Spatio-temporal dynamics of *gambiense* human African trypanosomiasis infection

by

Christopher Neil Davis

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

MathSys CDT

June 10, 2020

Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to thank my supervisors Matt Keeling and Kat Rock for all their guidance in constructing this thesis. I am very grateful for their continued support. I would also like to thank the other members of SBIDER, and in particular those involved with the NTD Modelling Consortium or HAT MEPP, for many helpful comments and advice throughout my time at the University of Warwick.

I have very much enjoyed my time here, and so must also thank the staff of the MathSys CDT for providing an excellent centre to complete my PhD. And many thanks to all the friends I have made on the CDT as office mates, pub quiz team members, swimming–cycling–running buddies and good people to have a conversation with at lunchtimes.

Particular thanks to former housemates Cameron, Sophie, Sami and Joe. And, of course, special thanks to Susie and my family.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. Parts of this thesis have previously been published or submitted to a journal.

Substantial sections of Chapter 2 have been published as:

- Davis, CN, Rock, KS and Keeling, MJ, 2020. Human African trypanosomiasis: Current status and eradication efforts. CAB Reviews. 2020;15(032):1-7.

The work from Chapter 3 has been published as, and included in:

- Davis, CN, Rock, KS, Miaka, EM and Keeling, MJ. 2019. Village-scale persistence and elimination of *gambiense* human African trypanosomiasis. PLoS Neglected Tropical Diseases, 13(10): e0007838.

- NTD Modelling Consortium Discussion Group on *Gambiense* Human African Trypanosomiasis. 2019. Insights from quantitative and mathematical modelling on the proposed 2030 goal for gambiense human African trypanosomiasis (gHAT) [version 2; peer review: 2 approved]. Gates Open Res, 3:1553.

The work from Chapter 4 is soon to be submitted as:

- Davis, CN, Rock, KS, Miaka, EM and Keeling, MJ. 2020. Optimising active screening for *gambiense* human African trypanosomiasis.

University of Warwick, June 10, 2020

Christopher N Davis

Abstract

Gambiense human African trypanosomiasis (gHAT) is a disease targeted for elimination by the World Health Organization. In this thesis, I develop a stochastic mathematical model for gHAT infection dynamics that can explore the infection on the level of individual villages. This allows us to make predictions on the same scale at which the active screening interventions occur. Since regional elimination is dependent on the local elimination of transmission in villages, we then scale up the model to consider the interaction of villages in relation to the infection.

The models show that infection is able to persist for long periods in relatively small isolated populations due to the long time scales of infection, and that not observing any cases is an indicator of local elimination, but not proof. Further analysis demonstrates that using a net monetary benefit framework, active screening has the most benefit when carried out at high coverage approximately annually. Modelling predictions also show that elimination is unlikely to be achieved in all areas with active screening at the current coverages alone.

These results provide quantitative support for some current intervention strategies, while the insights given also highlight some potential difficulties for achieving and measuring a 2030 elimination goal.

1 Introduction

1.1 Introduction to modelling infectious diseases

Infectious diseases are illnesses caused by pathogenic microorganisms, which can include viruses, bacteria, fungi and other parasites. These diseases are transmissible and can be spread from person to person, either directly, or indirectly, passing between hosts through the environment, or by secondary host or vector. The invasion of an organism by a pathogen can cause any number of symptoms, of varying severity, if the host's immune mechanisms are compromised. Indeed, a large percentage of human deaths are caused by infectious diseases; of the total number of deaths in 2017, approximately 18.6% were caused by communicable, maternal, neonatal, and nutritional (CMNN) causes [1]. Deaths by infectious disease also disproportionately affect low-income countries, with CMNN conditions causing more than 50% of deaths, compared to less than 7% in high-income countries [2].

The benefit of seeking to reduce the number of cases of people infected with infectious diseases, and, for specific infections, striving for elimination of the disease, is primarily the alleviation of suffering for the individuals inflicted. However, the health problems of individuals are also an obstacle to economic development[3]. As such, investments in healthcare can be economically beneficial. Therefore, in order to both limit disease burden on individuals and to not impair human development, we need an understanding of the spread of infectious diseases to be able to limit their impact. The potential outcome of any intervention strategy, in terms of a reduction of case numbers, should be recognised, along with which strategies are more likely to represent value for money, if dealing with a limited budget. To gain this level of understanding, we can utilise mathematical models.

Mathematical models can be designed to provide a precise description of a biological process, and in particular, be used to simulate population-level infection dynamics. Much of this work is development from the early twentieth century works of Ross [4] and Kermack & McKendrick [5]. These models can be used for both predicting future infection and for developing understanding of processes such as invasion dynamics or the effects of interventions. Good mathematical models should provide robust insights, either showing quantitative results or qualitative behaviour depending on their purpose. The mechanisms within models should be clear, such that they provide intuition of the processes involved, as well as the models being flexible, so they can be adapted to multiple scenarios [6].

Mathematical models also need to be well-communicated if they are to have the intended influence on policy decisions [7]. This requires a clear question to be stated that is a known priority for policymakers, which the model can be used to address. To increase the impact of mathematical modelling, the NTD Modelling Consortium has recently released a systematic review that identifies five key principles for communication with stakeholders [8]. These principles consist of: stakeholder engagement, complete model documentation, complete description of data used, communicating uncertainty, and testable model outcomes. Additionally, it is noted that models can have more impact when accompanied by an interactive interface that allows policymakers and stakeholders to change some initial assumptions of the modelling to suit their expert opinions. There are examples of work that does this, although some challenges remain [9, 10].

In this thesis, I use mathematical modelling with a focus on a specific infectious disease — *gambiense* human African trypanosomiasis (gHAT). gHAT is caused by the parasite *Trypanosoma brucei gambiense* and transmitted by an insect vector, the tsetse. Our mathematical modelling aims to extend the current literature, with a focus understanding the dynamics of gHAT infection and the potential effects of interventions at the village-scale, with a broader focus on the elimination of this disease, in line with targets set by the World Health Organization (WHO) [11].

1.2 Thesis structure

This thesis is divided into six chapters, including this introduction, which cover the research I have conducted throughout my PhD.

Chapter 2 provides a background to all the work, giving an overview of the biology of *gambiense* human African trypanosomiasis (gHAT), the historical context and current status of the infection, along with the existing modelling literature, both briefly of mathematical epidemiology in general, and of work specifically done for gHAT.

Chapter 3 considers the implementation of a mathematical model for gHAT infection on the scale of single village populations in a region of the Democratic Republic of Congo (DRC). This work investigates the long term persistence of the infection and assesses the risks of recrudescence within a village after the infection has been locally eliminated. I also explore how the probability of elimination of transmission is informed by the absence of detection of cases when village populations are screened for gHAT.

Chapter 4 uses the village-scale model, re-parameterised for a different region, to analyse the implementation of different strategies for screening and treatment interventions. By evaluating this model with the addition of costs, I am also able to examine the intervention strategies that maximise monetary benefits and to determine the cost-effectiveness of these strategies.

Chapter 5 explores the effects of simplifying our model structure to use a master equation implementation for predicting infection dynamics. Using this approach, I consider the interactions between different villages, and scale up the village-level model to make predictions at a health zone level.

Finally, **chapter 6** summarises and contrasts the differing modelling approaches used throughout and provides a discussion and of the whole thesis.

2 Background

In this chapter, we give an overview of the disease human African trypanosomiasis (HAT) and discuss its current status along with the efforts to reduce levels of infection, where the ultimate goal is disease eradication. In addition, we review the modelling literature for HAT, both in terms of the types of models and methods used for such vector-borne diseases and implications of these models. For more details on HAT, a selection of recent reviews from a range of viewpoints include Büscher *et al.* [12], Aksoy *et al.* [13] and Rock *et al.* [14].

2.1 The biology of human African trypanosomiasis

To understand the results of modelling human African trypanosomiasis, we first consider the cause of the disease, its transmission, and how it presents itself in terms of clinical signs and symptoms.

2.1.1 A fatal disease

Human African trypanosomiasis (HAT), commonly known as sleeping sickness, is a vector-borne disease affecting human populations in sub-Saharan Africa. It is caused by parasitic protists of the species *Trypanosoma brucei*, of which there are two subspecies that can infect humans: *gambiense* and *rhodesiense*. Both types of the parasite are typically transmitted to humans by infected tsetse (genus *Glossina*), a large biting fly that inhabits affected regions. The diseases caused by these infections are almost invariably

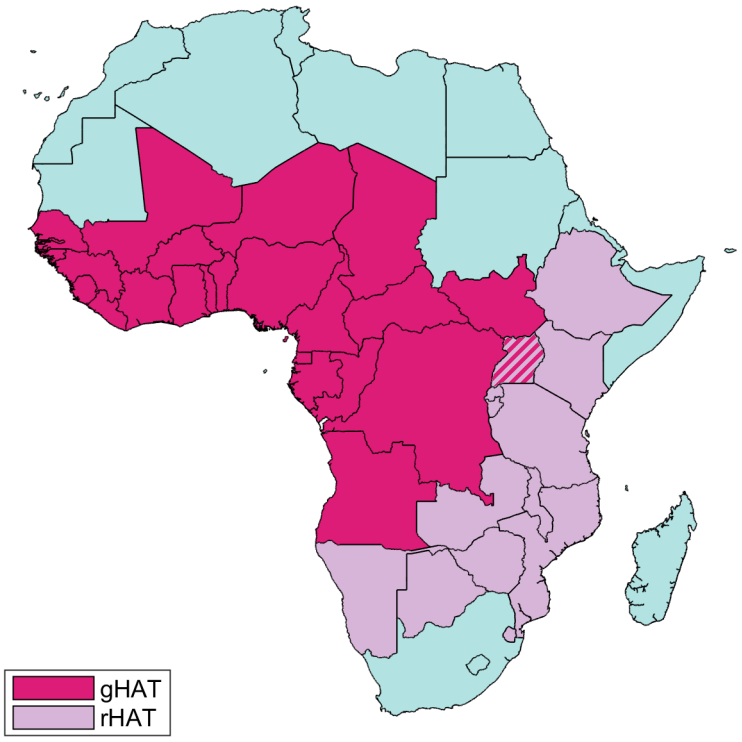


Fig 2.1: Map of the historically 36 HAT-endemic countries. 24 countries are gHAT endemic and 13 are rHAT endemic, including Uganda, which is endemic with both.

fatal without treatment [15]; in the last 100 years HAT has killed millions of people [12].

HAT is historically endemic in 36 countries of sub-Saharan Africa, with most *gambiense* human African trypanosomiasis (gHAT) cases occurring in West and Central Africa and most *rhodesiense* human African trypanosomiasis (rHAT) cases occurring in East Africa (see Figure 2.1); 24 of these counties are endemic with gHAT and 13 with rHAT [16]. The disease distribution across these countries is highly focal [12, 17], where we define a focus of infection as a place containing the epidemiological factors needed for transmission of an infection. This heterogeneous clustering of incidence across space is predominantly attributed to the habitat of the vector (tsetse), but is notable since disease prevalence can vary greatly over short distances and even between neighbouring villages [18]. Outside of the 36 endemic countries, a small number of cases have also been diagnosed, mainly due to travellers and migrants from these countries [19].

Gambiense HAT is the more widespread of the two diseases, causing 98% of all reported infections in the last 10 years of data (2009–2018) [20]; the Democratic Republic of the

Congo (DRC) is the country with the highest proportion of gHAT cases with 82% in this period [20]. And for much of this thesis, we will focus on gHAT in the DRC.

However, we note that trypanosomiasis is neither exclusive to Africa nor humans. Chagas disease (sometimes known as American trypanosomiasis) is an infectious disease of humans caused by *Trypanosoma cruzi*, while animal trypanosomiasis affects wild and domestic animals across Africa. Indeed, animals can be infected by both the *gambiense* and the *rhodesiense* forms of the trypanosome and can act as a reservoir to infection [21–23].

2.1.2 Introduction to *Trypanosoma brucei*

Trypanosoma brucei is a species of parasitic kinetoplastid that can cause forms of trypanosomiasis. Kinetoplastids are flagellated protists (eukaryotic organisms that are not animal, plant or fungus) with a kinetoplast (a dense mass of DNA within the large mitochondrion) [24]. *T. brucei* are unicellular and between 3 and 30 μm long, depending on life cycle stage, formed in a streamlined tapered shape [25]. Within the cell membrane of *T. brucei*, there are organelles including the nucleus, mitochondria, endoplasmic reticulum, Golgi apparatus, ribosomes and kinetoplast.

