed cases (as low as 18 cases reported in 2012) and resulting concerns surrounding



(a) Malawi (b) Uganda, gHAT (c) Uganda, rHAT

(d) South Sudan (e) DRC

Figure 3.4: National maps with administrative areas represented in the study data highlighted in red (a) Malawi: traditional authority (administrative level 3); (b-d) Uganda and South Sudan: county (administrative level 2); (e) DRC: territory (administrative level 2)

Table 3.1: Descriptive statistics, Malawi, Uganda, and DRC

Variable	Mean (sd)			
	Malawi	Uganda, gHAT	Uganda, rHAT	DRC
Cattle density	0.09 (0.31)	0.31 (0.92)	0.16 (0.42)	0.06 (0.06)
Pig density	0.09 (0.19)	0.24 (0.72)	0.06 (0.10)	0.06 (0.03)
Elevation (meters)	1,125 (280)	890 (166)	1,099 (79)	524 (194)
HAT cases	0.01 (0.14)	0.04 (0.35)	0.01 (0.15)	0.02 (0.35)
LST (K)	307 (4.7)	281 (96)	277 (95)	306 (3.67)
NDVI	0.15 (0.14)	0.28 (0.13)	0.26 (0.21)	0.22 (0.18)
Wealth	1.30 (1.04)	0.84 (0.03)	0.85 (0.03)	0.25 (0.22)
Conflict	0 (0%)*	743 (68%)*	1,824 (71%)*	177,468 (9%)*
Disaster				
Flood	19,548 (33%)*	413 (38%)*	2,290 (0.9%)*	46,921 (2.37%)*
Storm	1,213 (2%)*	0 (0%)*	0 (0%)*	0 (0%)*
Epidemic	0 (0%)*	2,065 (1.9%)*	6,307 (2.5%)*	28,343 (1.43%)*
Landslide	0 (0%)*	0 (0%)*	127 (0.05%)*	0 (0%)*
Drought	0 (0%)*	0 (0%)*	8 (<0.01%)*	0 (0%)*
Earthquake	0 (0%)*	0 (0%)*	0 (0%)*	410 (0.02%)*
Wildfire	0 (0%)*	0 (0%)*	0 (0%)*	16,575 (0.84%)*

Descriptive statistics over study clusters and period (2000-2014 for Malawi, 2000-2018 for Uganda and DRC). sd: standard deviation. *n(%)

Table 3.2: Descriptive statistics, South Sudan

Variable	Mean (sd)
Cattle density	0.77 (1.15*)
Pig density	0.02 (1.25*)
Elevation (meters)	693 (158)
HAT cases	34 (52)
Number screened	2,953 (4,200)
LST (K)	312.54 (3.22)
NDVI	0.28 (0.07)
Wealth	0.19 (0.004*)
Conflict	8 (47%)**

Descriptive statistics over study counties. 2008 data: cattle density, pig density, HAT cases, number screened, WorldPop population, LandScan population, wealth. 2007 data: LST, NDVI. 2006 data: conflicts. sd: standard deviation. *Mean of design-based standard error.**n, (%)

Table 3.3: Parametric g-formula results

	RR (95% CI)		
	Uganda	Malawi	DRC
Cattle			
rHAT	1.61 (0.90, 2.99)	0.88 (0.71, 1.10)	-
gHAT	0.88 (0.50, 1.77)	-	1.17 (1.04, 1.32)
Pigs			
rHAT	2.07 (1.15, 2.75)	0.42 (0.23, 0.67)	-
gHAT	2.02 (0.87, 3.94)	-	1.23 (1.10, 1.37)

Parametric g-formula implemented such that effect estimates correspond to a 50% increase in livestock density. Malawi results are for rHAT only, and DRC results are for gHAT only. RR: relative risk analog; 95% CI: credible interval over 100 iterations of the parametric g-formula

model over-fitting. Namely, location within a protected area, elevation, and disasters were removed from the Malawi models. LST was furthermore dropped from all analyses due to very large levels of missingness; this is reflected in the white areas in Figure C.4b.

Results from the parametric g-formula are presented in Table 3.3. We found a 50% increase in cattle density was associated with a 61% higher risk of rHAT in Uganda (95% CI 0.90, 2.99) and a 12% lower risk of rHAT in Malawi (95% CI 0.71, 1.10). For gHAT, a 50% increase in cattle density was associated with and a 12% lower risk of gHAT in Uganda (95% CI 0.50, 1.77) and a 17% higher risk of gHAT in DRC (95% CI 1.04, 1.32). For pigs, a 50% increase in density was associated with a 107% higher rHAT risk in Uganda (95% CI 1.15, 2.75) and a 58% lower rHAT risk in Malawi (95% CI 0.23, 0.67). For gHAT, a 50% increase in pig density was associated with and a 102% higher risk of gHAT in Uganda (95% CI 0.87, 3.94) and a 23% higher risk of gHAT in DRC (95% CI 1.10, 1.37).

Table 3.4: Total effect results, South Sudan

Denominator	Model	RR (95% CI)
Cattle		
WorldPop	Naive	0.59 (0.37, 0.83)
WorldPop	ME	0.63 (0.54, 0.77)
Number sampled	Active surveillance	0.79 (0.35, 1.20)
Pigs		
WorldPop	Naive	0.63 (0.31, 1.18)
WorldPop	ME	0.66 (0.50, 0.98)
Number sampled	Active surveillance	0.59 (0.36, 0.91)

Posterior mean rate ratios and 95% credible intervals for livestock density in South Sudan, after adjustment for wealth, NDVI (lagged 1 year), LST (lagged 1 year), elevation, and armed conflict. Density is parameterized such that rate ratios correspond to a 50% increase in density. RR: rate ratio; CI: credible interval; Naive: models which do not account for measurement error in wealth or livestock density; ME: measurement error models

3.3.3 Regression

Total effect estimates are presented in Table 3.4. While both results are presented, we will only discuss results from the measurement error model as the naive model estimates are biased. After adjustment for confounders {wealth, NDVI, LST, elevation, armed conflict}, the rate ratio (RR) was 0.63 (95% CI 0.54, 0.77) for cattle, and 0.66 (0.50, 0.98) for pigs. Cattle effect estimates were slightly attenuated and no longer significant when restricted to active surveillance data (RR 0.79, 95% CI 0.35, 1.20), while pig estimates were not appreciably changed (RR 0.59, 95% CI 0.36, 0.91).

Table 3.5: E-value results

	Uganda	Malawi	DRC	South Sudan
Cattle				
rHAT	2.6	1.53	-	-
gHAT	1.53	-	1.62	2.55
Pigs				
rHAT	3.56	4.19	-	-
gHAT	3.46	-	1.76	2.4

The strength, on the relative risk or rate ratio scale, of the relationship between an unmeasured confounder and exposure (livestock density) and between an unmeasured confounder and outcome (HAT risk) that would be required to fully explain the estimated effect. South Sudan results are presented for measurement error model results

3.3.4 E-values

E-values are presented in Table 3.5. In all cases, strong unmeasured confounding would be required to explain our findings. This is particularly true for the cattle-rHAT association in Uganda, the pig-rHAT association in Uganda and Malawi, the pig-gHAT association in Uganda and South Sudan, and the cattle-gHAT association in South Sudan.

3.4 Discussion

In Uganda we found cattle and pigs both have a positive effect on rHAT risk. For gHAT, we found cattle have negative (protective) effect, while pigs have a positive (harmful) effect. In DRC, we detected a positive effect on gHAT for both cattle and pigs, being marginally stronger for pigs than for cattle.

In Malawi and South Sudan, we found both cattle and pigs have protective (negative)

effects, with this result being stronger in pigs than in cattle in Malawi. In South Sudan, restricting analyses to the active surveillance data did not appreciably change our results.

These results indicate the zooprophylactic effect—whereby livestock exert a protective effect on HAT risk due to tsetse preference for animal hosts over humans—outweighs the reservoir effect for both cattle and pigs in Malawi and South Sudan and for cattle in gHAT foci in Uganda, however the reservoir effect outweighs the zooprophylactic effect for pigs in gHAT foci in both DRC and Uganda, for cattle in gHAT foci in DRC, and for pigs and cattle in rHAT foci in Uganda. Comparison of results across cattle versus pigs should also consider our choice of a relative parameterization for exposure; that is, a 50% increase in pig density is not equivalent, in absolute terms, to a 50% increase in cattle density, within and across foci.

Our findings initially appear discordant with mechanistic knowledge about HAT transmission, i.e., cattle and pigs are reservoirs for rHAT, and as the gHAT focus in Uganda is contiguous with that in South Sudan, transmission dynamics should be nearly equivalent. However, it is important to remember these results reflect complex epidemiological factors that extend beyond transmission mechanisms to local population dynamics and control measures. For instance our Malawi results do not conflict with a scenario in which cattle and pigs are theoretical reservoirs of rHAT, however the wildlife reservoir so markedly dominates these domestic animal reservoirs that the zooprophylactic effect prevails. This could arise if the predominant vector species (*Glossina morsitans morsitans*) exhibits a strong preference for wildlife, or if domestic animal reservoirs are adequately controlled through use of insecticides or trypanocides, or environmental measures such as brush clearing. Similarly, differences in host abundance and control measures between Uganda and South Sudan could explain the discrepancies between the results observed in these two countries. This is particularly relevant for the case of pigs; as South Sudan is a predominantly Muslim country, very few pigs are kept, and the sociocultural characteristics of higher pig density clusters are likely highly divergent from those in low pig density clusters, potentially driving similar divergence in HAT epidemiology. Presence of a zooprophylactic effect in South Sudan and for cattle in

gHAT foci in Uganda is consistent with evidence from southeastern Uganda that *Glossina fuscipes fuscipes*, the predominant gHAT vector in South Sudan and Uganda, exhibits strong zoophilic behavior [130]. Note that while there are several possible mechanisms which could give rise to this biting preference (i.e., relative abundance of host species, preferences inherent to fly species, etc.), pointing to different opportunities for intervention, all are expected to result in a net negative effect of livestock density on HAT risk. The finding of a positive cattle-gHAT effect in DRC but not Uganda or South Sudan is surprising, and of uncertain significance.

However, our results could, of course, be subject to bias. With regards to confounding, in Uganda, DRC, and South Sudan we feel we have achieved reasonable control for all confounders except vector control efforts. We would expect vector control to have a positive effect on livestock density (as AAT control would increase livestock density) and a negative effect on HAT risk (through decreased cases), thus unmeasured confounding by vector control would bias effect estimates towards zero, making it harder to detect positive (> 1) effects, and easier to detect negative (< 1) effects. Thus, it is possible failure to adjust for vector control explains our the negative effects we are attributing to zooprophylaxis, however the zooprophylactic effect was most pronounced in foci with limited top-down vector control efforts (Malawi and South Sudan). This, along with our calculated E-values, indicates unmeasured confounding by vector control may explain the rHAT effect for cattle in Malawi and the gHAT effect for cattle in Uganda, but is unlikely to be responsible for the other effect estimates. Furthermore, we approximated wealth, a latent construct, using factor analysis (detailed in Appendix B), which may not capture the dimension of wealth which is most relevant to our study, driving residual confounding by this variable. Finally, in Malawi we did not adjust for several known confounders due to overfitting concerns. Again, our E-values indicate this unmeasured confounding may explain the cattle results in Malawi, but are unlikely to explain the pig results.

A second major source of bias is selection bias. While no participation in our study was required *per se*, our definition of the study area may drive selection bias. In Malawi,

Uganda, and DRC, study area was defined by proximity to a fixed health facility capable of HAT diagnosis, however we did not have longitudinal data for this variable, possibly biasing our findings. In South Sudan, we defined study counties by appearance in the WHO Atlas of HAT dataset, and thus underreporting of the outcome (discussed below) may drive selection bias. In both cases, selection bias arises if excluded clusters or counties differ systematically in their distribution of livestock density or confounders from the included counties or clusters.

As with all research, measurement error threatens the validity of these findings. We attempted to account for measurement error arising from estimation of wealth and livestock density in South Sudan, however the measurement error model we employed assumes the errors are normally distributed and without structure in space or with respect to other key variables (that is, independent and non-differential, however heteroscedasticity was allowed through a scale factor) [129]. The assumption of normality is reasonable (while livestock density is a positive-valued real number, its error term could be negative) and there is no evidence that measurement error is differential (e.g., areas with higher HAT risk have higher variance for estimated livestock density), however we can be fairly certain there is spatial structure in these errors. Furthermore, these measurement error models do not account for errors in the raw data used for livestock or wealth mapping—for instance a systematic under-reporting of livestock ownership due to concerns about increased taxation—and we did not implement measurement error models in the remaining study countries, however we did restrict analyses to years bounded by those with available input data used to generate livestock density estimates (see Chapter 2 for detail). We expect such measurement error to be non-differential with respect to HAT risk, biasing our effect estimates towards the null in expectation. With regards to LST, NDVI, index events, and elevation, measurement error either in the attribute itself, or in its geolocation, will drive residual confounding and bias our effect estimates.

We also expect measurement error in the outcome. As with all surveillance data, reporting of HAT cases is likely to be incomplete, and some HAT foci have no effective surveillance systems [102]. This is particularly true for data collected via passive surveillance, which

accounts for all of the Malawi data and Uganda rHAT data. Under-reporting in the passive surveillance data arises from cases who do not seek care, seek care but are not correctly diagnosed, or are diagnosed but not reported [131]. Under-reporting in active surveillance data arises from lack of surveillance in remote or politically unstable foci [132], and incomplete coverage of populations in which active surveillance occurs [133]. If under-reporting is random with regards to space and either HAT risk or livestock density, power will be compromised but effect estimates will be unbiased in expectation. However this is unlikely, as surveillance capacity and uptake are not likely to be independent of HAT risk, nor are they likely to be independent of variables upstream of livestock density, in particular wealth and political stability. While complete control for wealth and political stability will close backdoor paths between livestock density and surveillance coverage, complete control is unlikely, and association between underreporting and HAT risk remains unaddressed. As we expect comparatively less underreporting in the active surveillance data than the passive surveillance data, the fact that our findings in South Sudan were not markedly different when restricted to active surveillance cases alone provides some comfort. In short, while our outcome of interest was HAT risk, the measured outcome was risk of reported HAT, and these measures diverge.

Ecological bias may also compromise the validity of our findings. Cross-level (ecological) bias may result from (1) an unmeasured variable which is either (a) an individual-level risk factor whose association with exposure differs across groups, or (b) an individual-level effect modifier whose distribution or effect differs across groups [134]; or (2) aggregation of variables (e.g., from the household to the cluster) [135, 136], termed "pure specification bias" [137]. While wealth was the only confounder measured at the household-level, elevation, NDVI and LST were aggregated from the resolution at which they were defined in the source data to our unit of analysis. Furthermore, analysis of contextual (inherently group-level) variables is not immune to ecological bias [138]. As the size of the ecological association is sensitive to the level of aggregation (being attenuated at higher levels of aggregation), we expect our South Sudan results to be more sensitive to this source of bias than those from Malawi,

Uganda, and DRC [139]. Furthermore, findings from ecological studies are sensitive to the grouping definition used [138]; in Malawi, Uganda, and DRC grouping was uniform across space (defined by grid cells), however in South Sudan we utilized administrative boundaries to define groups, which are uneven and do not have meaning to our research question.

Finally, our results are subject to the g-null paradox. When the parametric g-formula is implemented using non-saturated models and the vector of confounders has any discrete components, as is the case in our study, not all models can be correctly specified. While this may result in the null hypothesis being falsely rejected in theory, in practice the g-formula has been able to estimate null effects, indicating the bias induced by the g-null paradox is relatively small compared with random variance [140].

In summary, our greatest concerns for the validity of these findings are residual confounding by vector control, overfitting of the Malawi models, ecological bias in the South Sudan results, and underreporting in the outcome data. Despite these weaknesses, our study has taken a novel approach to studying the impact of livestock density on HAT risk in a real-world context in which transmission mechanisms exert their effects in the face of spatially-dynamic intervention deployment and vector ecology and behavior, indicating domestic animals are important reservoirs of rHAT in Uganda but not Malawi, and pigs—and possibly cattle—may be reservoirs for gHAT. While achieving 2020 targets for EPHP had previously appeared within reach, lapses in control activities due to COVID-19, and long-standing concerns regarding persistent foci and cryptic animal reservoirs, render the need to close HAT knowledge gaps addressed by this research as pressing as ever.

Chapter 4

LIVESTOCK, PATHOGENS, VECTORS, AND THEIR ENVIRONMENT: A CAUSAL INFERENCE-BASED APPROACH TO ESTIMATING THE PATHWAY-SPECIFIC EFFECT OF LIVESTOCK ON HUMAN AFRICAN TRYPANOSOMIASIS RISK

4.1 Background

Domestic and wild animals—in particular bovids such as cattle, bushbuck, and African buffalo, and suids such as domestic pigs and warthogs—are known to be important reservoirs for the acute form of human African trypanosomiasis, caused by *Trypanosoma brucei rhodesiense*(rHAT) [18, 57, 100, 101], and there is some evidence that domestic pigs may also be reservoirs for the chronic form, caused by *T. b. gambiense* (gHAT).

However, livestock may also influence the distribution of HAT beyond their role as reservoir hosts. Several important vectors of HAT have been documented to prefer animal hosts, and to only bite humans during times of stress [141]. Also appreciated with the malaria vector *Anopheles arabiensis* [60, 142, 143], this preference may shield humans from tsetse bites and thus HAT exposure when favored animal hosts are nearby, termed zoopropylaxis. The fly biting behaviors that favor a zooprophylactic role versus a reservoir role are distinct, with strongly zoophilic tsetse being unlikely to be competent HAT vectors, and animal reservoirs unlikely to be important where the predominant HAT vector is strongly anthrophilic. Furthermore, and zoonotic transmission requires host switching—whereby a given fly feeds on a human host, followed by an animal host, then subsequently on a human host—while zooprophylaxis does not [108].

Tsetse preference for animal hosts is thus harmful (up to a threshold) for HAT control in the presence of animal reservoir, and helpful in its absence. Consistent with this, gHAT

models have found that in the absence of animal reservoirs tsetse biting preference for animals had a negative effect on HAT risk, while in the presence of animal reservoirs biting preference for humans became more influential [108].

Livestock grazing can bring about ecosystem degradation, in the form of desertification (the conversion of fertile land to desert) in arid climates, increased woody plant cover in semi-arid rangelands, and deforestation in humid climates [144]. In the late 20th century, scholars observed that cattle increases (largely limited to tsetse-free areas at the time) resulted in overstocking, overgrazing, and habitat destruction, harming fragile, dry savannah ecosystems. As early as 1988, it was noted that this overgrazing could potentially change local and global weather systems [16]. Insect vectors of public health importance are sensitive to environmental factors, with modeling work demonstrating that gHAT distribution is influenced by habitat fragmentation and that environmental suitability is a function of vegetation activity, precipitation and humidity (positive); and land surface temperature and air temperature (negative) [145]. In Burkina Faso and Ghana, where gHAT was of historical importance and AAT remains so, cattle keeping, crop farming, and other climactic and anthropogenic factors have altered riverine landscapes, fragmenting gallery forests and influencing tsetse distribution [146]. Similarly, in Ethiopia, where AAT is endemic but HAT is not, the distribution of tsetse flies was observed to change following expansion in agricultural land in response to regulatory changes [15]. Furthermore, the clearing of bush by subsistence livestock agriculture in pre-colonial Africa is credited with small-scale control of tsetse, termed "agricultural prophlyaxis," and the resurgence of bush following catastrophic livestock losses to rinderpest ("cattle plague") is credited with the sleeping sickness outbreaks of the early 20th century [16].

Africa is currently undergoing a "livestock revolution," in which population growth and economic development are driving a rapid increase in the demand for livestock products such as meat, milk, and eggs [59, 67]. As agricultural prophylaxis and rinderpest outbreaks in pre-colonial and colonial Africa demonstrated, and decades of subsequent field and modeling studies have confirmed, livestock have an important role in in the control of tsetse habitat,

and it is likely that the livestock revolution will alter the distribution of tsetse flies in Africa.

Livestock may thus come to bear on HAT distribution through three hypothesized pathways: disease reservoir, zooprophylaxis, and environmental change. Neither the environmental change pathway nor the zooprophylaxis pathway are explicitly leveraged for HAT control currently, however the environmental change pathway was successfully employed for HAT and AAT control in pre-colonial Africa through agricultural prophylaxis, and in colonial Africa through environmentally-traumatic approaches such as brush clearing and widespread use of persistent organochlorine insecticides. Furthermore, vector control strategies that target livestock, such as insecticide treatment of cattle (ITC), do exploit the frequent proximity of cattle to humans and their need to travel daily to the riverine habitats favored by gHAT vectors for water, and thus zooprophylaxis. ITC, which been documented to effectively reduce tsetse density at less than US$2 per head annually while also contributing to AAT control [35, 147], is largely farmer-led in gHAT foci with few top-down operations. In contrast, in rHAT foci ITC—a means to control both the vector and the reservoir—and trypanocidal treatment of cattle (TT) are mainstays of HAT control, in particular in Uganda.

Estimation of pathway-specific effects holds potential to narrow knowledge gaps surrounding the epidemiology of HAT, facilitate the targeted deployment of control measures, and provide key parameters for mathematical modeling. Previous authors have estimated elements of these effects in select foci, including the predictive ability of remote-sensed environmental data for tsetse distribution in Nuguruman, Kenya [148], the effect of trypanocide mass-treatment of cattle on rHAT risk in Uganda [24], the effect of habitat fragmentation on rHAT risk in Eastern Zambia [149], and the effect of animal husbandry and climate on AAT risk in Burkina Faso [150]. However, we are aware of no study that seeks to estimate the totality of the effect of cattle and pig density on HAT risk, and further to disentangle the mechanisms driving this effect. Thus, significant knowledge gaps remain surrounding the role of pigs as reservoirs of gHAT and the ability of both pigs and cattle to bring about environmental changes that in turn affect the distribution of HAT.

In this study, we apply the mediational g-formula [111] to estimate the magnitude of

the direct and indirect effect of cattle and pig density on HAT risk in three high-burden countries: Malawi (rHAT), Uganda (rHAT and gHAT), and Democratic Republic of Congo (DRC, gHAT), using remote-sensed data, HAT surveillance data collated by the WHO Atlas of HAT, and a time-series of high resolution livestock density maps. We define livestock density as the ratio of livestock (separately for cattle and pigs) to humans, and fit all models separately for each country, for gHAT and rHAT, and for two mediators: normalized difference vegetation index (NDVI), a measure of vegetation cover, and land surface temperature (LST). Additionally, in South Sudan (gHAT) we use a spatial extension of the regression approach to mediation analyses described by VanderWeele [111], and a single county-level map of livestock density produced for 2008.

4.2 Methods

As detailed in Chapter 3, our study is a retrospective ecological cohort study with cluster-year (0.017° pixel) as the unit of analysis in Malawi, Uganda, and DRC, and county (administrative level 2) as the unit of analysis in South Sudan.

4.2.1 Study population

In Uganda, DRC, and Malawi, our study population included all individuals residing within five hours' travel time of a fixed health facility capable of HAT diagnosis [1]. Study areas were defined separately for gHAT and rHAT. In South Sudan, we restricted the study area to counties which reported one or more HAT cases or conducted active surveillance during the study period. The geographic distribution of this population is shown in Figure 3.4.

The study period was defined separately for each country, on the basis of WHO Atlas of HAT data access provided to the authors (2000-2018 for Uganda and DRC, 2000-2014 for Malawi and South Sudan), and then further limited to the bounds of available input data for livestock mapping to mitigate bias due to uncertainty (measurement error) in the exposure data: 2006-2018 for Uganda, 2003-2014 for Malawi, and 2010-2013 for DRC.

In South Sudan we further restricted the study period to reflect limited exposure data

availability (2008 alone). With confounder selection and the temporal relationship between variables determined by our directed acyclic graph (DAG), detailed in the following section, in South Sudan outcome was 2012 cases and exposure was 2008 livestock density; time-varying confounders were 2007 conflicts, LST, and NDVI, and 2008 wealth as this was the only year for which we had wealth data; and mediators were 2010 NDVI and LST. Note that while both 2007 NDVI and 2007 LST were adjusted for as confounders, mediation models for which 2010 NDVI was the mediator did not adjust for 2010 LST, and vice versa; i.e., we did not test for multiple mediators simultaneously.

4.2.2 DAG formulation

In all study countries we adopted the counterfactual approach to mediation analysis, identifying confounders by the DAG depicted in Figure 4.1. Time-varying variables are indexed by t, while non-time-varying variables are not. In South Sudan, we applied the regression approach [151], while in the remaining countries we used the mediational g-formula [151, 152].

Our DAG encodes our beliefs about the underlying temporal structure of these causal relationships. We believe both elevation and designation as a protected area precede LST and NDVI, index events precede vector control and surveillance (active and passive) coverage, and so on. For variables which exhibit feedback loops (i.e., variable A at time t is a cause of variable B at time t, but variable B at time t is a cause of variable A at time $t+1$), our DAG also encodes our beliefs regarding the structure of these loops. We believe wealth measured at time t precedes livestock density and environmental variables (LST and NDVI) measured at time t, index events at time t cause wealth at time t, wealth and index events at time t causes index events at time $t+1$, wealth at time t causes wealth at time $t+1$, and so on. We assume that livestock density at time t has a direct effect on HAT at time $t+2$, however all time-varying confounders at time t only exert their effects on HAT at time $t+2$ though their effect on the same variable at time $t+1$ (i.e., in the case of wealth, there is no lagged effect of wealth on HAT that is not fully mediated by wealth at the intervening time point).

Note that while ethnicity, educational attainment, and religion are generally not time-

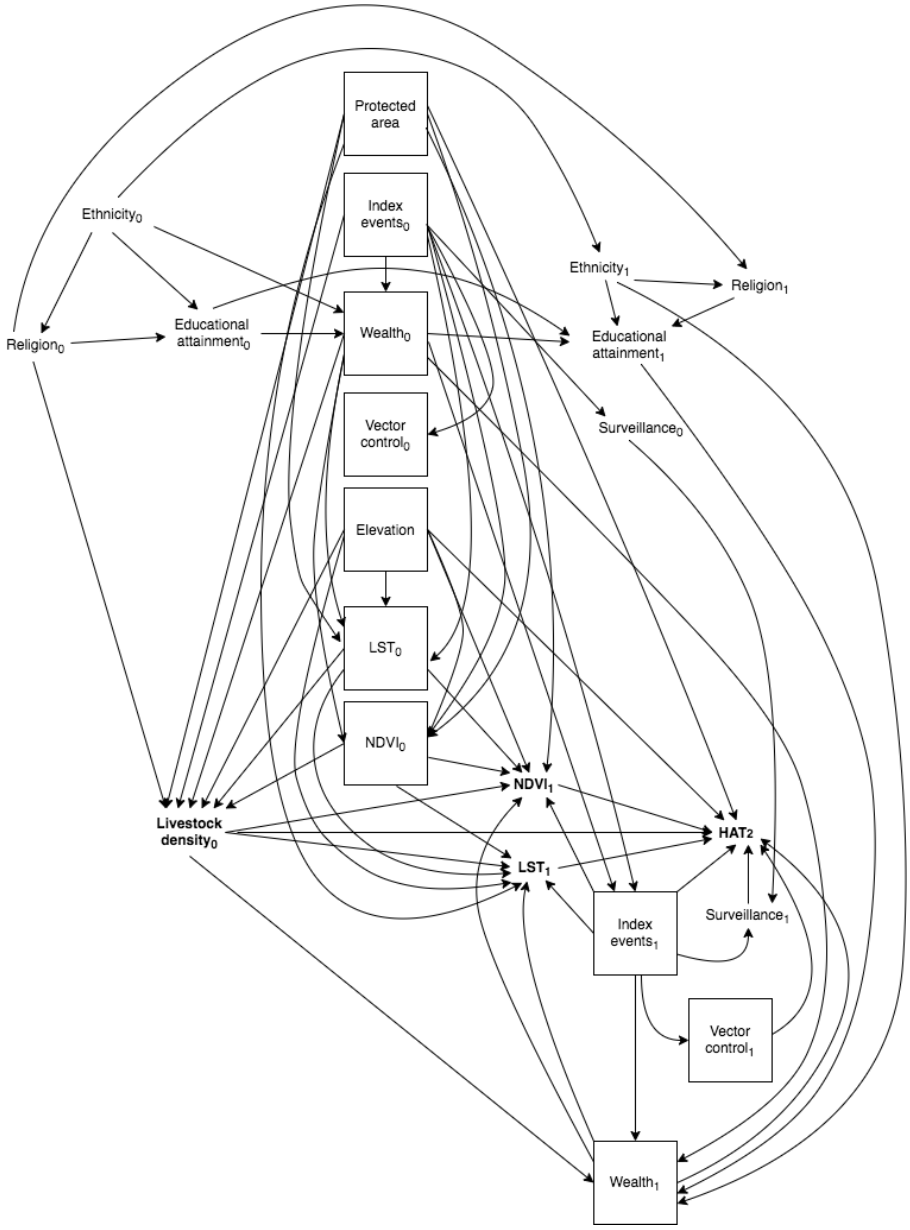


Figure 4.1: Final time-varying directed acyclic graph (DAG). Protected areas pertains to rHAT models only. Exposure and outcome of interest are bolded, and confounders in the minimally-sufficient set are denoted by solid boxes

varying at an individual-level, in our application these represent group-level variables (i.e., proportion receiving higher than a given level of education, proportion belonging to the majority ethnic group, and proportion belonging to Muslim, Christian, and other religions) and thus do vary in time.

4.2.3 Causal estimands

The parameters of interest will be the natural direct and natural indirect effects each for LST and NDVI, however our implementation of the mediational g-formula will generate a randomized interventional analog of the natural direct effect and indirect effects. We will assume partial interference for both the mediator M and outcome Y, in which all units' potential outcomes are independent conditional on a specified interference set, as detailed in Chapter 3. This is an adaptation of estimands proposed by VanderWeele to the setting of spatial autocorrelation [152].

The natural direct effect is the difference in counterfactual outcomes with the mediator held fixed at the value (M^{a*}) it would take under exposure value $A = a*$ (for a binary exposure, this is nominally $A = 0$, or unexposed), and the exposure varied from $A = a*$ to $A = a$. In counterfactual notation, the population average natural direct effect is expressed as $E[Y^{aM^{a*}} - Y^{a*M^{a*}}]$. The natural indirect effect is the difference in counterfactual outcomes with the exposure fixed at $A = a$, but the mediator varied from the value it would take under exposure $A = a$ vs. $A = a*$: $E[Y^{aM^{a}} - Y^{aM^{a*}}]$. These effects are not identifiable when there is exposure-confounder feedback, however their randomized interventional analogs, which compare counterfactual outcomes under mediator values derived from random draws of the mediator distribution under exposure $A = a$ vs. $A = a*$ (G^a and G^{a*}, respectively), can be identified [111]. The resultant natural direct effect analog is rNDE = $E[Y^{aG^{a*}} - Y^{a*G^{a*}}]$, and the natural indirect effect analog is rIDE = $E[Y^{aG^{a}} - Y^{aG^{a*}}]$. For notational simplicity, we will continue to use the Y^{aM^a} notation.

These counterfactual quantities are identifiable from the observed data when the following assumptions hold, in addition to positivity (every cluster-year has a non-zero probability of

receiving any level of treatment, i.e., livestock density), consistency (each cluster-year has one potential outcome for a given treatment level), and partial interference: (1) no exposure-outcome confounding, conditional on the past: $Y^{\overline{a}M^{\overline{a}}} \perp\!\!\!\perp A_t|\overline{A}_{t-1}, \overline{L}_t$,(2) no mediator-outcome-confounding, conditional on the past: $Y^{\overline{a}M^{\overline{a}}} \perp\!\!\!\perp M_t|\overline{A}_t, \overline{M}_{t-1}, \overline{L}_t$, and (3) no exposure-mediator confounding, conditional on the past: $M^{\overline{a}} \perp\!\!\!\perp A_t|\overline{A}_{t-1}, \overline{M}_{t-1}, \overline{L}_t$ [153].

As for our total effect estimate implementation (Chapter 3), we defined $A = a^*$ as mean livestock density, and $A = a$ as mean $\times$ 1.5.

4.2.4 Mediational g-formula

Implementation of the parametric g-formula is detailed in Chapter 3. Briefly, models are fit for each time-varying confounder (including the two mediators in question) and for outcome, separately for each year and ensuring only the past is controlled for. These models are then tied together using the law of total probability and marginalized to obtain the causal parameter $E[Y^{aM^a}]$. We assumed conditional exchangeability with regards to exposure-outcome, exposure-mediator, and mediator-outcome confounding, and again assumed partial interference, defining the interference set as all other clusters within a 5km radius and again stratifying this on administrative level 2 areas in DRC.

We implemented the mediational g-formula as for our total effect estimates, i.e., through fitting a series of `glm()` models in R. We simulated from each model in turn using the `predict()` function and the level of exposure (livestock density) specified, however for the mediation models we fixed exposure density at the level defined by the causal parameter in question. For instance, to estimate the natural direct effect, we need the distribution of $Y^{aM^{a*}}$, which requires $A = a$ for the outcome and confounder models, but $A = a^*$ for the mediator models. We produced uncertainty bounds through non-parametric bootstrapping (100 simulations).

Our targets of inference are:

$$\text{rNDE} = E_{(s)}\left[E_t\left[E_i\left[Y_{t(s)i}^{\bar{a}M^{\bar{a}*}} \;/\; Y_{t(s)i}^{\bar{a}*M^{\bar{a}*}}\right]\right]\right]$$

$$\text{rNIE} = E_{(s)}\left[E_t\left[E_i\left[Y_{t(s)i}^{\bar{a}M^{\bar{a}}} \;/\; Y_{t(s)i}^{\bar{a}M^{\bar{a}*}}\right]\right]\right]$$

where s indexes simulation (100 in total), t indexes year, i indexes cluster, and $Y_{t(s)i}^{\bar{a}M^{\bar{a}}}$ is the one-year cumulative incidence of HAT for a given livestock species, under $A = a$.

As for the total effect estimate, the mediational g-formula is validated using the natural course model.

4.2.5 Regression approach

Due to the absence of longitudinal data in South Sudan (Chapter 2), we applied a regression approach to our mediation analysis in this country. Under this approach, two models are fit—one for outcome, and one for mediator—and causal estimands are derived from linear combinations of parameters from these models.

Consistent with our DAG formulation, we assume that wealth measured in the 2008 South Sudan census is upstream of livestock density ascertained in the same census. Thus, wealth at time t_1 constitutes a mediator-outcome confounder which is downstream of exposure (as livestock density is a cause of wealth), precluding a simple regression approach to mediation analysis [151]. However, wealth was only measured at time 0 (2008), thus we proceeded with the regression approach and acknowledge our results will not have a causal interpretation.

The outcome (Poisson) model was as follows, where c indexes county:

$$Y_c|\lambda_c \sim Poisson(P_c\lambda_c)$$

$$\lambda_c = exp\left(\theta_0 + \theta_1\mu_c + \theta_2 m_c + \boldsymbol{\alpha}\mathbf{z}(\mathbf{k}_c) + S(\mathbf{s}_c) + \epsilon_c\right)$$

where

- Y_c is the number of cases in county c in 2012
- m_c is the mediator value in county c in 2010, $M \in$ {NDVI, LST}
- μ_c is estimated livestock density in cluster i in 2008
- $\boldsymbol{\alpha}$ is a vector of coefficients
- $\mathbf{z}(\mathbf{k}_c)$ is a vector of exposure-outcome, exposure-mediator, and mediator-outcome confounders, $k \in$ {2008 wealth; 2007 NDVI, LST, conflicts, disasters; 2009 conflicts, disasters}
- $S(\mathbf{s}_c)$ are structured county-level spatial random effects, assumed to follow the ICAR model detailed in Chapter 2
- ϵ_c are unstructured (iid) county-level spatial random effects
- P_c is the offset, given as population in county c

and the mediator model was:

$$M_c | x_c \sim_{iid} N(x_c, \sigma_M^2)$$

$$x_c = \beta_0 + \beta_1 \mu_c + \boldsymbol{\gamma} \mathbf{z}(\mathbf{k}_c) + S(\mathbf{s}_c) + \epsilon_c$$

where

- M is the mediator value in county c, $m \in$ {2010 NDVI, 2010 LST}
- μ_c is estimated livestock density in county c in 2008
- $\boldsymbol{\gamma}$ is a vector of coefficients

- $\mathbf{z}(\mathbf{k}_c)$ is a vector of exposure-mediator confounders, $k \in$ {2008 wealth; 2007 NDVI, LST, conflicts, disasters}

- $S(\mathbf{s}_c)$ are structured county-level spatial random effects, assumed to follow the ICAR model detailed in Chapter 2

- ϵ_c are unstructured (iid) county-level spatial random effects

As these models do not contain an exposure × mediator interaction term, the direct effect is then calculated as:

$$RR^{NDE} = exp[\theta_1 \times (\mu - \mu^*)]$$

and the indirect effect is calculated as:

$$RR^{NIE} = exp[\theta_2\beta_1 \times (\mu - \mu^*)]$$

where $\mu - \mu^*$ are the two values of livestock density we are comparing [151]. We have parameterized exposure as for the total effect models, where the contrast corresponds to a 50% increase in livestock density. Due to the absence of an exposure-mediator interaction term, which implies linearity in the direct and indirect effects of exposure on outcome across levels of the mediator, the direct effect can be interpreted as either the controlled direct effect or the natural direct effect, with the controlled direct effect being the expected change in outcome if the mediator was fixed at level m, but exposure changed from μ^* to μ [151]. No off-the-shelf software is available to calculate standard errors for these estimands in the presence of spatial structure, thus we derived uncertainty estimates by taking 1,000 draws from the posterior distribution of each parameter in the above models, and taking the 2.5% and 97.5% percentiles of this distribution as the posterior 95% credible interval.

4.2.6 Data and code availability

HAT outcome data can be requested from the WHO (`https://www.who.int/trypanosomiasis_african/country/foci_AFRO/en/`) and livestock density maps can be downloaded from `https://github.com/JulianneMeisnerUW/LivestockMaps`. Mediator and confounder data are available for download from the websites linked in our Bibliography. All analyses were performed in R, and all code are available in the same GitHub repository.

4.3 Results

Descriptive statistics are presented in Tables 3.1-3.2 and Appendix C.

4.3.1 Mediational g-formula

Validation of the parametric g-formula is presented in Chapter 3; results indicated very strong agreement between the natural course model and observed livestock density with the exception of 2010 and 2012 for the pig-gHAT models in Uganda. We removed these years from the Uganda analyses.

Results are presented in Tables 4.1 - 4.3. In Uganda, we found no evidence of an indirect effect for either mediator or livestock species, for both forms of HAT. Direct effects were positive for cattle in rHAT foci and for pigs in gHAT and rHAT foci (being strongest for pigs in rHAT foci), and negative for cattle in gHAT foci, however these effects only reached statistical significance for the direct effect of pigs on rHAT.

DRC results were similar to the Uganda gHAT results (i.e., no evidence of an indirect effect, positive direct effects for pigs), however we also detected a positive direct effect for cattle—albeit a weaker effect than the direct effect for pigs—and all direct effects reached statistical significance.