The parasite is grouped into three subspecies: the previously mentioned *T. b. gambiense* and *T. b. rhodesiense*, both of which can cause human African trypanosomiasis (HAT), and *T. b. brucei*, which only rarely infects humans [26]. *T. b. gambiense* and *T. b. rhodesiense* are able to infect humans because they are able to express proteins resistant to the trypanolytic factors of a human's innate immune system response [27]. These trypanolytic factors are serum complexes (TLF-1 and TLF-2) that contain apolipoprotein L1 (ApoL1) and haptoglobin related protein (HPR) that can cause lysis in other trypanosome subspecies [28, 29]. Trypanolytic factors are found in relatively few species, such as humans, gorillas, mandrills, baboons and sooty mangabeys [30], hence *T. b. brucei* is able to infect many other animals, and so all three forms (and other forms, such as *T. vivax* and *T. congolense*) can cause animal trypanosomiasis [31].

Trypanosoma brucei have different classes of cellular organisation, or morphological forms, which are presented in different parts of their life cycle. When *Trypanosoma brucei* is in a mammalian host, it is a trypomastigote, an elongated cell, circulating in the blood. The kinetoplast is posterior of the nucleus and the flagellum arises from the posterior end. When *Trypanosoma brucei* is in the tsetse, it becomes a epimastigote, with its kinetoplast anterior to the nucleus and a long flagellum that starts at the centre of the body is attached to the cell body [32].

2.1.3 The vector — tsetse

Tsetse are large biting flies that live by feeding on the blood of vertebrate animals. They can carry *Trypanosoma brucei* and hence act as a facilitator for the transmission of trypanosomiasis to humans and other animals. There are about 31 different species of tsetse, which are part of the genus *Glossina*, and grouped as either 'savannah', 'forest', or 'riverine' flies [33]. *T. b. gambiense* is typically transmitted by 'riverine' tsetse, particularly the subspecies of *Glossina fuscipes* and *Glossina palpalis*, which inhabit the vegetation around the rivers and lakes in West and Central Africa [34]. On the other hand, *T. b. rhodesiense* are transmitted by 'savannah' tsetse, particularly *Glossina morsitans* and *Glossina pallidipes*, which occupy the woodlands of East Africa [35].

Adult tsetse are large for flies, approximately 8–17 mm. They have large heads, separated eyes and a long proboscis extending forward from a bulb attached to the bottom of the head [36]. The thorax is large and the abdomen wide, shorter than the wings, which can fold completely to cover the top of the abdomen, distinguishing them from other fly species [37]. These properties mean tsetse are easier to analyse in a laboratory setting than other smaller flies [38–40].

Tsetse are also an unusual fly species due to their life cycle. The flies reproduce by adenotrophic viviparity [41], whereby a female fly fertilises only one egg at a time, from the age of about 6 days old, which remains and matures in the uterus for 7–12 days during the first three larval stages [42]. In the third larval stage, the larva is deposited by the female on loose soil and crawls into the ground to develop a hard outer shell called the puparial case [43]. During this time, which lasts approximately 20–40 days,

the morphological transformation into an adult fly occurs. This change occurs without feeding and so the larva relies on the nutritional resources provided by the parent tsetse while in its uterus. The intensive nature of this process means that the tsetse rate of reproduction is very low [44]. Males can inseminate a female fly every 2–3 days [45]. The life expectancy of a female adult tsetse is only 20–40 days, while a male is 14–21 days [46]; although in a laboratory tsetse have been known to live for up to 120 days. To ensure that a female fly produces two adult progeny and sustain the tsetse population, the female fly needs to live for at least 25 days. Thus, it is expected that the mean adult daily mortality needs to be less than 3.5% for tsetse populations to persist [44]. In addition, predation, parasites and extreme temperatures can induce mortality in the pupal stage [47].

Tsetse typically feed every 2–5 days [48] and these blood-meals satisfy all the tsetse's requirements. Hosts therefore need to be found frequently in order to avoid starvation, which many flies will die from. Tsetse may also die by attempting to feed on high-risk hosts, such as humans, that kill them [49]. Tsetse populations can continue to exist at densities as low as 1 tsetse per square kilometre [50], while densities are rarely seen higher than 10,000 per square kilometre [51]. Tsetse are also highly mobile, known to travel up to 1 kilometre per day [52].

2.1.4 The life cycle of *Trypanosoma brucei*

The life cycle of *Trypanosoma brucei* is split between the tsetse and a mammalian host (humans, domestic animals, livestock, or wild animals). When an infected tsetse takes a blood-meal on a host, the fly will inject metacyclic trypomastigotes into the skin tissue, which then enter the lymphatic system and the bloodstream [53]. While initially short and stumpy, the trypomastigotes will become long and slender and multiply. Daughter cells will be short and stumpy. The long slender forms are able to spread to various organs and tissues, eventually the central nervous system and brain [54]. The latent period between inoculation and becoming infectious is typically 7–14 days [55]. In some animals this process will not see the exhibition of severe adverse effects, yet the parasite can still be transmitted onwards. Thus, the animals act as a reservoir of infection. Animals that are able to maintain chronic infections, however, will predominantly have

lower parasitaemia and so be less infectious [38]. This trypanotolerance may also occur in some humans, who never show symptoms and may even self-cure, even after 5–15 years [15].

In order to evade a host's immune system, the surface of *Trypanosoma brucei* is coated in variant surface glycoprotein (VSG) [56]. A specific VSG coat will provoke a particular immune response to attempt to kill the trypanosome, but on a cell division it is possible that progeny will switch the VSG that is expressed and so a population of trypanosomes in a host can have very diverse VSG coats. It is estimated that between 10^{-6} and 10^{-3} coat switches occur per doubling of the trypanosome population [57]. Host immunity to specific VSG coats take time to develop and to infections can be prolonged as trypanosomes multiply and this process can cause waves of trypanosomes in the blood [12].

When a mammal has successfully been infected with trypanosomes, and a susceptible tsetse takes a blood-meal on this mammal, there is a chance that the short and stumpy trypomastigotes are consumed by the fly. These enter the midgut of the fly and become procyclic trypomastigotes before dividing to become epimastigotes [58]. These travel to the salivary glands, where they become attached to the salivary gland epithelium [59]. Some will transform back into short and stumpy trypomastigotes that can be injected back into another mammal upon biting for further transmission. Trypanosomes in tsetse do negatively affect the salivary gland function, but the overall impact is limited; and once infected, tsetse stay infected for life [60].

Tsetse also have immune defences to resist infection, such as the production of hydrogen peroxide, which damages DNA, and are in fact relatively resistant to *Trypanosoma brucei* [61]. In addition, the probability of a tsetse will become infected will depend on the age, sex and species of fly. In particular, tsetse susceptibility will be affected by the teneral effect, whereby tsetse that have not yet consumed their first blood-meal will be more susceptible to infection than older flies [62, 63].

Other than tsetse bites, *Trypanosoma brucei* can be transmitted congenitally [64, 65], sexually [66], by blood transfusion or organ transplantation [67], and by laboratory accident [68]. However, all these forms of transmission are considered rare.

2.1.5 Clinical signs and symptoms of human African trypanosomiasis

Upon infection, there are two distinct stages of human African trypanosomiasis (HAT) [69]. Initially, a person with HAT will experience a fever, headaches, joint pain and itching. This is Stage 1 (or the haemolymphatic stage [70]). In this stage, the fever is intermittent, lasting between a day and a week, with intervals of days or months in between. Trypanosomes invading the circulatory and lymphatic systems can cause swelling of lymph nodes. A red sore may also develop around the initial tsetse bite, which is more common in the *rhodesiense* form of the disease [71]. Other symptoms such as anemia, endocrine, cardiac and kidney dysfunctions can occur.

Stage 2 of HAT is characterised by when the parasite has crossed the blood–brain barrier and so invaded the central nervous system [70]. This is the neurological stage. While fever is less common, the leading symptom is disruption of the sleep cycle. The infected human will suffer drowsiness in the daytime and urges to sleep, coupled with nocturnal insomnia [72]. Short, frequent periods of rapid eye movement sleep can also occur preceded by wakefulness, which are equally likely to occur in the day or night [73, 74]. These effects on the sleep cycling is where HAT gets its colloquial name of 'sleeping sickness.' Additional symptoms in Stage 2 include confusion, aggressive behaviour and weakness or paralysis of limbs. The disruption of circadian rhythms also affects hormonal secretion [75]. The neurological phase can cause irreversible damage, with the neurological and psychiatric symptoms increasing significantly with disease progression and will ultimately result in death for the vast majority of those infected and not treated, typically from meningoencephalitis [76].

Gambiense HAT and *rhodesiense* HAT demonstrate different timescales and intensities of these symptoms. *Gambiense* HAT is a chronic condition, where symptoms do not

always present themselves immediately. The untreated duration of infection before death is approximately 3 years, split roughly equally between Stage 1 and Stage 2 [77]. Conversely, in *rhodesiense* HAT symptoms can emerge within weeks, followed by death in a period of months; this form of the disease is both faster to develop and more virulent [78].

2.2 A brief history of human African trypanosomiasis

Trypanosomes have been transmitted between animals and tsetse, since the emergence of tsetse about 35,000,000 years ago [79]. This long coexistence explains why many wild animal species are unaffected by the infection, since they have evolved to become trypanotolerant [80]. Indeed, the fact humans are resistant to trypanosomes, other than *T. b. gambiense* and *T. b. rhodesiense*, suggests that trypanosomes have long infected humans, although human African trypanosomiasis (HAT) is a more recent development. Sustained transmission of *T. b. gambiense* between tsetse and humans has led to the parasite being less virulent, while in contrast, *T. b. rhodesiense*, which is mainly zoonotic with few human infections, remains more virulent [80, 81].

2.2.1 Discovery and early history

While there is some evidence that animal African trypanosomiasis (AAT) was reported in 2000 BC in ancient Egypt [82], some of the first written accounts of human African trypanosomiasis (HAT) came from ship doctors and medical officers working for slave-trade companies. The first accurate medical report of HAT was published in 1734 [83], however this focused on the more pronounced Stage 2 symptoms and it was not until 1803 that work was published on Stage 1 of the disease [83].

In 1852, it was first noticed by the explorer David Livingston that cattle died from AAT after being bitten by tsetse, hence establishing the link between the disease and the flies [84]. Yet, the cause of trypanosomiasis, the trypanosome was not known until

microbiologist David Bruce identified that *T. brucei* resulted in AAT in cattle in 1895 [85]. Several discoveries about trypanosomiasis quickly followed this, when surgeon Robert Michael Forde observed trypanosomes in human blood in Gambia in 1901 [86], which were identified as trypanosomes in 1902 by physician Joseph Everett Dutton and given the name *Trypanosoma gambiense*, now *T. b. gambiense* [87]. Pathologist Aldo Castellani was the first to suggest that trypanosomes were the cause of HAT, when, in 1902, he observed them in the cerebrospinal fluid of patients [88]. In 1902, Émile Brumpt made the first link between tsetse and HAT when he noticed the geographic distribution of the flies and the disease was the same [84]. Building upon this in 1903, David Bruce suggested that trypanosomes were actually transmitted by tsetse [79], and in 1909, Fredrich Karl Kleine, proved the cyclical transmission from tsetse [89], which prompted Bruce to describe the full life cycle of trypanosomes between tsetse and animal host [83]. *T. b. rhodesiense* was only identified as a different human pathogenic species of trypanosoma in 1910 [90].

2.2.2 Infection in the twentieth century

In the nineteenth century, there were a relatively small number of human African trypanosomiasis (HAT) cases [84]. This changed at the end of the century, with the invasion of European explorers, missionaries and soldiers. In particular, large numbers of Belgian colonial personnel travelled to the present day Democratic Republic of the Congo (DRC), such that King Leopold II of Belgium was allowed to formally acquire the rights to the Congo territory in 1885, which became known as the Congo Free State [91].

The colonial expansion caused movement of people into new territories. The imposition of taxes, in particular, pushed people to expand crop farming into new tsetse-infested lands. Colonial personnel with no knowledge of the dangers of tsetse, also forced the establishment of new villages in many high-risk locations for HAT and continued to explore these areas with the construction of roads and railways. The increased transport links and movement, combined with forced labour, introducing further interactions of people both with each other and tsetse flies, increased the probability of disease transmission dramatically. This was further exacerbated by severe droughts driving

both people and tsetse to converge of water sources, and a panzootic of rinderpest killing much livestock and wildlife, forcing tsetse to feed more frequently upon a more dangerous food source — humans [84].

These factors led to a severe epidemic of HAT in 1896–1906, which caused an estimated 800,000 deaths across the Congo basin and Uganda [92]. In total, it was thought up to half the population living along the lower river Congo died of either HAT or smallpox in this time [93]. These deaths forced the colonial administrations to invest in medical research into HAT.

In 1905, the first drugs discovered to be effective in treating HAT were found. These were organic arsenicals called atoxyl, which had mixed success, since while trypanocidal and effective in reducing parasitaemia, they were less beneficial for patients in Stage 2 of the disease and were also relatively toxic themselves [79]. By 1919, a less toxic and more effective compound called tryparsamide was discovered, which was also able to treat Stage 2 and was particularly effective in combination with another drug called suramin. These drugs helped to treat the disease and also to reduce case numbers when a second HAT epidemic for the twentieth century occurred in 1920–1945 [79].

One of the biggest events in the pursuit of reducing HAT cases was the development of systematic case detection and treatment. This simple principle was devised by Eugène Jamot in the 1920s and was based on for the idea that for elimination of the parasite, cases detection and treatment should be carried out by mobile teams that should cover the highest possible percentage of the population at risk of HAT using only specifically trained people with autonomy from other HAT services. In Cameroon, the special services implemented by Jamot lead to a reduction in HAT prevalence from 60% in 1919 to 0.2–4.1% in 1930 [94]. This encouraged other colonial powers to introduce these methods for HAT control.

In addition to systematic case detection, systematic vector control was also introduced from as early as 1911. This was performed with tsetse traps and carrying cloths coated in the sticky substance, birdlime, as well as the culling of animals to reduce the potential reservoir of animals that could carry the disease [84]. Later, post-World War II,

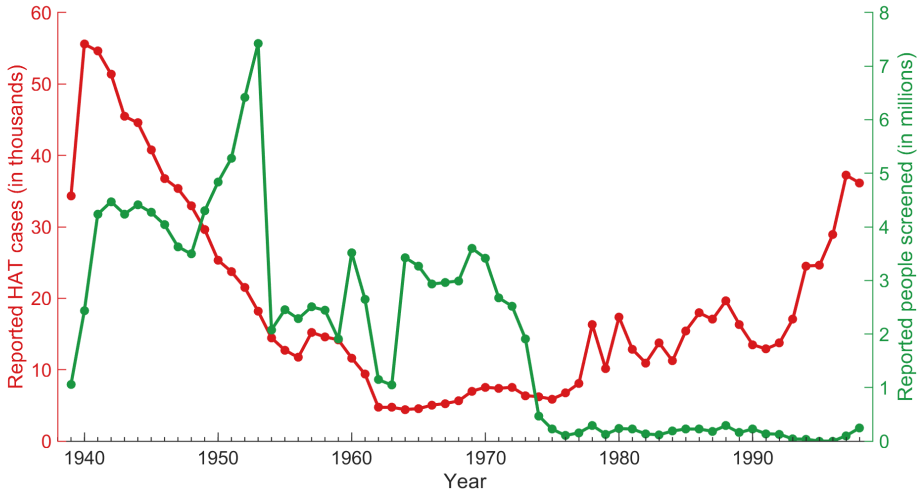


Fig 2.2: Cases and number of people screened for human African trypanosomiasis reported to the World Health Organization between 1939 and 1998. Data taken from a WHO report [95].

insecticide spraying was also used. These activities had some success in reducing the tsetse population. Hence, helped by other HAT control activities, the transmission of HAT reduced, resulting in a steady reduction of reported HAT cases after the end of the second HAT epidemic, which ended in the 1940s (Figure 2.2).

While in the nineteenth and early twentieth centuries, colonialism had caused much disruption and displacement of people and in many cases inducing famine, which meant a rise in HAT cases, by the 1960s there had been progress in both medical practices, disease control and general hygiene such that HAT case numbers were low and no longer considered a priority. When endemic countries became independent (the current DRC became independent and was no longer given the name Belgian Congo in 1960), the lack of funding for health services and political instability meant that case detection programmes were dramatically scaled back. The screened populations returned to the previous very low numbers (Figure 2.2). In 1991, the Belgian government withdrew all bilateral aid for the DRC, which included support for the national control programme (Programme National de Lutte contre la Trypanosomiase Humaine Africaine, PNLTHA), as part of sanctions against the President of the DRC, Mobutu Sese Seko [96].

Thus, with the lack of control, HAT started to have a resurgence from around 1970, and particularly after 1991, with the third epidemic of the twentieth century declared

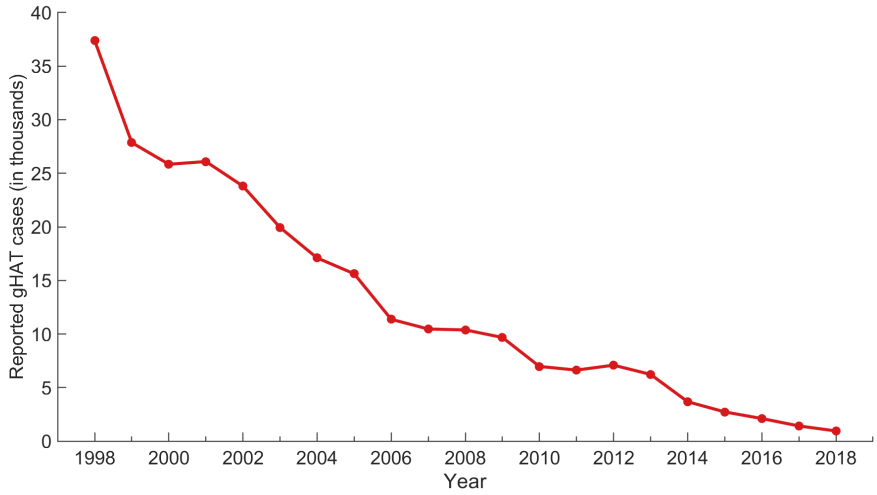


Fig 2.3: Cases of *gambiense* human African trypanosomiasis (gHAT) reported to the World Health Organization between 1998 and 2018. Data taken from WHO's Global Health Observatory data repository [20].

.

from 1970 until the late 1990s. By just before the turn of the century, case numbers had returned to levels not seen since the 1940s (Figure 2.2).

2.2.3 Recent history

While the number of annual reported HAT cases has varied dramatically in the last century as a consequence of different levels of investment in control [95] (Figure 2.2), by end of the 1990s, it was clear that HAT cases would not return to the lows of the early 1960s unless control programmes were reintroduced. In 1997, 29 mobile teams resumed screening for HAT in the villages of the DRC [97]. Thus, HAT again became a priority with coordination between the World Health Organization, national HAT control programmes, funding agencies, industrial partners, and non-governmental organisations [13]. This has improved the support of control activities within HAT-endemic countries with better surveillance and access to diagnostic tools and treatments [12]. This reinvestment, coupled with advancements in diagnostics and drugs to treat the infection, as well as plausible elimination strategies, has led to a steep decline in gHAT cases (Figure 2.3).

2.3 Current status and eradication efforts

The current drive for a reduction in human African trypanosomiasis (HAT) cases stems from the epidemic at the end of the twentieth century, when there were approximately 38,000 reported cases of HAT in 1998 and a further 300,000 estimated to be additionally unreported and hence untreated [98]. This initial reinvestment was largely funded by the resumption of Belgian bilateral cooperation [96].

With new case numbers for *gambiense* HAT (gHAT) now reported as below 1,000, eventual elimination of the diseases appears to be more of a possibility than ever [99] (Figure 2.3).

2.3.1 Current status

Annual newly reported gHAT cases have been steadily falling since 1998, in large part due to the case detection and treatment systems implemented by national HAT control programmes (PNLTHAs, Programme National de Lutte contre la Trypanosomiase Humaine Africaine) with support from the World Health Organization (WHO), among other stakeholders (Figure 2.3). Due to this initial success, the WHO set goals such that this progress could be sustained, building towards the ultimate goal of eradication.

The Neglected Tropical Diseases (NTD) Roadmap, published in 2012, identified HAT as a candidate for elimination as a public health problem [100]. The was formalised as a goal in 2013, with the elimination definition, comprising of two global indicators, updated in 2017 to: (i) fewer than 2,000 reported cases per year, and (ii) reducing the area at risk of reporting more than 1 case per 10,000 people per year by 90% as compared to the baseline for 2000–2004. The Roadmap is due to be updated in 2020 and is proposed to include the goal of zero reported gHAT cases by 2030 [11, 16, 101].

The first indicator for elimination as a public health problem is very likely to be met by 2020, since it is currently being achieved, with 977 cases reported to WHO (of which 953 were gHAT) in 2018, well below the target of 2,000 [20]. The second indicator is

more challenging to assess; in 2014–2018, 195,000 km^2 of land was estimated to be at moderate risk or higher of HAT, a reduction of 72% compared to the baseline 790,000 km^2 from 2000–2004 [99]. Therefore, the 90% reduction was not met by 2018, but the progress is encouraging, with a continued decline in the at-risk area; the size of this area was close to the milestone aim of 151,000 km^2 for 2014–2018 [100]. It is expected that these downward trends are indeed reflecting a real decline in transmission, rather than simply under-reporting, since the number of health facilities providing HAT surveillance, diagnosis and treatment is increasing [102].

There is less evidence that the 2030 elimination of gHAT transmission target [101] will be met, but it remains an important aspirational goal to ensure that progress is sustained and that the previous mistakes of early cessation of HAT programmes are not repeated in the twenty-first century [103].

2.3.2 Interventions

To attain the targets set by WHO and break the transmission cycle, HAT interventions need to be applied effectively at all levels, with an understanding of the geography, the community and field workers, the technology, and the governance [104]. However, to be able to implement intervention strategies successfully, there also needs to be adequate surveillance; this allows both for long-term disease monitoring and early identification of outbreaks [105]. HAT surveillance is recorded as screening and incidence data by the WHO in the Atlas of Human African Trypanosomiasis [106, 107]. This is a systematic approach of collating the number of new cases in villages across the endemic areas in each year, as well as the number of people screened for HAT and a census estimate. This data allows for the production of disease-risk maps, monitoring, planning of future surveillance and data for modelling and making future predictions [108].

Active screening

Since there is no vaccine or chemoprophylaxis for HAT [109], case control for gHAT is primarily through direct case detection by mass screening, followed by confirmation and

treatment [110]. This has widely been considered to be the most effective method of control, even since the early twentieth century [111–113].

Active screening is implemented by the operations of small mobile teams including microscopists, secretaries, drivers, messengers, guards, and health workers, who travel directly to villages in the HAT-endemic areas in four-wheel drive vehicles (or boats) and aim to test full populations for the infection through mass screening [97]. The region a mobile team can cover may include a population of up to 800,000, and teams typically travel for twenty days each month to conduct active screening, staying for multiple days in some large villages to ensure the available population is screened [97]. The choice of villages visited is dependent on the history of screening and cases in the village and area, the cases from nearby health centres, and local information [97]. Other predictive methods to identify at-risk villages are being devised [114].

For active screening programmes to be effective, a high screening coverage is important in all villages to ensure the detection and treatment of those infected and prevent onward transmission. This requires a detailed knowledge of the area with notification ahead of an active screening, to ensure that the village is ready for the maximum possible number of people willing and available to be screened. It should not be on an inconvenient day, such as when there is a market [115]. Furthermore, there also needs to be low drug toxicity, low cost to the patients and some level of privacy to achieve high attendances and treatment uptakes [113, 116, 117].

There is limited evidence for how frequently these active screenings should occur and when they should stop if no cases are found. Van Nieuwenhove [118] recommends three repeated screening rounds with one year intervals, while Simarro *et al.* [119] used six month intervals. The current WHO recommendation is for yearly screening with three years of zero case reporting before stopping active screening in the village [110]. There have been many calls for the need to maintain active screening, even when no cases are observed [120, 121], particularly given the feedback between surveillance and control [122].

In low-prevalence settings, due to consistent under-representation of certain demographics [117], targeted door-to-door screening can be more cost-effective and less labour-intensive if there is knowledge of suspected cases, alongside the availability of diagnostic tests and treatments for individuals suffering symptoms. Indeed, door-to-door screening has been found to detect significantly more HAT cases than standard active screening [123]. In locations where the terrain is difficult to traverse, active screening is also carried out by light mobile teams using motorbikes [124].