In Malawi, we did not run LST models due to large amounts of missingness, and we removed location within a protected area, elevation, and disaster from all models as very few HAT cases were observed in Malawi during the study period, driving concerns regarding

Table 4.1: Mediation analysis results, Uganda

		RR (95% CI)	
	Mediator	Direct effect	Indirect effect
Cattle			
rHAT	NDVI	1.62 (0.85, 2.85)	1.01 (0.88, 1.21)
	LST	1.68 (0.84, 2.82)	1.00 (0.87, 1.15)
gHAT	NDVI	0.88 (0.55, 1.65)	1.00 (0.89, 1.19)
	LST	0.79 (0.53, 1.35)	1.01 (0.90, 1.11)
Pigs			
rHAT	NDVI	2.15 (1.33, 2.98)	0.96 (0.87, 1.09)
	LST	2.16 (1.33, 3.05)	0.98 (0.89, 1.07)
gHAT	NDVI	2.18 (0.86, 1.72)	0.99 (0.90, 1.13)
	LST	1.41 (0.85, 1.58)	1.01 (0.94, 1.11)

Mediational g-formula implemented such that effect estimates correspond to a 50% increase in livestock density. RR: relative risk analog; 95% CI: credible interval over 100 iterations of the parametric g-formula

overfitting. For NDVI, we detected positive but non-significant and very small indirect effects, and negative direct effects (significant for pigs, not for cattle). The direct effect for pigs was particularly strong.

4.3.2 Regression

As the main models are expected to be biased, we will only interpret the measurement error (MEC) models here. For cattle, direct effect estimates were generally stronger than indirect effect estimates. While indirect effects were strong for LST, in all cases uncertainty bounds

Table 4.2: Mediation analysis results, Malawi

	RR (95% CI)	
Mediator	Direct effect	Indirect effect
Cattle		
NDVI	0.84 (0.66, 1.01)	1.03 (0.82, 1.31)
Pigs		
NDVI	0.37 (0.20, 0.67)	1.02 (0.72, 1.32)

Mediational g-formula implemented such that effect estimates correspond to a 50% increase in livestock density. LST not run for Malawi due to large amounts of missingness in the source data. RR: relative risk analog; 95% CI: credible interval over 100 iterations of the parametric g-formula

Table 4.3: Mediation analysis results, DRC

	RR (95% CI)	
Mediator	Direct effect	Indirect effect
Cattle		
NDVI	1.10 (1.00, 1.22)	1.04 (0.99, 1.14)
LST	1.18 (1.05, 1.30)	0.99 (0.94, 1.06)
Pigs		
NDVI	1.21 (1.07, 1.33)	1.02 (0.96, 1.08)
LST	1.25 (1.11, 1.42)	0.99 (0.95, 1.04)

Mediational g-formula implemented such that effect estimates correspond to a 50% increase in livestock density. RR: relative risk analog; 95% CI: credible interval over 100 iterations of the parametric g-formula

were very wide and no effects reached statistical significance. For pigs, indirect effects were generally stronger than direct effects, however direct and indirect effects were not in the same direction for NDVI (positive for the direct effect, negative for the indirect effect). As for cattle, uncertainty bounds were wide in all cases. Results are presented in Table 4.4.

4.4 Discussion

In Malawi, Uganda, and DRC, we did not find evidence of an indirect effect. In South Sudan we did detect a strong indirect effect for LST for both cattle and pigs, and for NDVI for pigs. Direct effects were positive for cattle and pigs in rHAT foci in Uganda and for pigs in gHAT foci in Uganda and DRC, indicating a reservoir effect. In gHAT foci in South Sudan, the direct effect of pigs not mediated by NDVI was positive, but the direct effect of pigs not mediated by LST was negative, providing mixed evidence of a reservoir effect for pigs. Direct effects were negative for cattle and pigs in rHAT foci in Malawi, and for cattle in gHAT foci in Uganda and South Sudan. As for total effects (Chapter 3), direct effects for cattle in gHAT foci in DRC were positive.

These results indicate the environmental pathway mediated by NDVI and LST is weak—if present—in rHAT and gHAT foci in Malawi, Uganda, and DRC, but strong in South Sudan. They also suggest that, for rHAT, livestock are an important reservoir in Uganda but not Malawi. For gHAT, our results point to a reservoir role for pigs in both Uganda and DRC and provide some evidence that pigs are a reservoir in South Sudan. These results also provide equivocal evidence that cattle may serve as a gHAT reservoir: while there is evidence of a slight reservoir effect for gHAT in DRC, Uganda and South Sudan provide moderate and strong evidence of a zooprophylactic effect, respectively.

As for our total effect estimates (Chapter 3), our findings in Malawi are not necessarily discordant with what is known mechanistically—i.e., cattle and pigs are rHAT reservoirs—if this reservoir poses little risk to humans, allowing the reservoir effect to be dominated by the zooprophylactic effect. This could arise if wildlife are the dominant source of human infection—due to host abundance, tsetse fly preference, contact patterns between hosts,

Table 4.4: Mediation analysis results, South Sudan

		RR (95% CI	
Mediator	Model	Direct effect	Indirect effect
Cattle			
NDVI	Naive	0.34 (0.09, 1.28)	0.99 (0.46, 1.60)
LST	Naive	0.32 (0.12, 0.80)	0.73 (0.09, 2.88)
NDVI	MEC	0.22 (0.13, 0.41)	0.98 (0.38, 1.55)
LST	MEC	0.28 (0.17, 0.45)	0.50 (0.09, 1.54)
Pigs			
NDVI	Naive	0.96 (0.09, 8.85)	0.75 (0.13, 3.59)
LST	Naive	0.93 (0.26, 3.05)	0.81 (0.13, 3.50)
NDVI	MEC	1.55 (0.52, 3.68)	0.62 (0.09, 1.35)
LST	MEC	0.86 (0.44, 1.68)	0.57 (0.13, 2.15)

Estimates of direct (natural or controlled) effect and natural indirect effect, using 2012 HAT data as outcome, 2008 livestock density data as exposure, and 2010 NDVI and LST data as mediators, fit in two separate models. Adjustment performed for 2008 wealth, 2007 NDVI, 2007 LST, and 2007 and 2009 armed conflicts. Density is parameterized such that rate ratios correspond to a 50% increase in density. RR: rate ratio; CI: credible interval, taken over 1,000 draws from the posterior distribution of each parameter; Naive: model fit without accounting for measurement error in livestock density or wealth; MEC: measurement error model.

adequate control of the domestic reservoir through insecticides or trypanocides, or other means—and domestic animals exert an important zooprophylactic effect. The spatial distribution of HAT in Malawi, where most active transmission occurs near protected areas, indicates this scenario may be consistent with HAT epidemiology in this setting [5].

As with any imperfect study—interventional or observational, randomized or not—our results may also be the result of bias. We discuss bias due to confounding, selection bias, measurement error, ecological bias, and the g-null paradox in detail in Chapter 3. Briefly, unmeasured confounding due to inability to control for vector control, incomplete control by wealth as our parameterization might not have captured the dimension most relevant to our research question, and the need to drop several confounders from our Malawi model due to overfitting concerns, may lead to residual confounding. In Chapter 3, we calculate E-values for total effect estimates, which indicate a strong degree of unmeasured confounding would be required to account for our cattle-rHAT findings in Uganda, our pig-rHAT findings in Uganda and Malawi, our pig-gHAT findings in Uganda and South Sudan, and our cattle-gHAT findings in South Sudan.

Selection bias may arise if clusters or counties excluded from the study are not actually not at risk of reporting a HAT case, and differ systematically in their distribution of the exposure (livestock density), mediators (LST and NDVI) or confounders. In South Sudan, this could arise if underreporting in the outcome data (discussed below) results in exclusion of an at-risk county from our analyses completely.

One of the major sources of bias in our study is underreporting in the outcome data. If such underreporting is non-differential with respect to HAT risk (i.e., clusters or counties counties at high risk of HAT are no more likely to underreport than those at low risk, even if high-risk individuals within each cluster or county are less likely to enter the reporting system than low-risk individuals), effect estimates will be biased to the null in expectation. While weak surveillance infrastructure may drive an association between cluster- or county-level HAT risk and HAT reporting, we expect this mechanism to largely arise via civil unrest, which we have attempted to control for. Furthermore, with the exception of South Sudan,

we have not addressed uncertainty in livestock density and wealth in our estimates, and in South Sudan our implementation of the measurement error model assumes errors in wealth and livestock density are not correlated in space, which is likely violated. Measurement error may also arise in our confounders, driving residual confounding, and in our mediators (NDVI and LST), biasing direct and indirect effects.

Finally, ecological bias arises from aggregation of variables, the presence of an unmeasured individual-level confounder whose association with exposure differs across groups, or the presence of an unmeasured individual-level effect modifier whose effect or distribution differs across groups [134, 135, 136]. As ecological bias is sensitive to grouping definition, we are particularly concerned about its effects in South Sudan, where spatial units were large and irregular.

Concerns about overfitting in the Malawi analyses, ecological bias in the South Sudan analyses, and underreporting in the outcome data for all countries aside, our study takes a novel approach to estimating the pathway-specific effects by which cattle and pigs come to bear on gHAT and rHAT risk. By conducting our study across four high-burden countries, our findings lend strong supporting evidence for a reservoir effect of pigs for gHAT. Our results also indicate that interventions on domestic animal reservoirs have the potential to improve rHAT control in Uganda but are unlikely to yield much benefit in Malawi, where efforts should instead be targeted to vector control and strengthening passive surveillance systems. In conjunction with high-resolution livestock maps these results can be interpreted on a policy-relevant scale, allowing policymakers and other stakeholders to identify priority areas for implementing domestic animal AAT control in a coodinated One Health framework.

Chapter 5

DOES A ONE HEALTH APPROACH TO HUMAN AFRICAN TRYPANOSOMIASIS CONTROL HASTEN ELIMINATION? A STOCHASTIC COMPARTMENTAL MODELING APPROACH TO STUDYING LINKED HUMAN AND ANIMAL AFRICAN TRYPANOSOMIASIS CONTROL

5.1 Background

Remarkable progress in the control of human African trypanosomiasis (HAT) in the early 2000s led the WHO to set targets for elimination as a public health problem (EPHP) by 2020, and elimination of transmission (EOT) by 2030. While EPHP goals include both the acute form of HAT, caused by *Trypanosoma brucei rhodesiense* (rHAT) and known to have important animal reservoirs, and the chronic form caused by *T. b. gambiense* (gHAT) and thought to be predominantly transmitted human-to-human, EOT goals target gHAT alone.

An obvious challenge to achieving EOT is that zero reported cases does not necessarily equal EOT: EOT could occur before zero detections due to the long natural history of gHAT, or after zero detections due to missed cases. Models fit to Bandundu Province, DRC indicate 63% of all gHAT infections are unreported, while only 20% of cases that escape active surveillance are identified by passive detection [2]. As decisions on when to stop active surveillance in favor of passive detection-only strategies rely on case reports, and HAT has been demonstrated to be remarkably persistent at low prevalence [7], establishing the positive predictive value of zero reported cases for EOT is critical.

Mathematical modeling efforts for HAT have largely been deterministic, requiring the adoption of proxy thresholds for EOT (e.g., <1 new infection per 100,000 or 1,000,000 per year) [2] as deterministic models represent populations with continuous variables that never reach zero [154]. Conversely, stochastic models evaluate the likelihood of disease elimination

through natural failure of transmission events, allowing for representation of uncertainty and distribution of time to elimination [9]. Recently, there have been several stochastic efforts to model HAT elimination using Gillespie-based simulation algorithms [2]. Davis et al. (2019) used this approach, and a Ross-Macdonald-type stochastic compartmental model extended from Rock et al.'s (2015) previous deterministic model [99], to model gHAT persistence under varying parameterizations of reported screening patterns and population sizes at the village-level. They found probability of persistence increased with population size, but that gHAT transmission can persist for long periods (15 years or more) in isolated populations as small as 2,000 (with yearly incidence of approximately 13 infections). This is in stark contrast with highly-transmissible disease such as measles, where local extinction is common in populations of up to 300,000. These findings suggest deterministic gHAT models at the level of the health district (approximately 100,000 population) are adequate, however it is less clear whether this threshold applies to the study of chance transmission events from animal reservoirs. Davis et al (2019) also found evidence of high spatial heterogeneity related to local environmental conditions, in particular proximity to large rivers. Citing only marginal improvement in model fit for the original deterministic model, and uncertainty being too great to model resurgence via this mechanism, animal reservoirs were not included in this model [155]. In an effort to evaluate whether Davis et al.'s (2019) findings scale-up to the health district level, Castaño et al. (2020) extended two previously deterministic models to a stochastic framework, exploring seven medical-only strategies. Defining EOT as the first five consecutive years of no new transmissions, they found no strong evidence that current medical interventions and decreasing case counts would be adequate to achieve EOT by 2030 in two health zones in DRC. As active surveillance coverage improved, so did both time to EOT and positive predictive value of zero reported cases per year for EOT, consistent with Davis et al.'s (2019) findings at the village level [7, 155].

While multiple animal species harbor trypanosomes, their role in the transmission cycle is unclear [9]. Using data from Uganda, Davis et al. (2011) demonstrated that proportion of bloodmeals taken from humans was the most important determinant of gHAT distribution

among those evaluated [156]. A model fit to animal prevalence data from Cameroon using the Next Generation Matrix approach and assuming a constant level of human infection suggested animals constitute a transmission reservoir for gHAT [109], while others fit to longitudinal human case data from Guinea, Chad, and DRC with decreasing reporting trends and time-varying interventions showed similar support for presence versus absence of an animal reservoir [6, 45, 99]. These discrepancies may reflect differences in model assumptions, or the foci-specific nature of reservoir potential based on human-tsetse-animal abundance and contact patterns [2].

Several authors have also used agent-based models (ABMs) to study the role of animal reservoirs on HAT transmission. Alderton et al. (2013) built an ABM that included tsetse, humans (stratified into farmer and non-farmer agents), and cattle in a riverine habitat, and found infection propagated faster among non-farmers as there were other potential hosts nearby when farmers encountered a tsetse fly, indicating cattle play a zooprophylactic role. However increases in cattle population were found to increase human infection up to a threshold of 1:7 [157]. Grebaut et al. (2016) later developed the HATSim model, which modeled the distribution of human, tsetse, and wild and domestic animal agents based on field data of tsetse and wild animal populations, while accounting for spatial complexity in the modeled focus by applying an environmental attribute to each cell. They found vector control (VC) offered the highest probability of EOT and that the importance of animal reservoirs is mainly limited by duration of animal infection, with four months being sufficient for maintaining an endemic situation in some simulations. No autonomous cycles in animals, which could be maintained without human cases and re-initiate an endemic situation, were observed [158].

Two simple and cost-effective methods of VC have emerged in the past decade: insecticide treatment of cattle (ITC) in the form of restricted application of synthetic pyrethroids (RAP) where cattle are present in sufficient numbers (density $>$ 10 animals/km^2) and provide a significant proportion of bloodmeals for HAT vectors, and deployment of pyrethroid-treated tiny targets along the banks of rivers, the preferred habitat for riverine tsetse, otherwise [120]. Application of synthetic pyrethroids to the belly and lower legs of older and larger

cattle, the preferred feeding sites and hosts of HAT and animal African trypanosomiasis (AAT) vectors, respectively [31, 32, 33], has been demonstrated to achieve VC at one tenth the cost of other ITC approaches (<US$2/head/year [35]). In addition to minimizing cost and environmental residues, this approach contributes to control of AAT, widely considered to be the single greatest constraint to increased livestock production in Africa and an important poverty-reinforcing disease [13, 14, 15, 16], increasing stakeholder engagement in HAT control [8]. In both gHAT and rHAT foci, ITC can lower the R_0 of HAT by reducing the average life expectancy and thereby proportion of tsetse flies old enough to harbor mature infection. In rHAT foci, ITC can further reduce R_0 by controlling the domestic animal reservoir of the disease and selectively killing tsetse infected by this reservoir, with potentially further gains being made through addition of trypanocide treatment (TT) of domestic animal reservoirs. Previous literature has demonstrated ITC is effective even when cattle are patchily distributed, with gaps of several kilometers wide [36].

In 2012, Hargrove et al. generalized Rogers (1988) [159] two-host deterministic compartmental model to study the effect of ITC and TT on the R_0 of rHAT in a scenario in which a single tsetse species can feed off of any finite number of vertebrate hosts. The authors found control of rHAT through TT alone, which has no effect on the age structure or abundance of the tsetse population, is unlikely, even if there are no wildlife reservoirs present: as long as such reservoirs contribute more than 10% of bites, the R_0 contribution from these hosts alone is greater than 1. In contrast, ITC could eliminate rHAT if cattle comprise at least 40% of non-human tsetse bloodmeals and 100% of cattle are treated, or if cattle comprise 100% of non-human bloodmeals and 25% are treated [8]. While studies of effectiveness of ITC in gHAT foci are limited, spot-on ITC was found to reduce the fly population from 54.2 flies/trap/day to 0.06 flies/trap/day in a gHAT focus in Burkina Faso, with the remaining flies (*Glossina palpalis palpalis*) mainly feeding on monitor lizards, which are not thought to harbor human-infective trypanosomes [147].

As of 2018, eight out of 26 gHAT endemic countries were eligible for EPHP validation. We focus here on four countries who do not meet the required criteria for EPHP validation

due to an excessive number of cases per health district (Malawi), inadequate control and surveillance activities (Uganda), or both (Democratic Republic of Congo (DRC) and South Sudan) [5]. In this study, we use a stochastic compartmental model implemented in the Institute for Disease Modeling's Compartmental Modeling Software (CMS) [160], to study the effect of including animal reservoirs in HAT control efforts, under a One Health approach, on time to EOT in these countries. We have parameterized this model to reflect the national epidemic in each country, fitting two models each for DRC and Uganda to reflect the presence of two gHAT vectors in DRC and the presence of both gHAT and rHAT in Uganda.

5.2 Methods

5.2.1 Structure

We constructed a stochastic compartmental of HAT transmission, implemented in CMS using the Gillespie algorithm [160], defining EOT as the first year after which no new transmission occurs in the modeled time horizon. For gHAT we fit a four-species model with humans, tsetse flies, reservoir animals, and non-reservoir tsetse hosts (i.e., animal species from which tsetse flies take bloodmeals but which do not harbor human-infective trypanosomes), with domestic swine defined as the animal reservoir. For rHAT we fit a seven-species model with humans, tsetse flies, non-reservoir tsetse hosts, and four reservoir animal species groups: domestic swine, domestic bovids (cattle), wild swine (warthog), and wild bovids (bushbuck and African water buffalo).

Model structure is detailed in Figure 5.1. The sub-model is S-E-I_1-I_2-R for humans, S-E-I-R for reservoir animals, and S-E-I for flies, with a non-susceptible compartment to reflect the teneral effect, whereby flies are most susceptible to transmission of trypanosomes during their first blood meal and within their first 24 hours of life. Humans with stage 2 illness are assumed to be inaccessible to flies and thus do not contribute to transmission. For the rHAT models, model structure is equivalent for the sub-model corresponding to each animal reservoir species group.

Time steps were day, with an overall time horizon of 2005-2045. Previous stochastic compartmental models of HAT have indicated varying step size from 0.01 days to 1 day does not appreciably change results [155]. We assumed all populations are closed, that is a death is replaced by a susceptible, and no migration into or out of foci by infected flies, humans, or reservoir hosts occurs. Again, previous stochastic modeling efforts for HAT have demonstrated that such assumptions have negligible effects on model findings [155].

5.2.2 Parameterization

We parameterized this model to the national gHAT epidemics in South Sudan and Uganda, the national rHAT epidemics in Uganda and Malawi; and the Bandundu and Sakuru foci, together, and Equateur Nord focus, separately, in DRC. This parameterization, which resulted in a total of 6 models, was chosen to reflect the differences in the epidemiology of the each country's HAT epidemics, with the gHAT and rHAT epidemics being distinct within Uganda, and the Bandundu/Sankuru foci being distinct from the Equateur Nord focus in DRC due to different vector populations.

We assumed *Glossina fuscipes fuscipes* was the vector species in the Uganda gHAT, Uganda rHAT, South Sudan, and DRC Equateur Nord models. We assumed *G. f. quanzensis* was the vector species in the DRC Bandundu/Sakuru model, and *G. morsitans morsitans* in the Malawi model. For the human sub-models, we assumed all detected cases are treated, treatment is always successful, and no recovery occurs without treatment (i.e., cases transition from stage 2 infection to death). We assumed the population of non-reservoir tsetse hosts was stable throughout the modeled period.

Base-case model parameters are detailed in Appendix D. We derived parameters from published literature or model fitting, with the exception of the human:wild animal (reservoir or non-reservoir) ratio, which we took to be 10 in all cases; ITC coverage, which we took to be 0% for all reservoirs; TT coverage, which we took to be 0% for wild animal reservoirs in rHAT foci and domestic animal reservoirs in gHAT foci, and 50% for domestic animal reservoirs in rHAT foci; and TT frequency, which we took to be every 3 months. Adapted

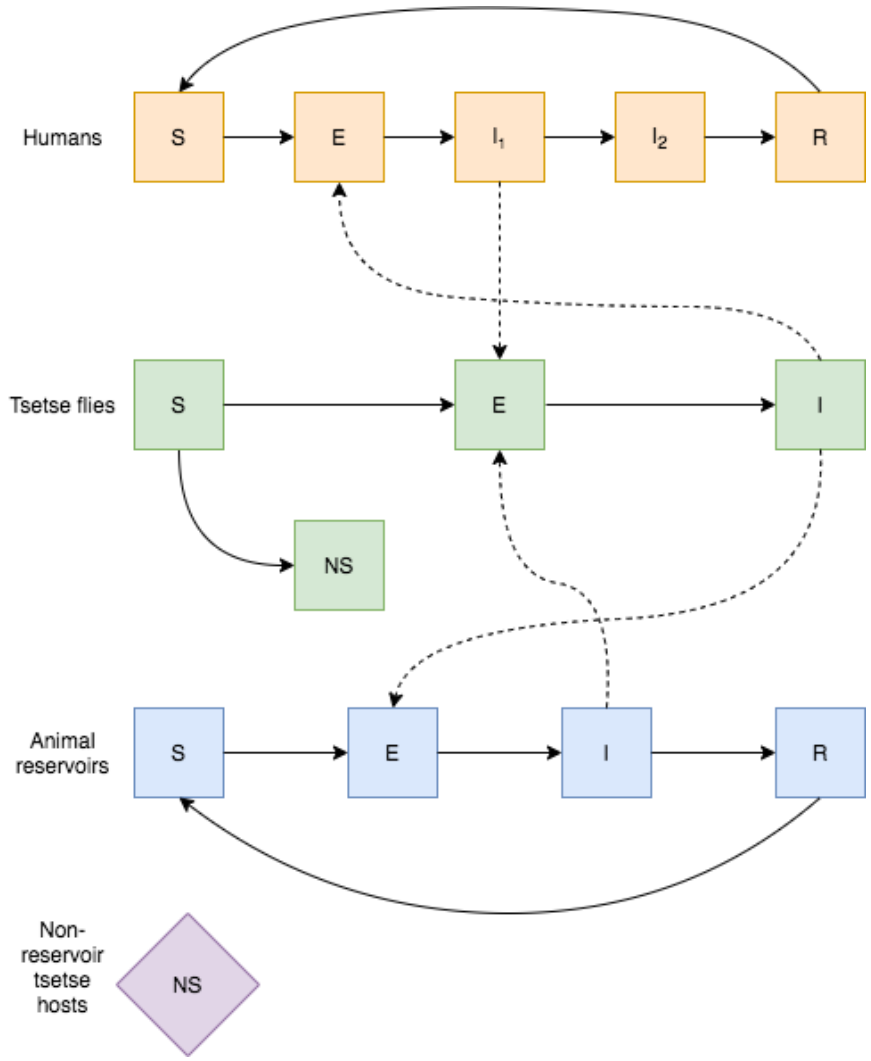


Figure 5.1: Schematic of stochastic compartmental model. Dashed lines correspond to transmission events; non-reservoir tsetse hosts do not contribute to transmission. S: susceptible; E: exposed; I_1: infected stage 1; I_2: infected stage 2; R: recovered; NS: non-susceptible.

from Davis et al. (2019), we defined effective tsetse density as the product of the fly:human ratio and β_H, the probability of human infection per single infective bite [155]. We fixed the former at 6.56, from Davis et al. (2019) in all models, but allowed β_H to vary.

5.2.3 Model equations

Humans

The rate of change in the susceptible compartment is given as:

$$\frac{dS_H}{dt} = \mu_H\Big(E_H(t) + I_{1H}(t) + I_{2H}(t) + R_H(t)\Big) + \omega_H R_H(t) - \lambda_H(t)S_H(t) \tag{5.1}$$

where μ_H is background mortality, $1 \,/\, \omega_H$ is the duration of immunity, and $\lambda_H(t)$ is the force of infection at time t, given as:

$$\lambda_H(t) = \beta_H \alpha p_H \frac{I_V(t)}{N_V(t)} \frac{N_V(t)}{N_H(t)} \tag{5.2}$$

where β_H is the probability of transmission from a tsetse fly to a human during a bloodmeal, α is the daily probability a fly takes a bloodmeal, p_H is the probability a given bloodmeal is taken from a human, $\frac{I_V(t)}{N_V(t)}$ is the probability a given fly is infectious at time t, and $\frac{N_V(t)}{N_H(t)}$ is the ratio of flies:humans at time t. Births are not explicitly modeled as they are assumed to have a 1:1 relationship with background mortality.

The rate of change in the exposed compartment is given as:

$$\frac{dE_H}{dt} = \lambda_H(t)S_H(t) - \sigma_H E_H(t) - \mu_H E_H(t) \tag{5.3}$$

where $1 \,/\, \sigma_H$ is the duration of the incubation period. The rate of change in the stage 1 infectious compartment is:

$$\frac{dI_{1H}}{dt} = \sigma_H E_H(t) - \phi_{H_1} I_{1H}(t) - \mu_H I_{1H}(t) \tag{5.4}$$

where $1 \,/\, \phi_{H_1}$ is the duration of the stage 1 infectious period. The rate of change in the stage 2 infectious compartment is:

$$\frac{dI_{2H}}{dt} = \phi_{H_1} I_{1H}(t) - \phi_{H_2} I_{2H}(t) - \mu_H I_{2H}(t) \tag{5.5}$$

where $1 / \phi_{H_2}$ is the duration of stage 2 infectious period. The rate of change in the recovered compartment is:

$$\frac{dR_H}{dt} = \phi_{H_1}\tau_{H_1}I_{1H}(t) + \phi_{H_2}\tau_{H_2}I_{2H}(t) - \omega_H R_H(t) - \mu_H R_H(t) \tag{5.6}$$

where τ_{H_1} is the probability of detection in stage 1 and τ_{H_2} is the probability of detection in stage 2. We assume detection perfectly predicts recovery, that is 100% of detected cases recover, and 100% of undetected cases transition to death, an absorbing state.

Reservoir animals With i indexing reservoir animal species group ($I = 1$ in gHAT models), rate of change in the susceptible compartment is given as:

$$\frac{dS_{Ri}}{dt} = \mu_{Ri}\Big(E_{Ri}(t) + I_{Ri}(t) + R_{Ri}(t)\Big) + \omega_{Ri}R_{Ri}(t) - \lambda_{Ri}(t)S_{Rit}(t) \tag{5.7}$$

where, for species i, μ_{Ri} is background mortality, $1 / \omega_{Ri}$ is the duration of immunity, and $\lambda_{Ri}(t)$ is the force of infection at time t, given as:

$$\lambda_{Ri}(t) = \beta_{Ri}\alpha p_{Ri}(t)\frac{I_V(t)}{N_V(t)}\frac{N_V(t)}{N_{Ri}(t)} \tag{5.8}$$

where, for species i, β_{Ri} is the probability of transmission from a tsetse fly to a reservoir animal during a bloodmeal, α is the daily probability a fly takes a bloodmeal, $\frac{I_V(t)}{N_V(t)}$ is the probability a given fly is infectious at time t, and $\frac{N_V(t)}{N_{Ri}(t)}$ is the ratio of flies:reservoir animals of species i at time t, and $p_{Ri}(t)$ is the probability a given bloodmeal is taken from species i at time t, given as:

$$p_{Ri}(t) = (1 - p_H) \times \frac{N_{Ri}(t)}{\left(\sum_{i=1}^{I} N_{Ri}(t)\right) + N_{nR}(t)} \tag{5.9}$$

where p_H is the probability a given bloodmeal is taken from a human, $N_{Ri}(t)$ is the number of reservoir animals of species i at time t, and $N_{nR}(t)$ is the number of non-reservoir tsetse hosts at time t. This formulation imposes the assumption that tsetse flies have an inherent level of anthrophily, with the remainder of bloodmeals being distributed among host species according to their abundance. Births are not explicitly modeled as they are assumed to have a 1:1 relationship with background mortality.

The rate of change in the exposed compartment is given as:

$$\frac{dE_{Ri}}{dt} = \lambda_{Ri}(t)S_{Ri}(t) - \sigma_{Ri}E_{Ri}(t) - \mu_{Ri}E_{Ri}(t) \tag{5.10}$$

where $1 \,/\, \sigma_{R_i}$ is the duration of the incubation period. The rate of change in the infectious compartment is:

$$\frac{dI_{Ri}}{dt} = \sigma_{Ri}E_{Ri}(t) - \phi_{Ri_{nt}}(1-\tau_{R_i})I_{Ri}(t) - \phi_{Ri_t}\tau_{R_i}I_{Ri}(t) - \mu_{Ri}I_{Ri}(t) \tag{5.11}$$

where $1 \,/\, \phi_{Ri_{nt}}$ is the duration of the infectious period without treatment,$1 \,/\, \phi_{Ri_t}$ is the duration of the infectious period with treatment, and τ_{R_i} is the probability an animal receives treatment. This parameterization imposes the assumption that all reservoir animals eventually recover, but the rate of this recovery is dependent on trypanocide treatment. This is consistent with observations that *T. brucei* infection, unlike *T. congolense* and *T. vivax*, the primary causative agents of AAT, generally cause no to mild disease in animals [18].

The rate of change in the recovered compartment is:

$$\frac{dR_{Ri}}{dt} = \phi_{Ri_{nt}}(1-\tau_{R_i})I_{Ri}(t) + \phi_{Ri_t}\tau_{R_i}I_{Ri}(t) - \omega_{Ri}R_{Ri}(t) - \mu_{Ri}R_{Ri}(t) \tag{5.12}$$

Flies The rate of change in the susceptible compartment is given as:

$$\frac{dS_V}{dt} = \mu_V\Big(E_V(t) + I_V(t) + NS_V(t)\Big) - \lambda_V(t)S_V(t) \tag{5.13}$$

where μ_V is background mortality and $\lambda_V(t)$ is the force of infection at time t, given as:

$$\lambda_V(t) = \alpha\left(\beta_{Vr}\sum_{i=1}^{I} p_{Ri}(t)\frac{I_{Ri}(t)}{N_{Ri}(t)} + \beta_{Vh}p_H\frac{I_{1H}(t)}{N_H(t)}\right) \tag{5.14}$$

where α is the daily probability a fly takes a bloodmeal, β_{Vr} is the probability of transmission from a reservoir animal to a tsetse fly during a bloodmeal, $p_{Ri}(t)$ is the probability a given bloodmeal is taken from species i at time t, $\frac{I_{Ri}(t)}{N_{R_i}(t)}$ is the probability a given individual of reservoir species i is infectious at time t, β_{Vr} is the probability of transmission from a human to a tsetse fly during a bloodmeal, p_H is the probability a given bloodmeal is taken from a

human, and $\frac{I_{1H}(t)}{N_H(t)}$ is the probability a given human is in the stage 1 infectious period, as it is assumed those with stage 2 disease are hospitalized or too debilitated to serve as a bloodmeal source.

The rate of change in the exposed compartment is given as:

$$\frac{dE_V}{dt} = \lambda_V(t)S_V(t) - \sigma_v E_V(t) - \mu_V E_V(t) \tag{5.15}$$

where $1 \,/\, \sigma_v$ is the duration of the infectious period. The rate of change in the infectious compartment is given as:

$$\frac{dI_V}{dt} = \sigma_V E_V(t) - \mu_V I_V(t) \tag{5.16}$$

imposing the assumption that tsetse flies do not recover once infected. Finally, the rate of change in the non-susceptible compartment is given as:

$$\frac{dNS_{Vt}}{dt} = \Big(1 - \lambda_V(t)\Big)S_V(t) - \mu_V NS_V(t) \tag{5.17}$$

reflecting the teneral effect.

5.2.4 Interventions

ITC While ITC refers specifically to insecticide treatment of cattle, here we will model insecticide treatment of both cattle and pigs without modification to this term. As in Hargrove et al. (2012) [8], we assume ITC exerts its effect by decreasing the probability a fly survives a given feed, q_f. If q_n is the probability a fly survives a non-feeding day, and d is the feeding cycle length, then the probability a fly survives a complete feeding cycle is ${q_f q_n}^d$. As daily mortality rate is $\approx -log({q_f q_n}^d) \,/\, d$, our parameterization of tsetse fly lifespan (26 days for *G. m. morsitans* and *G. f. fuscipes*, and 29 days for *G. f. quanzensis*) and feeding cycle length (3 days) yields ${q_f q_n}^d = 0.89$ for *G. f. fuscipes* and *G. m. morsitans*, and ${q_f q_n}^d = 0.91$ for *G. f. quanzensis*.

If itc_{Ri} is the proportion of reservoir species i treated with insecticides, $p_{Ri}(t)$ is the probability a given bloodmeal is taken from reservoir species i at time t, and flies feed on

individual members of a given reservoir species at random, assuming efficacy of treatment is 100% the probability a fly survives a complete feeding cycle of d days is now:

$$\left[1 - \sum_{i=1}^{I} \left(itc_{Ri} p_{Ri}(t) \right) \right] q_f q_n^d$$

TT Trypanocide treatment of domestic animal reservoirs was implemented through allowing a proportion of animals to receive trypanocidal treatment at three month intervals, thereby shortening their duration of infection. This is distinct from Hargrove's implementation of TT, which assumed continuous prophylactic use and therefore removal of a proportion of reservoirs from the reservoir population [8].

Scenarios modeled We evaluated the probability of and time to elimination under the following scenarios:

- gHAT foci:
  - ITC of pigs at 12.5%, 25% and 50% coverage
  - ITC and TT of pigs at 25% coverage of each
- rHAT foci:
  - ITC of cattle and pigs at 25%, 50%, and 75% coverage
  - ITC of cattle and pigs at 50% combined with TT of cattle and pigs at 75% coverage
  - ITC and TT of cattle and pigs at 100% coverage of each

Our base-case model assumed 0% TT or ITC coverage in gHAT foci, and 50% TT coverage (both cattle and pigs) in rHAT foci. Note "cattle and pigs" refers to domestic bovids and domestic swine, respectively.

5.2.5 Calibration

We calibrated our models to 2000-2005 annual surveillance data from the WHO Atlas of HAT [161] using Optuna [162]. We collapsed observed data by country, keeping rHAT and gHAT foci separate in Uganda, and assumed 65% of gHAT cases are reported [163] and 8.3% of rHAT cases are reported [103].

Fitting 10 trials with 3 model runs per trial and optimizing the mean over these runs, we defined the sum of the squared error terms as the objective function to be minimized. Under normally-distributed and errors with constant variance, this simple objective function yields the same estimate as a maximum likelihood approach [164]. We used a tau-leaping solver and specified trial parameters as $\beta_H \sim \text{Uniform}(0.001, 0.1)$ for gHAT foci in Uganda and South Sudan, where β_H is the probability an infected fly transmits to a human during a given bloodmeal; $\beta_H \sim \text{Uniform}(0.001, 0.01)$ for both gHAT foci in DRC, reflecting the markedly higher probability a bloodmeal is taken from a human (and thus force of infection) in these foci; and $\beta_H \sim \text{Uniform}(0.0001, 0.001)$ in rHAT foci.

5.2.6 Validation

We validated our model by comparing observed Atlas data to the predicted number of new cases per year in our base case model, adjusted for underreporting as detailed above. This was performed from 2005-2014 in Malawi and South Sudan, and from 2005-2018 in DRC and Uganda, reflecting the data made available to the authors.

5.2.7 Data and code availability

HAT outcome data can be requested from the WHO (`https://www.who.int/trypanosomiasis_african/country/foci_AFRO/en/`) and livestock density maps can be downloaded from `https://github.com/JulianneMeisnerUW/LivestockMaps`. Parameter data can be obtained from the references detailed in Appendix D.

CMS can be downloaded from `https://docs.idmod.org/projects/cms/en/latest/`

index.html#, and model code written in Python is available in the GitHub repository linked above.

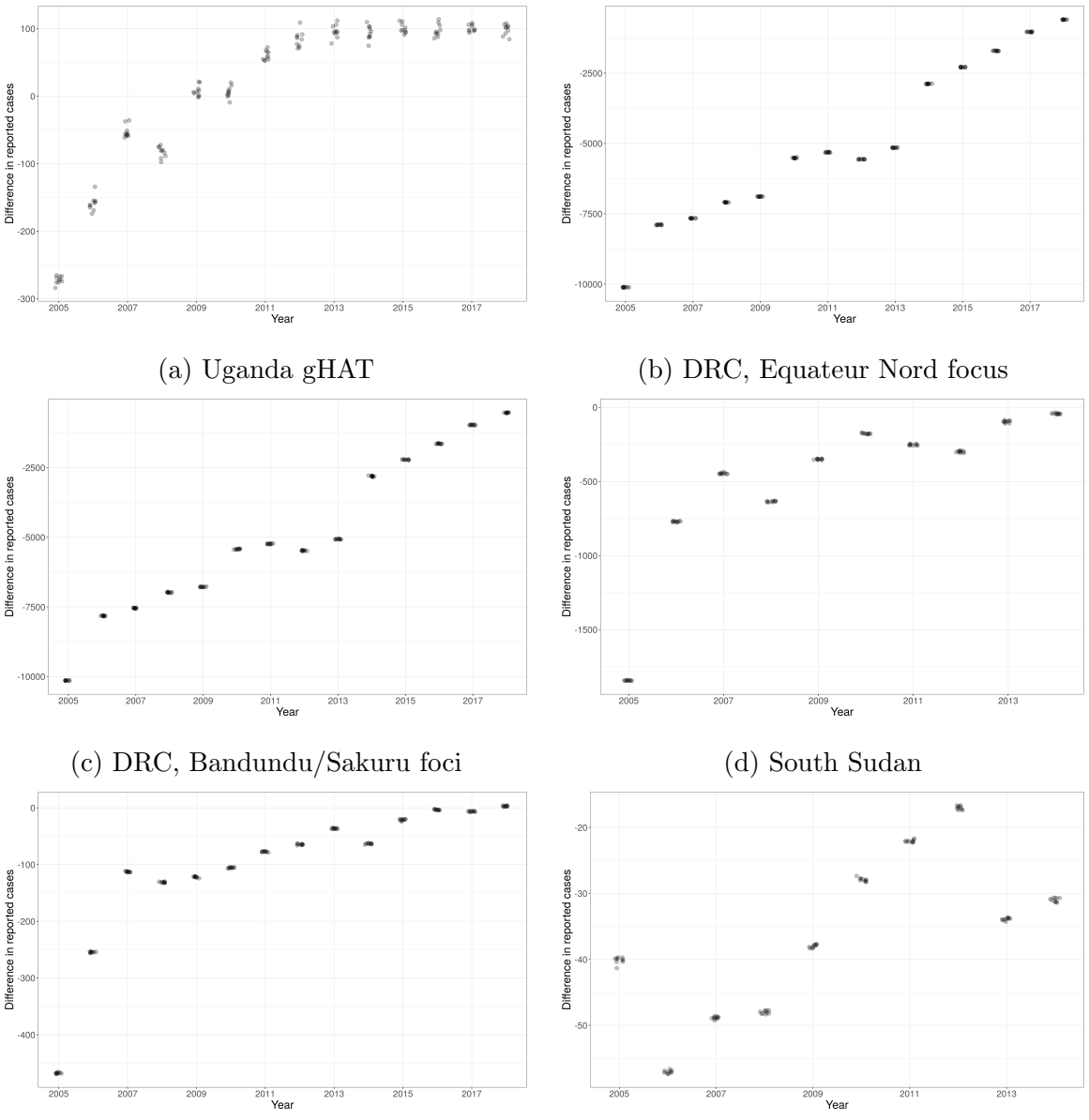
5.3 Results

Model trajectories for tsetse flies and humans are presented in Appendix E. While epidemiologic curves for humans vary in level across modeled scenarios, shape is relatively stable. In the rHAT models human infection is maintained at a remarkably low but steady level, consistent with epidemiologic evidence of HAT's ability to persist at low levels.

In contrast, the curves for tsetse flies vary markedly across modeled scenarios in the gHAT models, increasing (Uganda gHAT, DRC Bandundu/Sakuru) or reaching a steady-state (DRC Equateur Nord, South Sudan) in the base-case, versus decaying rapidly in all test scenarios. In the rHAT models, exposed and infected tsetse flip in the base-case versus test scenarios, with infected tsetse outnumbering exposed tsetse in the base-case, and the inverse being true in all test scenarios. This is consistent with ITC lowering the age distribution of tsetse flies.

5.3.1 Validation

Validation results are presented in Figure 5.2. On an absolute scale, the models appeared to perform better in later than earlier years, and performed well for Uganda in both gHAT and rHAT foci and for South Sudan and Malawi, but not very well in either DRC foci. However, the number of cases reported (observed) in each focus varies markedly (e.g., 608 total cases in Malawi and 16,756 total cases South Sudan over the period 2000-2014; and 4,219 total cases in Uganda gHAT foci, 3,281 total cases in Uganda rHAT foci, 68,414 total cases in the Bandundu/Sakuru foci in DRC, and 19,327 total cases in the Equateur Nord focus in DRC over the period 2000-2018), thus a difference between observed and predicted cases of 100 implies markedly different model performance in, for instance, Malawi versus either DRC foci.



(a) Uganda gHAT

(b) DRC, Equateur Nord focus

(c) DRC, Bandundu/Sakuru foci

(d) South Sudan

(e) Uganda rHAT

(f) Malawi

Figure 5.2: Observed minus predicted cases by year over 10 model runs in the base-case scenario

5.3.2 Elimination of transmission

EOT outcomes under all models and scenarios are presented in Tables 5.1-5.2; predicted number of number of cases in 2030 for the base-case scenario in each modeled focus is also presented in Figure 5.3, and for all model scenarios in the rHAT foci (Uganda rHAT, Malawi) in Figures 5.4-5.5. Corresponding results are not presented for the gHAT models (Uganda gHAT, South Sudan, and both DRC models) as EOT was observed by 2030 all gHAT foci for all test scenarios.

Under current conditions, reflected by the base-case scenario, EOT is not observed by 2045 in any modeled focus, with cases being particularly high in Uganda and DRC across the two epidemics modeled in each country. For the gHAT models (Uganda gHAT, both DRC models, and South Sudan), EOT was observed by 2030 in all test scenarios, with differences in time to elimination across scenarios occurring on the order of months, not years (Table 5.1, Figures 5.6-5.9). Time to elimination is shortest for South Sudan, followed by Uganda and the Equateur Nord focus in DRC, and lastly by the Bandundu/Sakuru foci in DRC.

For both rHAT models (Uganda rHAT and Malawi), EOT is not observed in any test scenario. Case counts reduce monotonically with increasing ITC coverage, however increasing TT coverage from 50% to 75% has inferior gains to the same increase in ITC coverage (Table 5.2, Figures 5.4-5.5).

5.4 Discussion

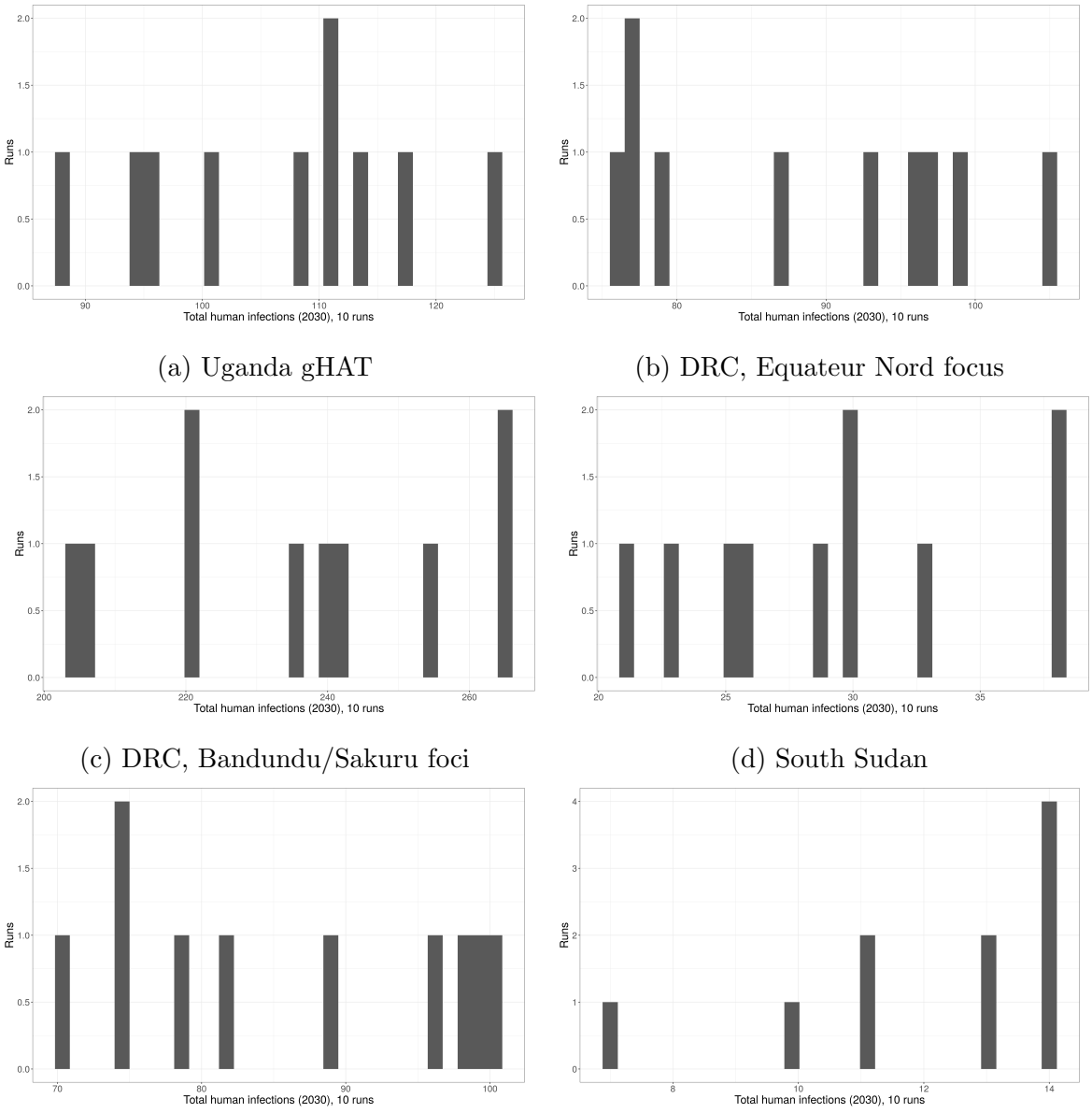
We present the results of a stochastic compartmental model fit to six distinct HAT epidemics. Our findings indicate WHO EOT goals will not be met by 2030 under current conditions, however in gHAT foci in Uganda, South Sudan, and DRC, EOT would have already been achieved if insecticide treatment of domestic cattle and pigs had been implemented at even 12.5% coverage starting in 2005. In rHAT foci in Uganda and Malawi, elimination of transmission is not achieved by the end of our modeled time horizon (2045) even if 100% of domestic cattle and pigs are constantly maintained on insecticide treatment with 100% effi-

Model	EOT Year, mean (sd)	Cases, 2030, mean (sd)
Uganda, gHAT		
Base-case	NA	106.50 (11.54)
12.5% ITC	2010 (1.89)	0 (0)
25% ITC	2010 (1.32)	0 (0)
25% TT, 25% ITC	2010 (1.89)	0 (0)
50% ITC	2009 (1.35)	0 (0)
DRC, Equateur Nord focus		
Base-case	NA	88.60 (10.77)
12.5% ITC	2010 (1.77)	0 (0)
25% ITC	2010 (2.02)	0 (0)
50% ITC	2010 (1.71)	0 (0)
25% TT, 25% ITC	2010 (1.99)	0 (0)
DRC, Bandundu/Sakuru foci		
Base-case	NA	235 (21.93)
12.5% ITC	2013 (1.23)	0 (0)
25% ITC	2013 (1.84)	0 (0)
50% ITC	2013 (1.17)	0 (0)
25% TT, 25% ITC	2013 (1.60)	0 (0)
South Sudan		
Base-case	NA	29.30 (5.81)
12.5% ITC	2009 (1.57)	0 (0)
25% ITC	2009 (1.83)	0 (0)
50% ITC	2009 (1.32)	0 (0)
25% TT, 25% ITC	2008 (1.52)	0 (0)

Table 5.1: Results for gHAT models across 10 runs per model. EOT: eliminaton of transmission; ITC: insecticide treatment of pigs; TT: trypanocide treatment of pigs; NA: eliminaton not observed. TT coverage is assumed to be 0% unless otherwise specified

Model	EOT Year, mean (sd)	Cases, 2030, mean (sd)
Uganda, rHAT		
Base-case	NA	86.20 (11.56)
25% ITC	NA	34.30 (5.96)
50% ITC	NA	19.10 (4.79)
75% ITC	NA	9.80 (1.69)
75% TT, 50% ITC	NA	16.60 (3.44)
100% TT, 100% ITC	NA	4.90 (2.33)
Malawi		
Base-case	NA	12.10 (2.33)
25% ITC	NA	3.50 (0.85)
50% ITC	NA	2.90 (1.66)
75% ITC	NA	1.80 (1.32)
75% TT, 50% ITC	NA	2.30 (1.42)
100% TT, 100% ITC	NA	0.80 (1.03)

Table 5.2: Results for rHAT models across 10 runs per model. EOT: eliminaton of transmission; ITC: insecticide treatment of cattle and pigs; TT: trypanocide treatment of cattle and pigs; NA: elimination not observed. TT coverage is assumed to be 50% unless otherwise specified



(a) Uganda gHAT

(b) DRC, Equateur Nord focus

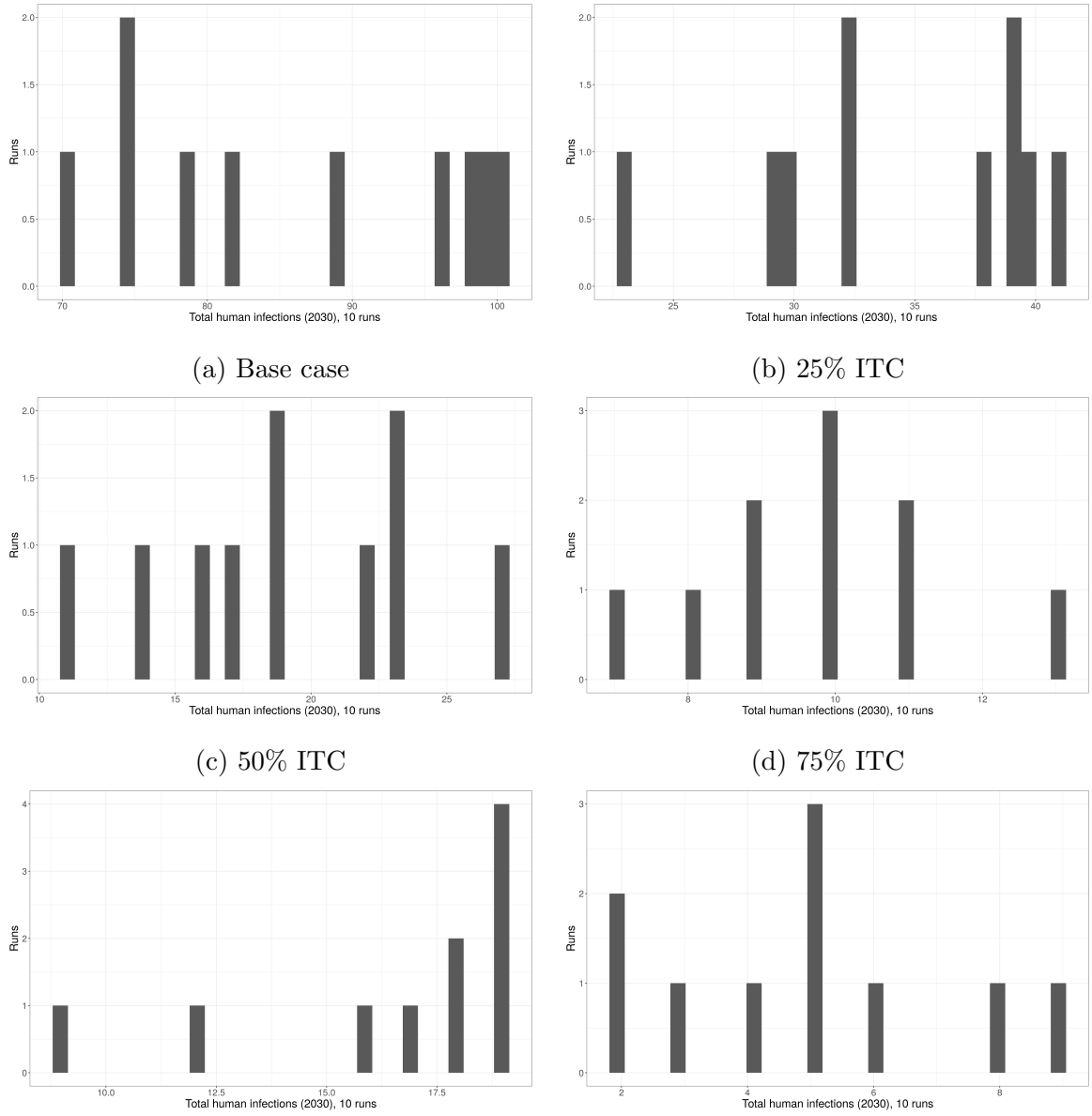
(c) DRC, Bandundu/Sakuru foci

(d) South Sudan

(e) Uganda rHAT

(f) Malawi

Figure 5.3: Predicted HAT cases in 2030 over 10 model runs in the base-case scenario



(a) Base case

(b) 25% ITC

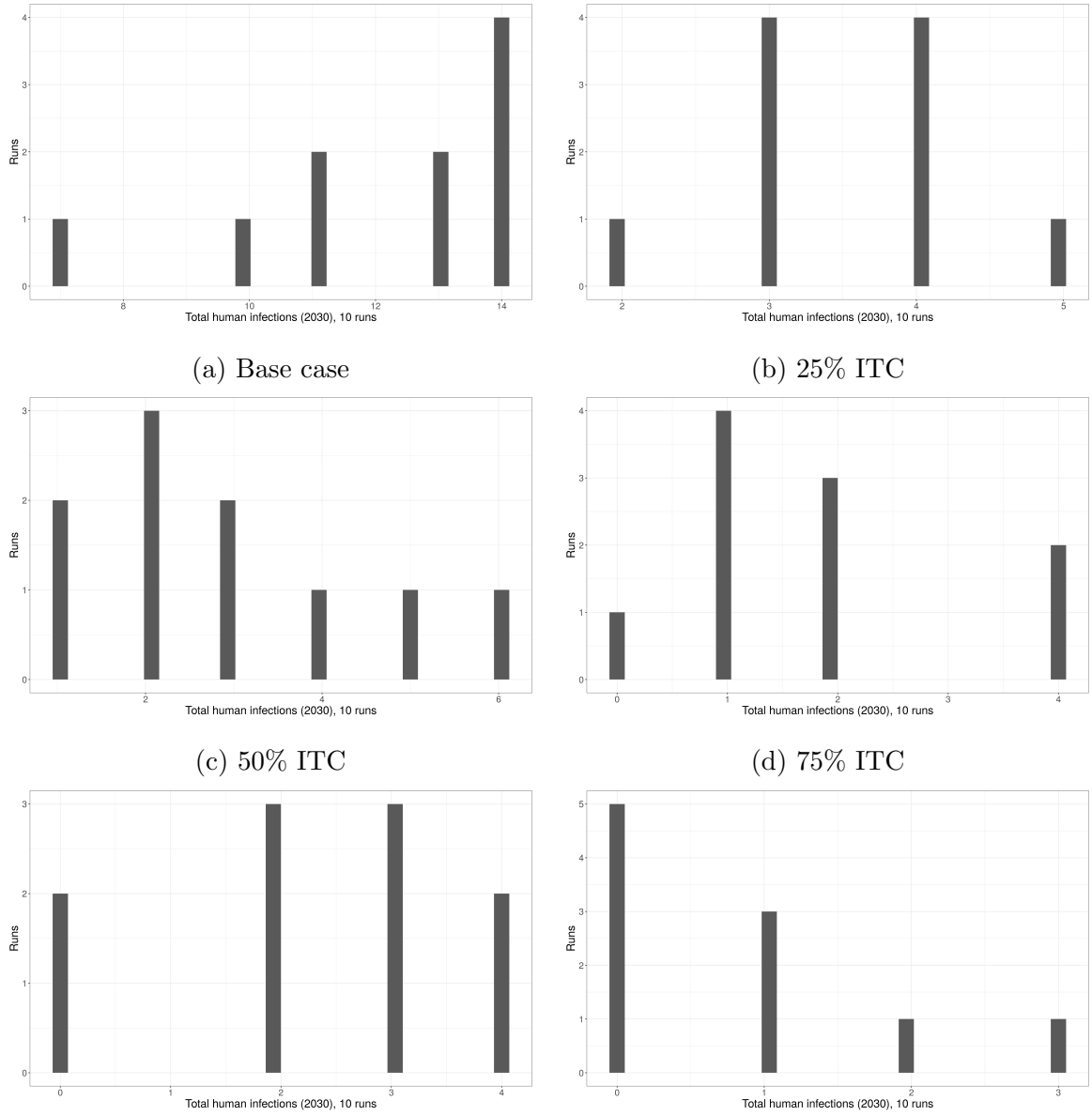
(c) 50% ITC

(d) 75% ITC

(e) 75% TT, 50% ITC

(f) 100% TT, 100% ITC

Figure 5.4: Predicted cases in 2030, Uganda rHAT, over 10 runs. ITC: insecticide treatment of cattle and pigs; TT: trypanocide treatment of cattle and pigs. TT coverage assumed to be 50% unless otherwise specified



(a) Base case

(b) 25% ITC

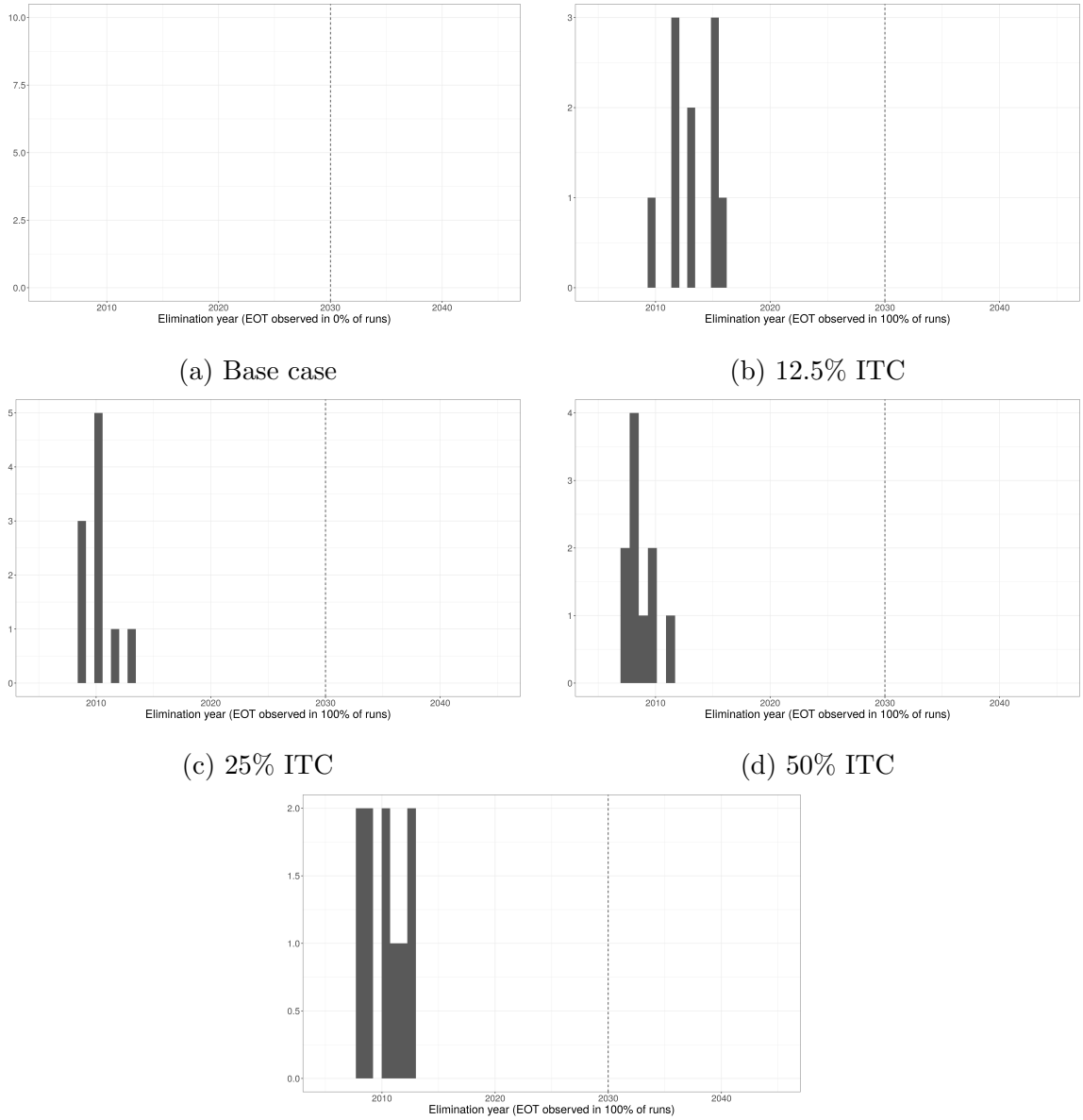
(c) 50% ITC

(d) 75% ITC

(e) 75% TT, 50% ITC

(f) 100% TT, 100% ITC

Figure 5.5: Predicted cases in 2030, Malawi, over 10 runs. ITC: insecticide treatment of cattle and pigs; TT: trypanocide treatment of cattle and pigs. TT coverage assumed to be 50% unless otherwise specified.



(a) Base case

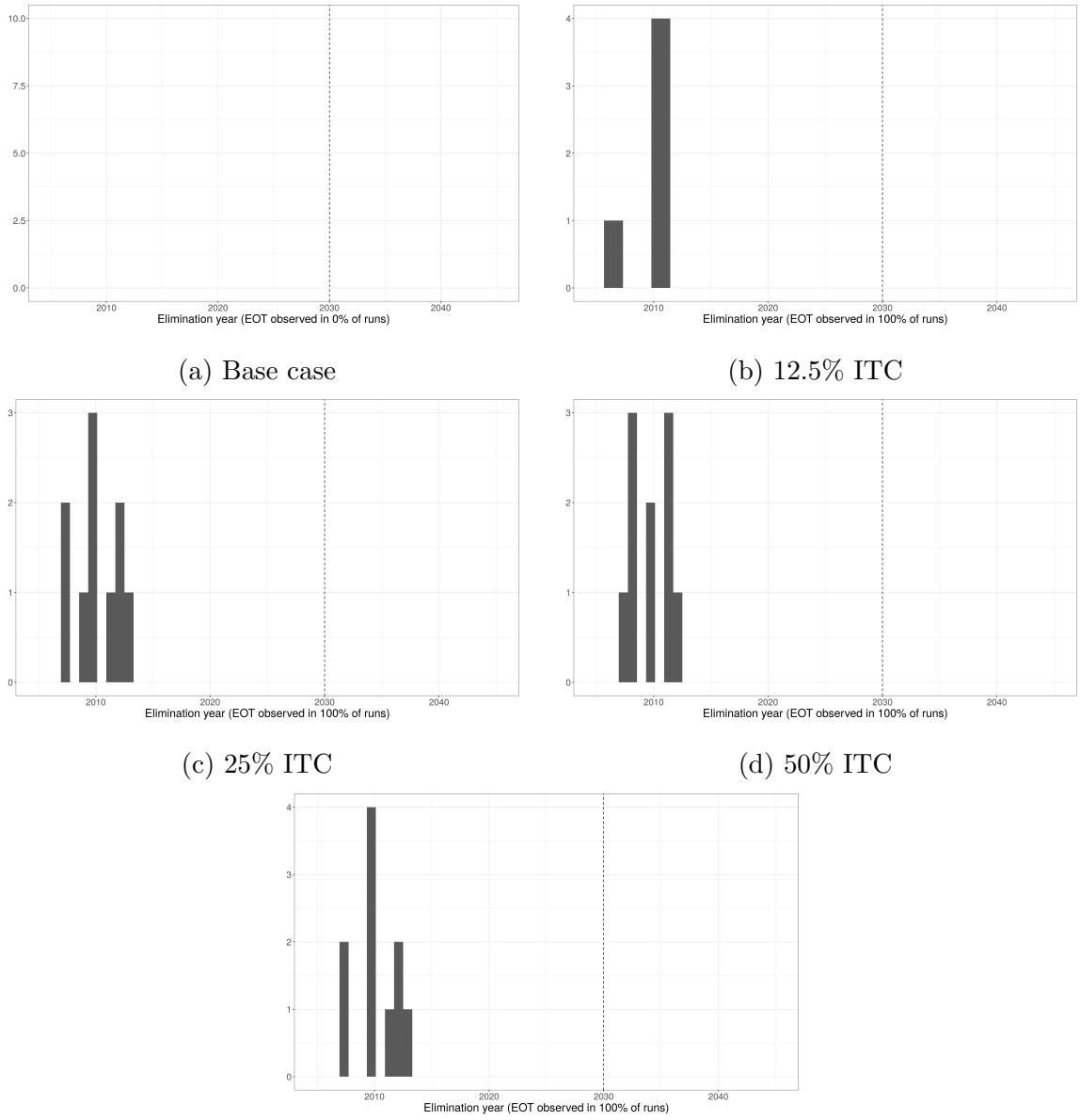
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure 5.6: Predicted year of gHAT elimination, Uganda, over 10 runs. ITC: insecticide treatment of pigs; TT: trypanocide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified. Dashed vertical line marks 2030 (WHO target year for gHAT EOT)



(a) Base case

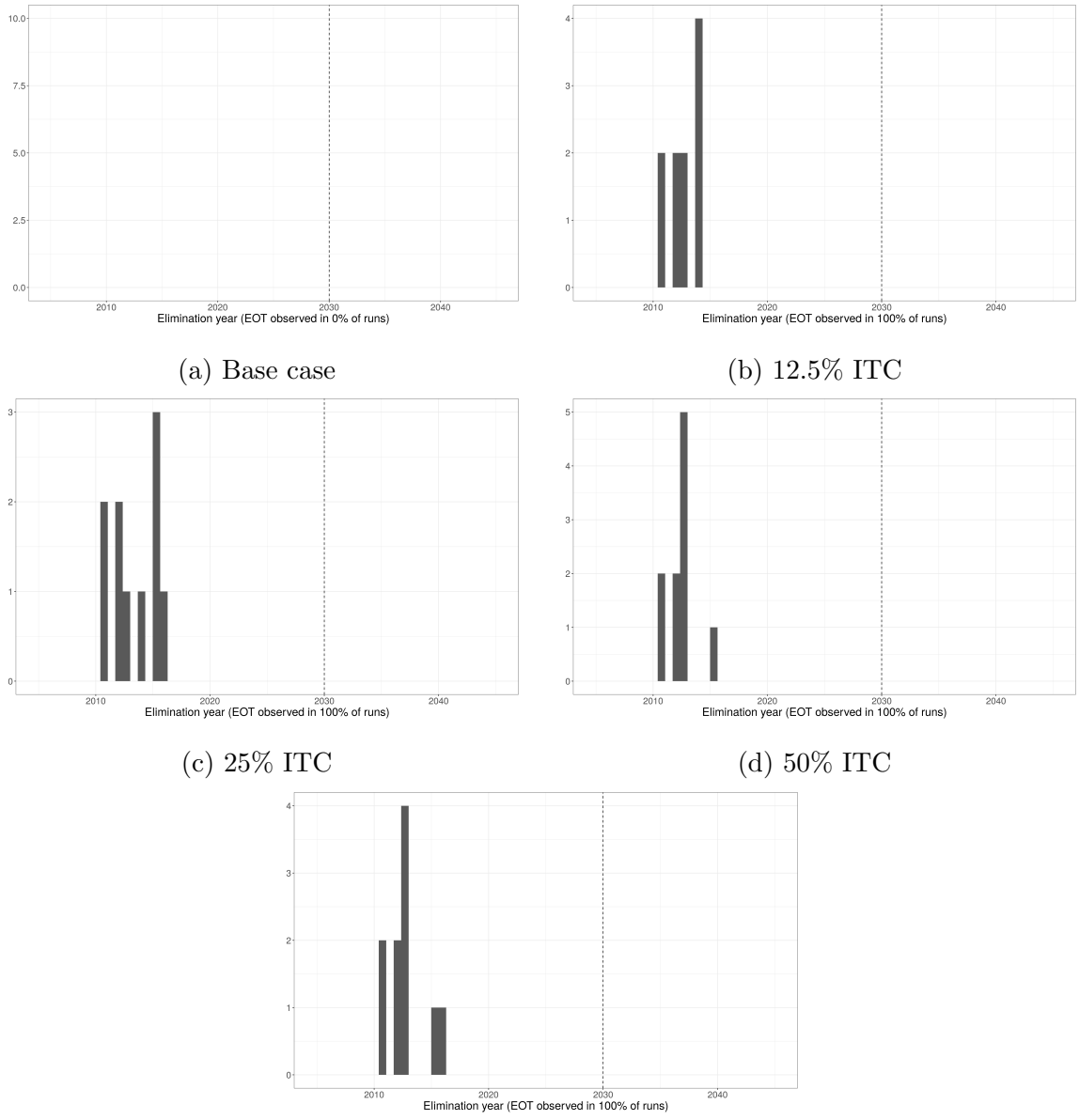
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure 5.7: Predicted year of gHAT elimination, DRC Equateur Nord focus, over 10 runs. ITC: insecticide treatment of pigs; TT: trypanocide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified. Dashed vertical line marks 2030 (WHO target year for gHAT EOT)



(a) Base case

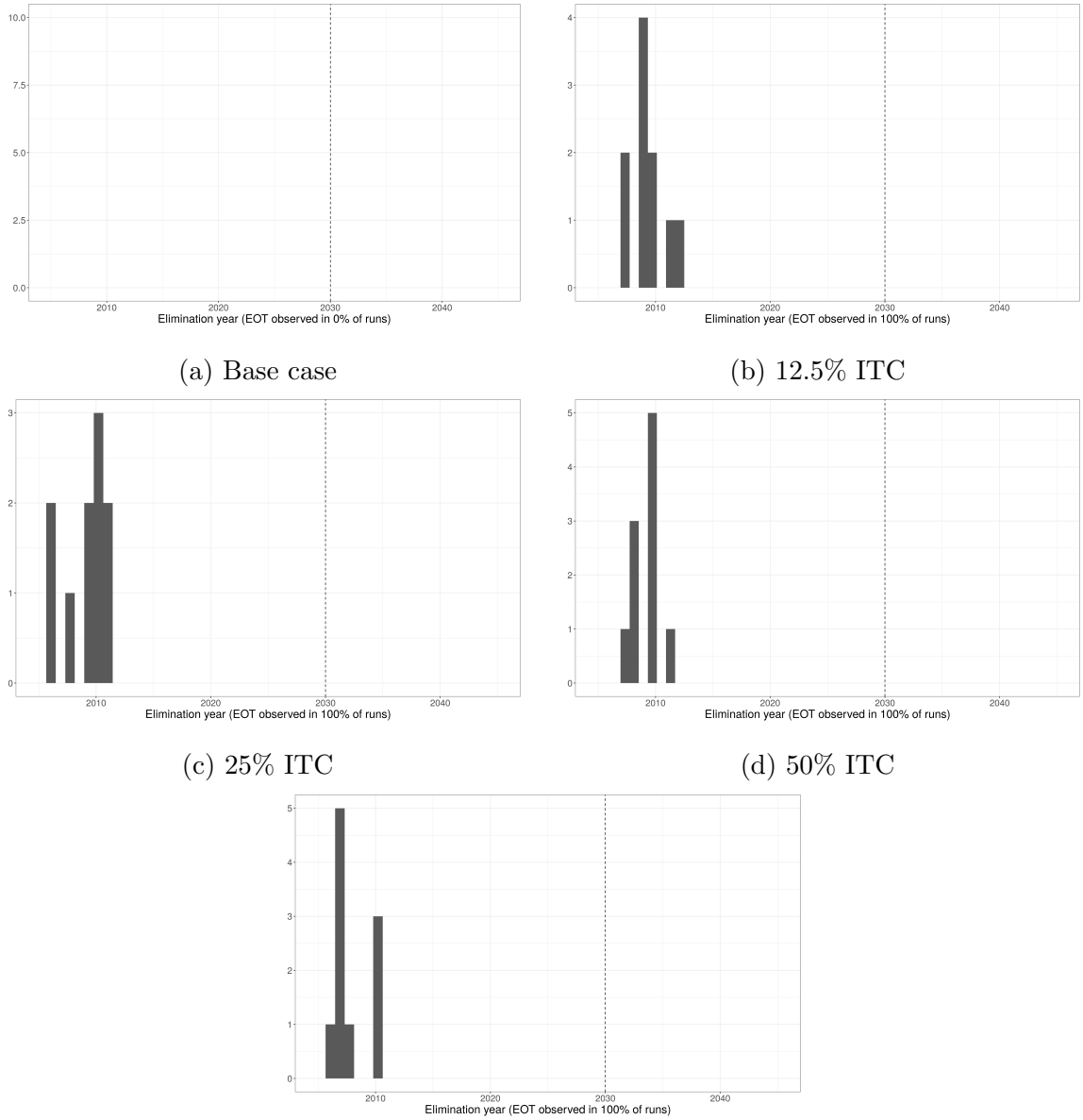
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure 5.8: Predicted year of gHAT elimination, DRC Bandundu/Sakuru foci, over 10 runs. ITC: insecticide treatment of pigs; TT: trypanocide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified. Dashed vertical line marks 2030 (WHO target year for gHAT EOT)



(a) Base case

(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure 5.9: Predicted year of gHAT elimination, South Sudan, over 10 runs. ITC: insecticide treatment of pigs; TT: trypanocide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified. Dashed vertical line marks 2030 (WHO target year for gHAT EOT)

cacy, and receive 100%-effective trypanocide treatment every three months. In both rHAT and gHAT foci, trypanocide treatment of domestic animal reservoirs hastens time to elimination (gHAT foci only) and reduces cases observed by 2030 (rHAT foci), but with inferior gains to increasing coverage with insecticide treatment of domestic animal resrevoirs.

Our model has several limitations. First, validation indicates our model predictions are a poor fit for both DRC models, suggesting results for this country should be interpreted in relative, rather than absolute, terms. Second, our results are sensitive to the assumptions made, namely detection perfectly predicts recovery in humans, TT and ITC are 100% effective, all populations are closed, and no in-migration of infected humans or tsetse occur (i.e., each modeled focus is sufficiently isolated from other HAT foci). Previous authors have demonstrated the closed population and no in-migration assumptions have negligible impact on results in other stochastic compartmental HAT models [155]. Furthermore, while most parameters were derived from the literature or from model fitting, density of non-reservoir tsetse hosts was assumed, as was the probability domestic animal reservoirs receive trypanocide or insecticide treatment and frequency of the former. We also assumed tsetse flies have an inherent (species-specific) level of anthropophily, and all "remaining" bloodmeals (i.e., those taken from animals) are distributed according to density of each animal species or species group. The results of our test scenarios may also be sensitive to the way we implemented ITC and TT. Our implementation of TT assumes therapeutic, rather than prophylactic, use, in contrast with Hargrove et al.'s implementation of TT [8]. With regards to ITC, our implementation assumes the tsetse population is constant, such that increasing ITC coverage reduces the mean tsetse age—and thus the proportion infected with *T. b. gambiense* or *T. b. rhodesiense*—rather than density, as in Hargrove et al. [8]. Finally, we did not model vector control in the base-case scenario nor use of stationary baits in the test scenarios, nor any longitudinal change in active or passive surveillance coverage.

Our results indicate insecticide treatment of cattle and pigs should be added to control strategies in both gHAT and rHAT foci, however delivery of insecticides to pigs is slightly more challenging than cattle. First, deltamethrin, which is widely-available in many HAT-

endemic areas and has a long duration of action, is not labeled for pigs, however it has been successfully used off-label to control biting flies and mosquitoes in Australia [165]. Second, while permethrin may be a suitable alternative, the efficacy of restricted application approaches have not been evaluated in pigs, and efficacy of alternative modes of delivery is uncertain. For instance, insecticide-impregnated ear tags are unlikely to work as well in pigs as in cattle as, unlike in cattle, pigs do not frequently move their ears or bend their neck to reach their flanks, limiting the protected area to the face and possibly neck. Possible delivery options include bi-weekly sprays, use of back rubbers (a rubbing surface treated with insecticides), or impregnated fabrics placed near pigs but out of their reach. Furthermore, the effectiveness of using pigs as live baits has not been established, however their proclivity for roaming in shady areas along the riverine habitats favored by HAT vectors [58] is a favorable indication in this regard.

Despite the limitations of our model, and potential challenges to implementation of the results it points to, our approach nonetheless represents an important contribution to the HAT modeling literature. This is, to our knowledge, the first effort use a stochastic compartmental model to study the utility of a One Health approach to HAT control across foci representing distinct epidemiologic, entomologic, and environmental conditions. By harmonizing model structure and assumptions, our study increases the comparability of results across these foci. Our results confirm the widely-held belief that elimination of rHAT transmission will not occur even at complete coverage of domestic animal reservoirs with trypanocides and insecticides, and indicate that if pigs are a reservoir of gHAT, EOT goals can only be obtained if insecticide or trypanocide treatment of this reservoir host is added to HAT control strategies, with insecticide treatment being superior. In addition to increasing the speed and probability of gHAT elimination, and contributing to control of rHAT, coordinated top-down implementation of joint HAT-AAT control strategies hold opportunity to retain donor engagement in HAT as gHAT elimination nears—preventing gHAT re-emergence due to animal reservoirs or latent human infections, and rHAT emergence as a major public health problem—and to contribute to the control of an important and poverty-reinforcing disease.