Passive detection

To better detect cases, there needs to be additional support for those that are not reached by active screening. As such, passive surveillance provides fixed health centres with the capacity and tools required to test and treat for HAT [125]. This is crucial in areas with low transmission intensity that will not be targeted in active screening [126], which will become more common as total case numbers fall [13]. It also ensures that individuals who miss active screening events or receive a false negative result in previous active screening, can still access diagnosis and treatment. These facilities need to be suitably equipped, such that infections are recognised promptly [127, 128]. Since passive surveillance relies on individuals to self-present to these health centres, a high proportion of them will be in the late stage of the disease, with significant symptoms [129]. The ability to access facilities where HAT can be rapidly diagnosed shortens the time between infection and treatment, reducing potential transmission opportunities. Indeed modelling has suggested there is great potential in improving rates of passive case detection [130].

Vector control

Since gHAT is largely considered an anthroponosis, control has heavily relied on active and passive surveillance, directly treating humans, rather than considering the tsetse [13]. However, if vector control can reduce the number of tsetse, there will be fewer flies able to become infected, and hence a reduction in HAT transmission [131]. Vector

control is frequently a staple component of other vector-borne disease intervention strategies, such as malaria and dengue, due to its potential to avert transmission [132, 133].

Tsetse control can be implemented by traps, targets, insecticide-treated cattle, aerial spraying, or sterile insect release [134]. One current strategy showing considerable potential is the use of 'tiny targets' [109]. These are small blue squares of cloth attached to a square of mesh impregnated with the insecticide, deltamethrin. They are attached to a frame and either planted in the ground or hung from vegetation. Tsetse are attracted to the blue colour, circle the cloth and come into contact with the insecticide, resulting in their death [135–139]. These targets are both highly effective and easier to deploy than traditional devices [131] in settings where livestock density is low. In regions with higher cattle ownership, restricted application of insecticides can also be a cost-effective approach to reduce tsetse populations [140].

The two main drawbacks of vector control until recently, were the expense [109] and the associated logistics of repeatedly deploying multiple controls. However, with developments in insecticide-treated targets and traps [109, 137, 141], tsetse control can now be considered more cost-effective [142]. The smaller size of the targets has helped reduce costs, while remaining effective [109]. Furthermore, tsetse control is species-specific to tsetse and does not negatively impact the environment, since tsetse are not a key species in the food chain. Therefore, it can be considered ethically defensible, as human deaths are averted [143]; the objective is for local reduction of tsetse in HAT foci to interrupt transmission, rather than global eradication of the fly [144].

'Tiny targets' have been introduced in several HAT foci, such as Guinea, Uganda and Chad where reductions in the tsetse population of 80% in 18 months [145], 90% in 12 months [144], and 99% in 4 months [146] have been observed respectively. Furthermore, no gHAT cases have been found in areas where 'tiny targets' were deployed in North West Uganda [147]. When challenges are presented to health services, tsetse control is often easier to maintain than traditional medical interventions. For example, when active screening was postponed in Guinea due to the 2014–2016 Ebola outbreak, a rise in gHAT prevalence was observed; however, in the area where tsetse control had been

implemented, no cases were found [148]. Vector control is also now part of HAT control strategy in some high-burden areas in the DRC, the country with the highest HAT burden [20, 149].

2.3.3 Diagnostics

Medical treatment of HAT patients can cure them of the infection and so prevent suffering and potential death, however, early detection of the infection will also reduce the duration a person is infectious and able to transmit the infection to biting tsetse. Therefore, accurate diagnostic tools are essential to identify early stages of the infection and both prevent the severe symptoms for the individual and reduce further transmission to the population. Different diagnostics are available as field-applicable and laboratory-bound tests.

The most commonly used and reliable test for gHAT infection in the field is the card agglutination test for trypanosomiasis (CATT) [150]. This is a serological test developed in the 1970s, which uses blood collected from a finger prick, plasma, or serum [151]. The test is most suited to be carried out by mobile teams in active screening since it is relatively quick, inexpensive and reliable. However, the test does require an electricity supply, a cold chain and trained personnel [152]. A positive CATT test requires additional parasitological validation to visibly detect the presence of parasites by microscopy for HAT confirmation [153].

More recently, rapid diagnostic tests (RDTs) have been available to screen for gHAT. These tests have an important role in the fixed passive detection health centres, since they do not require electricity and are instrument-free [154]. This means rural hospitals, that are often ill-equipped, can still screen for HAT. Hence, RDTs have been widely distributed in remote endemic areas [154–156]. While these tests are being developed to have both high sensitivity and specificity (comparable to CATT) [157–162], in areas where the infection numbers are low, the number of false positives from RDTs can far outweigh the number of true positives, resulting in a very low positive predictive value [12, 163]. Cost-effectiveness analysis has suggested RDTs could be more cost-effective than CATT in both mobile and fixed health facilities [154].

For laboratory-bound tests, the trypanolysis test is a confirmatory test with extremely high specificity, such that positives from other tests can be verified and thus the patients treated. However, this test is expensive to perform and can only be done in selected laboratories in Europe and Africa [110]. Notwithstanding this, the trypanolysis test is particularly useful in the context of elimination since its high specificity means it can be used as a surveillance tool to identify areas which are disease-free [164, 165]. Enzyme-linked immunosorbent assays (ELISAs) can also be used as confirmatory tests with high specificity, but are time-consuming, expensive and need to be performed in large batches [166]. Molecular tests have also been developed to detect *T. b. gambiense* [167] and exhibit high sensitivities and specificities. The fact these tests are not directly applicable in the field yet, however, means the direct benefit remains limited [155, 168].

Rhodesiense HAT currently has no field-applicable serodiagnostic test [12], however the more obvious symptoms and high levels of parasitaemia make this less crucial for detection of the infection [110].

2.3.4 Treatments

Classically, because of the very different severity of symptoms and location of trypanosomes in the two stages of HAT, treatments are generally stage specific [169]. The earlier HAT is treated, the better prospects for the patient; drugs for Stage 1 will not cure a patient in Stage 2, and drugs for Stage 2 are unnecessarily toxic for patients in Stage 1. Hence staging is traditionally an important first step in determining whether the parasite has passed the blood–brain barrier into the central nervous system. This relies on a lumbar puncture to collect cerebrospinal fluid for the counting of white blood cells and to ascertain whether trypanosomes are present [170].

Until recently, the drugs used to treat Stage 1 infection were pentamidine or suramin [171]. Pentamidine has a high efficacy in treating gHAT and is administered intramuscularly for seven days, with generally minimal ill-effects [12]. Suramin is effective too, but is only used for rHAT as the slow intravenous infusion is more difficult to manage and the side-effects more frequent [12].

For Stage 2, the first-line treatment is nifurtimox–eflornithine combination therapy (NECT) (nifurtimox is delivered orally and eflornithine delivered intravenously) [172, 173]. This is an aggressive treatment with common side-effects including abdominal pain, vomiting and headaches, but a high probability of a successful treatment [12]. The alternative, melarsoprol, is now restricted to Stage 2 rHAT, due to the frequency of life-threatening reactions it can induce [174].

All five of the drugs are donated by manufacturers to WHO, who is able to freely distribute them across HAT-endemic countries [12].

In addition to these drugs, fexinidazole [175], an oral drug that is taken for ten days, was recently included in WHO guidelines for gHAT treatment [176] and approved for use in the DRC in December 2018 [177]. This drug is effective in treating both stages of gHAT, when the symptoms are not overly severe [178]; so eliminating the need for a painful lumbar puncture to determine the infection stage, and simplifying the treatment process whilst improving access to care [149, 177]. However, it is notable that a lack of stage determination provides less information for subsequent surveillance and can reduce the accuracy of recommendations from predictive models [179]. Fexinidazole appears less effective than NECT in treating late Stage 2 patients however [176], and the effect on parasites in the skin is still unknown [180]. The safety profile of fexinidazole is not sufficient to consider treatment without parasite confirmation as part of the diagnostic algorithm.

Another drug, acoziborole [181] is currently being trialled as a one-day, one-dose oral treatment for all gHAT patients. This could potentially revolutionise treatment due to the ease with which it would be delivered and has the potential to be administered to all at-risk populations based on RDT results, or even given to all high-risk individuals if suitable safety standards are met [180].

2.3.5 Considering elimination

There are many reasons to be optimistic about the eventual elimination of HAT: the declining trend in reported cases; the availability of accurate diagnostics; effective drugs

that are freely donated; new diagnostics and drugs being developed; and continuing operations to reduce infection numbers through both active and passive surveillance and tsetse control. However, as the case numbers decrease to very low levels, there will be more competition for funding with other diseases [182, 183] and activities will have to continue to avoid resurgence [13]. In addition, other factors may emerge that were undetectable at high prevalences but could pose problems for elimination and eventual eradication:

Firstly, all figures for HAT infections are based on reported case numbers and it is expected that the true number infected will be much higher. For rHAT in particular, with very low case numbers, there has been a decrease in HAT-skilled staff, causing a decrease in awareness and hence reporting as a consequence [102]. There is also an issue with systematic non-participation in screening for gHAT, where sections of the population are likely to avoid being screened [117]. Data on the age and gender of screened participants could be used to determine which groups are not attending screening, although this is not routinely collated in an electronic format. Anecdotal evidence suggests working age individuals are the least likely to participate, as they may be away from the village working when active screening teams visit. From the perspective of elimination, this is particularly troubling since this group is also more likely to be working in the tsetse habitat, close to the vegetation surrounding rivers. Hence, there could be a high-risk (core) group for infection never being tested. If screening stops in areas where there are no identified cases, transmission could be sustained by such a core maintenance population, which could reinfect those who have partaken in active screening [184, 185].

Without active surveillance that can reach high proportions of the at-risk populations, there is also the danger that gHAT could sustain itself in low numbers due to a possible asymptomatic reservoir of humans [186]. It has been observed that some individuals infected with gHAT do not present symptoms for a long time and so will not seek medical attention or be detected, as they are unaware of the infection [15]. These individuals may have the trypanosomes surviving in their skin with no blood parasitaemia, which is difficult to screen for in large numbers [180]. However, the parasite can still be ingested by tsetse and so transmitted [184, 187, 188]; modelling has suggested treatment of these

asymptomatic cases should be considered [189]. Gaps in active screening coverage for at-risk populations may also hinder elimination programmes [190], with high coverage needed to be maintained to prevent a decrease in detected cases being due to a decrease in screening effort [191].

Movement of infected people into disease-free areas should also be considered in intervention planning, as this can lead to recrudescence [192]. This is especially important in former-endemic areas, where HAT control is no longer considered a priority and high influxes of refugees could be an ideal environment for parasite transmission [193].

Finally, even if the *gambiense* form of the infection was eliminated from humans, there is the possibility that the transmission cycle could be preserved through animal reservoirs [194]. This is certainly the case for rHAT [12], but while *T. b. gambiense* infection exists in animals, it remains unclear if animal hosts are able to sustain infection or are likely to reinfect human populations [195]. Spraying livestock could prevent some transmission in domestic animals, but pockets of infected wild animals could still pose a problem. The existence of a *T. b. gambiense* infected animal on the island of Luba, where there have been no reported human cases since 1995 [196], also provokes wider questions about persistence in the absence of human cases and potential reintroduction from the animal reservoir [184, 197]. To achieve rHAT elimination, there will need to be multisectoral (One Health) cooperation, with impetus for improved surveillance of infection in both humans and animals [198].

HAT cases have declined substantially in the twenty first century due to considerable efforts to eliminate the diseases [20]. Elimination of transmission of gHAT has also been shown to be cost-effective, with economic benefits greater than the costs [182, 199]. Efforts need to be maintained to sustain the current decline in cases, with continued investment in diagnostics and treatment, as well as their implementation in active and passive surveillance, and tsetse control; even recent interruption of interventions has been known to lead to an increase in cases [200]. For rHAT, there are now only tens of cases, but completely eliminating transmission could be less achievable due to substantial zoonotic transmission. Despite over a century of study and data, there still

remain key unknowns concerning the biology and epidemiology which influence the likely success of the proposed elimination of these diseases [102, 184].

2.4 Mathematical modelling of human African trypanosomiasis

Mathematical models are an important part of efforts for HAT elimination and have been instrumental in identifying the likely effectiveness of different control strategies, along with providing evidence for the timescales on which elimination could occur, and answering many other questions, accounting for uncertainties [201]. Mechanistic models can be attuned to be an accurate representation of reality with expert information used to inform the implementation, such that the underlying biological processes are considered. They provide a framework for the spread of infection and help understanding of the processes involved.