Chapter 6
CONCLUDING REMARKS

In this dissertation, we use spatial epidemiologic methods, modeling approaches rooted in the potential outcomes framework for causal inference, and a stochastic compartmental model to study the role of livestock in the epidemiology and control of human African trypanosomiasis. Our results indicate that pig are likely a gHAT reservoir in DRC and Uganda, and that treatment of this reservoir with trypanocides or, preferably, insecticides, will hasten time to elimination of transmission. In rHAT foci, our results suggest the cattle and pig reservoir are an important source of human infection in Uganda but not Malawi, and elimination of transmission will not be achieved in either country even at complete coverage of domestic cattle and pigs with trypanocides and insecticides. Finally, our results indicate livestock play a zooprophylactic role in some settings—Malawi and South Sudan for both pigs and cattle, and in gHAT foci in Uganda for cattle—however environmental changes exerted by livestock and approximated by NDVI and LST do not sufficiently modify tsetse distribution to have a strong effect on HAT epidemiology.

6.1 Livestock mapping

In Chapter 2 we present a new contribution to the livestock mapping literature, a time series of high-resolution livestock (cattle and pig) density maps for Malawi, Uganda, and DRC, and a single county-level map for South Sudan. Complimenting the Gridded Livestock of the World 3 maps, our approach uses a definition of density—the ratio of livestock to humans—that is particularly well-suited to epidemiologic research and development applications. Our product is furthermore available as a time series in three out of four countries, in contrast with GLW-3 results. Finally, by using modeling approaches and data sources distinct from

those implemented by GLW-3, good agreement between the two sets of results indicates the validity of both.

6.2 Total effect of livestock on HAT risk

In Chapter 3 we use the parametric g-formula—employing a spatial extension of the partial interference assumption—in Uganda, Malawi, and DRC, and spatial regression models accounting for measurement error in South Sudan, to estimate the total effect of cattle and pig density, separately, on rHAT and gHAT risk. We found evidence that the zooprophylactic effect outweighs any reservoir effect in rHAT foci in Malawi and gHAT foci in South Sudan (both cattle and pigs), and in gHAT foci in Uganda (cattle alone). We found the inverse—reservoir effect outweighs any zooprophylactic effect—in rHAT foci in Uganda (both cattle and pigs), in gHAT foci in Uganda (pigs alone), and in gHAT foci in DRC (both cattle and pigs).

Our calculated E-values indicate unmeasured confounding is an unlikely explanation of most of our results however underreporting in the outcome data, overfitting in Malawi, and ecological bias in South Sudan are remaining concerns.

6.3 Mediation of this total effect by environmental change

In Chapter 4, we use the mediational g-formula in Malawi, Uganda, and DRC, and a spatial regression approach—again accounting for measurement error in livestock density and wealth, variables we modeled—to decompose the total effect estimated in Chapter 3 into that mediated by environmental variables, and a remaining direct effect which we hypothesize arises from either a reservoir (positive) or zooprophylactic (negative) effect. Our findings support the conclusions drawn from our total effect estimates overall, with the exception of a positive pig-gHAT direct effect in South Sudan indicating a reservoir role, however with high uncertainty in this estimate. We did not detect any appreciable effect of livestock density on HAT mediated by NDVI, a remote-sensed indicator of vegetation coverage, or LST, a remote-sensed indicator of surface temperature, with the exception of South Sudan where

this effect is negative. Given our concerns about ecological bias in South Sudan, and the absence of similar findings in the other study countries, we recommend caution in interpreting these results.

6.4 Investigation of a One Health approach to HAT control using a stochastic compartmental model

Using the Compartmental Modeling Software developed by the Institute for Disease Modeling, we built and parameterized models to six distinct HAT epidemics: the national gHAT epidemics in South Sudan and Uganda, the national rHAT epidemics in Uganda and Malawi, and in, DRC, the Bandundu and Sakuru gHAT foci (together) and the Equateur Nord gHAT focus (separately). We implemented a four-species model in gHAT foci, containing humans, tsetse flies, reservoir animals (domestic pigs), and non-reservoir tsetse hosts (i.e., animals which serve as bloodmeals for humans but do not harbor human-infective trypanosomes). In rHAT foci we implemented a seven-species model through the addition of 3 more reservoir hosts: domestic cattle, wild swine (warthog), and wild bovids (bushbuck, African buffalo).

In all models we found insecticide treatment of domestic animal reservoirs to be superior to trypanocide treatment of the same species, with even 12.5% coverage of domestic pigs with insecticides achieving EOT by 2030 in gHAT foci in all model runs, compared with no EOT observed under current control strategies. In rHAT foci EOT was not observed under any modeled scenario, however the epidemiologic situation was markedly improved by increasing levels of insecticide coverage.

6.5 Limitations

Our findings are sensitive to limitations in our data and assumptions made in our modeling approaches. Namely, error may be introduced into our livestock maps through errors in the source data or in geolocation, including that derived from jittering implemented to protect participant privacy. While we performed leave-one-out cross-validation to select our final models and compared our product against GLW-3 estimates and national density estimates,

it is not practical nor feasible to ground-truth a longitudinal product. Our livestock mapping approach is further limited by challenges to appropriately parameterizing urbanicity. Finally, data availability limitations precluded inclusion of a space-time interaction term in the DRC models—enforcing the assumption that temporal trends are constant in space, and spatial trends are constant in time—and limited our South Sudan product to a single county-level map for 2008 alone.

In both our total effect and mediation analyses we were unable to adjust for vector control, resulting in residual confounding by this variable. This is expected to bias effect estimates towards zero and thus could explain the negative effects detected (i.e., those we attributed to zooprophylaxis), but would attenuate, rather than strengthen, the positive effects detected (i.e., those we attributed to the reservoir effect). Residual confounding may also arise from incomplete control of wealth through measurement error in this modeled variable—which we addressed in our models for South Sudan alone—and challenges to parameterizing this latent construct, and to the need to drop several confounders from the Malawi models due to overfitting concerns. Measurement error in livestock density may also result in bias, as well as underreporting in the outcome data if this underreporting is associated with cluster- or county-level HAT risk or livestock density and not completely addressed through our control for political stability and wealth. Finally, ecological bias may compromise the validity of our findings in South Sudan, where the unit of analysis was administratively-defined counties, which are large and irregular and reflect a grouping which has no relationship to HAT epidemiology.

Validation of our stochastic compartmental model was favorable for Uganda, South Sudan, and Malawi; while performance was less favorable for the DRC foci on the absolute scale, case reports were far higher for the DRC foci than the Malawi and Uganda foci. Regardless, results for the DRC foci should perhaps be considered in relative, rather than absolute, terms. As for all models, our results are sensitive to the parameters chosen and the assumptions made, however several assumptions (closed populations, no in-migration of infected tsetse and humans) have been demonstrated to have little bearing on model results

in other stochastic compartmental HAT models, and nearly all parameters were derived from published literature.

6.6 Impact

The primary goals of this dissertation were to (a) narrow knowledge gaps surrounding the role of pigs as a gHAT reservoir, (b) examine the feasibility of rHAT elimination with control of domestic animal reservoirs alone, and (c) evaluate whether the environmental impacts of livestock keeping modify vector distribution sufficiently to have impact on the epidemiology of diseases transmitted by those vectors.

While our total effect results initially appear discordant with what is known about HAT epidemiology—cattle and pigs are important reservoirs of rHAT, in contrast with the protective effect detected in Malawi, and the gHAT epidemics in Uganda and South Sudan are essentially one transboundary focus and thus should have equivalent epidemiologic features—deeper examination of this apparent discordance underscores the very utility of our approach. That is, in contrast with tightly-controlled mechanistic studies or mathematical models, our results reflect the real-world epidemiologic context of HAT, in which transmission dynamics are modified by local factors such as tsetse host abundance and preference, contact patterns between hosts, implementation and efficacy of HAT interventions, and the presence of distinct sociocultural forces. Regardless, we cannot rule out bias as the driver of these findings.

In addition to our livestock maps, an important addition to the only existing large-scale livestock mapping effort (i.e., GLW-3), our results indicate that pigs indeed may be a gHAT reservoir, that the environmental impact of livestock approximated through NDVI and LST is not sufficient to modify tsetse distribution and thus HAT risk, and that while rHAT elimination is unlikely with control of the domestic animal reservoir alone, a One Health approach to HAT control holds promise for improving the epidemiologic situation of rHAT, in particular in Uganda. In concert with our findings that a One Health approach also reduces time to gHAT elimination, and agreement in the HAT community that such an

approach alleviates concerns surrounding donor disengagement near gHAT elimination and emergence of rHAT as a public health problem following gHAT elimination, this dissertation represents an important contribution to the evidence that, indeed, coordination of AAT and HAT control should be pursued.

BIBLIOGRAPHY

[1] J. R. Franco, G. Cecchi, G. Priotto, M. Paone, A. Diarra, L. Grout, R. C. Mattioli, and D. Argaw. Monitoring the elimination of human African trypanosomiasis: Update to 2014. *PLoS Negl Trop Dis*, 11(5):e0005585, 2017.

[2] Open Peer Review and Reviewer Status. Insights from quantitative and mathematical modelling on the proposed 2030 goal for gambiense human African trypanosomiasis (gHAT) [version 2; peer review: 2 approved] NTD Modelling Consortium Discussion Group on Gambiense Human African Gates Open R. pages 1–16, 2020.

[3] G. Cecchi, M. Paone, J.R. Franco, E.M. Fevre, A. Diarra, J.A. Ruiz, R.C. Mattioli, and P.P. Simarro. Towards the Atlas of human African trypanosomiasis. *International Journal of Health Geographics*, 8(15), 2009.

[4] World Health Organization. Mapping the distribution of human African trypanosomiasis: Uganda, 2000-2014. Accessed 10 Apr 2020.

[5] J.R. Franco, G. Cecchi, G. Priotto, M. Paone, A. Diarra, L. Grout, P.P. Simarro, W. Zhao, and D. Argaw. Monitoring the elimination of human African trypanosomiasis at continental and country level: Update to 2018. *PLoS Neglected Tropical Diseases*, 14(5):1–18, 2020.

[6] A. Pandey, K. E. Atkins, B. Bucheton, M. Camara, S. Aksoy, A. P. Galvani, and M. L. Ndeffo-Mbah. Evaluating long-term effectiveness of sleeping sickness control measures in Guinea. *Parasites and Vectors*, 8:550, 2015.

[7] M.S. Castaño, Ma. Aliee, E. Mwamba Miaka, M.J. Keeling, N. Chitnis, and K.S. Rock. Screening Strategies for a Sustainable Endpoint for Gambiense Sleeping Sickness. *Journal of Infectious Diseases*, 221:S539–S545, 2020.

[8] J.W. Hargrove, R. Ouifki, D. Kajunguri, G.A. Vale, and S.J. Torr. Modeling the control of trypanosomiasis using trypanocides or insecticide-treated livestock. *PLoS Neg Trop Dis*, 6(5):e1615, 2012.

[9] K.S. Rock, M.L. Ndeffo-Mbah, S. Castaño, C. Palmer, A. Pandey, K.E. Atkins, J.M. Ndung'U, T.D. Hollingsworth, A. Galvani, C. Bever, N. Chitnis, and M.J. Keeling. Assessing strategies against gambiense sleeping sickness through mathematical modeling. *Clinical Infectious Diseases*, 66(Suppl 4):S286–S292, 2018.

[10] P. Buscher, J. M. Bart, M. Boelaert, B. Bucheton, G. Cecchi, N. Chitnis, D. Courtin, L. M. Figueiredo, J. R. Franco, P. Grébaut, E. Hasker, H. Ilboudo, V. Jamonneau, M. Koffi, V. Lejon, A. MacLeod, J. Masumu, E. Matovu, R. Mattioli, H. Noyes, A. Picado, K. S. Rock, B. Rotureau, G. Simo, S. Thévenon, S. Trindade, P. Truc, N. Van Reet, and Informal Expert Group on Gambiense HAT Reservoirs. Do cryptic reservoirs threaten gambiense-sleeping sickness elimination? *Trends Parasitol*, 34(3):197–207, 2018.

[11] F. Checchi, S. Funk, D. Chandramohan, F. Chappuis, and D.T. Haydon. The impact of passive case detection on the transmission dynamics of gambiense human African trypansomiasis. *PLoS Neg Trop Dis*, page 17, 2018.

[12] F. Giordani, L.J. Morrison, T.G. Rowan, H.P. De Koning, and M.P. Barrett. The animal trypanosomiases and their chemotherapy: a review. *Parasitology*, 143(14):1862–1889, 2017.

[13] M.J.B. Verysen. Prospects for area-wide integrated control of tsetse flies (Diptera: Glossinidae) and trypanosomosis in sub-Saharan Africal. *Rev Soc Entomol Argent*, 65:1–21, 2006.

[14] M. Alsan. The effect of the tsetse fly on African development. *American Economic Review*, 105(1):382–410, 2015.

[15] R. Duguma, S. Tasew, A. Olani, D. Damena, D. Alemu, T. Mulatu, Y. Alemayehu, M. Yohannes, M. Bekana, A. Hoppenheit, E. Abatih, T. Habtewold, V. Delespaux, and L. Duchateau. Spatial distribution of *Glossina sp.* and *Trypanosoma sp.* in southwestern Ethiopia. *Parasites and Vectors*, 8(430), 2015.

[16] D.J. Rogers and S.E. Randolph. Tsetse flies in Africa: bane or boon? *Conservation Biology*, 2(1):57–65, 1988.

[17] B. Tsegaye, S. Dagnachew, and G. Terefe. Review on drug resistant animal trypanosomes in Africa and overseas. *African Journal of Basic and Applied Sciences*, 7(2):73–83, 2015.

[18] E.M. Fevre, P. Coleman, M. Odiit, J.W. Magona, S.C. Welburn, and M.E.J. Woolhouse. The origins of a new *Trypanosoma brucei rhodesiense* sleeping sickness outbreak in eastern Uganda. *Lancet*, 358(625-628), 2001.

[19] H.J. Barclay and M.J.B. Vreysen. A dynamic population model for tsetse (Diptera: Glossinidae) area-wide integrated pest management. *Popl Ecol*, 153:89–110, 2011.

[20] W. Mulugeta, J. Wilkes, W. Mulatu, P.A. Majiwa, R. Masake, and A.S. Peregrine. Long-term occurrence of *Trypanosoma congolense* resistant to diminazene, isometamidium, and homidium in cattle at Ghibe, Ethiopia. *Acta Tropica*, 64(3-4):205–17, 1997.

[21] P.H. Clausen, I. Sidibe, I. Kabore, and B. Bauer. Development of multiple drug resistance to *Trypanosoma congolense* in Zebu cattle under high natural tsetse fly challenge in the pastoral zone of Samorogouan, Burkina Faso. *Acta Tropica*, 51(3-4):229–36, 1992.

[22] S. C. Welburn, P. G. Coleman, I. Maudlin, E. M. Fevre, M. Odiit, and M. C. Eisler. Crisis, what crisis? Control of Rhodesian sleeping sickness. *Trends Parasitol*, 22(3):123–8, 2006.

[23] E. M. Fevre, P. G. Coleman, S. C. Welburn, and I. Maudlin. Reanalyzing the 1900-1920 sleeping sickness epidemic in Uganda. *Emerging Infectious Diseases*, 10(4):567–73, 2004.

[24] L. Hamill, K. Picozzi, J. Fyfe, B. von Wissmann, S. Wastling, N. Wardrop, R. Selby, C. A. Acup, K. L. Bardosh, D. Muhanguzi, J. D. Kabasa, C. Waiswa, and S. C. Welburn. Evaluating the impact of targeting livestock for the prevention of human and animal trypanosomiasis, at village level, in districts newly affected with *T. b. rhodesiense* in Uganda. *Infectious Diseases of Poverty*, 6(1):16, 2017.

[25] J. Fyfe, K. Picozzi, C. Waiswa, K.L. Bardosh, and S.C. Welburn. Impact of mass chemotherapy in domestic livestock for control of zoonotic *T. b. rhodesiense* human African trypanosomiasis in Eastern Uganda. *Acta Tropica*, 165:216–229, 2017.

[26] R. Du Toit. Trypanosomiasis in Zululand and the control of tsetse flies by chemical means. *Ondestepoort J Vet Res*, 26:317–387, 1954.

[27] U. Spielberger, B.K. Ma'isa, and U. Abdurrahim. Tsetse (Diptera: Glossinidae) eradication by aerial (helicopter) spraying of persistent insecticides in Nigeria. *Bull Entomol Res*, 67:589–598, 1977.

[28] The Pan African Tsetse and Trypanosomiasis Eradication Campaign- PATTEC. http://www.who.int/trypanosomiasis_african/partners/pattec/en/. Accessed 24 October 2018.

[29] Use of tiny targets to control tsetse flies in Gambian HAT foci: standard operating procedures (edited 27 October 2016). https://www.tsetse.org/sites/lstm-tsetse/files/content/attachments/2018-05-09/FinalisedSOPEnglishVersion-v01_0.pdf. Accessed 24 Oct 2018.

[30] S.J. Torr, I. Maudlin, and G.A. Vale. Less is more: restricted application of insecticide to cattle to improve the cost and efficacy of tsetse control. *Medical and Veterinary Entomology*, 21:53–64, 2007.

[31] G.A. Vale, G. Mutika, and D.F. Lovemore. Insecticide-treated cattle for controlling tsetse flies (Diptera: Glossinidae): some questions answered, many posed. *Bulletin of Entomological Research*, 89:569–578, 1999.

[32] S.J. Torr and W. Hargrove. Factors affecting the landing and feeding responses of the tsetse fly *Glossina pallidipes* to a stationary ox. *Medical and Veterinary Entomology*, 12:196–207, 1998.

[33] S.J. Torr, P.J. Wilson, S. Schofield, T.N.C. Mangwiro, S. Akber, and B.N. White. Application of DNA markers to identify the individual-specific hosts of tsetse feeding on cattle. *Medical and Veterinary Entomology*, 15:78–86, 2001.

[34] M.C. Eisler, S.J. Torr, P.G. Coleman, N. Machila, and J.F. Morton. Integrated control of vector-borne diseases of livestock–pyrethroids: panacea or poison? *Trends Parasitol*, 19(8):341–5, 2003.

[35] A.P.M. Shaw, S.J. Torr, C. Waiwa, G. Cecchi, G.R.W. Wint, R.C. Mattioli, and T.P. Robinson. Estimating the costs of tsetse control options: an example for Uganda. *Preventative Veterinary Medicine*, 110:290–303, 2013.

[36] S.J. Torr and G.A. Vale. Is the even distribution of insecticide-treated cattle essential for tsetse control? Modelling the impact of baits in heterogeneous environments. *PLoS Neg Trop Dis*, 5(10):e1360, 2011.

[37] J.W. Hargrove, S. Omolo, J.S.I. Msalilwa, and B. Fox. Insecticide-treated cattle for tsetse control: the power and the problems. *Medical and Veterinary Entomology*, 14:123–130, 2000.

[38] J.W. Hargrove, S.J. Torr, and H.M. Kindness. Insecticide-treated cattle against tsetse (Diptera: Glossinidae): what governs success? *Bulletin of Entomological Research*, 93:203–217, 2003.

[39] D. Muhanguzi, K. Picozzi, J. Hatendorf, M. Thrusfield, S.C. Welburn, J.D. Kabasa, and C. Waiswa. Collateral benefits of restricted insecticide application for control of African trypanosomiasis on *Theileria parva* in cattle: a randomized controlled trial. *Preventative Veterinary Medicine*, 110:290–303, 2013.

[40] K. Bardosh, C. Waiswa, and S.C. Weburn. Conflict of interest: use of pyrethroids and amidines against tsetse and ticks in zoonotic sleeping sickness endemic areas of Uganda. *Parasites and Vectors*, 6(204), 2013.

[41] A. Shaw, S. Torr, C. Waiswa, and T. Robinson. Comparing costings of alternatives for dealing with tsetse: estimates for Uganda. *Pro-Poor Livestock Policy Initiative*, Working Paper No. 40:51, 2007.

[42] J.M. Lindh, S.J. Torr, G.A. Vale, and M.J. Lehane. Improving the cost-effectiveness of artificial visual baits for controlling the tsetse fly *Glossina fuscipes fuscipes*. *PLoS Neg Trop Dis*, 3(7):e474, 2009.

[43] J.B. Rayaisse, J. Esterhuizen, I. Tirados, D. Kaba, E. Salou, A. Diarrassouba, G.A. Vale, and M.J. Lehane. Towards an optimal design of target for tsetse control: Comparisons of novel targets for the control of palpalis group tsetse in West Africa. *PLoS Neg Trop Dis*, 5(9):e1332, 2011.

[44] J. Esterhuizen, J.B. Rayaisse, I. Tirados, S. Mpiana, P. Solano, G.A. Vale, M.J. Lehane, and S.J. Torr. Improving the cost-effectiveness of visual devices for the control of riverine tsetse flies, the major vectors of human African trypanosomiasis. *PLoS Neg Trop Dis*, 5(8):e1257, 2011.

[45] M.H. Mahamat, M. Peka, J.B. Rayaisse, K.S. Rock, M.A. Toko, J. Darnas, G.M. Brahim, A.B. Alkatib, W. Yoni, I. Tirados, F. Courtin, S.P.C. Brand, C. Nersy, I.O. Alfaroukh, S.J. Torr, M.J. Lehane, and P. Solano. Adding tsetse control to medical activities contributes to decreasing transmission of sleeping sickness in the Mandoul focus (Chad). *PLoS Neg Trop Dis*, 11(7):e0005792, 2017.

[46] J.J. McDermott and P.G. Coleman. Comparing apples and oranges: model-based assessment of different tsetse-transmitted trypanosomosis control strategies. *International Journal for Parasitology*, 31:603–609, 2001.

[47] Human African trypanosomiasis: vector control. http://www.who.int/trypanosomiasis_african/vector_control/en/. Accessed 24 Oct 2018.

[48] R. De Deken and J. Bouyer. Can sequential aerosol technique be used against riverine tsetse? *PLoS Neg Trop DIs*, 2018.

[49] J.W. Hargrove. Extinction probabilities and times to extinction for populations of tsetse flies Glossina spp. (Diptera: Glossinidae) subjected to various control measures. *Bulletin of Entomological Research*, 95:13–21, 2005.

[50] M.J.B. Verysen, J.M. Saleh, M.Y. Ali, A.M. Abdulla, Z.R. Zhu, K.G. Juma, V.A. Dyck, A.R. Msangi, P.A. Mkonyi, and H.U. Feldmann. *Glossina austeni* (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. *Veterinary Entomology*, 93(1):123–135, 2000.

[51] S.C. Welburn, I. Maudlin, and P.P. Simarro. Controlling sleeping sickness- a review. *Parasitology*, 136:1943–1949, 2009.

[52] Alere and FIND launch second-generation rapid test for sleeping sickness. https://www.finddx.org/news/alere-find-launch-2nd-gen-hat-rapid-test/. Accessed 26 Oct 2018.

[53] C. Lubmala, P.R. Bessell, P. Lutumba, S. Baloji, S. Bieler, and J.M. Ndung'u. 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*, 12(7):e0180555, 2017.

[54] E.D. Deeks. Fexinidazole: First global approval. *Drugs*, 79:215–220, 2 2019.

[55] P.G.E. Kennedy. Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *Lancet Neurology*, 12(2):186–194, 2013.

[56] G. Simo and J.B. Rayaisse. Challenges facing the elimination of sleeping sickness in west and central Africa: sustainable control of animal trypanosomiasis as an indispensable approach to achieve the goal. *Parasites and Vectors*, 8:640, 2015.

[57] L.C. Hamill, M.T. Kaare, S.C. Welburn, and K. Picozzi. Domestic pigs as potential reservoirs of human and animal trypanosomiasis in northern Tanzania. *Parasites and Vectors*, 6:322, 2013.

[58] M. K. N'Djetchi, H. Ilboudo, M. Koffi, J. Kaboré, J. W. Kaboré, D. Kaba, F. Courtin, B. Coulibaly, P. Fauret, L. Kouakou, S. Ravel, S. Deborggraeve, P. Solano, T. De Meeûs, B. Bucheton, and V. Jamonneau. The study of trypanosome species circulating in domestic animals in two human African trypanosomiasis foci of Côte d'Ivoire identifies pigs and cattle as potential reservoirs of *Trypanosoma brucei gambiense*. *PLoS Negl Trop Dis*, 11(10):e0005993, 2017.

[59] M. Herrero, P. Havlik, J. McIntire, A. Palazzo, and H. Valin. African livestock futures: Realizing the potential of livestock for food security, poverty reduction and the environment in sub-Saharan Africa. *Office of the Special Representative of the UN Secretary General for Food Security and Nutrition and the United Nations System Influenza Coordination (UNSIC), Geneva, Switzerland*, 188p, 2014.

[60] T.M. Lunde and B. Lintjorn. Cattle and climate in Africa: How climate variability has influenced national cattle holdings from 1961–2008. *PeerJ*, 2013.

[61] P.K. Thornton, J. van de Steeg, A. Notenbaert, and M. Herrero. The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know. *Agricultural Systems*, 101:113–127, 2009.

[62] J. Meisner, R. Barnabas, and P. Rabinowitz. A mathematical model for evaluating the role of cattle treatment in the epidemiology and control of *Trypanosoma brucei rhodesiense* and *T. b. gambiense* sleeping sickness in Uganda. *[In Preparation]*, 2018.

[63] D. Grace, F. Mutua, P. Ochungo, R. Kruska, K. Jones, L. Brierley, L. Lapar, M. Said, M. Herrero, Duc P. P., Bich T. N., I. Akuku, and F. Ogutu. Mapping of poverty and likely zoonoses hotspots. *International Livestock Research Institute*, 2012.

[64] B. Hermesh, A. Rosenthal, and N. Davidovitch. The cycle of distrust in health policy and behavior: Lessons learned from the Negev Bedouin. *PloS One*, 15(8):e0237734, 2020.

[65] B. Hermesh, A. Rosenthal, and N. Davidovitch. Boundaries and Politics. *Monash Bioethics Review*, 37(1):22–37, 2019.

[66] P.K. Thornton. Livestock production: recent trends, future prospects. *Phil. Trans. R. Soc. B*, 365:2853–2867, 2010.

[67] The Livestock Revolution. http://www.fao.org/WAIRDOCS/LEAD/X6115E/x6115e03.htm. Accessed 2 Nov 2018.

[68] H . Steinfeld, P. Gerber, T. D . Wassenaar, V . Castel, and C de Haan. Livestock's long shadow: environmental issues and options. *FAO*, 5:7, 2006.

[69] G.R.W. Wint and T.P. Robinson. Grided livestock of the world 2007. Technical report, FAO, Rome, 2007.

[70] M. Gilbert, G. Nicolas, G. Cinardi, T.B. Van Boeckel, S.O. Vanwambeke, G.R. Wint, and T.P. Robinson. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Scientific Data*, 5:1–11, 10 2018.

[71] B.R. Hankerson, F. Schierhorn, A. V. Prishchepov, C. Dong, C. Eisfelder, and D. Müller. Modeling the spatial distribution of grazing intensity in Kazakhstan. *PLoS ONE*, 14, 1 2019.

[72] C. Jahel, M. Lenormand, S. Seck, A. Apolloni, I. Toure, C. Faye, B. Sall, M. Lo, C.S. Diaw, R. Lancelot, and C. Coste. Mapping livestock movements in Sahelian Africa. *Scientific Reports*, 10, 12 2020.

[73] Global Health Data Exchange. Institute for Health Metrics and Evaluation. http://ghdx.healthdata.org/. Accessed 24 Aug 2017.

[74] Central Data Catalog. Integrated Household Survey Network. https://catalog.ihsn.org/index.php/catalog. Accessed 24 Aug 2017.

[75] M. Heger Boyle, M. King, and M. Sobek. IPUMS-Demographic and Health Surveys: version 7 [dataset]. Minnesota Population Center and ICF International, 2019.

[76] Minnesota Population Center. Integrated Public Use Microdata Series, International: version 7.2 [dataset]. Minneapolis, MN: IPUMS, 2019.

[77] J. Wakefield, D. Simpson, and J. Godwin. Comment: Getting into space with a weight problem. *J Am Stat Assoc*, 111(515):1111–1118, 2016.

[78] P.J. Diggle and P.J. Ribiero Jr. *Model-based geostatistics*. Springer Series in Statistics, New York, 2007.

[79] WorldPop: Population/Individual countries 2000-2020. https://www.worldpop.org/geodata/listing?id=29. Accessed 25 March 2020.

[80] Version 4 DMSP-OLS Nighttime Lights Time Series. National Oceanographic and Atmospheric Administration. https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html. Accessed 30 March 2020.

[81] IUCN: World Database on Protected Areas. https://www.iucn.org/theme/protected-areas/our-work/world-database-protected-areas. Accessed 09 Apr 2018.

[82] Biodiversity A-Z. https://www.biodiversitya-z.org/content/protected-area. Accessed 01 Apr 2020.

[83] World Wildlife Foundation. Global Lakes and Wetlands Database. https://www.worldwildlife.org/pages/global-lakes-and-wetlands-database. Accessed 09 Apr 2018.

[84] U.S. Geological Survey. Digital Eelevation - Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010). Accessed 01 Apr 2020.

[85] F. Lindgren and H. Rue. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19):1–25, 2015.

[86] N. Asmarian, S.M.T. Ayatollahi, Z. Sharafi, and N. Zare. Bayesian spatial joint model for disease mapping of zero-inflated data with R-INLA: A simulation study and an application to male breast cancer in Iran. *Int J Environ Res Public Health*, 16(22):4460, 2019.

[87] J. Wakefield. Multi-level modeling, the ecologic fallacy, and hybrid study designs. *International Journal of Epidemiology*, 38:330–336, 2009.

[88] G-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Constructing priors that penalize the complexity of Gaussian random fields. *J Am Stat Assoc*, 114(525):445–452, 2019.

[89] D. Simpson, H. Rue, A. Riebler, T.G. Martins, and S.H. Sorbye. Penalising model component complexity: A principled, practical approach to constructing priors. *Stat Sci*, 32(1):1–28, 2017.

[90] M. Gilbert, G. Nicolas, G. Cinardi, T.B. Van Boeckel, S.O. Vanwambeke, G.R. Wint, and T.P. Robinson. Global cattle distribution in 2010 (5 minutes of arc), 2018.

[91] M. Gilbert, G. Nicolas, G. Cinardi, T.B. Van Boeckel, S.O. Vanwambeke, G.R. Wint, and T.P. Robinson. Global pigs distribution in 2010 (5 minutes of arc), 2018.

[92] T. Lumley. survey: analysis of complex survey samples, 2016. R package version 3.32.

[93] L. Mercer, J. Wakefield, C. Chen, and T. Lumley. A comparison of spatial smoothing methods for small area estimation with sampling weights. *Spatial Statistics*, 8:69–85, 2014.

[94] FAOSTAT. Live Animals. http://www.fao.org/faostat/en/data/QA. Accessed 16 Feb 2021.

[95] The World Bank. Population, total. https://data.worldbank.org/indicator/SP.POP.TOTL. Accessed 16 Feb 2021.

[96] Demographic and Health Survey: Methodology - collecting geographic data. https://dhsprogram.com/What-We-Do/GPS-Data-Collection.cfm. Accessed 22 Oct 2018.

[97] European Commission. Global Human Settlement Layer. https://ghsl.jrc.ec.europa.eu/index.php. Accessed 09 Apr 2018.

[98] M. Odiit, F. Kansiime, and J. C. Enyaru. Duration of symptoms and case fatality of sleeping sickness caused by *Trypanosoma brucei rhodesiense* in Tororo, Uganda. *East African Medical Journal*, 74(12):792–5, 1997.

[99] K.S. Rock, S.J. Torr, C. Lumbala, and M.J. Keeling. Quantitative evaluation of the strategy to eliminate human African trypanosomiasis in the Democratic Republic of Congo. *Parasites and Vectors*, 8(1):1–13, 2015.

[100] R. Selby, K. Bardosh, K. Picozzi, C. Waiswa, and S.C. Welburn. Cattle movements and trypanosomes: restocking efforts and the spread of *Trypanosoma brucei rhodesiense* sleeping sickness in post-conflict Uganda. *Parasites and Vectors*, 6:281, 2013.

[101] K. Picozzi, E. Fevre, M. Odiit, M. Carrington, M.C. Eisler, I. Maudlin, and S. Welburn. Sleeping sickness in Uganda: a thin line between two fatal diseases. *Briths Medical Journal*, 331:1238–1241, 2005.

[102] P.P. Simarro, J.R. Franco, A. Diarra, J.A. Ruiz Postigo, and J. Jannin. Diversity of human African trypanosomiasis epidemiological settings requires fine-tuning control strategies to facilitate disease elimination. *Research and Reports in Tropical Medicine*, 4:1–6, 2013.

[103] M. Odiit, P. G. Coleman, W. C. Liu, J. J. McDermott, E. M. Fèvre, S. C. Welburn, and M. E. Woolhouse. Quantifying the level of under-detection of *Trypanosoma brucei rhodesiense* sleeping sickness cases. *Trop Med Int Health*, 10(9):840–9, 2005.

[104] Fexinidazole for T.b. rhodesiense. DNDi. https://dndi.org/research-development/portfolio/fexinidazole-tb-rhodesiense/. Accessed 1 Mar 2021.

[105] P. Babokhov, A. O. Sanyaolu, W. A. Oyibo, A. F. Fagbenro-Beyioku, and N. C. Iriemenam. A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis. *Pathogens and Global Health*, 107(5):242–52, 2013.

[106] L. Berrang-Ford, C. Wamboga, and A. S. Kakembo. *Trypanososma brucei rhodesiense* sleeping sickness, Uganda. *Emerg Infect Dis*, 18(10):1686–7, 2012.

[107] P. Capewell, C.C. Cren-Travaille, F. Marchesi, P. Johnston, C. Clucas, R.A. Benson, T. Gorman, E. Calvo-Alvarez, A. Crouzois, G. Jouvion, V. Jammoneau, W. Weir, M. Lynn Stevenson, K. O'Neill, A. Cooper, N. Kuispond Swar, B. Bucheton, D. Mumba Ngoyi, P. Garside, B. Rotureau, and A. MacLeod. The skin is a significant but overlooked anatomical reservoir for vector-borne African trypanosomes. *eLife*, 5, 2016.

[108] C.M. Stone and N. Chitnis. Implications of heterogeneous biting exposure and animal hosts on *Trypanosomiasis brucei gambiense* transmission and control. *PLoS Comp Bio*, 2015.

[109] S. Funk, H. Nishiura, H. Hesterbeek, W. J. Edmunds, and F. Checchi. Identifying transmission cycles at the human-animal interface: the role of animal reservoirs in maintaining gambiense human African trypanosomiasis. *PLoS Comput Biol*, 9(1):e1002855, 2013.

[110] V. Jamonneau, S. Ravel, M. Koffi, D. Kaba, D. G. Zeze, L. Ndri, B. Sane, B. Coulibaly, G. Cuny, and P. Solano. Mixed infections of trypanosomes in tsetse and pigs and their epidemiological significance in a sleeping sickness focus of Côte d'Ivoire. *Parasitology*, 129(Pt 6):693–702, 2004.

[111] T. J. VanderWeele and E. J. Tchetgen Tchetgen. Mediation analysis with time varying exposures and mediators. *J R Stat Soc Series B Stat Methodol*, 79(3):917–938, 2017.

[112] J. Textor, B. van der Zander, M.K. Gilthorpe, M. Liskiewicz, and G.T.H. Ellison. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology*, 45(6):1887–1894, 2016.

[113] J. Robins and M. Hernan. *Estimation of the causal effects of time-varying exposures*, page 553 – 599. Chapman and Hall, Boca Raton, FL, 2009.

[114] F.R. Stevens, A.E. Gaughan, C. Linard, and A.J. Tatem. Disaggregating census data for population mapping using random forests with remote-sensed and ancillary data. *PLOS ONE*, 10(2):e0107042, 2015.

[115] Wealth Index Construction. https://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm. Accessed 29 Oct 2018.

[116] EM-DAT: The International Disaster Database. http://www.emdat.be. Accessed 12 May 2018.

[117] LAADS DAAC. Products N_19AVH13C1, N_18AVH13C1, and N_16AVH13C1. https://ladsweb.modaps.eosdis.nasa.gov. Accessed 10 Apr 2020.

[118] MOD21A2 v006. MODIS/Terra Land Surface Temperature/3-Band Emissivity 8-Day L3 Global 1 km SIN Grid. https://lpdaac.usgs.gov/products/mod21a2v006/. Accessed 10 Apr 2020.

[119] MYD21A2 v006. MODIS/Aqua Land Surface Temperature/3-Band Emissivity 8-Day l3 Global 1 km SIN Grid. https://lpdaac.usgs.gov/products/myd21a2v006/. Accessed 10 Apr 2020.

[120] Inaki Tirados, Andrew Hope, Richard Selby, Fabrice Mpembele, Erick Mwamba Miaka, Marleen Boelaert, Mike J. Lehane, Steve J. Torr, and Michelle C. Stanton. Impact of tiny targets on *Glossina fuscipes quanzensis*, the primary vector of human African trypanosomiasis in the Democratic Republic of the Congo. *PLOS Neglected Tropical Diseases*, 14:e0008270, 10 2020.

[121] J.M. Ndung'u, A. Boulangé, and A. et al. Picado. Trypa-no! contributes to the elimination of gambiense human african trypanosomiasis by combining tsetse control with "screen, diagnose and treat" using innovative tools and strategies. *PLOS Neglected Tropical Diseases*, 14:e0008738, 11 2020.

[122] T.J. VanderWeele and P. Ding. Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Medicine*, 4(167):268–274, 2017.

[123] M.B. Mathur, P. Ding, C.A. Ridell, and T.J. VanderWeele. Website and R package for computing E-values. *Epidemiology*, 6(29):e45–e47, 2018.

[124] L. Liu, M.G. Hudgens, and S. Becker-Dreps. On inverse probability-weighted estimators in the presence of interference. *Biometrika*, 103(4):829–842, 2016.

[125] B.C. Saul, M.G. Hudgens, and M.A. Mallin. Upstream causes of downstream effects. *ArXiv e-prints*, May 2017.

[126] M.D. Turner and E. Schlecht. Livestock mobility in sub-Saharan Africa: A critical review. *Pastoralism*, 9:13, 12 2019.

[127] S. J. Torr and G. A. Vale. Is the even distribution of insecticide-treated cattle essential for tsetse control? Modelling the impact of baits in heterogeneous environments. *PLoS Negl Trop Dis*, 5(10):e1360, 2011.

[128] J.M. Robins and M.A. Hernan. *Estimation of the causal effects of time-varying exposures*, book section 23. CRC Press, Boca Raton, 2009.

[129] X. Wang, Y.R. Yue, and J.J. Faraway. *Errors-in-Variables Regression*, book section 10. Chapman & Hall/CRC. Computer Science and Data Analysis Series, Boca Raton, FL, 2018.