Mathematical models should have some predictive power, with results able to be extrapolated to different scenarios and problems; the current process should also be described as accurately as possible, such that intuition for the process can be attained. As such, they should also give a solution. In addition, they must be useful in terms of the available data, such that parameter values should be identifiable in the fitting process or biologically justifiable. Models can provide conflicting advice on occasion, since the predictions are entirely dependent on the specific modelling assumptions and inherent uncertainties. However, they are a useful conceptual tool for how reality might behave and can be used to simulate scenarios before they are actually applied in real life.

Specific to human African trypanosomiasis, we want to use mathematical models to gain a greater understanding of the disease and related processes and provide evidence towards answering many questions the disease poses. These questions can be as simple as forecasting future prevalence and when we might expect disease elimination to occur given particular intervention strategies, or the lack of them, but also simulating the effects of reservoirs of infection (animal or human asymptomatic) as to hypothesise their biological importance. We could also predict how different diagnostics or treatments

would impact our objectives, or movement of people or influxes of refugees. Using an age-structured model, we could uncover which demographics are under-represented in different parts of the infection and treatment process. Or we could examine why in different locations HAT seems particularly persistent and so identify key areas for future interventions. We can also optimise intervention processes to project what ideal coverages, frequencies and cessation criterions of interventions might be. By considering economic and health costs, we can also proffer which strategies might be the most impactful or cost-effective subject to constraints.

Many of these questions have been considered in the pre-existing modelling literature and these thesis will also attempt to address additional questions.

2.4.1 Modelling infectious diseases

The mathematical modelling literature on infectious diseases is vast and we only provide a limited summary of some basics here. For more details see works such as Anderson & May [202], Diekmann & Heesterbeek [203], Keeling & Rohani [6], or Andersson & Britton [204]. As stated in Chapter 1, much of this work modelling infectious diseases stems from Ross [4] and Kermack & McKendrick [5] in the early twentieth century and has typically involved the use of compartmental models.

In a compartmental model, there are a number of distinct states and the number of units in each state is tracked over time. One of the most common models is the SIR model, where the number of people in three classes, susceptible (S), infected (I) and recovered (R), is tracked over time t. In its most simplistic form, we assume there are only two transitions between these classes: S to I, which describes the transmission of disease, such that a susceptible person becomes infected, and I to R, whereby an infected person recovers from the infection. The recovery transition occurs with recovery rate γ, which is the inverse of the infectious period, the infection transmission is given by the force of infection λ, the per capita rate at which a susceptible person becomes infected. The force of infection can be further decomposed to its constituent parts, depending on the assumption. Either $\lambda = \beta I/N$ (frequency dependence) or $\lambda = \beta I$ (density dependence),

where β is the product of the contact rate and the transmission probability, I is the number of infected (and so infectious) individuals, and N is the total population size.

Frequency dependence assumes the number of contacts is independent of population size, and more to do with social constraints than the size of a village, town or city. Alternatively, density dependence assumes the contact rate will increase with population size, as if there are more people in a small area, they will have more contacts. Typically, frequency dependence is used for vector-borne pathogens and where there are heterogeneous contacts, while density dependence is used for plant and animal diseases [6].

Deterministic formulation

These compartmental models were originally formulated as systems of ordinary differential equations (ODEs), where the equations can be written as:

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t) = -\frac{\beta}{N}S(t)I(t), \tag{2.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I(t) = \frac{\beta}{N}S(t)I(t) - \gamma I(t), \tag{2.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}R(t) = \gamma I(t), \tag{2.3}$$

noting the $S(t)$, $I(t)$ and $R(t) \in [0, N]$ are the numbers of susceptible, infected and recovered people at time $t > 0$. Since we do not consider any demographic events and the population size is constant, N, $S(t) + I(t) + R(t) = N$ for all $t > 0$ and so we can reduce the dimensionality of the system by ignoring $R(t)$, which equals $N - S(t) - I(t)$.

This ODE model is deterministic in that there is no randomness in the system and for a specified initial state, the model will always evolve in time in exactly the same way.

Stochastic formulation

In contrast, in a stochastic model there is randomness and repeated simulation of the system from the same initial condition will produce entirely different realisations of the

process, building up an ensemble of different solutions. In the ODE formulation you also have non-integer numbers of people in each infection class, whereas a stochastic formulation can be implemented that only deals with integer values. This is useful in that exact numbers of people are tracked and with the additional benefit that extinction of the disease can occur, while the number of infecteds in an ODE model can only get arbitrarily close to zero, but never completely die out.

One method of implementing a stochastic infectious disease SIR model is to use a continuous time Markov process. This uses a discrete state space for the states S, I and R, with events occurring at the points of independent Poisson processes. The rates are defined by the current state of the system and are as follows:

$$s \to s-1, i \to i+1 \qquad \text{at rate } \beta si/N, \tag{2.4}$$

$$i \to i-1, r \to r+1 \qquad \text{at rate } \gamma i. \tag{2.5}$$

This describes the process of a susceptible individual meeting infected individuals at points of a Poisson process with rate $\beta i/N$, and becoming infected for an exponentially distributed time with mean $1/\gamma$. Like the ODE model for a fixed population, we only need two dimensions for the system as $R(t) = N - S(t) - I(t)$.

One way of exploring the properties of this Markov chain is to directly simulate the process. A common method for simulating a stochastic epidemic is the Gillespie algorithm [205]. This algorithm works by firstly specifying the initial conditions of the system, possible events that can occur, and rates these events happen. Then, to simulate the epidemic, the time to any event is determined; this is exponentially distributed with a rate equal to the sum of the rates of all possible events. The event that occurs is determined by a probability in proportion to the rate this event occurs. These two steps are then repeated until a maximum time is reached, giving one realisation of the epidemic. The whole process can then be repeated multiple times to get a distribution of possible epidemics. For a more formal algorithm, see Algorithm 2.1.

Algorithm 2.1: The Gillespie algorithm.

Specify model run time T_{max};
Set initial conditions $t = t_0$ and $\mathbf{x}(t_0) = \mathbf{x_0}$;
while $t < T_{\text{max}}$ **do**
  Compute the event rates $R_i(\mathbf{x}(t))$,...,$R_n(\mathbf{x}(t))$;
  Define $R(\mathbf{x}(t)) = \sum_{i=1}^{n} R_i(\mathbf{x}(t))$;
  $\delta t \sim \exp(R(\mathbf{x}(t)))$;
  $r \sim \text{Unif}(0, 1)$;
  $a = rR(\mathbf{x}(t))$;
  if $\sum_{i=1}^{j-1} R_i(\mathbf{x}(t)) < a < \sum_{i=1}^{j} R_i(\mathbf{x}(t))$ **then**
    Event j occurs;
  end
  Update $\mathbf{x}(t)$;
  $t = t + \delta t$
end

2.4.2 Vector-borne disease models

Basic mathematical models, such as the presented SIR model (see Equations 2.1–2.3), assume that disease transmission occurs either by direct contact with an infected individual, or by contact with them through the air (such as a cough or sneeze). However, since this is not the case for human African trypanosomiasis (HAT), as the primary method for transmission is through a vector, an infected tsetse, we need to consider this biology in our models.

To consider this problem in a simple way, we can take our SIR model and remove the recovered (R) class, such the model becomes an SIS model, whereby upon recovery, an individual returns to the susceptible class and can become reinfected later. We can now add these two classes again for the vector population. As such, a simple vector-borne infection model would have four classes: S_H and S_V for susceptible humans and vectors respectively, and I_H and I_V for infected humans and vectors.

The force of infection for both humans and vectors is given by $\lambda_i = \alpha p_i I_j / N_H$, where $i \neq j$ and the subscripts can take the values of H or V. α is the biting rate of the vector, p_H and p_V are the probability that a vector biting a human will lead to infected in that

human or vector respectively, and N_H is the population size of the humans. Thus, all transmission is vector to human or vice versa. This gives the following equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}S_H(t) = -\frac{ap_H}{N_H}S_H(t)I_V(t) + \gamma_H I_H, \tag{2.6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I_H(t) = \frac{ap_H}{N_H}S_H(t)I_V(t) - \gamma_H I_H, \tag{2.7}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}S_V(t) = -\frac{ap_V}{N_H}S_V(t)I_H(t) + \gamma_V I_V, \tag{2.8}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I_V(t) = \frac{ap_V}{N_H}S_V(t)I_H(t) - \gamma_V I_V, \tag{2.9}$$

where we now have separate recovery/mortality rates for infected humans and vectors, γ_H and γ_V respectively.

Early work in mathematical epidemiology by both Ross and MacDonald [4, 206–208], has led to the standard Ross–MacDonald model in usage for vector-borne disease transmission today [202]. Many different modifications to this model have been used, considering additions such as human demography [209], vector life expectancy [210], latency periods [211], spatial heterogeneity [212], and many other variants.

2.4.3 Previous modelling work on human African trypanosomiasis

The first vector-borne disease transmission models were focused on the modelling of malaria [213], but these model, which are based on what is now known as the Ross–MacDonald model, have been adapted to serve as a model for human African trypanosomiasis (HAT). Some of the first HAT models were developed by Rogers [214] in 1988 for *T. vivax*, *T. congolense* and *T. brucei*. This extended the malaria model and made it more applicable to HAT by including two host species with variable efficiencies of transmission between host and vector and allowing for incubation and immune periods. Parameter values were taken for a typical village in West Africa, with some data obtained from East Africa [14, 214]. A subsequent model from Milligan & Baker [215], also in 1988, included enhanced susceptibility in teneral tsetse, disease-induced mortality and multiple host types, such as distinguishing between cattle and wild animals. Other early examples of HAT modelling include Artzrouni & Gouteux [216] and Baker *et al.* [217].

More recently, there have emerged several different and independent models for *gambiense* HAT [186, 218–221].

The model of Funk *et al.* [218] is a deterministic model composed of ODEs similar to the Ross–MacDonald model presented in Equations 2.6–2.9. However, instead of a single host species (humans), the model considers n different host species, with tsetse having a preference and some rate of switching that preference. This study gave one of the first estimates of the basic reproductive number for gHAT, as the average number of secondary cases generated by a single infected individual in an otherwise susceptible population [6], at approximately 1.1. The study also provided evidence for the existence of an animal reservoir by indicating that with humans and vectors alone becoming infected, transmission could not be sustained [218]. There is ongoing modelling work to assess the potential for recrudescence of gHAT from animal reservoirs [184] and some suggestion that elimination cannot be achieved or maintained without control of non-human reservoirs [222].

Four other models for gHAT were published more recently (post-2015), and have also been compared in a 2018 publication [221]. These models are referred to as the Warwick Model [146, 219, 223–225], the Yale Model [186, 224, 226], the Swiss TPH Model [220, 225] and the IDM Model. All four models are defined using ODEs, and so are deterministic (although in some cases use a stochastic framework [225]), and all are applied in the comparison publication of Rock *et al.* [221] (and later [179]) for the purpose of evaluating intervention strategies. As such, they all include simulation of passive detection and treatment of cases, active screening and and treatment at discrete events, and vector control to reduce the tsetse population. However, the underlying assumptions and implementation of model details are different, capturing different parts of the biology and so provide different quantitative predictions.

The IDM Model defines four classes for humans (susceptible, infected Stage 1, infected Stage 2, and recovered), while three for tsetse (susceptible, exposed and infected) and then fit the model using a collection of data from HAT-endemic foci using the WHO HAT Atlas [106, 107]. Applying this model to different endemicity settings was done simply by scaling the value of R_0.

The Swiss TPH Model [220] considers two separate sets of populations of tsetse and non-human animals to simulate the effects of humans moving between a village and a 'high-transmission area', such as around the bank of a river or a plantation. Humans are then divided into two groups as well, those that are low-risk and thus remain in the village at all times, and those that are high-risk and commute from the village to the 'high transmission area'. In this model, humans are classified into five compartments: susceptible, incubating (exposed), infective (Stage 1), removed (Stage 2) and treated (recovered). High-risk humans in the infective (Stage 1) class cannot be detected in passive and active surveillance. Tsetse are simply susceptible, exposed or infected; the flies have an intrinsic preference for biting animals or humans (and animal bites are assumed not to contribute to transmission).

This model suggests that if animal reservoirs do not contribute to transmission, the attractiveness of these animals to tsetse may act as a sink for their bites, reducing human transmission. In addition, increasing the intensity of screening and treatment interventions reduces the time to gHAT elimination, particularly if targeting the high-risk group. The model also suggests intensive vector control is likely to achieve elimination [220].