[130] C. Waiswa, K. Picozzi, W. Katunguka-Rwakishaya, E. Olaho-Mukani, R.A. Musoke, and S.C. Welburn. *Glossina fuscipes fuscipes* in the trypanosomiasis endemic areas of south eastern Uganda: apparent density, trypanosome infection rates and host feeding preferences. *Acta Tropica*, 99(1):23–29, 2006.

[131] C. Acup, Bardosh K.L., Picozzi K., Waiswa C., and S. Welburn. Factors influencing passive surveillance for *T. b. rhodesiense* human African trypanosomiasis in Uganda. *Acta Tropica*, 165:230–239, 2017.

[132] F. Chappuis, M.A. Lima, L. Flevaud, and K. Ritmeijer. Human African trypanosomiasis in areas without surveillance. *Emerging Infectious Diseases*, 16(2):354–355, 2010.

[133] K.S. Rock, S.J. Torr, C. Lumbala, and M.J. Keeling. Predicting the impact of intervention strategies for sleeping sickness in two high-endemicity health zones of the Democratic Republic of Congo. *PLoS Negl Trop Dis*, 11(1):e0005162, 2017.

[134] S. Greenland and H. Morgenstern. Ecological bias, confounding, and effect modification. *International Journal of Epidemiology*, 18(1):269–274, 1989.

[135] S. Greenland and J.M. Robins. Invited commentary: ecological studies– biases, misconceptions, and counterexamples. *Am J Epidemiol*, 139:45–59, 1994.

[136] S. Greenland and J. Robins. Accepting the limits of ecologic studies: Drs. Greenland and Robins reply to Drs. Piantadosi and Cohen. *Am J Epidemiol*, 139(8):769–771, 1994.

[137] T.F. Webster. Bias magnification in ecologic studies: a methodological investigation. *Environ Health*, 6(17), 2007.

[138] S. Greenland. Ecologic versus individual-level sources of bias in ecological estimates of contextual health effects. *International Journal of Epidemiology*, 30:1343–1350, 2001.

[139] W.S. Robinson. Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3):351–357, 1950.

[140] M.A. Hernan and J.M. Robins. *Causal Inference*. Champan & Hall/CRC, Boca Raton, forthcoming edition, 2018.

[141] D.J. Rogers. Satellites, space, time and the African trypanosomiases. *Adv Parasitol*, 47:129–71, 2000.

[142] T.M. Lunde, M. Balkew, D. Korecha, T. Gebre-Michael, F. Massebo, A. Sorteberg, and B. Lintjorn. A dynamic model of some malaria-transmitting anopheline mosquitoes of the Afrotropical region. i. Model description and sensitivity analysis. *Malaria Journal*, 12(28), 2013.

[143] T.M. Lunde, M. Balkew, D. Korecha, T. Gebre-Michael, F. Massebo, A. Sorteberg, and B. Lintjorn. A dynamic model of some malaria-transmitting anopheline mosquitoes of the Afrotropical region. ii. Validation of species distribution and seasonal variations. *Malaria Journal*, 12(78), 2013.

[144] G.P. Asner, A.J. Elmore, L.P. Olander, R.E. Martin, and A.T. Harris. Grazing systems, ecosystem responses, and global change. *Annu. Rev. Environ. Resour.*, 29:261–99, 2004.

[145] J. Bouyer, A.H. Dicko, G. Cecchi, S. Ravel, L. Guerrini, P. Solano, M.J.B. Vreysen, T. De Meeus, and R. Lancelot. Mapping landscape friction to locate isolated tsetse populations that are candidates for elimination. *Proceedings of the National Academy of Sciences*, 112(47):14575–14580, 2015.

[146] A.H. Dicko, L. Percoma, A. Sow, Y. Adam, C. Mahama, I. Sidibe, G. Dayo, S. Thevenon, W. Fonta, S. Sanfo, A. Djiteye, E. Salou, V. Djohan, C. Cecchi, and J. Bouyer. A spatio-temporal model of African animal trypanosomiasis risk. *PLoS Neg Trop Dis*, 9(7):e0003921, 2015.

[147] B. Bauer, S. Amsel-de Lafosse, P.H. Calusen, I. Kabore, and J. Petrich-Bauer. Successful application of deltamethrin pour on to cattle in a campaign against tsetse flies (Glossina spp.) in the pastoral zone of Samorogouan, Burkina Faso. *Trop Med Parasitol*, 36(3):183–9, 1995.

[148] S. Lin, M. H. DeVisser, and J. P. Messina. An agent-based model to simulate tsetse fly distribution and control techniques: a case study in Nguruman, Kenya. *Ecol Modell*, 314:80–89, 2015.

[149] C. Mweempwa, T. Marcotty, C. De Pus, B. L. Penzhorn, A. H. Dicko, J. Bouyer, and R. De Deken. Impact of habitat fragmentation on tsetse populations and trypanosomosis risk in Eastern Zambia. *Parasit Vectors*, 8:406, 2015.

[150] S. Pagabeleguem, M. Sangaré, Z. Bengaly, M. Akoudjin, A. M. Belem, and J. Bouyer. Climate, cattle rearing systems and African animal trypanosomosis risk in Burkina Faso. *PLoS One*, 7(11):e49762, 2012.

[151] T.J. VanderWeele. *Mediation: Introduction and Regression-Based Approaches*, book section 2. Oxford University Press, New York, NY, 2015.

[152] T.J. VanderWeele. Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Social Methods Res*, 38(4):515–544, 2010.

[153] S. H. Lin, J. Young, R. Logan, E. J. Tchetgen Tchetgen, and T. J. VanderWeele. Parametric mediational g-formula approach to mediation analysis with time-varying exposures, mediators, and confounders. *Epidemiology*, 28(2):266–274, 2017.

[154] M. Aliee, K.S.. Rock, and M.J.. Keeling. Estimating the distribution of time to extinction of infectious diseases in mean-field approaches. *Journal of the Royal Society, Interface*, 17(173):20200540, 2020.

[155] Christopher N. Davis, Kat S. Rock, Erick Mwamba Miaka, and Matt J. Keeling. Village-scale persistence and elimination of gambiense human African trypanosomiasis. *PLoS Neglected Tropical Diseases*, 13(10):1–15, 2019.

[156] S. Davis, S. Aksoy, and A. Galvani. A global sensitivity analysis for African sleeping sickness. *Parasitology*, 138(4):516–526, 2011.

[157] S. Alderton, J. Noble, and P. Atkinson. *ECAL 2013: The Twelfth European Conference on Artificial Life*, pages 27–34, 2013.

[158] P. Grébaut, K. Girardin, V. Fédérico, and F. Bousquet. Simulating the elimination of sleeping sickness with an agent-based model. *Parasite*, 23:1247–1405, 2016.

[159] D.J. Rogers. A general model for the African trypanosomiases. *Parasitology*, 97(1):193–212, 1988.

[160] C.W. Lorton, J.L. Proctor, M.K. Roh, and Philip A. Welkhoff. Compartmental Modeling Software: A fast, discrete stochastic framework for biochemical and epidemiological simulation. *bioRxiv*, 2019.

[161] P.P. Simarro, G. Cecchi, M. Paone, J.R. Franco, A. Diarra, J.A. Ruiz, E.M. Fevre, F. Courtin, R.C. Mattioli, and J.G. Jannin. The Atlas of human African trypanosomiasis: a contribution to global mapping of neglected tropical diseases. *International Journal of Health Geographics*, 9:57, 2010.

[162] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[163] F. Checchi, A.P. Cox, F. Chappuis, G. Priotto, D. Chandramohan, and D.T. Haydon. Prevalence and under-detection of gambiense human African trypanosomiasis during mass screening sessions in Uganda and Sudan. *Parasites & Vectors*, 5(157), 2012.

[164] A. King. Introduction to inference: parameter estimation. https://kingaa.github.io/short-course/parest/parest.html. Accessed 1 Mar 2021.

[165] Control of biting insects on pigs. The Pig Site. https://www.thepigsite.com/articles/control-of-biting-insects-on-pigs. Accessed 26 Feb 2021.

[166] The DHS Program. https://dhsprogram.com. Accessed 8 Apr 2020.

[167] Malawi. Demographic and Health Survey 2004-2005. National Statistical Office. Zomba, Malawi. ORC Macro. Calverton, Maryland, USA. December 2005. https://dhsprogram.com/pubs/pdf/FR175/FR-175-MW04.pdf. Accessed 8 Apr 2020.

[168] Malawi. Demographic and Health Survey 2010. National Statistical Office. Zomba, Malawi. ORC Macro. Calverton, Maryland, USA. September 2011. https://dhsprogram.com/pubs/pdf/FR247/FR247.pdf. Accessed 8 Apr 2020.

[169] Malawi. Demographic and Health Survey 2015-2016. National Statistical Office. Zomba, Malawi. The DHS Program, ICF International, Rockville, Maryland, USA. February 2017. https://dhsprogram.com/pubs/pdf/FR319/FR319.pdf. Accessed 8 Apr 2020.

[170] Malawi. Demographic and Health Survey 2000. National Statistical Office. Zomba, Malawi. ORC Macro. Calverton, Maryland, USA. August 2001. https://dhsprogram.com/pubs/pdf/FR123/FR123.pdf. Accessed 8 Apr 2020.

[171] World Health Survey Results - Report of Malawi. https://www.who.int/healthinfo/survey/whsmwi-malawi.pdf. Accessed 8 Apr 2020.

[172] Malawi Second Integrated Household Survey (IHS-2). 2004-2005. Basic Information Document. National Statistical Office. Zomba, Malawi. October 2005. https://microdata.worldbank.org/index.php/catalog/2307/download/35042. Accessed 8 Apr 2020.

[173] Malawi Third Integrated Household Survey (IHS-3). 2010-2011. Basic Information Document. National Statistical Office. Zomba, Malawi. March 2012. https://microdata.worldbank.org/index.php/catalog/1003/download/40803. Accessed 8 Apr 2020.

[174] Malawi Fourth Integrated Household Survey (IHS-4). 2016-2017. Basic Information Document. National Statistical Office. Zomba, Malawi. November 2017.

https://microdata.worldbank.org/index.php/catalog/2936/download/47116. Accessed 8 Apr 2020.

[175] Malawi Integrated Household Panel Survey 2013. Basic Information Document. September 2014. National Statistical Office. Zomba, Malawi. https://microdata.worldbank.org/index.php/catalog/2939/download/48123. Accessed 8 Apr 2020.

[176] Malawi Integrated Household Panel Survey 2016. Basic Information Document. National Statistical Office, Zomba, Malawi. November 2017. https://microdata.worldbank.org/index.php/catalog/2939/download/48116. Accessed 8 Apr 2020.

[177] Technology Adoption and Risk Initiative Household Baseline Survey 2006. The World Bank Microdata Library. https://microdata.worldbank.org/index.php/catalog/1541/study-description. Accessed 8 Apr 2020.

[178] US Board on Geographical Names. National Geospatial-Intelligence Agency. https://www.nga.mil/ProductsServices/Pages/Geographic-Names.aspx. Accessed 9 Apr 2020.

[179] Alexandria Digital on Library Gazetteer. http://adl-gazetteer.geog.ucsb.edu/. Accessed 9 Apr 2020.

[180] Getty Thesaurus of Geographic Names Online. The Getty Research Institute. https://www.getty.edu/research/tools/vocabularies/tgn/. Accessed 9 Apr 2020.

[181] Goggle Maps World Gazetteer. http://www.maplandia.com/. Accessed 9 Apr 2020.

[182] Malaria Indicator Survey Overview. The DHS Program. https://dhsprogram.com/What-We-Do/Survey-Types/MIS.cfm. Accessed 9 Apr 2020.

[183] Malawi. Malaria Indicator Survey 2012. National Malaria Control Programme, Ministry of Health, Lilongwe, Malawi. MEASURE DHS, ICF International, Rockville, Maryland, USA. November 2012. https://dhsprogram.com/pubs/pdf/MIS13/MIS13.pdf. Accessed 9 Apr 2020.

[184] Malawi. Malaria Indicator Survey 2014. National Malaria Control Programme, Ministry of Health, Lilongwe, Malawi. The DHS Program, ICF International, Rockville, Maryland, USA. February 2015. https://dhsprogram.com/pubs/pdf/MIS18/MIS18.pdf. Accessed 9 Apr 2020.

[185] Malawi. Malaria Indicator Survey 2017. National Malaria Control Programme, Ministry of Health, Lilongwe, Malawi. The DHS Program, ICF International, Rockville, Maryland, USA. January 2018. https://dhsprogram.com/pubs/pdf/MIS28/MIS28.pdf. Accessed 10 Dec 2020.

[186] Multiple Indicator Cluster Survey (MICS). UNICEF. https://mics.unicef.org/. Accessed 9 Apr 2020.

[187] MDG Endline Survey 2013-2014. International Household Survey Network. https://catalog.ihsn.org/index.php/catalog/6494/study-description. Accessed 9 Apr 2020.

[188] Malawi MDG Endline Survey 2014. Main Report. National Statistical Office of Malawi. June 2015. https://mics-surveys-prod.s3.amazonaws.com/MICS5/Eastern Accessed 9 Apr 2020.

[189] Uganda. Demographic and Health Survey 2001. Uganda Bureau of Statistics. Kampala, Uganda. ORC Macro. Calverton, Maryland, USA. December 2001. https://dhsprogram.com/pubs/pdf/FR128/FR128.pdf. Accessed 22 Dec 2020.

[190] Uganda. Demographic and Health Survey 2006. Uganda Bureau of Statistics. Kampala, Uganda. ORC Macro. Calverton, Maryland, USA. August 2007. https://dhsprogram.com/pubs/pdf/FR194/FR194.pdf. Accessed 10 Dec 2020.

[191] Uganda. Demographic and Health Survey 2011. Uganda Bureau of Statistics. Kampala, Uganda. ORC Macro. Calverton, Maryland, USA. August 2012. https://dhsprogram.com/pubs/pdf/FR264/FR264.pdf. Accessed 22 Dec 2020.

[192] Uganda. Demographic and Health Survey 2016. Uganda Bureau of Statistics. Kampala, Uganda. ORC Macro. Calverton, Maryland, USA. January 2018. https://dhsprogram.com/pubs/pdf/FR333/FR333.pdf. Accessed 22 Dec 2020.

[193] Uganda. Malaria Indicator Survey 2009. Uganda Bureau of Statistics, Uganda Malaria Surveillance Project Molecular Laboratory, and National Malaria Control Programme, Kampala, Uganda. ICF Macro, Calverton,, Maryland, USA. August 2010. https://dhsprogram.com/pubs/pdf/MIS6/MIS6.pdf. Accessed 22 Dec 2020.

[194] Uganda. Malaria Indicator Survey 2014-15. Uganda Bureau of Statistics, Uganda Malaria Surveillance Project Molecular Laboratory, and National Malaria Control Programme, Kampala, Uganda. ICF Macro, Calverton,, Maryland, USA. October 2015. https://dhsprogram.com/pubs/pdf/MIS21/MIS21.pdf. Accessed 22 Dec 2020.

[195] Uganda. Malaria Indicator Survey 2018-19. Uganda Bureau of Statistics, Uganda Malaria Surveillance Project Molecular Laboratory, and National Malaria Control Programme, Kampala, Uganda. ICF Macro, Calverton,, Maryland, USA. March 2020. https://dhsprogram.com/pubs/pdf/MIS34/MIS34.pdf. Accessed 22 Dec 2020.

[196] The Uganda National Panel Survey (UNPS) 2009/10. Basic Iinformation Document. Uganda Bureau of Statistics. https://microdata.worldbank.org/index.php/catalog/1001/related-materials. Accessed 22 Dec 2020.

[197] Uganda. Malaria Indicator Survey 2011. Uganda Bureau of Statistics, Uganda Malaria Surveillance Project Molecular Laboratory, and National Malaria Control Programme, Kampala, Uganda. ICF Macro, Calverton,, Maryland, USA. August 2012. https://dhsprogram.com/publications/publication-AIS10-AIS-Final-Reports.cfm. Accessed 22 Dec 2020.

[198] République Démocratique du Congo. Enquéte Démographique et de Santé (EDS-RDC). Ministère du Plan and Ministère de la Santé, and Macro International Inc. Kinshasa, DRC and Calverton, Maryland. August 2008. https://dhsprogram.com/pubs/pdf/FR208/FR208.pdf. Accessed 16 Feb 2021.

[199] République Démocratique du Congo. Deuxiéme Enquéte Démographique et de Santé (EDS-RDC). Ministère du Plan et Suivi de la Mise en oeuvre de la Révolution de la Modernité, Ministère de la Santé Publique, and Measure DHS International. Rockville, Maryland. September 2014. https://dhsprogram.com/pubs/pdf/FR300/FR300.pdf. Accessed 16 Feb 2021.

[200] Multiple Indicator Cluster Survey MICS-2010: monitoring the situation of women and children, Democratic Republic of Congo summary report. Ministry of Planning and National Institute of Statistics in collaboration with the United Nations Children's Fund. May, 2011. https://www.unicef.org/drcongo/en/reports/multiple-indicator-cluster-survey-2010. Accessed 16 Feb 2021.

[201] The Fifth population census in Sudan: a census with a full coverage and high accuracy. https://catalog.ihsn.org/index.php/catalog/4216/download/55707. Accessed 9 Apr 2020.

[202] Republic of the Sudan. Fifth Population and Housing Census. Census Enumerator's Manual Long Questionnaire. https://microdata.worldbank.org/index.php/catalog/1631/download/26122. Accessed 9 Apr 2020.

[203] W. Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2019. R package version 1.9.12.

[204] A.J. Tatem, N. Campiz, P.W. Gething, R.W. Snow, and C Linard. The effects of spatial population dataset choice on estimates of population at risk of disease. *Population Health Metrics*, 9(4), 2011.

[205] Life expectancy at birth, total (years). The World Bank. https://data.worldbank.org/indicator/SP.DYN.LE00.IN. Accessed 3 Feb 2021.

[206] J.R. Franco, G. Cecchi, G. Priotto, M. Paone, D. Diarra, G. Grout, R.C. Mattioli, and D. Argaw. Monitoring the elimination of human African trypanosomiasis: Update to 2014. *PLoS Neglected Tropical Diseases*, 11, 5 2017.

[207] F. Checchi, J.A.N. Filipe, D.T. Haydon, D. Chandramohan, and F. Chappuis. Estimates of the duration of the early and late stage of gambiense sleeping sickness. *BMC Infectious Diseases*, 8(16), 2008.

[208] K.S. Rock, S.J. Torr, C. Lumbala, and M. Keeling. Quantitative evaluation of the strategy to eliminate human African trypanosomiasis in the Democratic Republic of Congo. *Parasites & Vectors*, (532):8, 2015.

[209] FAO. Pig Sector Kenya. FAO Animal Production and Health Livestock Country Reviews. No. 3. Rome. Technical report, FAO, Rome, 2012.

[210] B.C. Omeke and D.O. Ugwu. Pig trypanosomiasis: comparative anaemia and histopathology of lymphoid organs. *Rev Elev Med Vet Pays Trop*, 44(3):267–72, 1991.

[211] L. Penchenier, D. Alhadji, S. Bahébégué, G. Simo, C. Laveissière, and G. Cuny. Spontaneous cure of domestic pigs experimentally infected by Trypanosoma brucei gambiense: Implications for the control of sleeping sickness. *Veterinary Parasitology*, 133(1):7–11, 2005.

[212] R.J. Phelps and G.A. Vale. Studies on populations of *Glossina morsitans morsitans* and *G. pallidipes* (Diptera: Glossinidae) in Rhodesia. *Journal of Applied Ecology*, 15:743–760, 1978.

[213] C. Dale, S. C. Welburn, I. Maudlin, and P. J.M. Milligan. The kinetics of maturation of trypanosome infections in tsetse. *Parasitology*, 111(2):187–191, 1995.

[214] G. Simo, B. Silatsa, N. Flobert, P. Lutumba, P. Mansinsa, J. Madinga, E. Manzambi, R. De Deken, and T. Asonganyi. Identification of different trypanosome species in the mid-guts of tsetse flies of the Malanga (Kimpese) sleeping sickness focus of the Democratic Republic of Congo. *Parasit Vectors*, 5(201), 2012.

[215] J.M.B. Harley. Comparison of the susceptibility to infection with *Trypanosoma rhodesiense* of *Glossina pallidipes, G. morsitans, G.fuscipes* and *G. brevipalpis*. *Annals of Tropical Medicine & Parasitology*, 65(2):185–9, 1971.

[216] J.P. Gouteux, J.C. Kounda Gboumbi, L. Noutoua, F. D'Amico, C. Bailly, and J.B. Roungou. Man-fly contact in the Gambian trypanosomiasis focus of Nola-Bilolo (Central African Republic). *Trop Med Parasitol*, 44(3):213–8, 1993.

[217] L.M. MacLean, M. Odiit, J.E. Chisi, P.G.E. Kennedy, and J.M. Sternberg. Focus–Specific Clinical Profiles in Human African Trypanosomiasis Caused by Trypanosoma brucei rhodesiense. *PLoS Negl Trop Dis*, 4(12):e906, 2010.

[218] C. Waiswa, W. Olaho-Mukani, and E. Katunguka-Rwakishaya. Domestic animals as reservoirs for sleeping sickness in three endemic foci in south-eastern uganda. *Annals of Tropical Medicine and Parasitology*, 97:149–155, 3 2003.

[219] Warthog. African Wildlife Foundation. https://www.awf.org/wildlife-conservation/warthog.

[220] A.P. Cox, O. Tosas, A. Tilley, K. Picozzi, P. Coleman, G. Hide, and S.C. Welburn. Constraints to estimating the prevalence of trypanosome infections in East African zebu cattle. *Parasites and Vectors*, 3:82, 2010.

[221] N.E. Anderson, J. Mubanga, E.M. Fevre, K. Picozzi, M.C. Eisler, R. Thomas, and S.C. Welburn. Characterisation of the wildlife reservoir community for human and animal trypanosomiasis in the Luangwa Valley, Zambia. *PLoS Neglected Tropical Diseases*, 5, 6 2011.

[222] M.J. Otte and P. Chilonda. Cattle and small ruminant production systems in sub-Saharan Africa. A systematic review. Technical report, FAO, Rome, 2002.

[223] V. M. Nantulya, A. J. Musoke, F. R. Rurangirwa, and S. K. Moloo. Resistance of cattle to tsetse-transmitted challenge with Trypanosoma brucei or Trypanosoma congolense after spontaneous recovery from syringe-passaged infections. *Infection and Immunity*, 43(2):735–738, 1984.

[224] Bushbuck. African Wildlife Foundation. https://www.awf.org/wildlife-conservation/bushbuck. Accessed 1 Mar 2021.

[225] African Buffalo. African Wildlife Foundation. https://www.awf.org/wildlife-conservation/african-buffalo. Accessed 1 Mar 2021.

[226] M. Marsela, K. Hayashida, R. Nakao, E. Chatanga, A.K. Gaithuma, K. Naoko, J. Musaya, C. Sugimoto, and J. Yamagishi. Molecular identification of trypanosomes in cattle in Malawi using PCR methods and nanopore sequencing: Epidemiological implications for the control of human and animal trypanosomiases. *Parasite*, 27, 2020.

Appendix A

SAMPLING METHODS FOR LIVESTOCK AND WEALTH MAPPING DATA SOURCES

A.1 Malawi

Demographic and Health Survey (2000, 2004-2005, 2010, 2015-2016) The Demographic and Health Surveys (DHS) program, funded by the US Agency for International Development and other donors, has provided technical assistance to over 400 surveys in 90 countries since 1984. These surveys generally perform stratified two-stage sampling. For DHS rounds conducted in Malawi between 2000 and 2016, the Malawi Census of Population and Housing was used as the sampling frame: the 2000 and 2004-2005 rounds used the 1998 census, while the 2010 and 2015-2016 rounds used the 2008 census [166].

The smallest geographical unit contained in these sampling frames are enumeration areas, which are referred to as clusters for DHS sampling. Generally the sampling frame is stratified by urban/rural status and domain (district, district group, or region; see detail below). In the first stage, clusters are randomly selected within each stratum, typically with probability proportional to the size of each cluster (defined by the number of households per the sampling frame), and complete household listings are performed in the sampled clusters. Large clusters (e.g., more than 250 households) may be broken into two segments prior to selection, and one segment selected with probability proportional to size. Generally the number of clusters selected is equivalent across domains, however some domains may be oversampled if they are particularly large, or if warranted by programmatic interests [166].

In the second stage, households within each cluster are randomly selected, typically with an equal number of households selected per cluster, generating a self-weighted sample: an average of 29 households per cluster in the 2004-2005 DHS [167]; 20 households per urban

cluster and 35 households per rural cluster in the 2010 DHS [168]; and 30 households per urban cluster and 33 households per rural cluster in the 2015-2016 DHS [169]. All women ages 15-49 in selected households are eligible for interview, and in a subsample of households (one in three to one in four, depending on the survey), additional data collection is performed, including interview of men 15-54 years old, anemia testing and morphometric measurements of women and children, and HIV testing of women [166].

In the 2000 DHS, a total of 560 clusters (449 rural, 111 urban) and 15,421 households (2,868 urban, 12,553 rural) were selected [170]. In the 2004-2005 DHS, 552 clusters (458 rural, 64 urban) and 15,410 households were selected (the 2004-2005 DHS report does not break this figure down by urban/rural status) [167]. For the 2010 round, 849 clusters (691 rural, 158 urban) and 27,345 households (24,185 rural, 3,160 urban) were selected [168], and for the 2015-2016 DHS, 850 clusters (677 rural, 183 urban) and 27,531 households (22,341 rural, 5,190 urban) were selected[169].

The goal of these surveys is to provide estimates of indicators which are representative at the national and regional levels, for both urban and rural populations, and for designated districts (administrative level 2). In 2000, there were 11 designated districts which were oversampled to meet this goal: Lilongwe, Blantyre, Zomba, Mzimba, Mangochi, Kasungu, Salima, Machinga, Mulanje, Thyolo, and Karonga. In 2004-2005 the same districts were designated, with the exception of Karonga, and with the additional goal of generating representative samples from three regional groups comprised of the remaining 17 districts. In 2010, 26 out of 28 districts were designated, while Nkhata Bay and Likoma were combined. Finally, in 2015-2016, all 28 districts were designated.

To protect participant privacy, the GPS coordinates provided in the DHS data are randomly displaced, known as "jittering." This produces an error of 0-2 kilometers for urban clusters and 0-5 kms for rural clusters, with an additional 1% of rural clusters displaced 0-10kms. The displacement is restricted so that clusters stay within DHS survey region for surveys released prior to 2009, or within administrative level 2 for surveys released after 2009 where possible [96].

World Health Survey (2003) The World Health Survey was developed and implemented by the World Health Organization (WHO), and has been completed in 90 countries. The sample is all adult (over 18 years of age) members of the national *de facto* population; that is, all people, including guest workers, immigrants, and refugees. The sampling plan for Malawi is not detailed in its report, however general guidance on sample design is provided [171].

As with DHS, in most countries WHS sapling is done via stratified multi-stage cluster sampling. Only probability sampling (vs. quota, convenience sampling, or random walk) is allowed, such that every individual in the sampling frame ha s a known and non-zero probability of selection. Stratification variables vary by country but may include administrative boundaries, urban vs. rural status, socioeconomic zone, and presence of a health facility.

The most recent population census is recommended as the sampling frame. Clusters, defined as naturally occurring groupings within the population for which administrative units have clear, non-overlapping boundaries, are selected proportional to size. If the sampling frame was assembled over three years ago or is known to be unreliable, a complete enumeration of households within selected clusters is recommended.

Households within clusters are selected by systematic sampling, in which sampling is performed at fixed intervals on a list, starting from a randomly chosen point. This could be linear (where the start of the list is not returned to once the end is reached), or circular (where it is). Within households, a random household member is selected using a Kish grid. The maximum respondents per cluster is set as 50, with the ideal range being 20-30. In Malawi, 5,880 households were selected, and 5,490 households and 5,287 individuals interviewed. No details are provided in survey documentation on jittering of cluster locations.

Integrated Household Survey (2004, 2010-2011, 2016-2017) and Integrated Household Panel Survey (2010-2013-2016) The Integrated Household Survey (IHS) is conducted by the Government of Malawi through the National Statistical Office roughly every 5 years. IHS-1 (1997-1998) and IHS-2 (2004) were implemented with technical as-

sistance from the World Bank, and IHS-3 (2010) and IHS-4 (2016-2017) were implemented under the umbrella of the World Bank Living Standards Measurement Survey - Integrated Surveys on Agriculture Initiative (LSMS-ISA) [172]. The Integrated Household Panel Survey (IHPS) also received technical and financial assistance from LSMS-ISA, in addition to DFID and the Government of Norway. The first IHPS wave was launched in 2010 from a subset of IHS-3 clusters, detailed below. Second and third waves were conducted in 2013 and 2016 [173].

All IHS rounds used the most recent Malawi Census of Population and Housing as the sampling frame; for IHS-2 this was the 1998 census, and for IHS-3 and IHS-4 this was the 2008 census. Sampling was two-stage stratified sampling, with stratification being urban/rural × geographic domain. For all IHS rounds four urban strata were established (Lilongwe, Blantyre, Mzuzu, and the Municipality of Zomba) and each of the 27 districts were considered a separate sub-stratum of the main rural stratum. In IHS-2 and IHS-3 Likoma was excluded; in IHS-4 Likoma was included, but combined with Nkhata Bay [172, 173, 174].

In stage 1 sampling, enumeration areas (clusters) were selected with probability proportional to size, defined by the number of households. For IHS-3 and IHS-4, enumeration areas (clusters) within each district were sorted by administrative area prior to selection to provide implicit geographic stratification within districts, and IHS-3 enumeration areas were removed from the sampling frame before stage 1 sampling for IHS-4. IHS-2 sampled 564 enumeration areas (492 rural, 72 urban), IHS-3 sampled 768 EAs (628 rural, 140 urban), and IHS-4 sampled 779 EAs (637 rural, 142 urban). Complete household listings were performed in selected enumeration areas after stage 1 sampling [172, 173, 174].

In stage 2 sampling, households were randomly selected in each enumeration area: 20 households per enumeration area for IHS-2, and 16 households per enumeration area for IHS-3 and IHS-4, to lower the design effect seen in IHS-2. In all rounds, 5 replacement households were selected in each enumeration area if any of the initial households could not be located or refused to participate. IHS-3 and IHS-4 also excluded institutionalized populations (hospitals, prisons, and military barracks). IHS-2 sampled 11,280 households

(9,120 rural, 2,160 urban) [172], IHS-3 sampled 12,271 households (10,038 rural, 2,223 urban) [173], and IHS-4 sampled 12.447 households (10,175 rural, 2,272 urban) [174].

IHPS was conducted on an IHS-3 (2010) sub-sample, selected prior to the start of IHS-3 fieldwork: 204 enumeration areas (3,247 households) were selected to be representative at the national, regional, and urban/rural levels. To reduce the variability of household weights (inverse of the probability of selection) in each stratum (urban/rural × region), as households in larger districts have higher weights in IHS-3, enumeration areas in each IHPS stratum were selected with probability proportional to $M_h \,/\, n_h$, where M_h is the number of households in district h, and n_h is the number of enumeration areas selected from district h. This results in an approximately proportionate allocation of the sample to districts within each region: 48 enumeration areas (786 households) in the North region (reflecting oversampling, as this region has the smallest population size); 72 enumeration areas (1,152 households) in the Central Region; and 84 enumeration areas (864 households) in the South region. Across regions, 54 urban enumeration areas (864 households, also reflecting oversampling), and 150 rural enumeration areas (2,400 households) were selected. We did not use data from the first wave of the IHPS as these are a subset of IHS-3 [173].

In the second (2013) and third (2016) waves of the IHPS, individuals who had moved away from baseline dwellings were tracked as long as they were neither servants nor guests in the 2010 wave, were projected to be at least 12 years old, were known to reside in mainland Malawi (not Likoma), and were not institutionalized. These new households were then brought into the IHPS sample. Out of the 3,247 baseline households, 3,104 were remaining in 2013, with the rest having died in entirety. After tracking individuals that had moved, the sample was comprised of 4,000 households; excluding those no longer eligible, individual-level attrition was 6.39%. Of the 4,000 household sample, 66.5% were located within 1km of their baseline location, 16.1% were 1-10km, and 17.5% were more than 10km from their baseline location; 83.1% were rural, 46.3% in the Southern Region, 45.1% in the Central Region, and 8.7% in the Northern Region [175].

Due to budget issues, the target sample for the third wave (2016) was adjusted to house-

holds associated with 102 out of 204 baseline enumeration areas. Stratification by region, urban, and rural strata was maintained, with proportional allocation across regions, and domain analysis was limited to national, urban, and rural (no longer regional). This left 1,990 wave 2 households in the target sample for wave 3, which increased to 2,507 by the end of tracking, with a household-level attrition rate of 4% and an individual-level attrition rate of 9% when restricted to eligible individuals. Of the 2,507 household sample, 51% of were located within 1km of baseline location, 21% were 1-10km, and 28% were more than 10km from their baseline location; 91% were rural, 50% in the Southern Region, 45% in the Central Region, and 5% in the Northern Region [176].

Technology Adoption and Risk Initiative Survey (2006) The Malawi Technology Adoption and Risk Initiative (MTARI) household survey represents baseline data collected for a cluster-randomized trial conducted by the World Bank and researchers at the University of Michigan, to examine whether provision of insurance against major drivers of production risk encourages produces (maize and groundnut) to take out loans to adopt new technologies [177].

Study recruitment was comprised of convenience sampling. The National Smallholder Farmers Association of Malawi recruited 159 farmers' clubs (10-20 producers each) from among its members in Lilongwe North, Mchinji, Kasungu, and Nkhotakota. From these clubs, 787 producers consented to participate. Out of 341 village names contained in the geographical data, GPS coordinates could be established for 143 clusters. Geolocation was performed by searching publicly-available shapefiles for Malawi, and common gazetteers including the GEOnet Names Server database of the United States National Geospatial-Intelligence Agency [178], the Alexandria Digital Library Gazetteer [179], the Getty Thesaurus of Geographic Names [180], and Google Maps World Gazetteer [181].

Malaria Indicator Survey (2012, 2014, 2017) The DHS Program is a major contributor to the development of the Malaria Indicator Survey (MIS), which was developed by the

Monitoring and Evaluation Working Group (MERG) of Roll Back Malaria [182].

The MIS sample is a nationally representative sample which aims to provide estimates at regional, urban, and rural levels. Sample design was stratified two-stage cluster sampling, with the 2008 National Population and Housing Census used as the sampling frame for all rounds relevant to our study. The number of enumeration areas per region was selected proportional to the population size of each region, with oversampling of urban areas. In the 2014 and 2017 MIS, the Northern Region was also oversampled. Within regions, enumeration areas were also selected with probability proportional to size (number of households). The final sample contained 140 enumeration areas and 3,500 households in both 2012 and 2014, and 150 enumeration areas and 3,750 households in 2017; the 2012 MIS was conducted in the same enumeration areas as the 2010 MIS to facilitate longitudinal analyses, with 96 rural and 44 urban enumeration areas. The 2014 MIS, in contrast, had 90 rural and 50 urban enumeration areas, and the 2017 MIS had 90 rural and 60 urban enumeration areas [183, 184, 185]. Jittering was performed as for DHS [96].

Multiple Indicator Cluster Survey (2013-2014) Multiple Indicator Cluster Surveys (MICS) are conducted by national statistical offices, assisted by UNICEF, to monitor the health of women and children [186]. The fifth round (MICS5), which collected data on more than 130 internationally agreed-upon indicators [187], was conducted between 2013 and 2016 globally, and between 2013-2014 in Malawi.

The goal of MICS5 in Malawi was to produce reliable estimates at the national- and district-level for urban and rural areas, with the exception of Likoma. The sampling frame was the 2008 census, and sampling was stratified two stage cluster sampling, where strata were defined by urban vs. rural status and district.

The target sample was 1,050 households per district, powered for the key indicator "children under five who received anti-malaria treatment." The number of households per cluster was determined to be 25 households, based on a design effect estimated from previous studies, budget, and time considerations. Dividing the total number of households by number of

households per cluster, this generates a sample size of 42 clusters per district. These clusters were allocated equally across districts, with the exception of Blantyre and Lilongwe which were allocated 45 clusters each. Within each district, clusters were distributed to urban and rural strata proportional to the urban and rural population sizes within each district.

In stage 1, clusters were selected in each stratum using systematic probability proportional to size sampling, with size defined by number of households. Clusters with more than 300 households were split into 2 or 3 segments of roughly equal size divided by clear boundaries (e.g., roads or rivers), with only one segment randomly selected. A new household listing was then conducted in selected clusters, and households were selected using random systematic sampling. A further one in three households were selected for the men's questionnaire, in which a questionnaire was administered to individual men. The final sample was 1,140 clusters (163 urban, 977 rural) and 28,500 households [188].

A.2 Uganda

Demographic and Health Survey (2001, 2006, 2011, 2016) The Uganda Population and Housing Census serves as the sampling frame for all DHS rounds, with the 1991 census used for the 2001 DHS, the 2002 census used for the 2006 and 2011 DHS, and the 2014 census for the 2016 DHS [166]. Jittering was conducted as for Malawi [96].

All DHS rounds used stratified two-stage cluster sampling. Within strata defined by geography and urban/rural status, enumeration areas (clusters) were selected with probability proportional to size. Following a household listing activity, a fixed number of households were selected within each enumeration area with equal probability.

The 2001 DHS round oversampled urban clusters, as well as districts included in two ongoing studies (Delivery of Improved Services for Health project, and the Community Reproductive Health Project). Geographic strata were Central, Eastern, Northern, and Western regions, with four (Kasese and Bundibugyo in the Western Region, and Gulu and Kitgum in the Northern Region) out of 45 districts excluded for safety reasons. The final sample was comprised of 298 clusters (102 urban, 196 rural) and 14,672 households (5,880 urban, 8,792

rural) [189].

In the 2006 round, portions of the northern region were oversampled to provide estimates for Karamoja and internally displaced persons (IDP) camps. 321 clusters were selected from those sampled in the Uganda National Household Survey, plus 17 additional clusters in Karamoja, and 30 IDP camps (from among those compiled by UN OCHA). Each IDP camp was divided into segments and one segment randomly selected. In the second stage, 15-20 households were selected from each cluster sampled from the Uganda National Household Survey, 17 households were selected from each cluster in Karamoja, and 30 households were selected from each IDP camp segment. In total, 9,864 households were sampled (1,637 urban, 8,227 rural) [190].

The 2011 DHS sample also included clusters sampled by the Uganda National Household Survey. The final sample was comprised of 405 EAs (119 urban, 286 rural) and 10,086 households (3,570 urban, 8,580 rural) [191].

In the 2016 round, three regions (South Central, North Central, and Busoga) were stratified into island and non-island sub-regions, then into urban and rural strata, resulting in 34 total strata. The sampling frame was then sorted according to lower-level administrative units, resulting in implicit stratification. Large enumeration areas were segmented, with one segment selected, and 30 households per cluster subsequently sampled. In total, 667 clusters (162 urban, 535 rural) and 20,910 households (4,860 urban and 16,050 rural) were sampled [192].