The Yale Model [186], does not partition the population into risk groups like the Swiss TPH Model, and instead achieves more heterogeneity through the use of an asymptomatic compartment. This means the human population is subdivided into classes: susceptible, exposed, asymptomatic, symptomatic Stage 1, symptomatic Stage 2 and recovered. After exposure, only a proportion of infected human hosts become symptomatic and progress through the infection stages, while a subset remain carriers and are less infectious and recover on average after the same time as the sum of the time spent in the two symptomatic classes. Tsetse are assumed to only be susceptible to infection for 24 hours after they emerge as an adult fly. Thus, tsetse are susceptible, exposed, then infected, or just susceptible and recovered for their lives.

This model has been applied to data from the Boffa focus in Guinea and predicts that for this focus, annual vector control, with complementary active case detection gives a high probability (90%) of achieving elimination as a public health problem in 2020,

but high coverage needs to be maintained to prevent re-emergence [186]. Subsequent analysis using this model has shown that in low-transmission areas, enhanced medical interventions could be enough to interrupt transmission if treatment coverage is over 70%, however in moderate and high transmission areas vector control is likely required for disease elimination with 10 years, with the probability of elimination decreasing with a higher tsetse migration rate, highlighting the importance of vector control [226].

The Warwick Model [219] uses five different compartments for classifying humans: susceptible, exposed, infected Stage 1, infected Stage 2 and recovered. However, like the Swiss TPH Model uses risk classes for humans. The human population is partitioned into high- and low-risk, whereby the high-risk population have a higher biting exposure, although the same probability of being infected per single infective bite as the low-risk group. High-risk humans do not participate in active screening, while low-risk humans participate randomly. Unlike the Swiss TPH Model, however, there is no division of the tsetse population between locations and the distinction between exposure to humans comes from the higher biting rate. Tsetse are divided into unfed; fed, yet uninfected; exposed; and infected classes. This is similar to the system in the Yale Model, however, rather than tsetse becoming recovered after 24 hours, they just have a reduced chance of becoming infected after their first blood-meal, simulating the teneral effect [63].

This model originally was used to be fitted using the WHO HAT Atlas [106, 107] for a model selection exercise on the impact of human risk classes and an animal reservoir. This found that variation in human risk and participation in active screening played a key role in disease transmission, but the role of an animal reservoir was unclear, while active screening campaigns were found to have been effective in reducing the number of new human infections [219]. Subsequently, this work has been expanded to many new contexts including projections for transmission in the Equator Province of the DRC [224], predicting the impact of improved surveillance and vector control in the Kwilu Province of the DRC [223], and modelling the reduction in infection from implementing tsetse control in the Mandoul focus in Chad [146].

In the comparison piece on these current HAT models, which were applied to both generic low- and high-risk gHAT settings, across the board vector control was shown

to be the most impactful intervention, with local elimination likely, even with only moderate efficacy [221]. The Swiss TPH and Warwick models predicted that targeted active screening averted the second most transmissions, while the IDM Model found enhanced passive surveillance to be more effective, with the Yale model displaying overlapping prediction intervals for these two interventions. Overall however, the four models, which were built independently with different assumptions, demonstrated qualitatively similar predictions [221].

Alternative modelling approaches, such as agent-based modelling have been considered [227, 228]. Grébaut *et al.* [227] implement an agent-based model on a village in Cameroon and suggest regular screening here could be enough to achieve elimination, while Alderton *et al.* [228] consider an rHAT transmission model that considers seasonality in the tsetse population. De Vries *et al.* [229] also consider the dynamics of HAT within a single village and present a method for forecasting future HAT prevalence using past observed prevalence levels and past screening activities in the village. The study suggests that annual active screening is only expected to lead to local elimination if at least half of the cases are detected during the screening rounds [229].

Additional to the direct modelling of gHAT, there is much existing modelling literature for tsetse. There have been several studies on the population dynamics of tsetse [51, 230, 231]. The studies are useful for modelling gHAT to ensure the tsetse components of the model are somewhat realistic, while the number of parameters is kept to a minimum. Furthermore, the effects of tsetse control have also been modelled, aside from considering HAT. This has involved considering how the population of tsetse might bounce back or crash after a certain level of control, the feasibility of local elimination and the cost-effectiveness of the intervention [50, 232]. Modelling of tsetse control using insecticide-treated cattle (in *rhodesiense* HAT-endemic areas) has also been considered [233, 234].

Developments in geo-statistics have also allowed maps of tsetse density to be produced to guide interventions [235, 236] and patterns to be observed in HAT distribution [237]. And since the climate is known to affect tsetse abundance [238], modelling has been used to predict where current HAT-free areas may become at risk, with Moore *et al.*

[239] predicting that 46–77 million additional people may be at risk of exposure by 2090.

Understanding of the dynamics of tsetse and infection by *T. brucei* is also pivotal in modelling for providing estimates of the various parameters that are required to make the modelling realistic. Estimates for the tsetse bite rate [240], tsetse incubation rate [241, 242], proportion of blood-meals taken on humans [243] can be found in the literature, along with the human mortality rate [244], incubation rate [214], progression from Stage 1 to Stage 2 rate [77, 245], and recovery rate [117].

3

Village-scale persistence and elimination of *gambiense* human African trypanosomiasis

Gambiense human African trypanosomiasis (gHAT) has recently seen a substantial decline in the number of globally reported cases, largely driven by an intensive process of screening and treatment. There continues to be pressure to reduce the number of new infections, as gHAT has been targeted for elimination by the World Health Organization. However, this infection is highly focal, continuing to persist at low prevalence even in small populations. Regional elimination, and ultimately global eradication, rests on understanding the dynamics and persistence of this infection at the local population scale.

In this chapter, we develop a stochastic model of gHAT dynamics, which is underpinned by screening and reporting data from one of the highest gHAT incidence regions, Kwilu Province, in the Democratic Republic of Congo. We use this model to explore the persistence of gHAT in villages of different population sizes and subject to different patterns of screening. In contrast to other commonly studied infections, our models demonstrate that infection is expected to persist for long periods even in relatively small isolated populations, despite very low prevalence. We further use the model to assess the risk of recrudescence following local elimination and consider how failing to detect cases during active screening events informs the probability of elimination. These quantitative results provide insights for public health policy in the region, particularly highlighting the difficulties in achieving and measuring the 2030 elimination goal.

3.1 Considering the village-scale

Mathematical models of gHAT have been beneficial in identifying the effectiveness of differing control strategies and predicting when elimination is likely to occur [201]. However, much of the modelling work on the gHAT infection dynamics has been done in large populations using deterministic models, either for an entire regional infection focus or at a health zone level (approximately 100,000 people) [186, 219, 220]. We focus on modelling gHAT infection within individual villages because we know that gHAT is such a focal disease, with hot-spots around the tsetse habitats and minimal incidence outside these areas; thus neighbouring villages can have entirely different experiences with the infection [18].

By modelling the dynamics of gHAT within villages, we are considering the infection dynamics at the level at which the interventions are being deployed, since active screening targets individual villages. To optimise deployment we need to understand the dynamics at this level. From this information, we hope to build understanding of how villages interact and so determine importation rates between villages, with the hope that the results can then be scaled up to an infection focus, or the national level to measure regional elimination, which is especially important as we approach zero transmission.

Some modelling work has been done on village level infections. De Vries *et al.* [229] presented a model for predicting HAT prevalence using past observed prevalence levels and past screening activities in the village. However, we have developed a new mechanistic model that captures details of the epidemiology and tsetse biology. This process is significantly more flexible and allows modification of model components, facilitating a number of different scenarios and control options.

3.1.1 *Gambiense* HAT in the villages of the Kwilu province

The Kwilu province in the Democratic Republic of the Congo (DRC) has many areas that are endemic with gHAT. This province (along with Mai-Ndombe to the north and Kwango to the south) used to form a larger province called Bandundu (from 1966–2015)

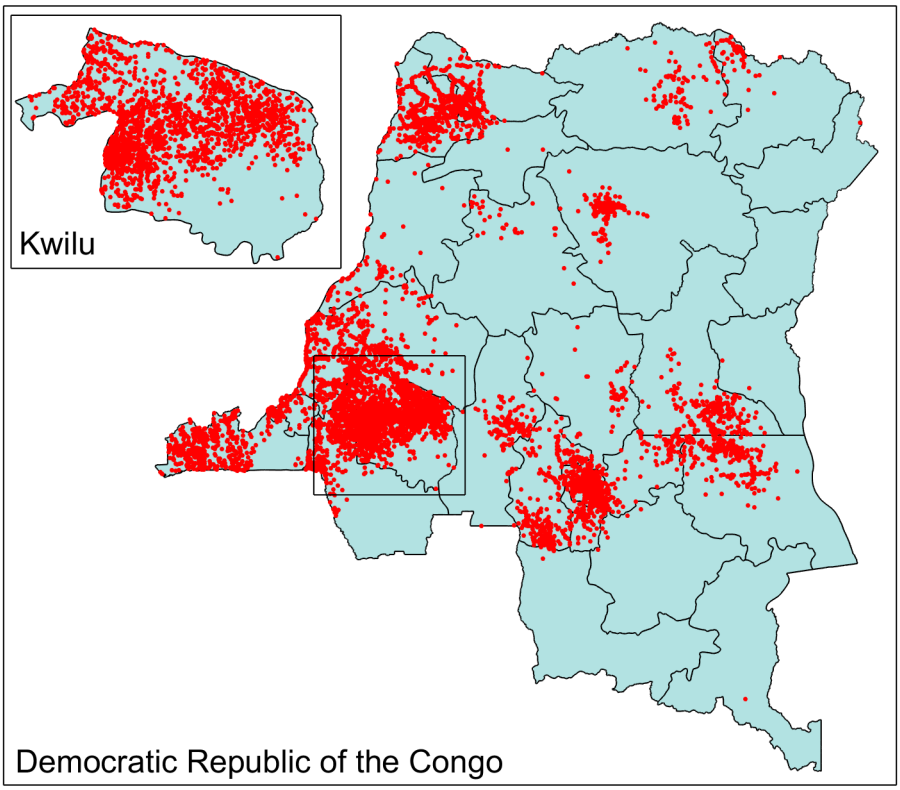


Fig 3.1: Locations of villages that have reported gHAT cases in the period 2000–2012. Inset map shows the Kwilu province, which has a particularly high endemicity. Data from the WHO HAT Atlas [106, 107].

and this province contained contained 46% of all reported gHAT cases in the DRC from the period 2000–2012 [246]. Indeed, there are a high proportion of gHAT cases still in this province (Figure 3.1).

For our model, we have chosen two health zones within the Kwilu province on which to base our analysis: Yasa-Bonga and Mosango. Health zones are subdivisions within a province that contain approximately 100,000 people; there are 519 health zones in the DRC [247]. There are many villages within Yasa-Bonga and Mosango with populations typically below 2,000 people, although there are a small number of larger villages with populations of up to around 12,000 (Figures 3.2 and 3.3A). Both Yasa-Bonga and Mosango contain rivers along which riverine tsetse inhabit, with villages along the rivers and further away.

These chosen health zones historically have a high incidence of gHAT and have had reported cases through both active and passive surveillance in the period 2000–2012.

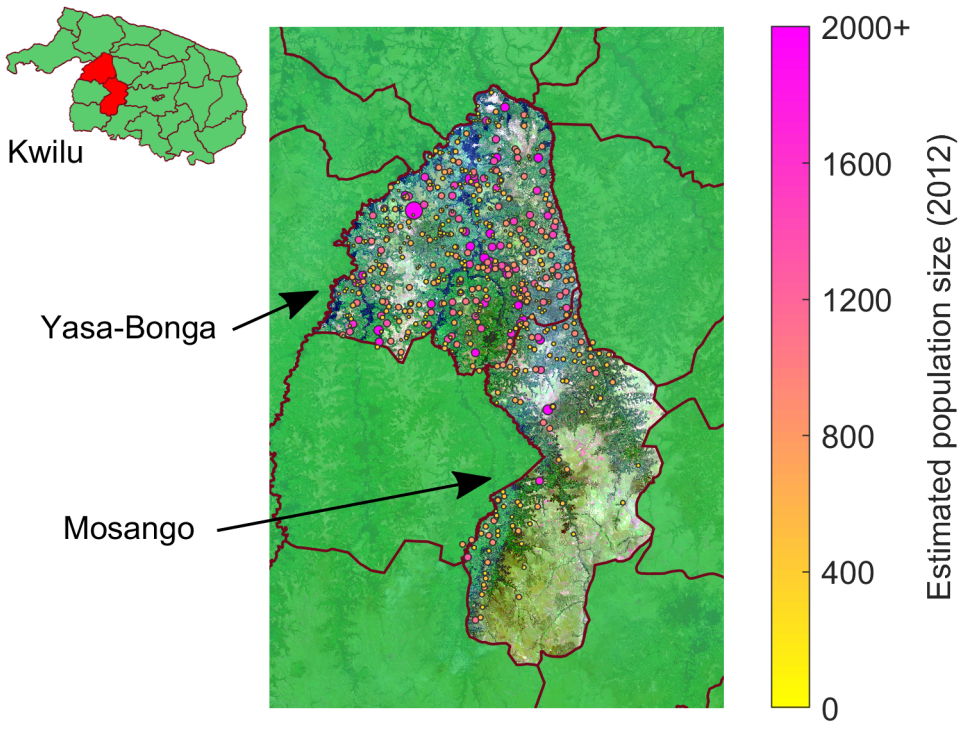


Fig 3.2: Locations of villages within Yasa-Bonga and Mosango at which an active screening has occurred [106, 107].