Malaria Indicator Survey (2009, 2014-2015, 2018) As for the DHS, the sampling frame for all MIS rounds is the Uganda National Population and Housing Census, and stratified two-stage cluster sampling is used with strata defined by geography and urban/rural status. For the 2009 and 2014-2015 rounds the 2002 census was used, while the 2014 census was used for the 2018 rounds. Enumeration areas are generally defined as natural villages in rural areas and city blocks in rural areas, containing on average 80-100 households.

For the 2009 round, Uganda was divided into nine regions comprised of 8-10 contigu-

ous administrative districts (North-eastern, Mid-northern, West Nile, Mid-western, South-western, Mid-eastern, Central 1, Central 2, and East-central) and Kampala. In each region 17 clusters were selected (170 in total, 26 urban and 144 rural), with 28 households selected in each cluster (4,760 total) [193].

The 2014-2015 round used the same 10 geo-regions as the 2009 round. Within each stratum, the sampling frame was sorted by smaller administrative levels (district, sub-district, parish, village, and enumeration area code) resulting in implicit stratification by location. In each of 210 selected clusters, 28 households were selected. Oversampling was conducted in Karamoja and in three study domains defined by malaria endemicity [194].

In the 2018 round Uganda was divided into 15 regions and refugee settlements, with a mapping exercise conducted by UBOS in 2018 used as the sampling frame for the latter. Implicit stratification by location was performed as for 2014-2015; 320 clusters (84 urban, 236 rural) were selected from the main sampling frame, and an additional 22 from refugee settlements. Within each cluster, 28 households were selected [195].

National Panel Survey (2009-2010, 2010-2011, 2011-2012) The Uganda National Panel Survey is carried out annually over a 12-month period on a nationally representative sample of households. Each household is interviewed twice, six months apart, to capture agricultural outcomes associated with the wet and dry cropping seasons.

The sampling frame is the 2005/2006 Uganda National Household Survey. All clusters in Kampala district were selected, and 72 clusters (58 rural, 14 urban) were selected in each region (Eastern, Northern, Western, and Central minus Kampala), resulting in a total of 322 clusters. Clusters and households are selected with equal probability [196].

AIDS Indicator Survey (2011) The Uganda Ministry of health conducts the AIDS Indicator Survey in collaboration with its Health Development Partners, with the goal of obtaining national and regional prevalence estimates for HIV and syphilis, their risk factors, program coverage, and indicators of behavior, knowledge, and attitudes. The Demographic

and Health Surveys division at ICF International provides financial and technical assistance for this survey through a contract with USAID/Uganda.

As for the DHS and MIS, the AIS uses the Uganda National Population and Housing Census as the sampling frame, with the 2002 census used for the 2011 round. Also as for DHS and MIS, AIS uses stratified multi-stage cluster sampling with probability proportional to size sampling in stage 1 and equal probability sampling in stage 2. Stratification was by urban/rural status, and region: Central 1, Central 2, Kampala, East-central, Mid-eastern, North-eastern, West Nile, Mid-northern, South-western, and Mid-western. The sample of 470 clusters (79 urban, 391 rural) was allocated equally across these regions. In each cluster, 25 households were sampled, with a total of 11,750 households (1,975 urban, 9,775 rural) sampled in total [197].

A.3 DRC

Demographic and Health Survey (2007, 2013-2014) The last census conducted in DRC was in 1984, however it has been partially updated on several occasions by administrative and electoral censuses. Imputed population data based on a 2003 update was used as the sampling frame for the 2007 DHS, and imputed data based on a 2010 update used as the sampling frame for the 2013-2014 DHS. Both rounds produced representative results for the entire country, for urban and rural areas, and for each of the eleven provinces. The 2013 round also provided representative results for each of the 26 new provinces formed. Both rounds adopted a stratified cluster sampling approach, implemented in two stages for statutory (major) cities and three stages otherwise. First, each province was stratified into three strata: statutory city, city, and rural. In the 2007 round, Kinshasa was instead stratified by its four districts as it has only one statutory city, resulting in 34 strata across 11 provinces. As not all of the provinces formed between the 2007 and 2013 rounds are represented by all three strata, in the 2013-2014 round there were 66 total strata across 26 provinces.

Primary sampling units (districts in statutory cities, sectors and chiefdoms in rural clusters, and cities otherwise) were selected with probability proportional to size. In statutory

cities, households were selected in the second stage. In cities and rural clusters, villages and neighborhoods in cities, respectively, were selected in the second stage with equal probability, and households in the third stage (30 per cluster in 2007, 34 per cluster in 2013-2014). In the 2013-2014 round, large neighborhoods (population >2,000) were segmented and one segment selected, and in very large sectors two to three villages were selected (rather than just one) per sector, resulting in a cluster corresponding to a neighborhood or neighborhood segment in urban strata, and a village or group of villages in rural strata.

Prior to selection of primary sampling units, the sampling frame was sorted within strata by municipality/territory, and then by population size to generate implicit stratification. Prior to household sampling, a household listing activity was conducted; due to challenges to fieldwork in DRC, this activity was done at the same time as data collection. In the 2007 round urban clusters (with the exception of Kinshasa) were oversampled.

In the 2007 round 300 sample clusters (123 urban, 177 rural) were sampled, comprised of 9,000 households (3,690 urban and 5,310 rural). In the 2013-2014 round 540 clusters were sampled, comprised of 18,360 households (5,474 urban and 12,886 rural) [198, 199].

Multiple Indicator Cluster Survey (2010) Sampling approaches implemented by MICS are detailed above as for Uganda. In the 2010 round conducted in DRC, the 2010 updated population data was used as the sampling frame (i.e., the base from which population data were imputed for the 2013-2014 DHS round). Provinces were stratified as for DHS, and samples were drawn from the 11 provinces with probability proportional to population size. A total of 11,490 households were sampled across 383 clusters, with 30 households per cluster (32 from cities, 115 from townships, and 236 from rural areas) [200].

A.4 South Sudan

Population and Housing Census (2008) The 2008 Population and Housing Census represents the fifth population census conducted in Sudan. Presidential Decree No (2) established the Population Census Council on 7th January 2006, which was comprised of

representatives from the Government of National Unity (GNU), the Government of South Sudan (GOSS), the Parliament, Senate, and Academics. A similar council, South Sudan Population Census Council (SSPCC) was also established in the south against the advice of the Council [201]..

British maps from 1930 were used to make basic reference maps for state boundaries. Geocoding levels were as follows: state; locality; administrative unit, town, city; populated administrative area; enumeration area; city quarter, block, village, farigue [201], with each enumerator assigned one enumeration area [202]. Detail on selection of enumeration areas is not provided in published reports.

The IPUMS International extract provides census data on 542,765 individuals residing in 72 counties in South Sudan; these were the data used for our analyses [76].

Appendix B

WEALTH MAPPING

B.1 Data collection and processing

In Malawi, Uganda, and DRC, we used several additional data sources for wealth mapping that contained data on the input variables needed for wealth mapping, but did not contain data on livestock ownership and thus could not be utilized for livestock mapping (Table 2.1). There were 10,330 clusters in the final dataset for Malawi, 4,361 in Uganda, and 1,164 in DRC; these are broken down by year in Table B.1.

For Malawi, Uganda, and DRC, as with data processing for our livestock maps we processed each survey separately. After reading in data, we harmonized variables so that the wealth index could be calculated across surveys, allowing for smoothing in time.

B.2 Generating the wealth index

Our process for wealth mapping followed that laid out in the DHS Wealth Index [115] handbook, with the exception of exclusion of livestock-related variables in our wealth index. The final variables included in our wealth models were, defined at the level of the household (M: Malawi, U: Uganda, S: South Sudan, D: DRC):

- Presence of a domestic servant *(M, U, D)*
- Ownership of agricultural land *(M, U, S, D)*
- Amount of land owned (converted to acres) *(M, U, S, D)*
- Ownership of dwelling *(M, U, S, D)*
- Type of fuel used for cooking *(M, U, S, D)*
- Water source *(M, U, S, D)*

Table B.1: Data availability (number of clusters) by year for wealth mapping

Year	Malawi	Uganda	DRC
2000	560	0	0
2001	0	266	0
2002	0	0	0
2003	3,404	0	0
2004	640	0	0
2005	0	0	0
2006	142	336	0
2007	0	0	293
2008	0	0	0
2009	0	776	0
2010	1,595	567	361
2011	0	1281	0
2012	140	0	0
2013	1,582	0	492
2014	140	222	0
2015	0	0	0
2016	1,978	685	0
2017	148	0	0
2018	0	227	0
2019	0	0	0
2020	0	0	0

- Toilet type *(M, U, S, D)*
- Whether toilet is shared with other households *(M, U, D)*
- Main floor material of dwelling *(M, U, D)*
- Main material of dwelling walls *(M, U, D)*
- Roof material *(M, U, D)*
- Presence of electricity *(M, U, D)*
- One or more household members has a bank account *(M, U, D)*
- Vehicle (none; animal cart or bicycle; motorcycle or scooter; car, mini-bus or truck/lorry) *(M, U, D)*
- Number of rooms per household member *(M, U, S, D)*
- Household owns one or more:
  - Chairs *(M, U, D)*
  - Table *(M, U)*
  - Clock or watch *(M, U, D)*
  - Bucket *(M)*
  - Clothes washing machine *(M)*
  - Dish washing machine *(M, U, D)*
  - Refrigerator *(M, U, S, D)*
  - Mobile phone *(M, U, S, D)*
  - Landline (non-mobile phone) *(M, U, S, D)*
  - Television *(M, U, S, D)*
  - Computer *(M, U, S, D)*
  - Radio *(M, U, S, D)*
  - Sewing machine *(M, D)*
  - Bed *(M, U, D)*
  - Upholstered couch, sofa, chairs, or set *(M, U)*
  - Paraffin or kerosene lamp *(M, U, D)*
  - Boat *(M, U, S, D)*
  - Fishing net *(M)*

  - Mortar and pestle *(M)*
  - Fan *(M, S*
  - Air conditioning *(M)*
  - Machine to play cassette tapes, CDs, DVDs, or Hi-Fi *(M, U)*
  - VCR *(Ms)*
  - Kerosene or paraffin stove *(M)*
  - Electric or gas stove/hot plate *(M, D)*
  - Beer-brewing drum *(M)*
  - Coffee table for sittingroom *(M)*
  - Cupboards, drawers, bureau *(M, U)*
  - Desk *(M)*
  - Iron for pressing clothes *(M)*
  - Satellite dish *(M, S)*
  - Solar panel *(M, U)*
  - Generator *(M, U, D)*
  - Plough *(M)*
  - Tractor *(S)*
  - Axe or hoe *(D)*

- Proportion of possessions enumerated by the survey in question that the household owns *(M, U, D)*

We coded all variables such that a lower level corresponded to lower wealth, and a higher lever corresponded to greater wealth (e.g., we used the variable number of rooms per household member rather than number of household members per room, as the latter would have a negative association with wealth). We assigned levels to categorical variables as detailed in Tables B.2-B.7; note apparent repetition within variable levels reflect different categorizations across surveys.

Factor analysis For Malawi, Uganda, and DRC, we performed exploratory factor analysis separately for each of three periods, to reflect the expectation that factor loadings would

Table B.2: Levels for cooking fuel

Level	Fuel types
	Malawi, Uganda., DRC
4	Electricity
3	Gas, kerosene, liquefied petroleum gas, biogas, natural gas, paraffin, oil, solar
2	Charcoal, coal, lignite
1	Wood, firewood (collected or purchased)
0	Dung, straw, shrub, grass, crop residue, saw dust
	South Sudan
4	Electricity
3	Gas
2	Charcoal
1	Firewood
0	Cow dung, Grass

Table B.3: Levels for water source

Malawi, Uganda, DRC	
Level	Water source
4	Piped into own dwelling, compound, yard or plot
3	Public or communal standpipe/tap, piped into neighbor's dwelling
2	Private (in own dwelling, yard or plot) or public protected well, tube well, borehole, protected spring, rain tank or rainwater, channeled by gravity flow scheme
1	Tanker truck/bowser, cart with small tank, bottled or sachet water, water vendor
0	Unprotected well or spring, private (in own dwelling, yard or plot) or public open well, river, dam, lake, pond, stream, canal, irrigation channel
South Sudan	
3	Water filtering stations with common standpipe, mechanical boreholes with common standpipe, sand filter with common standpipe,
2	Deep borehole with network, deep borehole without network, hand pump, water vendor from deep borehole
1	Shallow well, hafeer/dam with filter, water vendor from shallow well/pond/river/spring
0	Hafeer/dam without filter, still (turda/fula/river) or running (river/pond/tura'a) open water source

Table B.4: Levels for toilet type

Level	Toilet type
	Malawi, Uganda, DRC
3	Flush toilet (piped, pour flush, or unspecified)
2	VIP latrine, composting toilet, Ecosan toilet, latrine with slab and/or roof
1	Latrine without slab or roof, public pit toilet or latrine, bucket, hanging toilet
0	No facilities/open, bush, field
	South Sudan
5	Private flush toilet
4	Shared flush toilet
3	Private pit latrine
2	Shared pit latrine
1	Bucket
0	No facilities

Table B.5: Levels for dwelling floor

Level	Floor material
	Malawi, Uganda, DRC
2	Carpet, parquet/polished wood, ceramic or other tile or mosaic, cement/concrete, smooth cement, vinyl/asphalt strips, linoleum, stone, bricks
1	Palm/bamboo, wood planks, broken bricks
0	Earth, sand, dung, smoothed mud

Table B.6: Levels for dwelling walls

Level	Wall material
	Malawi, Uganda, DRC
3	Cement, concrete, stone with lime/cement or mud, finished/burnt brick with stone or mud, covered adobe
2	Wood, wood planks/shingles, reused wood, plywood, plastic sheet, metal sheet, corrugated iron
1	Mudbrick, unburnt bricks with plaster, cement, or mud, uncovered adobe
0	Thatch, straw, grass, cane/palm/trunks, bamboo or poles with mud, timber, mud, dirt, compacted earth, cardboard, none

Table B.7: Levels for dwelling roof

Level	Roof material
	Malawi, Uganda, DRC
3	Roofing shingles, ceramic tiles, clay tiles, cement, concrete, calamine/cement fiber, wood, asbsetos
2	Iron sheets, metal, tin
1	Wood planks, cardboard, palm/bamboo grass, rustic mat, plastic/polythene sheeting, tin cans
0	Thatch/palm leaf, grass, sod, mud/earth, none

change over time (e.g., in early periods ownership of a mobile phone would be much more strongly associated with wealth than in later periods): 2000-2006, 2006-2012, and 2012-2020. For South Sudan, as with livestock mapping, wealth mapping was only conducted for 2008.

First, we removed all variables with more than 5% missingness, and all variables with standard deviation equal to 0. We then created a correlation matrix using the `mixedCor` function in the `psych` package in R [203], and performed exploratory factor analysis using the `fa()` function in the same package, setting our arguments to those specified by the DHS Wealth Index manual: principal components extraction with one factor extracted, imputation of mean for missing data, and estimation of the factor scores using the regression method. We performed linear regression of the extracted scores on input variables as a sanity check. For all four countries, in general only variables with very large levels of missingness had negative associations with our final wealth score. In all countries, ownership of land was negatively associated with our final wealth score and had high levels of missingness (17% missing in Malawi, 25% in South Sudan, 15% in Uganda, and 31% in DRC). Similar results were found for ownership of dwelling (negative association in Malawi with 87% missing, South Sudan with 0% missing, DRC with 93% missing), area of land owned (negative association in South Sudan with 25% missing, Uganda with 92% missing, and DRC with 93% missing), and household sharing toilet with other households (negative association in Uganda with 12% missing, DRC with 15% missing). In DRC, negative associations were also found for ownership of a canoe, axe, or hoe, with 0% missing for these variables, reflecting absence of a strong association between ownership of these items and wealth. All other variables had positive associations with the final wealth score.

As the goal was to adjust for wealth score in final regression models, we did not perform factor analysis separately for urban and rural clusters in Malawi, Uganda, or DRC, however we use urban/rural status as a predictor in our SPDE models to reflect both the expected association between this variable and wealth score, and the stratified design of the surveys we used (Appendix A).

B.3 Wealth mapping

Malawi, Uganda and DRC After the wealth index was constructed, we collapsed over cluster by taking the mean wealth score in each cluster.

Our predictors included urban/rural status to reflect sampling strategy (Appendix A), as detailed above, and nighttime lights. Data on nighttime lights came from the National Oceanographic and Atmospheric Administration's Nighttime Lights Time Series, which are annual cloud-free composites made from archived DMSP-OLS data, available at a resolution of 30-arc-seconds. These data come from Operational Linescan System instruments onboard the Defense Meteorological Satellite Program satellites, which collect visible and infrared imagery to monitor the global distribution of cloud and cloud top temperatures twice daily [80]. These data are available from 1992-2013, as average visible lights, stable lights, and a normalized version of average lights. We used average visible lights, and for model fitting and prediction for 2014-2020 we used data from 2013.

As for livestock mapping, we used the SPDE approach to map wealth in Malawi, Uganda, and DRC. We used the same mesh and an equivalent random effects structure, but a Gaussian model rather than a ZIP. Accordingly, while we used the same priors, the interpretation is different as the link is now an identity link. This yields a posterior 95% credible interval of (-0.52, 0.52) for each standard deviation (σ) for the iid random effects, and of (-3,3) for the marginal standard deviation σ_s of the spatial random effect. Also as for the livestock maps, we fit a series of three models differing only in their fixed effects (model 1: intercept only; model 2: intercept and urban/rural; model 3: intercept, urban/rural, and nightlights) and a fourth model which replaced the RW2 in time with an RW1.

For prediction we used the same grid and the same approach, with the exception of exponentiation at the final step. We performed model selection via leave one out cross-validation, as for livestock mapping. For all three countries, model 4 performed best.

Finally, for external validation, we performed spatial regression to check for association between our final wealth index and the proportion of the population earning under $2 per day

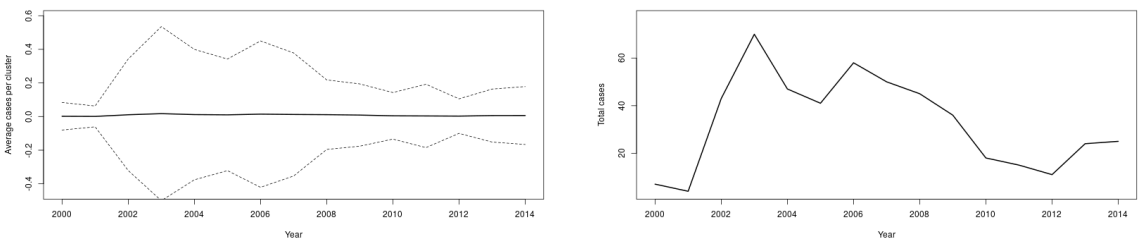
in 2010 produced by WorldPop, which use Bayesian model-based geostatistics and household survey data from the LSMS program, and are available at a resolution of 0.00833° [204]. We found that a one unit increase in proportion earning less than $2/day (i.e., going from 0 to 1) was associated with a 0.84 lower wealth score as predicted by our final Malawi model and a 0.87 lower wealth score as predicted by our final Uganda model, indicating good agreement between our wealth maps and the WorldPop poverty maps. WorldPop does not produce these estimates for DRC.

South Sudan In South Sudan, as for livestock mapping, we produced direct estimates of county-level wealth and design-based variance using the `svydesign()` and `svyby()` functions in the `survey` package. Again, we used household weights provided in the IPUMS extract. We did not perform smoothing as our primary goal was not to estimate wealth, thus stabilizing variance of this estimate was not critical.

Appendix C

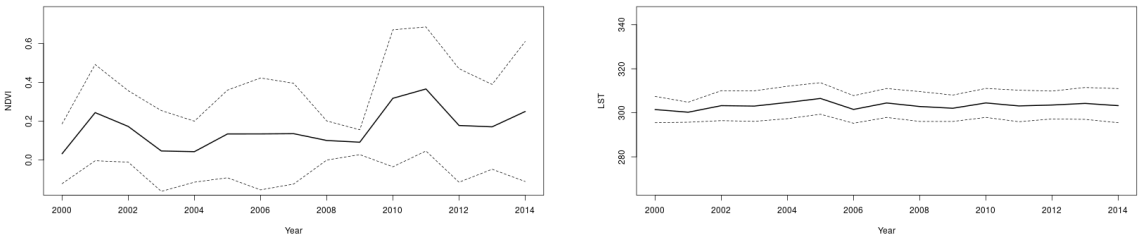
DESCRIPTIVE STATISTICS PLOTS

C.1 Malawi



(a) (b)

Figure C.1: Mean (a) and sum (b) of cases over time in study clusters, Malawi



(a) NDVI (b) LST

Figure C.2: NDVI (a) and LST (b) over time in study clusters, Malawi

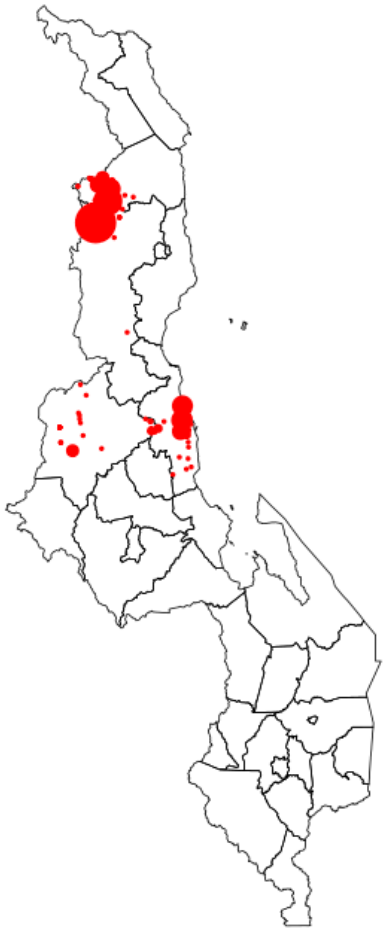
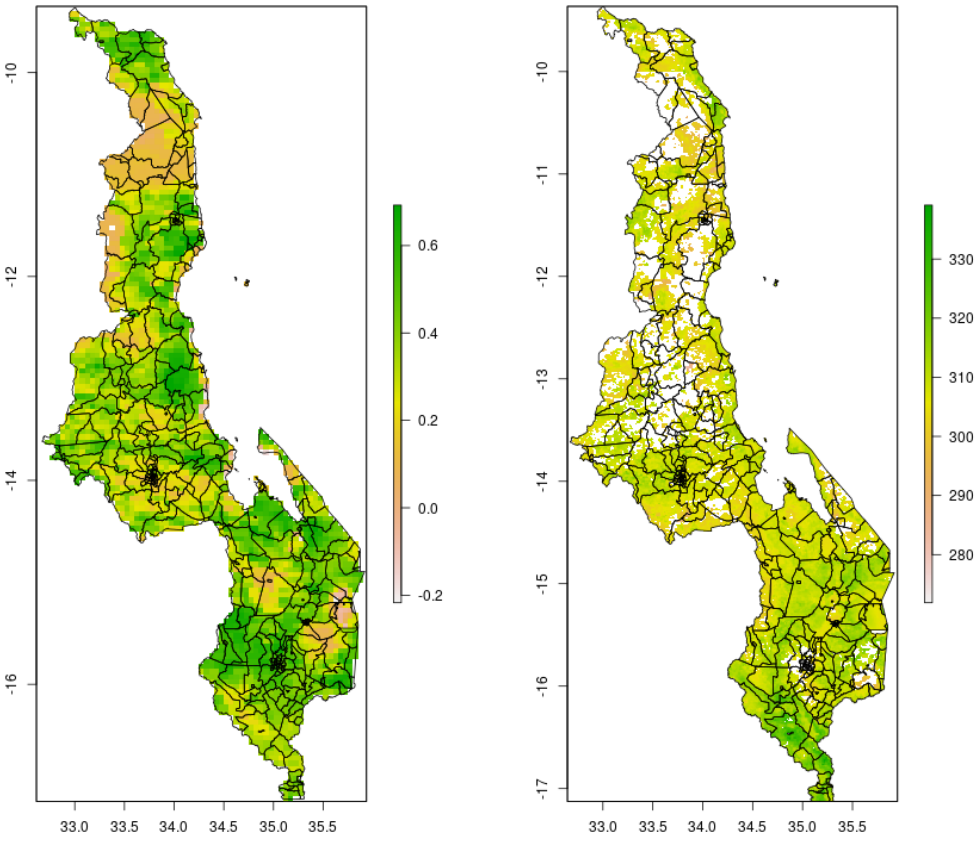


Figure C.3: HAT cases 2000-2014, Malawi



(a) (b)

Figure C.4: 2010 NDVI (a) and LST (b), Malawi

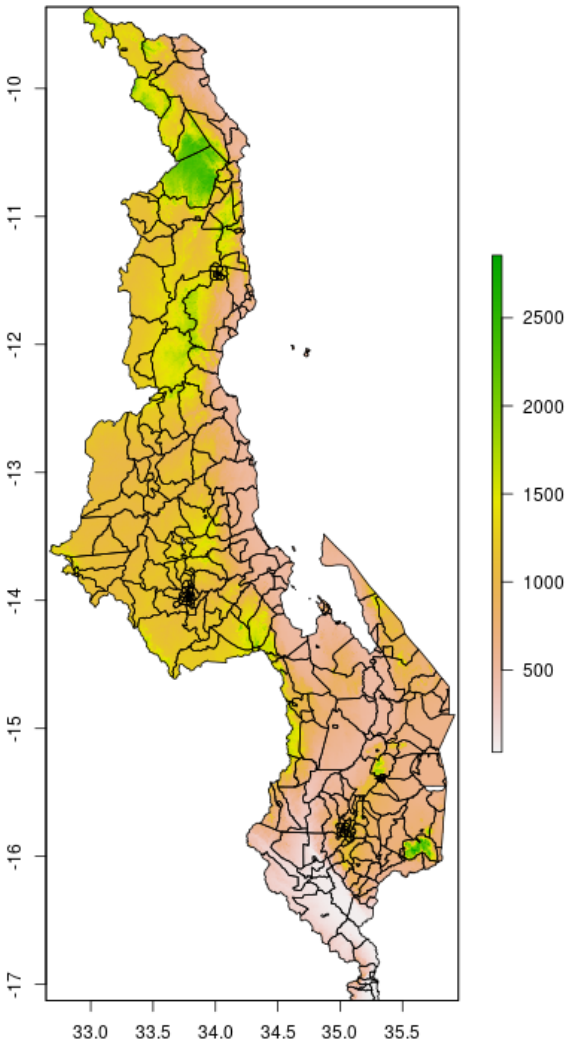
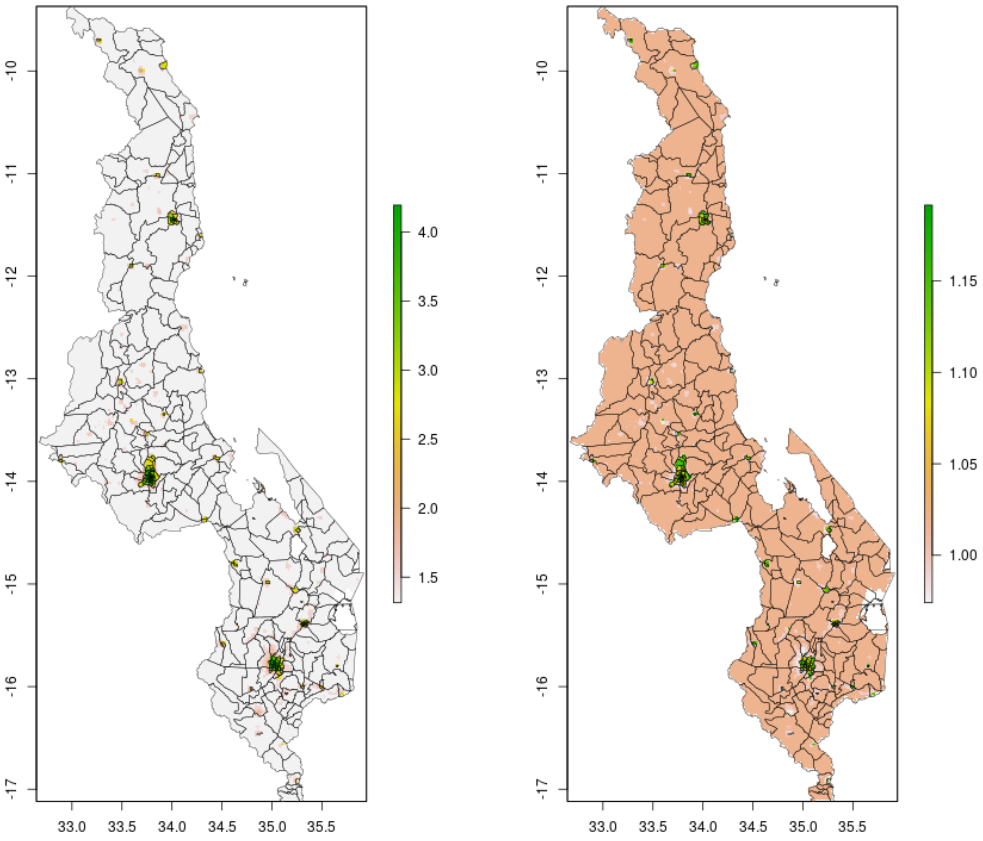


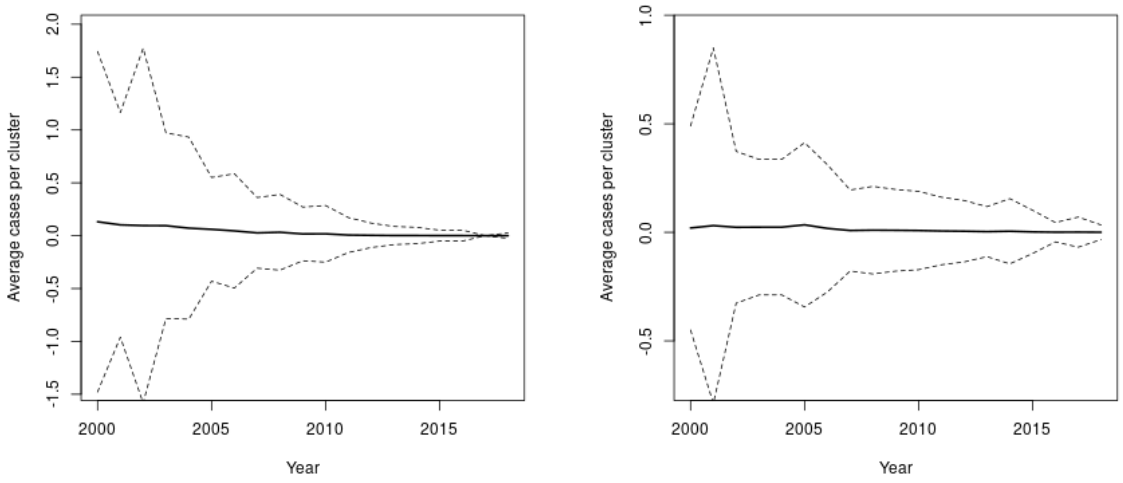
Figure C.5: Elevation, Malawi



(a) (b)

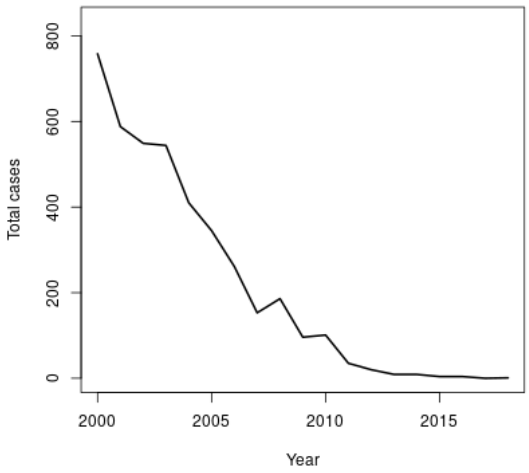
Figure C.6: 2010 wealth scores, mean (a) and posterior 95% credible interval (b), Malawi

C.2 Uganda

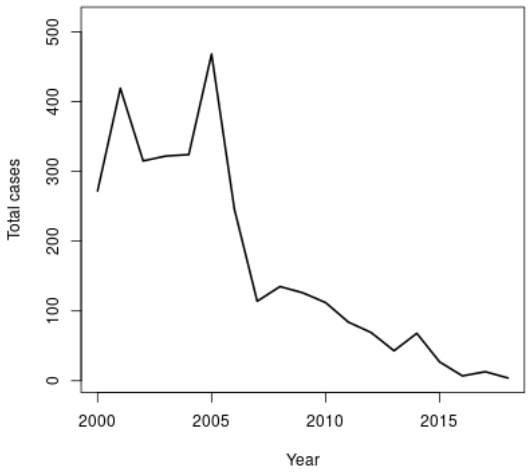


(a) gHAT

(b) rHAT

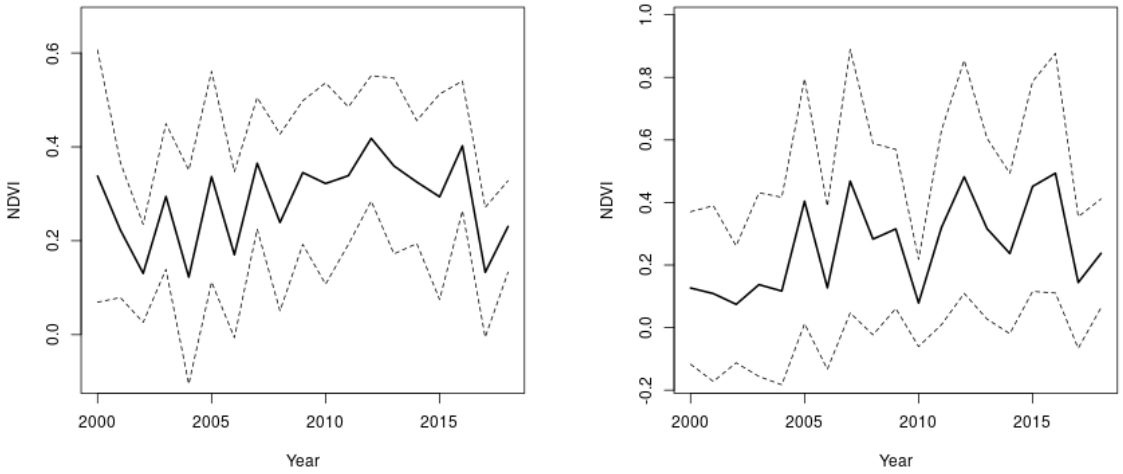


(c) gHAT



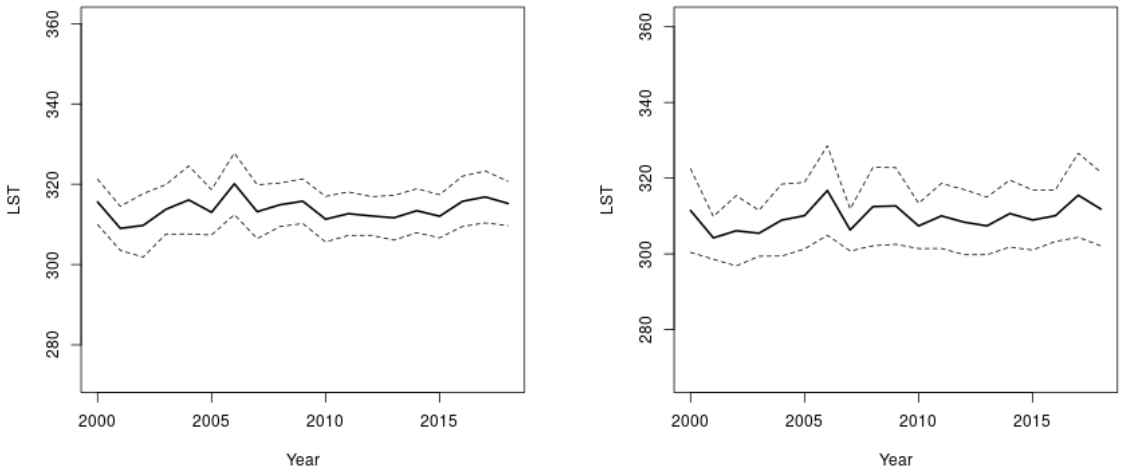
(d) rHAT

Figure C.7: Mean (a-b) and (c-d) of cases over time in study clusters, Uganda



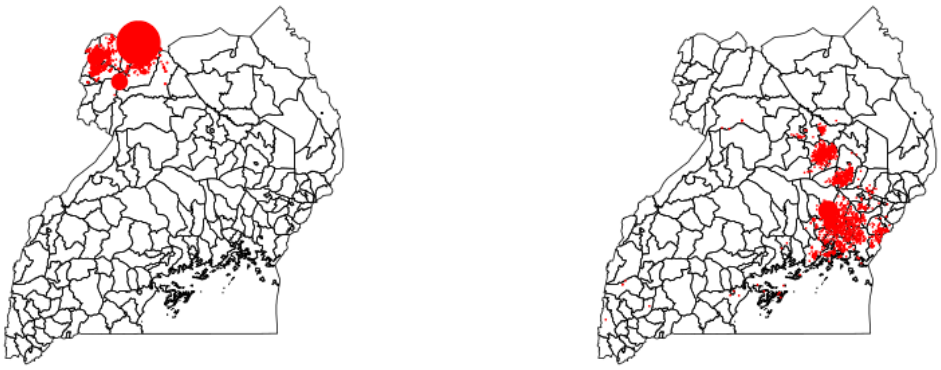
(a) gHAT (b) rHAT

Figure C.8: NDVI over time in study clusters, Uganda



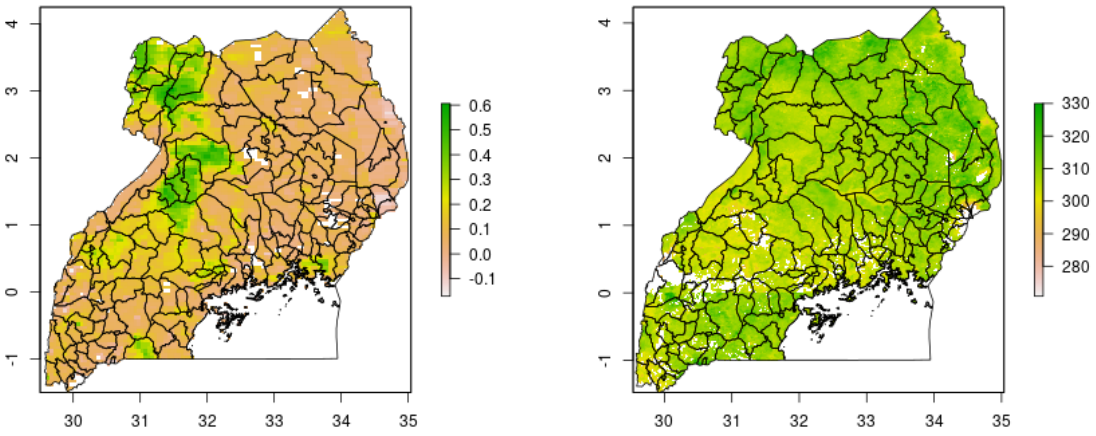
(a) gHAT (b) rHAT

Figure C.9: LST over time in study clusters, Uganda



(a) gHAT (b) rHAT

Figure C.10: HAT cases 2000-2018, Uganda



(a) (b)

Figure C.11: 2010 NDVI (a) and LST (b), Uganda

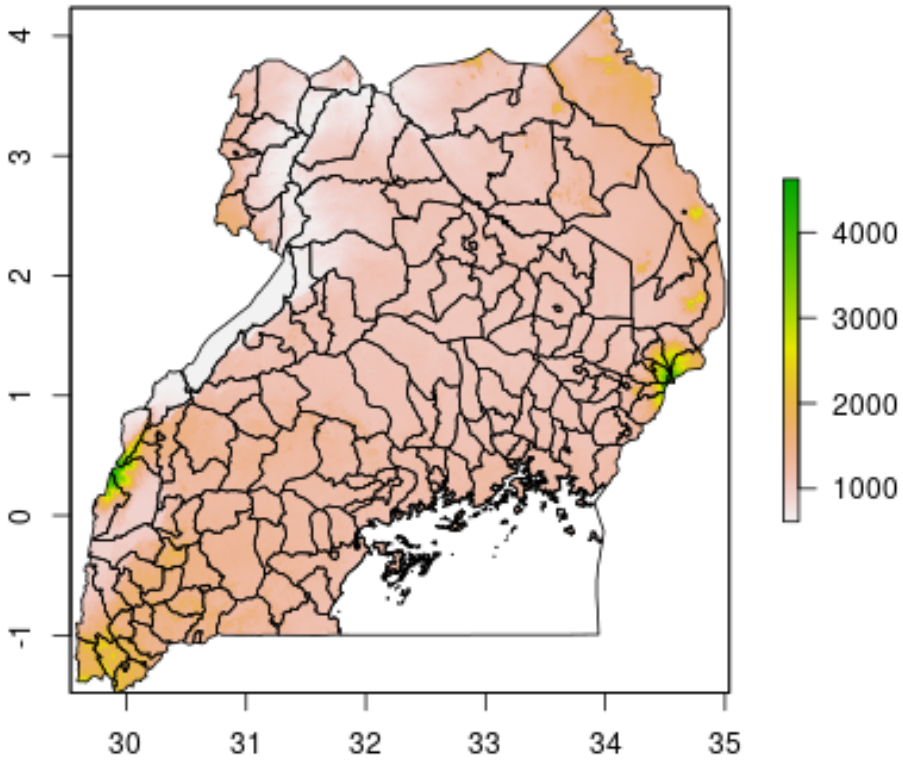
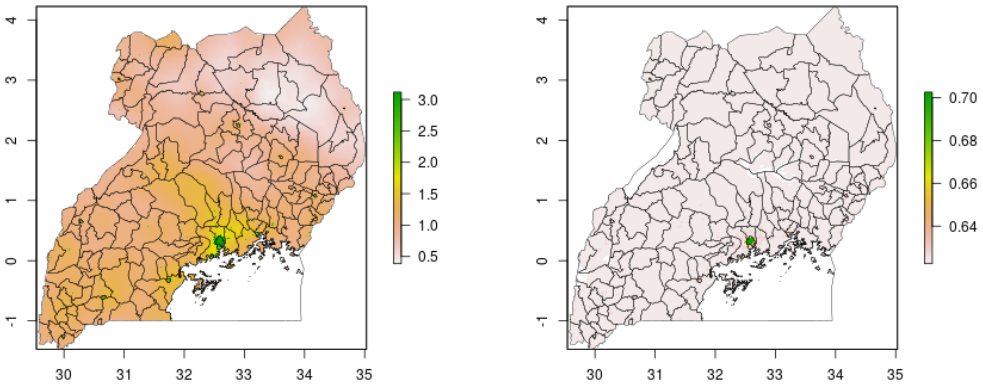


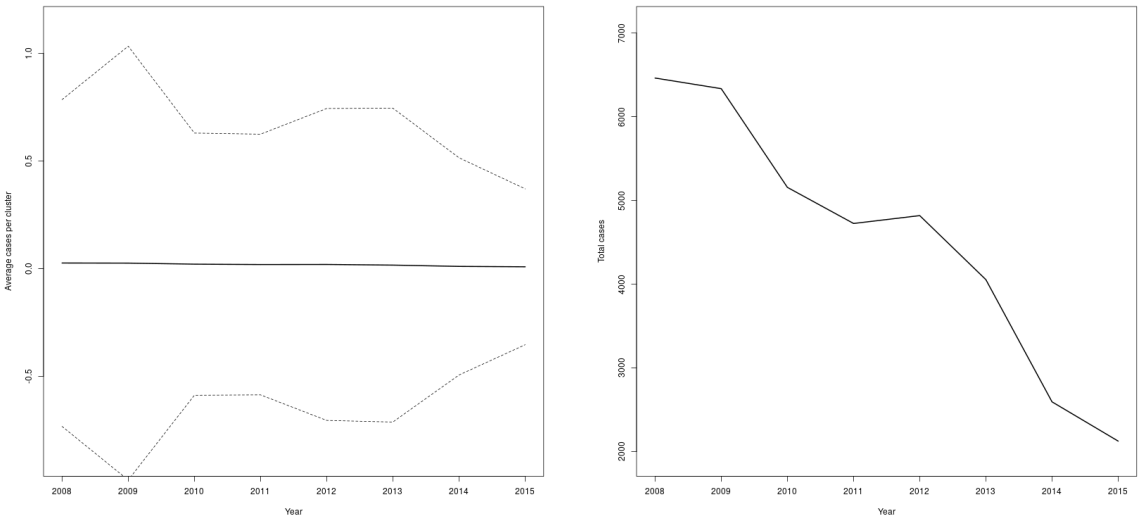
Figure C.12: Elevation, Uganda



(a) (b)

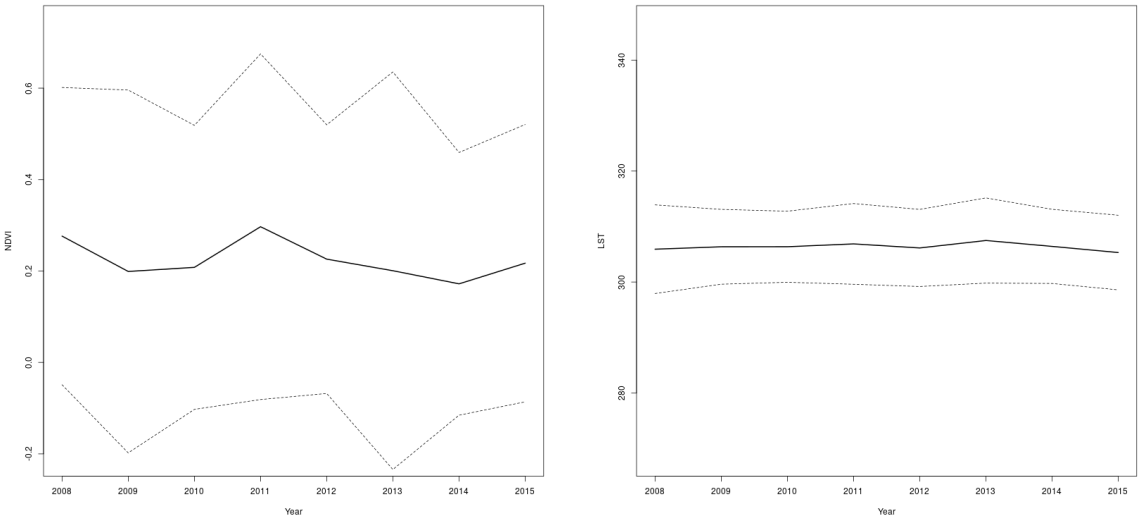
Figure C.13: 2010 wealth scores, mean (a) and posterior 95% credible interval (b), Uganda

C.3 DRC



(a) Mean (b) Sum

Figure C.14: Mean (a) and sum (b) of cases over time in study clusters, DRC



(a) NDVI (b) LST

Figure C.15: NDVI (a) and LST (b) over time in study clusters, DRC

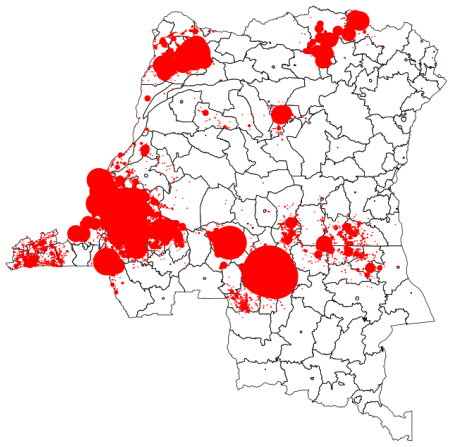
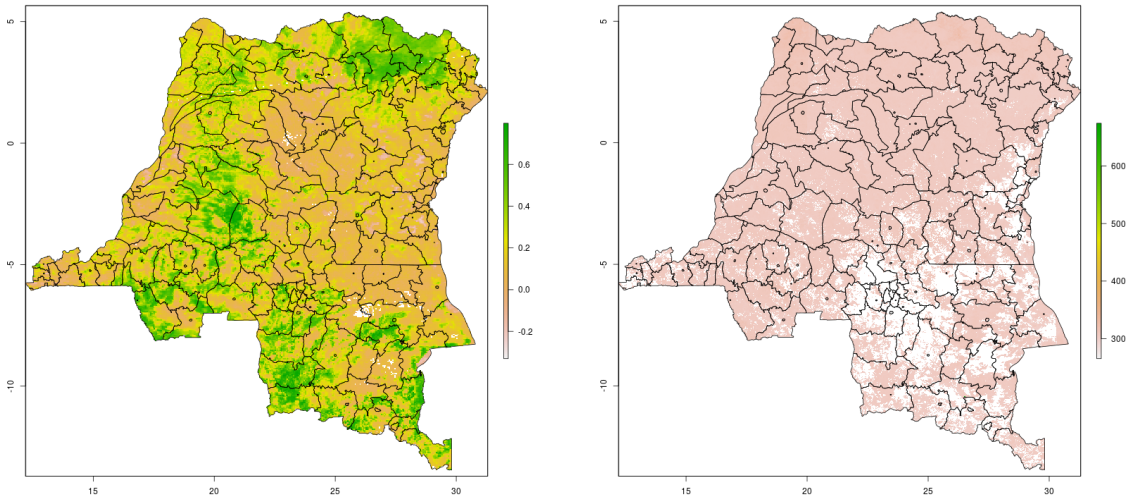


Figure C.16: HAT cases 2008-2015, DRC



(a) (b)

Figure C.17: 2010 NDVI (a) and LST (b), DRC

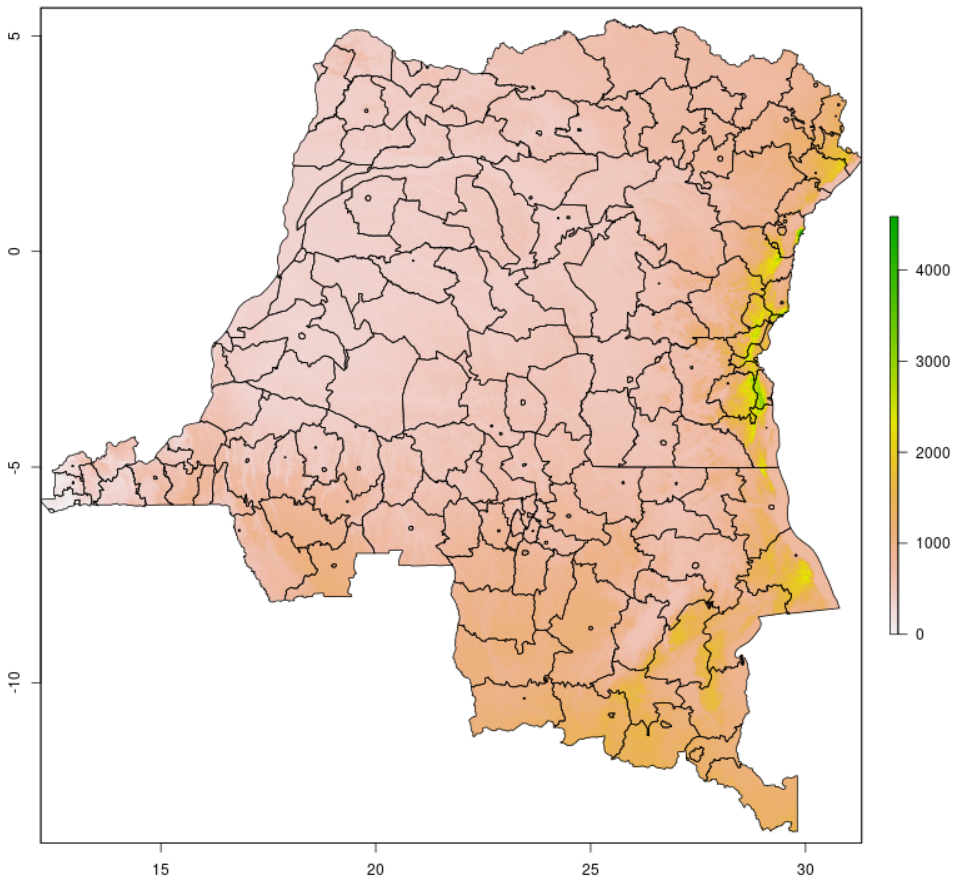
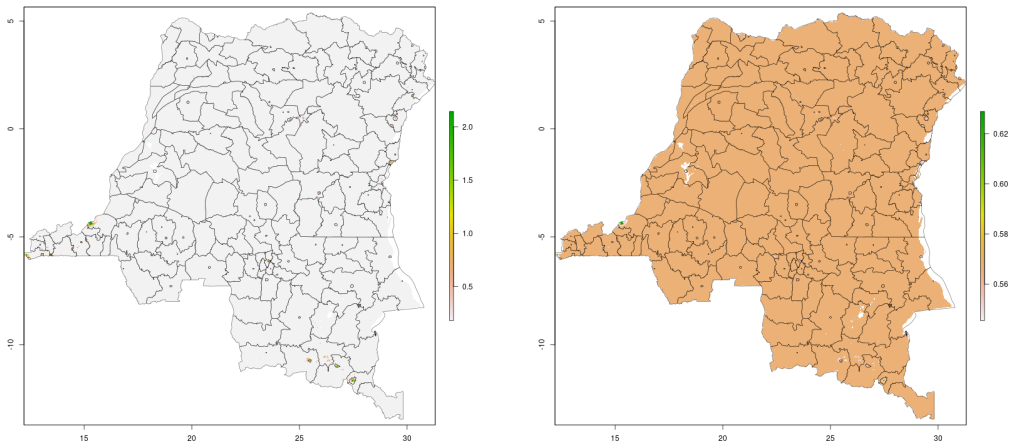


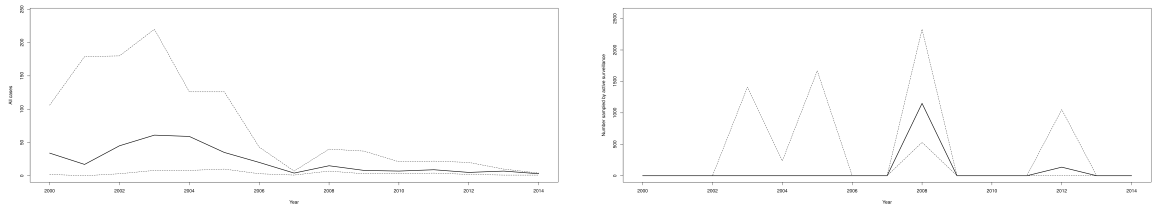
Figure C.18: Elevation, DRC



(a) (b)

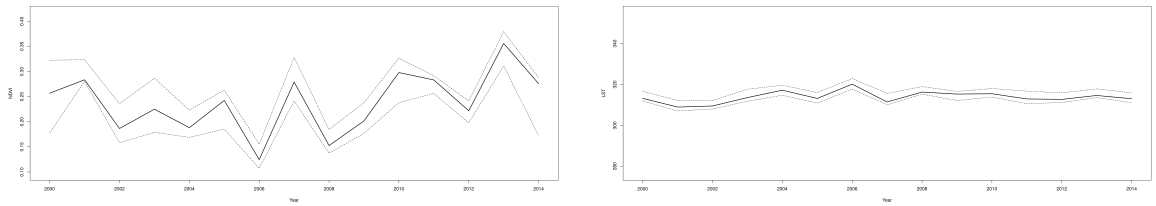
Figure C.19: 2010 wealth scores, mean (a) and posterior 95% credible interval (b), DRC

C.4 South Sudan



(a) (b)

Figure C.20: Mean number of cases (a) and people sampled by active surveillance (b) over time in study counties, South Sudan



(a) NDVI (b) LST

Figure C.21: NDVI (a) and LST (b) over time in study counties, South Sudan

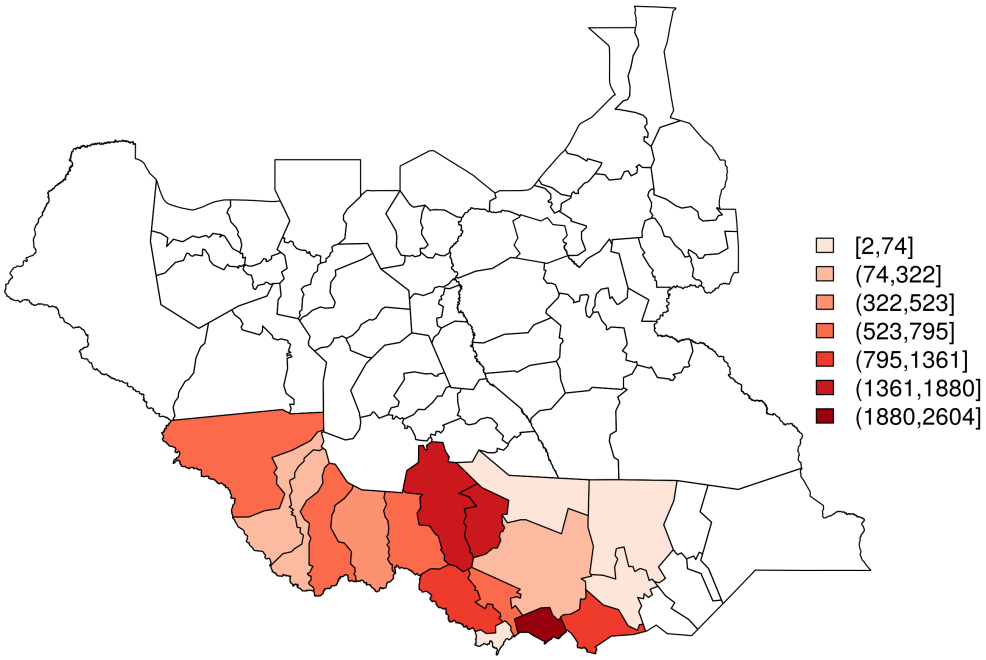
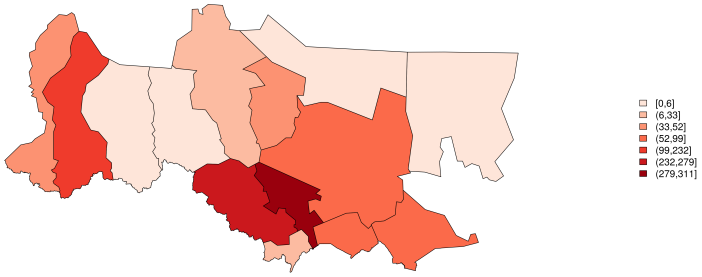
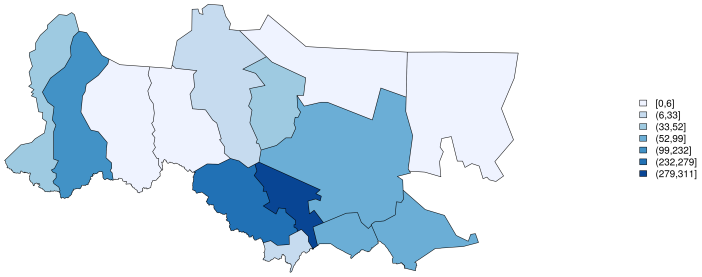


Figure C.22: HAT cases 2000-2014, South Sudan

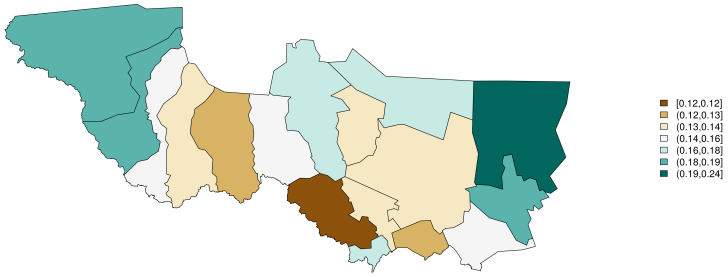


(a) Cases detected

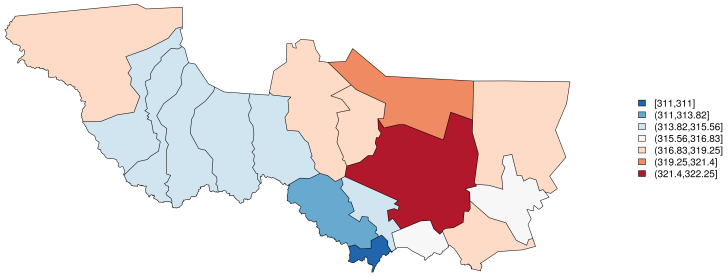


(b) Number sampled

Figure C.23: Total cases detected (a) and number sampled (b) by active surveillance in South Sudan, 2008 (restricted to active surveillance area)



(a) NDVI



(b) LST

Figure C.24: 2008 NDVI (a) and LST (b), South Sudan (study area)

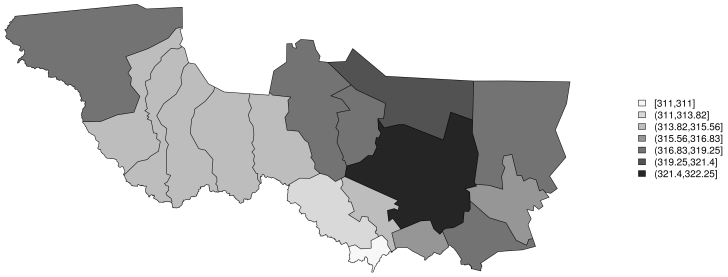
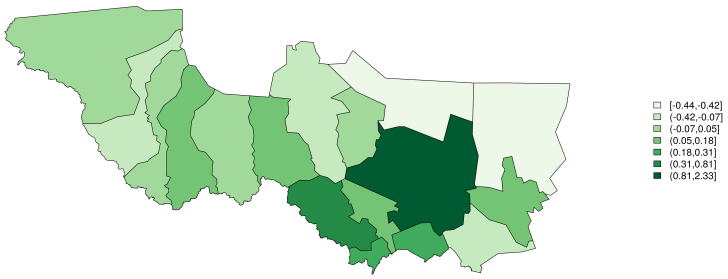
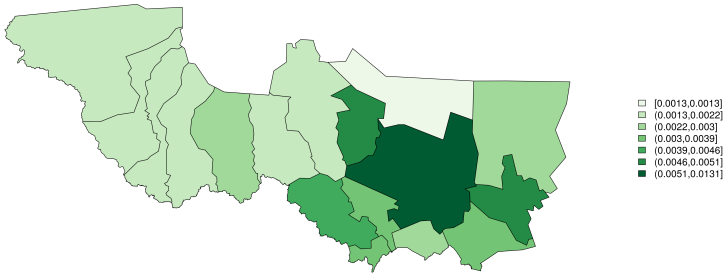


Figure C.25: Elevation, South Sudan (study counties)



(a) Median



(b) Standard error

Figure C.26: 2008 wealth, median (a) and standard error (b), South Sudan (study counties)

Appendix D

MODEL PARAMETERS

Table D.1: Model parameterization, DRC

Species	Parameter	Value
Humans	Life expectancy, 2005 ($1 / \mu_H$)	53.7 years [205]
	Population, 2005, B/S ($N_H(0)$)	15,082 [206][1]
	Population, 2005, EN ($N_H(0)$)	13,191 [206][1]
	Incubation period ($1 / \sigma_H$)	12 days [159]
	Duration of stage 1 ($1 / \phi_{H_1}$)	526 days [207][2]
	Duration of stage 2 ($1 / \phi_{H_2}$)	182.5 days [208][3]
	Duration of immunity ($1 / \omega_H$)	50 days [159]
	Stage 1 infecteds, 2005, B/S ($I_{1H}(0)$)	10,790 [161, 207, 208][4]
	Stage 1 infecteds, 2005, EN ($I_{1H}(0)$)	1,595 [161, 207, 208][4]
	Stage 2 infecteds, 2005, B/S ($I_{2H}(0)$)	3,744 [161, 208, 9][5]
	Stage 2 infecteds, 2005, EN ($I_{2H}(0)$)	553 [161, 208, 9][5]
	Probability of detection, stage 1 (τ_{H_1})	0.25 [9][6]
	Probability of detection, stage 2 (τ_{H_2})	0.25 [9]
	Probability of fly→human transmission, B/S (β_H)	0.01 (fitted)
	Probability of fly→human transmission, EN (β_H)	0.01 (fitted)
Reservoirs**	Lifespan ($1 / \mu_R$)	2 years [209]
	Reservoir:human ratio (N_{Rit} / N_{Ht})	0.06†
	Incubation period ($1 / \sigma_R$)	7 days [210]
	Duration, infectious stage w/o treatment ($1 / \phi_{R_{nt}}$)	182.5 days [211]

Table D.1: Model parameterization, DRC

Species	Parameter	Value
	Duration, infectious stage w/ treatment ($1 \,/\, \phi_{R_t}$)	91.25 days[*7]
	Duration of immunity ($1 \,/\, \omega_R$)	50 days [159]
	Initial prevalence ($I_R(0)$)	0.2847 [58][8]
	Probability of trypanocide treatment (τ_R)	0[*]
	Probability of insecticide treatment (itc_R)	0[*]
	Probability of fly→reservoir transmission (β_R)	0.1345 [6]
Flies	Lifespan, *G. f. fuscipes* ($1 \,/\, \mu_V$)	26 days [130]
	Lifespan, *G. f. quanzensis* ($1 \,/\, \mu_V$)	29 days [212]
	Fly:human ratio, B/S ($N_{Vt} \,/\, N_{Ht}$)	6.56 [155]
	Fly:human ratio, EN ($N_{Vt} \,/\, N_{Ht}$)	6.56 [155]
	Incubation period ($1 \,/\, \sigma_V$)	18 days [213]
	Initial prevalence, *G. f. fuscipes* ($I_V(0)$)	0.0068 [130, 214][9]
	Initial prevalence, *G. f. quanzensis* ($I_V(0)$)	0.024 [158]
	Probability of feeding in the first 24 hours (α)	1/3 [99]
	Probability of human→fly transmission (β_{Vh})	0.212 [215][10]
	Probability of reservoir→fly transmission (β_{Vr})	0.212 [215][10]
	Probability of human bite (p_H)	0.42 [216]
Non-resevoirs	Ratio to humans	0.1[*]

Table D.1: Model parameterization, DRC

Species	Parameter	Value

*Assumed; **Domestic swine; †Our SPDE maps; t indexes time; B/S: Bandundu/Sakuru; EN: Equateur Nord. [1]National population at risk weighted by proportion of population each focus ("census" variable in Atlas data). Used 2010-2014 population at risk estimates as otherwise reported cases would exceed population at risk in Bandundu and Sankuru; [2]Mean value used; [3]Parameterized as 6 months; [4]Taken as the product of the number of stage 2 cases and the stage 1 : stage 2 duration ratio; [5]Total reported cases adjusted to reported stage 2 cases using national proportions in stage 2 (Atlas data), then adjusted for probability of detection in stage 2, then derived cases on a given day using stage 2 duration data; [6]50% of cases detected, 50% of these via active surveillance; [7]Parameterized as 3 months; [8]Results from trypanolysis test; [9]Mean taken over the two cited papers; [10]Mean of male and female flies, overall % infected

Table D.2: Model parameterization, South Sudan

Species	Parameter	Value
Humans	Life expectancy, 2005 ($1 \,/\, \mu_H$)	51.7 years [205]
	Population, 2005 ($N_H(0)$)	2,397 [206][1]
	Incubation period ($1 \,/\, \sigma_H$)	12 days [159]
	Duration of stage 1 ($1 \,/\, \phi_{H_1}$)	526 days [207][2]
	Duration of stage 2 ($1 \,/\, \phi_{H_2}$)	182.5 days [208][2]
	Duration of immunity ($1 \,/\, \omega_H$)	50 days [159]
	Stage 1 infecteds, 2005 ($I_{1H}(0)$)	1,654 [161, 207, 208][3]
	Stage 2 infecteds, 2005 ($I_{2H}(0)$)	574 [161, 208, 9][4]
	Probability of detection, stage 1 (τ_{H_1})	0.066 [9][5]

Table D.2: Model parameterization, South Sudan

Species	Parameter	Value
	Probability of detection, stage 2 (τ_{H_2})	0.565 [163]
	Probability of fly→human transmission (β_H)	0.0757 (fitted)
Reservoirs**	Lifespan ($1 \,/\, \mu_R$)	2 years [209]
	Reservoir:human ratio ($N_{Rt} \,/\, N_{Ht}$)	0.02†
	Incubation period ($1 \,/\, \sigma_R$)	7 days [210]
	Duration, infectious stage w/o treatment ($1 \,/\, \phi_{R_{nt}}$)	182.5 days [211]
	Duration, infectious stage w/ treatment ($1 \,/\, \phi_{R_t}$)	91.25 days*[6]
	Initial prevalence ($I_R(0)$)	0.2847 [58][7]
	Duration of immunity ($1 \,/\, \omega_R$)	50 days [159]
	Probability of trypanocide treatment (τ_R)	0*
	Probability of insecticide treatment (itc_R)	0*
	Probability of fly→reservoir transmission (β_R)	0.1345 [6]
Flies‡	Lifespan ($1 \,/\, \mu_V$)	26 days [130]
	Fly:human ratio ($N_{Vt} \,/\, N_{Ht}$)	6.56 [155]
	Incubation period ($1 \,/\, \sigma_V$)	18 days [213]
	Initial prevalence ($I_V(0)$)	0.0068 [130, 214][8]
	Probability of feeding in the first 24 hours (α)	1/3 [99]
	Probability of human→fly transmission (β_{Vh})	0.212 [215][9]
	Probability of reservoir→fly transmission (β_{Vr})	0.212 [215][9]
	Probability of human bite (p_H)	0.089 [21][10]
Non-resevoirs	Ratio to humans	0.1*

Table D.2: Model parameterization, South Sudan

Species	Parameter	Value

*Assumed; **Domestic swine; †Our SPDE maps; ‡*G. f. fuscipes*; t indexes time. [1]Value from 2010-2014 used as number of cases reported in 2005 was greater than the estimated population at risk for 2005-2009; [2]Parameterized as 6 months; [3]Taken as the number of stage 2 cases times the stage 1 : stage 2 duration ratio; [4]Total reported cases adjusted to reported stage 2 cases using national proportions in stage 2 (Atlas data), adjusted for probability of detection in stage 2, then derivd cases on a given day using stage 2 duration data; [5]Mean of 42.2-84% of total cases detected in cited paper, with 10.5% of reported cases detected by active surveillance per Atlas data; [6]Parameterized as 3 months; [7]Results from trypanolysis test; [8]Mean taken over the two cited papers; [9]Mean of males and females, overall % infected; [10]Primates taken as human

Table D.3: Model parameterization, Uganda (gHAT)

Species	Parameter	Value
Humans	Life expectancy, 2005 ($1 \,/\, \mu_H$)	51.7 years [205]
	Population, 2005 ($N_H(0)$)	2,030 [206]
	Incubation period ($1 \,/\, \sigma_H$)	12 days [159]
	Duration of stage 1 ($1 \,/\, \phi_{H_1}$)	526 days [207][1]
	Duration of stage 2 ($1 \,/\, \phi_{H_2}$)	182.5 days [208][2]
	Duration of immunity ($1 \,/\, \omega_H$)	50 days [159]
	Stage 1 infecteds, 2005 ($I_{1H}(0)$)	546 [161, 207, 208][3]
	Stage 2 infecteds, 2005 ($I_{2H}(0)$)	189 [161, 208, 9][4]
	Probability of detection, stage 1 (τ_{H_1})	0.161 [9][5]
	Probability of detection, stage 2 (τ_{H_2})	0.538 [163]

Table D.3: Model parameterization, Uganda (gHAT)

Species	Parameter	Value
	Probability of fly→human transmission (β_H)	0.087 (fitted)
Reservoirs**	Lifespan ($1 \,/\, \mu_R$)	2 years [209]
	Reservoir:human ratio ($N_{Rt} \,/\, N_{Ht}$)	0.24†
	Incubation period ($1 \,/\, \sigma_R$)	7 days [210]
	Duration, infectious stage w/o treatment ($1 \,/\, \phi_{R_{nt}}$)	182.5 days [211]
	Duration, infectious stage w/ treatment ($1 \,/\, \phi_{R_t}$)	91.25 days*[6]
	Duration of immunity ($1 \,/\, \omega_R$)	50 days [159]
	Initial prevalence ($I_R(0)$)	0.2847 [58][7]
	Probability of trypanocide treatment (τ_R)	0*
	Probability of insecticide treatment (itc_R)	0*
	Probability of fly→reservoir transmission (β_R)	0.1345 [6]
Flies‡	Lifespan ($1 \,/\, \mu_V$)	26 days [130]
	Fly:human ratio ($N_{Vt} \,/\, N_{Ht}$)	6.56 [155]
	Incubation period ($1 \,/\, \sigma_V$)	18 days [213]
	Initial prevalence ($I_V(0)$)	0.0068 [130, 214][8]
	Probability of feeding in the first 24 hours (α)	1/3 [99]
	Probability of human→fly transmission (β_{Vh})	0.212 [215][9]
	Probability of reservoir→fly transmission (β_{Vr})	0.212 [215][9]
	Probability of human bite (p_H)	0.089 [21][10]
Non-resevoirs	Ratio to humans	0.1*

Table D.3: Model parameterization, Uganda (gHAT)

Species	Parameter	Value

*Assumed; **Domestic swine; †Our SPDE maps; †*G. f. fuscipes*; t indexes time. [1]Mean value used; [2]Parameterized as 6 months; [3]Taken as the number of stage 2 cases times the stage 1 : stage 2 duration ratio; [4]Total reported cases adjusted to reported stage 2 cases using national proportions in stage 2 (Atlas data), adjusted for probability of detection in stage 2, then derived cases on a given day using stage 2 duration data; [5]Mean of % reported in cited paper, with 23% of reported cases detected by active surveillance per Atlas data; [6]Parameterized as 3 months; [7]Results from trypanolysis test; [8]Mean taken over the two cited papers; [9]Mean of males and females, overall % infected; [10]Primates taken as human

Table D.4: Model parameterization, Uganda (rHAT)

Species	Parameter	Value
Humans	Life expectancy, 2005 ($1 / \mu_H$)	51.7 years [205]
	Population, 2005 ($N_H(0)$)	7,642 [206]
	Incubation period ($1 / \sigma_H$)	12 days [159]
	Duration of stage 1 ($1 / \phi_{H_1}$)	23 days [217][1]
	Duration of stage 2 ($1 / \phi_{H_2}$)	27 days [217][2]
	Duration of immunity ($1 / \omega_H$)	50 days [159]
	Stage 1 infecteds, 2005 ($I_{1H}(0)$)	204 [161, 207, 208][3]
	Stage 2 infecteds, 2005 ($I_{2H}(0)$)	240 [161, 208, 9][4]
	Probability of detection, stage 1 (τ_{H_1})	0*
	Probability of detection, stage 2 (τ_{H_2})	0.0833 [103]
	Probability of fly→human transmission (β_H)	0.001 (fitted)
Reservoirs	Domestic swine	

Table D.4: Model parameterization, Uganda (rHAT)

Species	Parameter	Value
	Lifespan ($1 \,/\, \mu_{Ri}$)	2 years [209]
	Reservoir:human ratio ($N_{Rit} \,/\, N_{Ht}$)	0.06**
	Incubation period ($1 \,/\, \sigma_{Ri}$)	7 days [210]
	Duration, infectious stage w/o treatment ($1/\phi_{Ri_{nt}}$)	182.5 days [211]
	Duration, infectious stage w/ treatment ($1 \,/\, \phi_{Ri_t}$)	91.25 days*[5]
	Duration of immunity ($1 \,/\, \omega_{Ri}$)	50 days [159]
	Initial prevalence ($I_{Ri}(0)$)	0.014 [218]
	Probability of trypanocide treatment (τ_{Ri})	0*
	Probability of insecticide treatment (itc_{Ri})	0*
	Probability of fly→reservoir transmission (β_{Ri})	0.1345 [6]
	Wild swine (warthog)	
	Lifespan ($1 \,/\, \mu_{Ri}$)	17.5 years [219]
	Reservoir:human ratio ($N_{Rit} \,/\, N_{Ht}$)	0.1*
	Incubation period ($1 \,/\, \sigma_{Ri}$)	7 days [210]
	Duration, infectious stage w/o treatment ($1/\phi_{Ri_{nt}}$)	182.5 days [211]
	Duration, infectious stage w/ treatment ($1 \,/\, \phi_{Ri_t}$)	91.25 days*[5]
	Duration of immunity ($1 \,/\, \omega_{Ri}$)	50 days [159]
	Initial prevalence ($I_{Ri}(0)$)	0.018 [220, 221][6]
	Probability of trypanocide treatment (τ_{Ri})	0*
	Probability of insecticide treatment (itc_{Ri})	0*
	Probability of fly→reservoir transmission (β_{Ri})	0.1345 [6]
	Domestic bovids	
	Lifespan ($1 \,/\, \mu_{Ri}$)	15.5 years [222]
	Reservoir:human ratio ($N_{Rit} \,/\, N_{Ht}$)	0.16**
	Incubation period ($1 \,/\, \sigma_{Ri}$)	12 days [159]

Table D.4: Model parameterization, Uganda (rHAT)

Species	Parameter	Value
	Duration, infectious stage w/o treatment ($1/\phi_{Ri_{nt}}$)	225 days [223]
	Duration, infectious stage w/ treatment ($1 / \phi_{Ri_t}$)	91.25 days[*5]
	Duration of immunity ($1 / \omega_{Ri}$)	50 days [159]
	Initial prevalence ($I_{Ri}(0)$)	0.012 [218]
	Probability of trypanocide treatment (τ_{Ri})	0[*]
	Probability of insecticide treatment (itc_{Ri})	0[*]
	Probability of fly→reservoir transmission (β_{Ri})	0.1345 [6]
	Wild bovids (bushbuck, buffalo)	
	Lifespan ($1 / \mu_{Ri}$)	15 years [224, 225]
	Reservoir:human ratio (N_{Rit} / N_{Ht})	0.1[*]
	Incubation period ($1 / \sigma_{Ri}$)	12 days [159]
	Duration, infectious stage w/o treatment ($1/\phi_{Ri_{nt}}$)	225 days [223]
	Duration, infectious stage w/ treatment ($1 / \phi_{Ri_t}$)	91.25 days[*5]
	Duration of immunity ($1 / \omega_{Ri}$)	50 days [159]
	Initial prevalence ($I_{Ri}(0)$)	0.022 [221]
	Probability of trypanocide treatment (τ_{Ri})	0[*]
	Probability of insecticide treatment (itc_{Ri})	0[*]
	Probability of fly→reservoir transmission (β_{Ri})	0.1345 [6]
Flies[†]	Lifespan ($1 / \mu_R$)	26 days [130]
	Fly:human ratio (N_{Vt} / N_{Ht})	6.56 [155]
	Incubation period ($1 / \sigma_V$)	18 days [213]
	Initial prevalence ($I_{Ri}(0)$)	0.0068 [130, 214][7]
	Probability of feeding in the first 24 hours (α)	1/3 [99]
	Probability of human→fly transmission (β_{Vh})	0.212 [215][8]
	Probability of reservoir→fly transmission (β_{Vr})	0.212 [215][8]