Screening coverages (the proportion of a village population that was subjected to active screening) varied between villages based on data from the WHO HAT Atlas [106, 107]. Typical screening coverages were in the reasonable zone of 20% to 90%, however there were some very high coverages, including some over 100%. For a full histogram of the coverages of all active screenings in Yasa-Bonga and Mosango in 2000–2012, see Figure 3.3B.

From the data, we have observed 2,701 active screenings in 2000–2012, across 559 different settlements, which resulted in 4,875 detected gHAT cases, with an additional 2,496 cases detected from passive surveillance. Population sizes for settlements were obtained from census estimates within the WHO HAT Atlas [106, 107] (Figures 3.2 and 3.3A); we account for an estimated yearly population growth of 2.6% [248]. Screening coverage from active surveillance is then calculated as the number of individuals screened divided by this population estimate in each year. Annual screenings larger than the estimated accessible population (yellow and red bars in Figure 3.3B) may indicate either multiple screenings in a calendar year, incorrect census estimates, or misreporting of individuals attending from neighbouring settlements. Low annual screening coverage (at

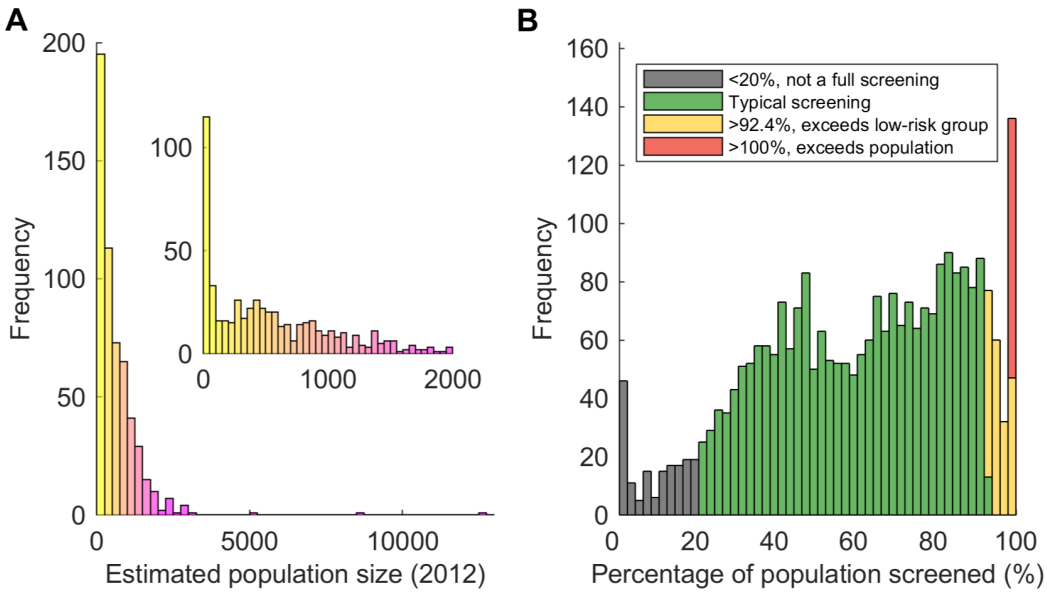


Fig 3.3: Histograms of population and screening data from the WHO HAT Atlas [106, 107]. (A) Estimated population size in 2012. (B) Active screening coverage for each active screening event 2000–2012.

less than 20% of the population size, gray bars in Figure 3.3B) is assumed to represent when someone has travelled from their village to a different village to be tested; in this instance, their home village has been recorded and represented in the data as a separate active screening. Therefore this is not considered as a complete active screening of a village. It is also notable that it may be more likely that someone who chooses to travel to an active screening may be more likely to be infected as they have noticed symptoms and hence this may also cause additional 100% screening coverages as well.

The full WHO HAT Atlas data [106, 107] includes information on the years active screenings occurred in each village, the full location of the village (geo-coordinates, health area, health zone, province), a census estimate, the number of people screened and the number of resulting detected cases. Information is also included on the number of passive cases detected in each year in a village.

3.2 A stochastic model of gHAT dynamics

To capture the village-scale infection dynamics for each individual population, we introduce a stochastic compartmental model. We choose to implement a stochastic

model rather than previous deterministic models because of the small population sizes we are considering. With a stochastic model we take into account the element of chance in infection extinction, rather than considering average behaviour; with small populations and the small number of infections, these events will be relatively magnified and so we consider different possibilities for the evolution of the dynamics. Since, the probability of local gHAT extinction is high within the time frames we want to consider, we can also model this directly and not consider thresholds below which the disease is effectively extinct. We adapt a model of Rock *et al.* [219].

3.2.1 Model selection

Model selection from Rock *et al.* [219] considered seven different model structures, all considered in an ODE framework, to determine the best fit to the gHAT incidence data from the WHO HAT Atlas [106, 107]. This was done using the deviance information criterion (DIC), which assigns low scores for models with high posterior mean log-likelihood and penalises models with more parameters [249]. The DIC was calculated by the following:

$$\mathrm{DIC} = -2\mathrm{LL}\left(\bar{\theta}\right) + 4\mathrm{Var}\left(\mathrm{LL}\left(\theta\right)\right),$$

where LL is the log-likelihood of unknown parameters θ, and where $\bar{\theta}$ is the expectation of these parameters.

The models considered were:

Model 1: Homogeneous human risk and behaviour model.

Model 2: Low- and high-risk populations for exposure to tsetse bites.

Model 3: Random participation with and non-participation in active screening.

Model 4: Random participation in active screening for low-risk populations and non-participation for high-risk.

Model 5: All combinations of low or high-risk and random or non-participation.

Model 6: Homogeneous human population and two animal populations – a reservoir able to acquire and transmit disease and a non-reservoir, which still receive tsetse bites.

Model 7: Heterogeneous human populations, as in Model 4, with the animal reservoir.

Model 4 and Model 7 were found to have the smallest DIC, with the best model being Model 7. The paper concluded that since the difference was negligible however, Model 4 should be used since DIC is known to over-favour over-fitted models and Model 7 has more parameters (associated with the animal component) [219]. Like this study, we therefore favour and use Model 4, without the animal component but with human risk structure, where high-risk individuals have a larger exposure to tsetse bites and do not present for active screening, while low-risk individuals have a lower exposure and present randomly for screening. We do not favour Model 7, since its role remains unclear [12, 146, 186, 218, 219] and its inclusion does not significantly improve the match between model outputs and currently available data in this setting [219].

Figure 3.4 shows the comparison of the models with the reported incidence data.

3.2.2 Model formulation

We adapted our new stochastic model using 'Model 4' from the work of Rock *et al.* [219]. The compartments are the same as this model, where humans are classified by: susceptible, exposed (but not infectious), Stage 1 infection, Stage 2 infection and hospitalised (or recovering at home). These variables are denoted by $S_{Hj}(t)$, $E_{Hj}(t)$, $I_{1Hj}(t)$, $I_{2Hj}(t)$ and $R_{Hj}(t)$, where $j = 1, 2$, with $j = 1$ for low-risk individuals, randomly participating in active screening, and $j = 2$ for high-risk individuals, never participating in active screening, each at time $t > 0$. We assume that there is natural mortality from all compartments, which leads to replacement of that individual as a susceptible in the population. Thus, the population size is assumed constant:

$$N_{Hj} = S_{Hj}(t) + E_{Hj}(t) + I_{1Hj}(t) + I_{2Hj}(t) + R_{Hj}(t), \quad j = 1, 2.$$

The given risk structure is used to capture the behaviour of the small proportion of individuals that both work in the habitat of many tsetse and so have a higher biting exposure and also do not partake in active screening, thereby acting as a human reservoir

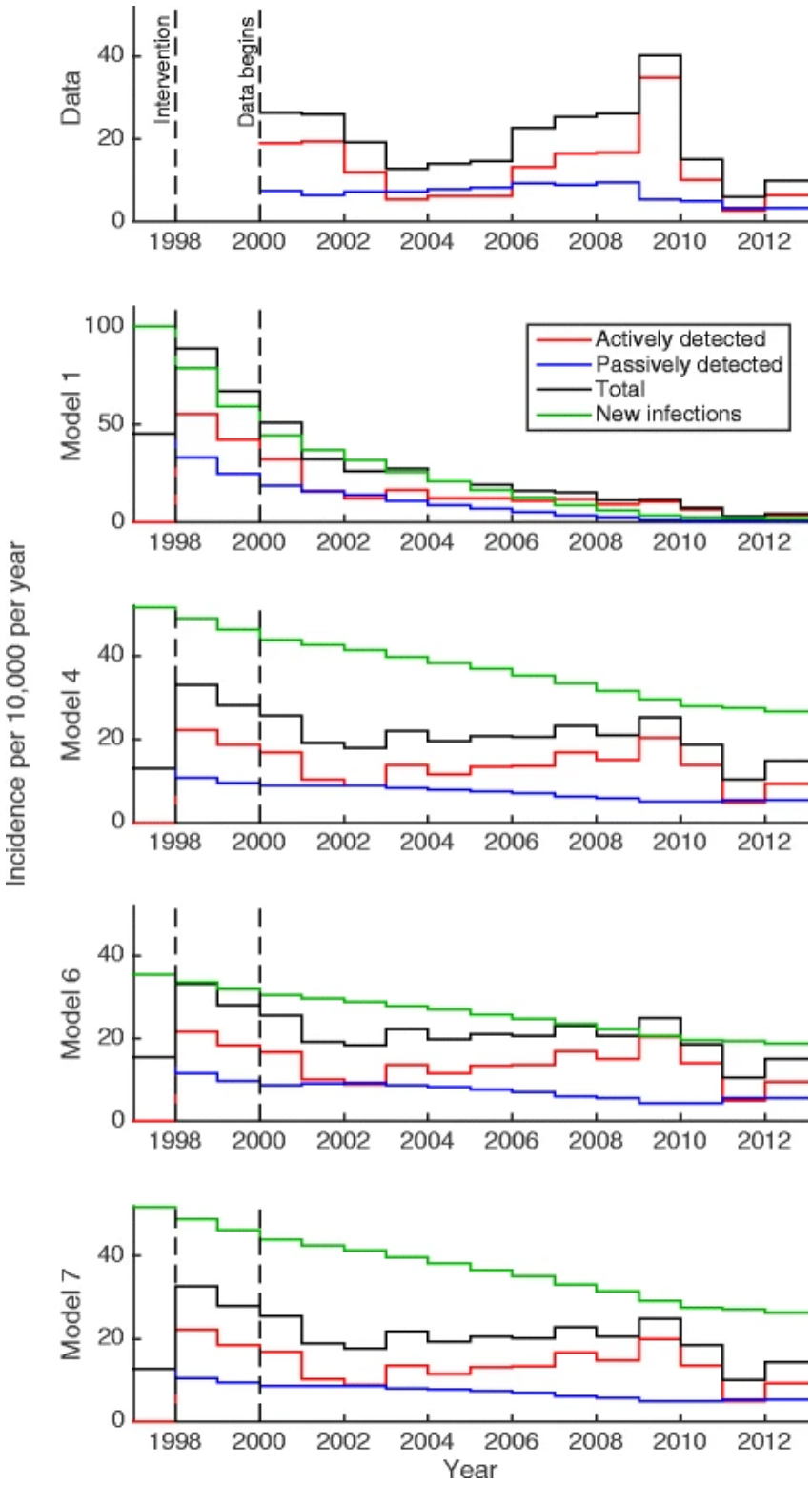


Fig 3.4: Deterministic model selection. Comparison of the reported incidence data aggregated across the study region and the corresponding model output over the years 1998–2012, for the deterministic model with different model structures. Model 4 is chosen as a good fit with the incidence data without unnecessary parameters. This figure is adapted from Rock et al. [219].