Table D.4: Model parameterization, Uganda (rHAT)

Species	Parameter	Value
	Probability of human bite (p_H)	0.05 [130]
Non-reservoirs	Ratio to humans	0.1*

*Assumed; **Our SPDE maps; †*G. f. fuscipes*; t indexes time, i indexes reservoir species group. [1]Median time from infection to stage 1; [2]Median time from infection to stage 2, minus median time from infection to stage 1; [3]Taken as the number of stage 2 cases times the stage 1 : stage 2 duration ratio; [4]Total reported cases adjusted to reported stage 2 cases using national proportions in stage 2 (Atlas data), adjusted for probability of detection in stage 2, then derived cases on a given day using stage 2 duration data; [5]Parameterized as 3 months; [6]Product of proportion that is *T. brucei s. l.*, and proportion of *T. brucei s. l.* that is *T. b. rhodesiense*; [7]Mean taken over the two cited papers; [8]Mean of males and females, overall % infected

Table D.5: Model parameterization, Malawi

Species	Parameter	Value
Humans	Life expectancy, 2005 ($1 / \mu_H$)	47.8 years [205]
	Population, 2005 ($N_H(0)$)	939 [206]
	Incubation period ($1 / \sigma_H$)	12 days [159]
	Duration of stage 1 ($1 / \phi_{H_1}$)	30 days [217][1]
	Duration of stage 2 ($1 / \phi_{H_2}$)	52 days [217][2]
	Duration of immunity ($1 / \omega_H$)	50 days [159]
	Stage 1 infecteds, 2005 ($I_{1H}(0)$)	26 [161, 207, 208][3]
	Stage 2 infecteds, 2005 ($I_{2H}(0)$)	70 [161, 208, 9][4]

Table D.5: Model parameterization, Malawi

Species	Parameter	Value
	Probability of detection, stage 1 (τ_{H_1})	0*
	Probability of detection, stage 2 (τ_{H_2})	0.0833 [103]
	Probability of fly→human transmission (β_H)	0.001 [215, 6][5]
Reservoirs	Domestic swine	
	Lifespan ($1 / \mu_{Ri}$)	2 years [209]
	Reservoir:human ratio (N_{Rit} / N_{Ht})	0.09**
	Incubation period ($1 / \sigma_{Ri}$)	7 days [210]
	Duration, infectious stage w/o treatment ($1/\phi_{Ri_{nt}}$)	182.5 days [211]
	Duration, infectious stage w/ treatment ($1 / \phi_{Ri_t}$)	91.25 days*[6]
	Duration of immunity ($1 / \omega_{Ri}$)	50 days [159]
	Initial prevalence ($I_{Ri}(0)$)	0.014 [218]
	Probability of trypanocide treatment (τ_{Ri})	0*
	Probability of insecticide treatment (itc_{Ri})	0*
	Probability of fly→reservoir transmission (β_{Ri})	0.1345 [6]
	Wild swine (warthog)	
	Lifespan ($1 / \mu_{Ri}$)	17.5 years [219]
	Reservoir:human ratio (N_{Rit} / N_{Ht})	0.1*
	Incubation period ($1 / \sigma_{Ri}$)	7 days [210]
	Duration, infectious stage w/o treatment ($1/\phi_{Ri_{nt}}$)	182.5 days [211]
	Duration, infectious stage w/ treatment ($1 / \phi_{Ri_t}$)	91.25 days*[6]
	Duration of immunity ($1 / \omega_{Ri}$)	50 days [159]
	Initial prevalence ($I_{Ri}(0)$)	0.018 [220, 221][7]
	Probability of trypanocide treatment (τ_{Ri})	0*
	Probability of insecticide treatment (itc_{Ri})	0*
	Probability of fly→reservoir transmission (β_{Ri})	0.1345 [6]

Table D.5: Model parameterization, Malawi

Species	Parameter	Value
	Domestic bovids	
	Lifespan ($1 / \mu_{Ri}$)	15.5 years [222]
	Reservoir:human ratio (N_{Rit} / N_{Ht})	0.09**
	Incubation period ($1 / \sigma_{Ri}$)	12 days [159]
	Duration, infectious stage w/o treatment ($1/\phi_{Ri_{nt}}$)	225 days [223]
	Duration, infectious stage w/ treatment ($1 / \phi_{Ri_t}$)	91.25 days*[6]
	Duration of immunity ($1 / \omega_{Ri}$)	50 days [159]
	Initial prevalence ($I_{Ri}(0)$)	0.061 [226][8]
	Probability of trypanocide treatment (τ_{Ri})	0*
	Probability of insecticide treatment (itc_{Ri})	0*
	Probability of fly→reservoir transmission (β_{Ri})	0.1345 [6]
	Wild bovids (bushbuck, buffalo)	
	Lifespan ($1 / \mu_{Ri}$)	15 years [224, 225]
	Reservoir:human ratio (N_{Rit} / N_{Ht})	0.1*
	Incubation period ($1 / \sigma_{Ri}$)	12 days [159]
	Duration, infectious stage w/o treatment ($1/\phi_{Ri_{nt}}$)	225 days [223]
	Duration, infectious stage w/ treatment ($1 / \phi_{Ri_t}$)	91.25 days*[6]
	Duration of immunity ($1 / \omega_{Ri}$)	50 days [159]
	Initial prevalence ($I_{Ri}(0)$)	0.022 [221]
	Probability of trypanocide treatment (τ_{Ri})	0*
	Probability of insecticide treatment (itc_{Ri})	0*
	Probability of fly→reservoir transmission (β_{Ri})	0.1345 [6]
Flies†	Lifespan ($1 / \mu_V$)	29 days [212]
	Fly:human ratio (N_{Vt} / N_{Ht})	6.56 [155]
	Incubation period ($1 / \sigma_V$)	18 days [213]

Table D.5: Model parameterization, Malawi

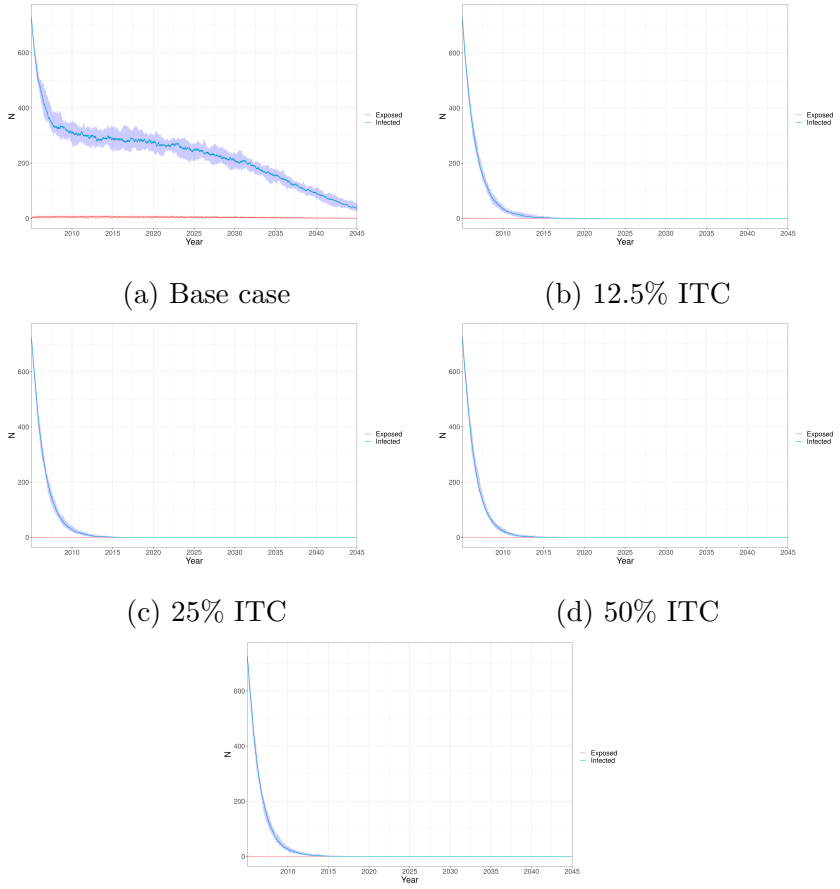
Species	Parameter	Value
	Initial prevalence ($I_R(0)$)	0.022 [157]
	Probability of feeding in the first 24 hours (α)	1/3 [99]
	Probability of human→fly transmission (β_{Vh})	0.258 [215][9]
	Probability of reservoir→fly transmission (β_{Vr})	0.258 [215][9]
	Probability of human bite (p_H)	0.05 [21][10]
Non-resevoirs	Ratio to humans	0.1*

*Assumed; **Our SPDE maps; †*G. m. morsitans*; t indexes time, i indexes reservoir species group. [1]Median time from infection to stage 1; [2]Median time from infection to stage 2, minus median time from infection to stage 1; [3]Taken as the number of stage 2 cases times the stage 1 : stage 2 duration ratio; [4]Total reported cases adjusted to reported stage 2 cases using national proportions in stage 2 (Atlas data), adjusted for probability of detection in stage 2, then derived cases on a given day using stage 2 duration data; [5]Ratio of the mean of males and females who acquired patent infection after feeding on infected rats for *G. m. morsitans* to *G. f. fuscipes*, times transmission probability for *G. f. fuscipes*. [6]Parameterized as 3 months; [7]Product of proportion that is *T. brucei s. l.*, and proportion of *T. brucei s. l.* that is *T. b. rhodesiense*; [7]Mean taken over the two cited papers;[8]Mean across the two foci; [9]Mean of males and females, overall % infected; [10]Primates taken as humans

Appendix E

MODEL TRAJECTORIES

E.1 Humans



(a) Base case

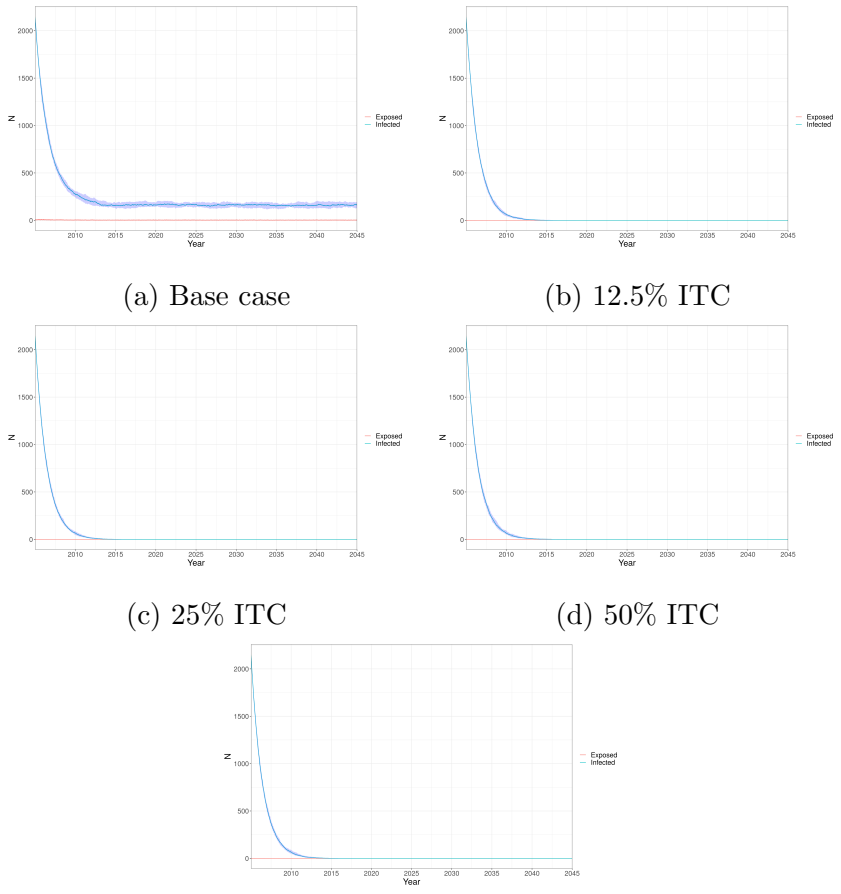
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure E.1: Number of exposed (red) and infected (blue) humans, Uganda gHAT. ITC: insecticide treatment of pigs; TT: trypanocide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified.



(a) Base case

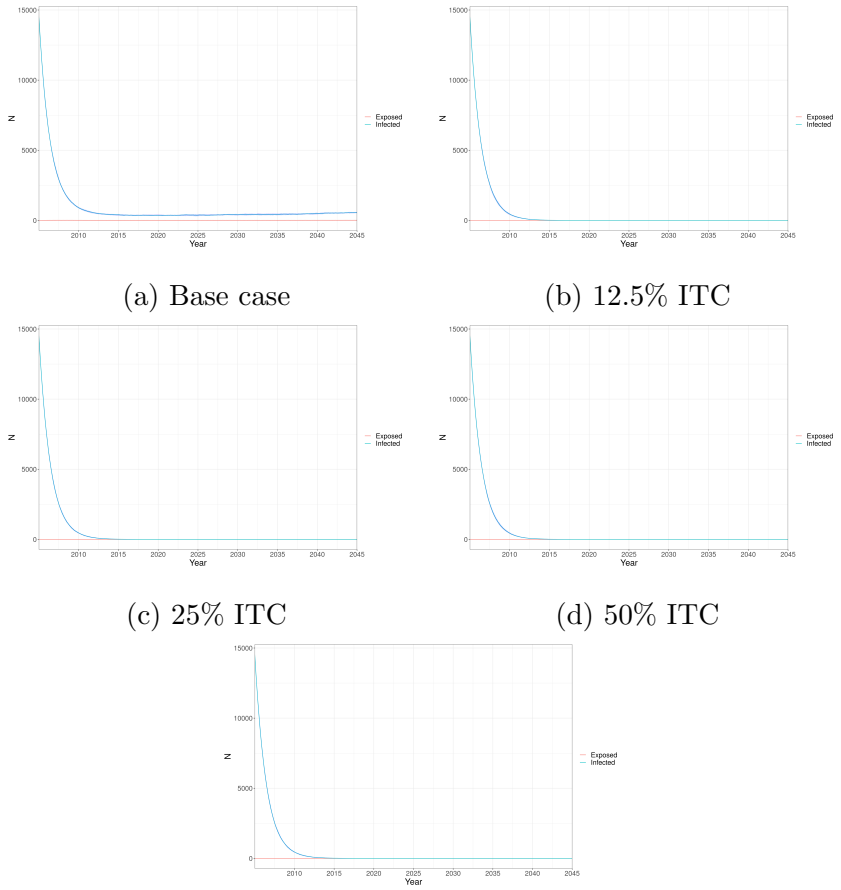
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure E.2: Number of exposed (red) and infected (blue) humans, DRC Equateur Nord focus. ITC: insecticide treatment of pigs; TT: trypanocide treatment of pigs.TT coverage assumed to be 0% unless otherwise specified.



(a) Base case

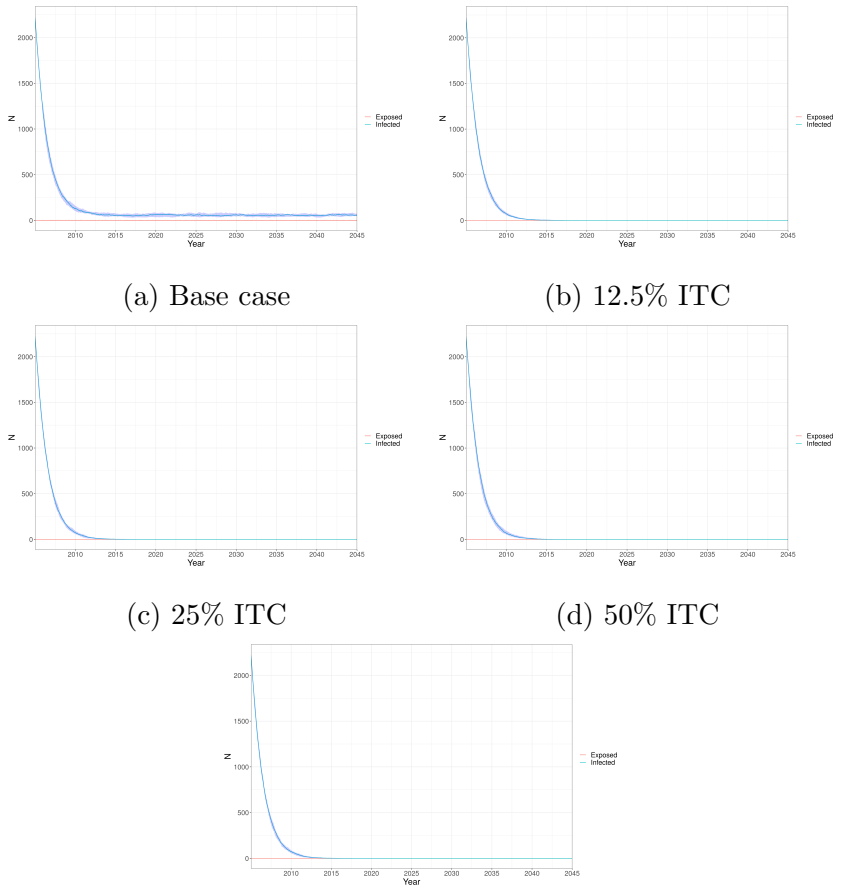
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure E.3: Number of exposed (red) and infected (blue) humans, DRC Bandundu/Sakuru foci. ITC: insecticide treatment of pigs; TT: trypanocide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified.



(a) Base case

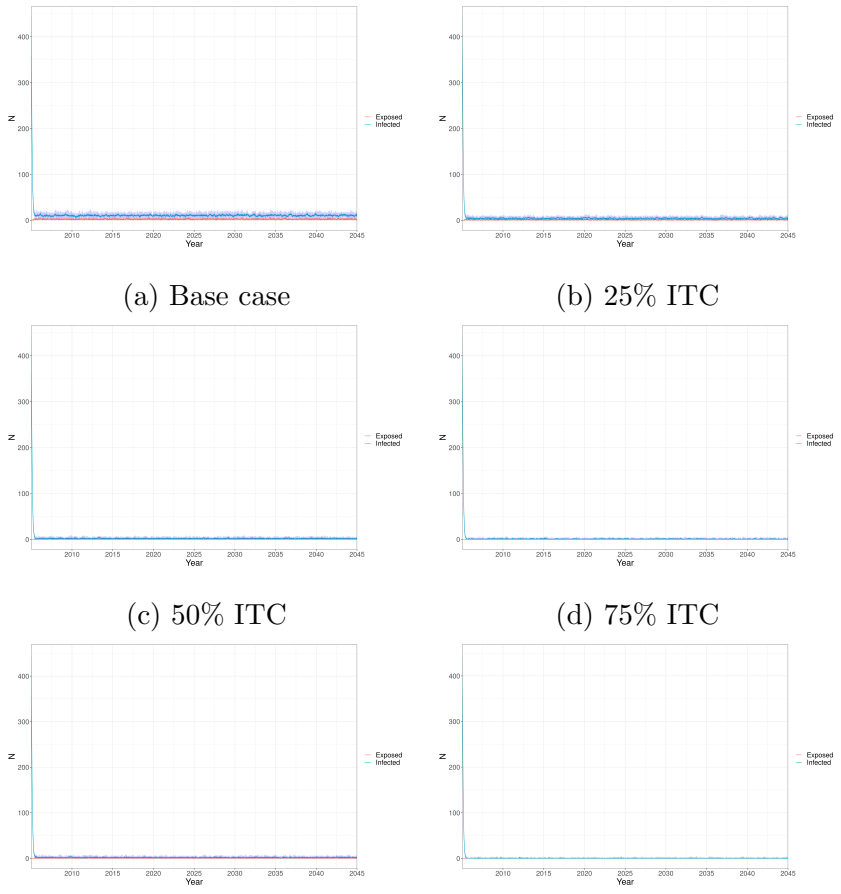
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure E.4: Number of exposed (red) and infected (blue) humans, South Sudan. ITC: insecticide treatment of pigs; TT: trypanocide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified.



(a) Base case

(b) 25% ITC

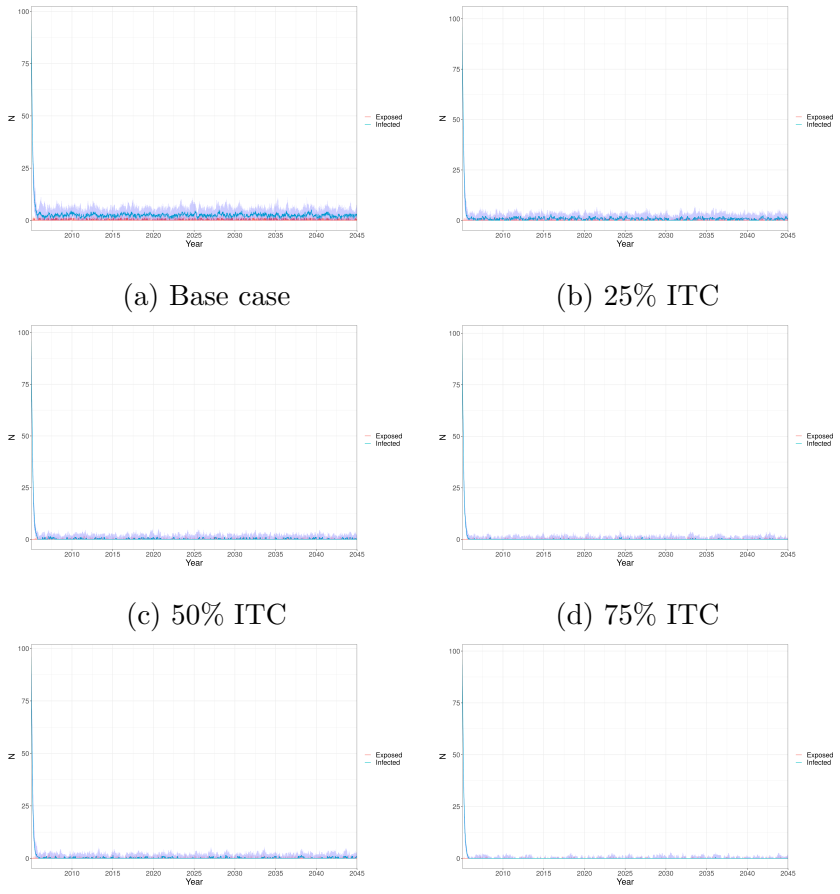
(c) 50% ITC

(d) 75% ITC

(e) 75% TT, 50% ITC

(f) 100% TT, 100% ITC

Figure E.5: Number of exposed (red) and infected (blue) humans, Uganda rHAT. ITC: insecticide treatment of cattle and pigs; TT: trypanocide treatment of cattle and pigs. TT coverage assumed to be 50% unless otherwise specified.



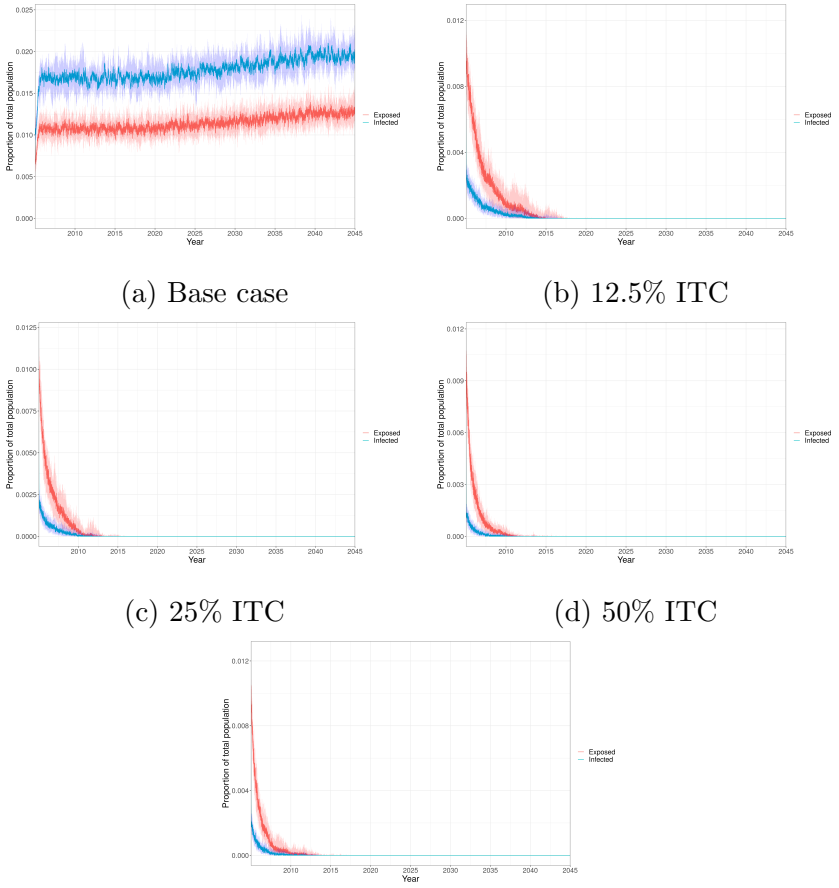
(a) Base case (b) 25% ITC

(c) 50% ITC (d) 75% ITC

(e) 75% TT, 50% ITC (f) 100% TT, 100% ITC

Figure E.6: Number of exposed (red) and infected (blue) humans, Malawi. ITC: insecticide treatment of cattle and pigs; TT: trypanocide treatment of cattle and pigs. TT coverage assumed to be 50% unless otherwise specified.

E.2 Tsetse flies



(a) Base case

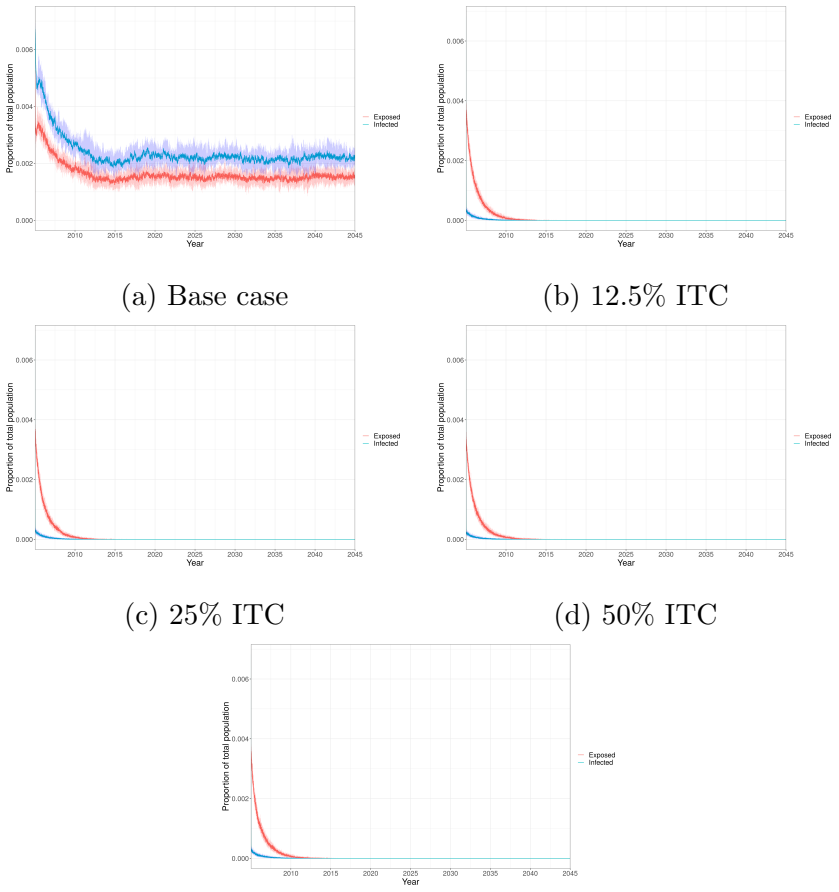
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure E.7: Proportion of total tsetse population exposed (red) and infected (blue), Uganda gHAT. ITC: insecticide treatment of pigs; TT: insecticide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified.



(a) Base case

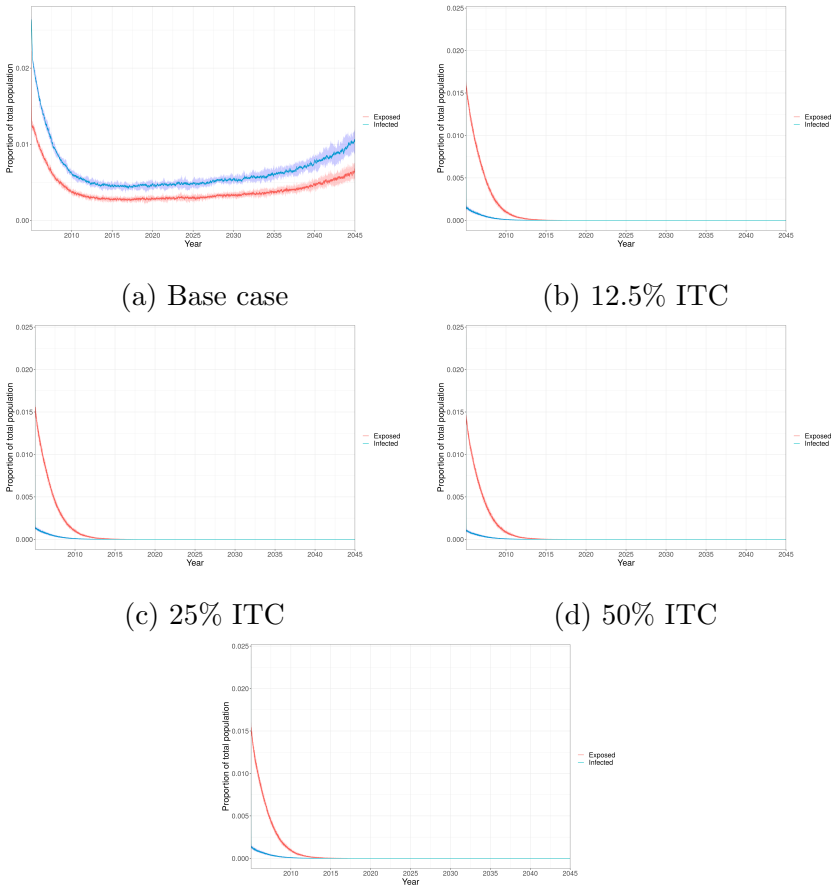
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure E.8: Proportion of total tsetse population exposed (red) and infected (blue), DRC Equateur Nord focus. ITC: insecticide treatment of pigs; TT: insecticide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified.



(a) Base case

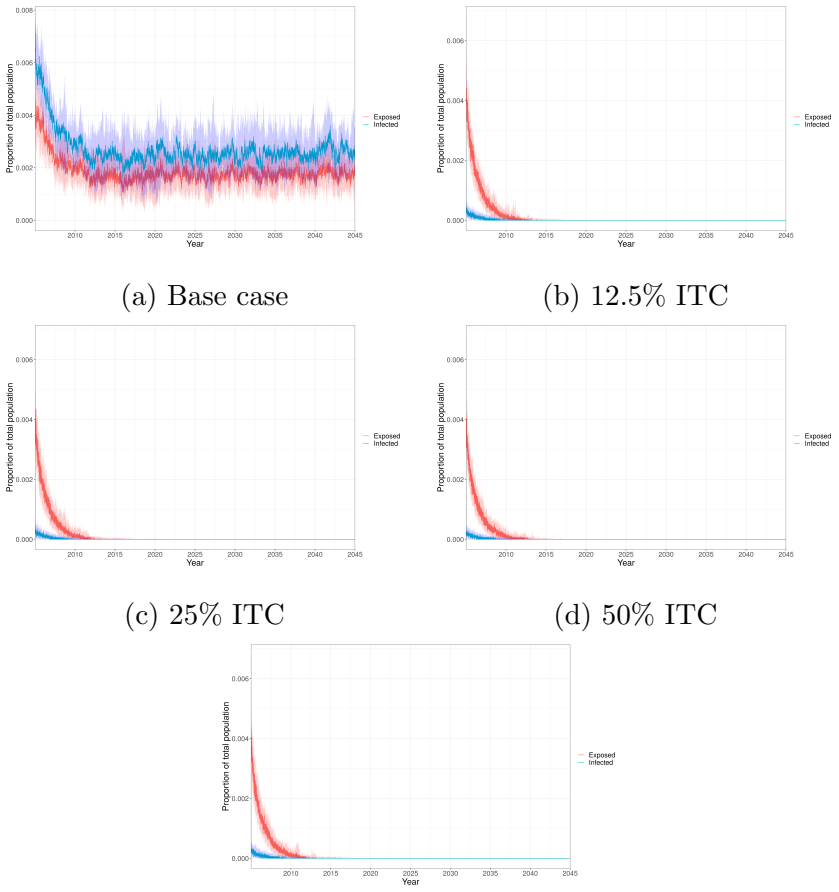
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure E.9: Proportion of total tsetse population exposed (red) and infected (blue), DRC Bandundu/Sankuru foci. ITC: insecticide treatment of pigs; TT: insecticide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified.



(a) Base case

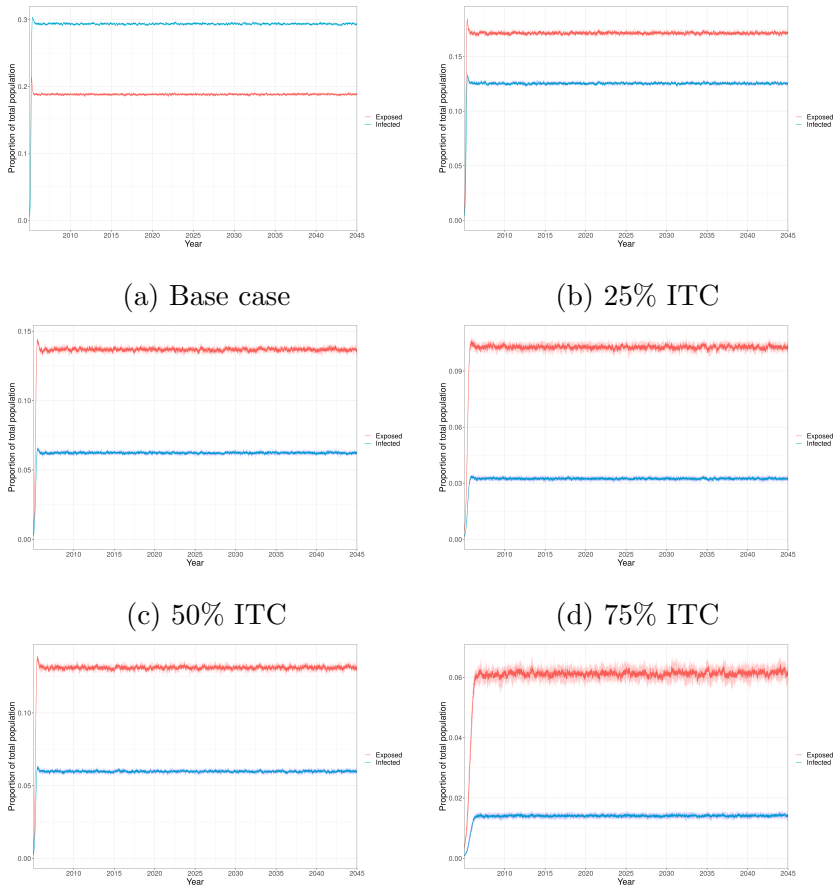
(b) 12.5% ITC

(c) 25% ITC

(d) 50% ITC

(e) 25% TT, 25% ITC

Figure E.10: Proportion of total tsetse population exposed (red) and infected (blue), South Sudan. ITC: insecticide treatment of pigs; TT: insecticide treatment of pigs. TT coverage assumed to be 0% unless otherwise specified.



(a) Base case

(b) 25% ITC

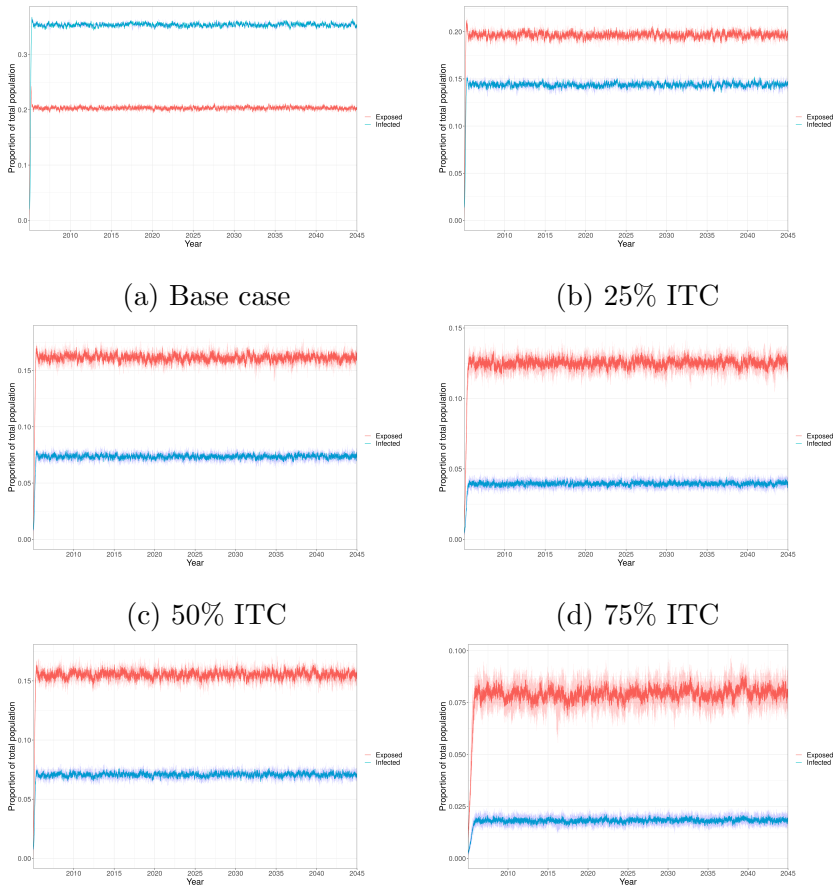
(c) 50% ITC

(d) 75% ITC

(e) 75% TT, 50% ITC

(f) 100% TT, 100% ITC

Figure E.11: Proportion of total tsetse population exposed (red) and infected (blue), Uganda rHAT. ITC: insecticide treatment of cattle and pigs; TT: insecticide treatment of cattle and pigs. TT coverage assumed to be 50% unless otherwise specified.



(a) Base case

(b) 25% ITC

(c) 50% ITC

(d) 75% ITC

(e) 75% TT, 50% ITC

(f) 100% TT, 100% ITC

Figure E.12: Proportion of total tsetse population exposed (red) and infected (blue), Malawi. ITC: insecticide treatment of cattle and pigs; TT: insecticide treatment of cattle and pigs. TT coverage assumed to be 50% unless otherwise specified.