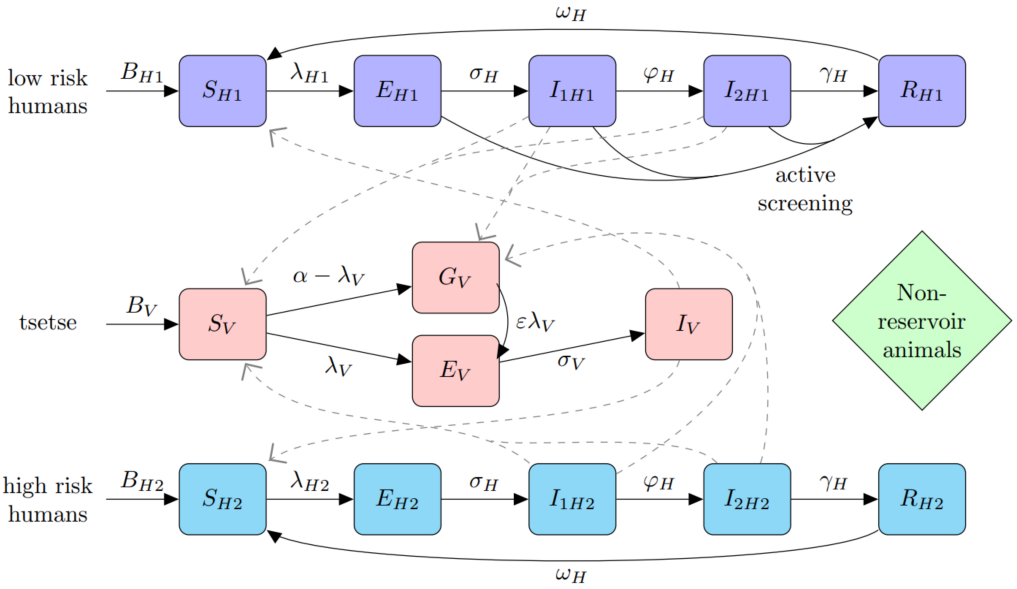


Fig 3.5: Compartmental model of gHAT infection dynamics in humans (low- and high-risk) and tsetse. The transmission of infection between humans and tsetse is shown by grey paths. Infected animals are not considered.

of infection [14, 117, 220]. The proportion of the individuals in the high-risk group is estimated through extensive model fitting to be 7.6% of the population for these health zones [219]. While the people in the high-risk group have a higher biting exposure, the probability of tsetse infection per single infective bite is the same for both risk groups.

Tsetse in the model are similarly compartmentalised into four epidemiological states: teneral (unfed); non-teneral yet uninfected; exposed (or latent); and infected. These classes are given by the variables S_V, E_{1V}, E_{2V}, E_{3V}, I_V and G_V, where the exposed class is divided into three classes, respectively. We make this distinction such that the time a tsetse remains in the exposed class is gamma-distributed [250]. Furthermore, the distinction between teneral and non-teneral yet uninfected is used to capture the observation that tsetse are far more susceptible to infection at their first blood-meal than at any subsequent blood-meals [63].

In considering the dynamics, humans can become exposed on a bite from an infective tsetse and progress through the infection stages, Stage 1 and Stage 2, before moving to the non-infectious class due to hospitalisation. The birth rates B_{Hi} are given by $\mu_H N_{Hi}$ for $i = 1, 2$ and B_V is equal to $\mu_V N_H$. Active screening moves exposed or infected people directly to the hospitalised class. Unfed tsetse can become exposed then infected on con-

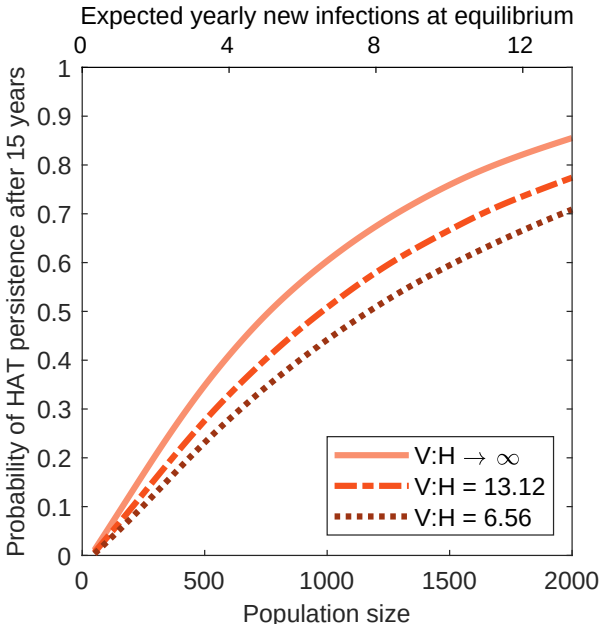


Fig 3.6: Impact of the parameters forming the effective density of tsetse, m_{eff}. When the probability of human infection per single infective bite is high the vector-to-host ratio, V:H, is around 6.56, whereas when the probability is low, V:H becomes large and deterministic models are valid for tsetse dynamics. The expected number of yearly new infections for each population size, if the system were at equilibrium, is also given by the top scale.

sumption of a blood-meal, of which a proportion are taken on humans. Alternatively, first blood-meals not resulting in exposure means tsetse are less susceptible to trypanosomes in future meals, where we define $\lambda_V = \alpha p_V \left(f_{H1} \frac{I_{1H1}+I_{2H1}}{N_{H1}} + f_{H2} \frac{I_{1H2}+I_{2H2}}{N_{H2}} \right)$. Parameters are defined fully in Table 3.2. $f_H = f_{H1} + f_{H2}$, where $f_{H1} = f_H R_1/(R_1 + rR_2)$ and $f_{H2} = f_H rR_2/(R_1 + rR_2)$.

The compartments of the model and possible transitions between them are shown graphically in Figure 3.5.

3.2.3 Tsetse density

One notable difference between the deterministic and stochastic formulation is the need to delineate the effective tsetse density term in the equations. This is the product of the relative tsetse density and the probability of human infection per single infective bite $m_{\text{eff}} = mp_H$. When using a deterministic model, Rock *et al.* [219] only needed to estimate the parameter value of m_{eff} since m and p_H are multiplied together in the

equations, however, for our stochastic model, m determines the number of tsetse being modelled for a particular village and so is needed explicitly. Therefore in this analysis, we show some sensitivity analysis to estimate the effect of varying p_H between 0 and 1 such that the value of m_{eff} remains fixed. For $p_H \to 0$, we take fixed $p_H = m_{\text{eff}}/m$ and $m \to \infty$. Hence we are assuming an effectively infinite population of vectors (i.e. the vector-to-host ratio, V:H $\to \infty$). From a modelling perspective, this reduces the equations down to the previously formulated deterministic approximation for the vector population. From the sensitivity analysis (Figure 3.6), we see there is minimal effect of varying the vector-to-host ratio, and hence for simplicity we choose to keep $p_H = m_{\text{eff}}/m$ and $m \to \infty$, to give the desired fixed value of $m_{\text{eff}} = 6.56$ (taken as the result from fitting the ODE Model 4 from Rock *et al.* [219] to case data from the WHO HAT Atlas [106, 107]).

In our analysis, the endemic disease equilibrium is calculated as the steady state of the deterministic model, excluding control measures other than basic passive surveillance [219]. Initial conditions for number of people in each infection class within a village are taken by sampling from a multinomial distribution, where the probability of being selected in a class is given by the proportion in that class in the steady state of the model. Using random initial conditions with the mean at endemic equilibrium removes the effects of rounding the initial conditions to the same integer values for each simulation.

3.2.4 Model equations

The model is formulated in two possible ways based on the choice of the vector-to-host ratio (Table 3.1). It is either entirely a Markov-chain if $p_H > 0$, or a Markov-chain for the human compartments with an additional ordinary differential equation component for the tsetse otherwise.

Tab 3.1: The transition rates of the Markov-chain gHAT-infection model and additional ordinary differential equation component of the model when $m \to \infty$ and $p_H \to 0$ (V:H $\to \infty$).

Class	Event	Transition	Rate
Humans, $j = 1, 2$	Recovery from hospitalisation	$s_{Hj} \to s_{Hj} + 1, r_{Hj} \to r_{Hj} - 1$	$\omega_H r_{Hj}$
	Natural death of hospitalised	$s_{Hj} \to s_{Hj} + 1, r_{Hj} \to r_{Hj} - 1$	$\mu_H r_{Hj}$
	Exposure of susceptibles	$s_{Hj} \to s_{Hj} - 1, e_{Hj} \to e_{Hj} + 1$	$f_{Hj}\alpha p_H s_{Hj} i_V / N_{Hj}$, $p_H > 0$ $f_{Hj}\alpha m_{\text{eff}} s_{Hj} I_V / N_{Hj}$, $p_H \to 0$
	Progression to Stage 1 infection	$e_{Hj} \to e_{Hj} - 1, i_{1Hj} \to i_{1Hj} + 1$	$\sigma_H e_{Hj}$
	Natural death of exposed	$s_{Hj} \to s_{Hj} + 1, e_{Hj} \to e_{Hj} - 1$	$\mu_H e_{Hj}$
	Progression to Stage 2 infection	$i_{1Hj} \to i_{1Hj} - 1, i_{2Hj} \to i_{2Hj} + 1$	$\phi_H i_{1Hj}$
	Natural death of Stage 1 infection	$s_{Hj} \to s_{Hj} + 1, i_{1Hj} \to i_{1Hj} - 1$	$\mu_H i_{1Hj}$
	Treatment or death from Stage 2 infection	$i_{2Hj} \to i_{2Hj} - 1, r_{Hj} \to r_{Hj} + 1$	$\gamma_H i_{2Hj}$
	Natural death of Stage 2 infection	$s_{Hj} \to s_{Hj} + 1, i_{2Hj} \to i_{2Hj} - 1$	$\mu_H i_{2Hj}$
	Infection importation	$s_{Hj} \to s_{Hj} - 1, e_{Hj} \to e_{Hj} + 1$	δs_{Hj}
Tsetse ($p_H = m_{\text{eff}}/m$, $m > 0$)	Exposure of teneral	$s_V \to s_V - 1, e_{1V} \to e_{1V} + 1$	$\alpha p_V (f_{H1}(i_{1H1} + i_{2H1})/N_{H1} + f_{H2}(i_{1H2} + i_{2H2})/N_{H2}) s_V$
	Exposure of non-teneral, uninfected	$e_{1V} \to e_{1V} + 1, g_V \to g_V - 1$	$\alpha p_V (f_{H1}(i_{1H1} + i_{2H1})/N_{H1} + f_{H2}(i_{1H2} + i_{2H2})/N_{H2}) \varepsilon g_V$
	Progression to exposed 2	$e_{1V} \to e_{1V} - 1, e_{2V} \to e_{2V} + 1$	$3\sigma_V e_{1V}$
	Death in exposed 1	$s_V \to s_V + 1, e_{1V} \to e_{1V} - 1$	$\mu_V e_{1V}$
	Progression to exposed 3	$e_{2V} \to e_{2V} - 1, e_{3V} \to e_{3V} + 1$	$3\sigma_V e_{2V}$
	Death in exposed 2	$s_V \to s_V + 1, e_{2V} \to e_{2V} - 1$	$\mu_V e_{2V}$
	Progression to infected	$e_{3V} \to e_{3V} - 1, i_V \to i_V + 1$	$3\sigma_V e_{3V}$
	Death in exposed 3	$s_V \to s_V + 1, e_{3V} \to e_{3V} - 1$	$\mu_V e_{3V}$
	Death in infected	$s_V \to s_V + 1, i_V \to i_V - 1$	$\mu_V i_V$
	Fed but not exposed	$s_V \to s_V - 1, g_V \to g_V + 1$	$\alpha(1 - p_V(f_{H1}(i_{1H1} + i_{2H1})/N_{H1} + f_{H2}(i_{1H2} + i_{2H2})/N_{H2})) s_V$
	Death in non-teneral uninfected	$s_V \to s_V + 1, g_V \to g_V - 1$	$\mu_V g_V$

Tsetse ($p_H = m_{\text{eff}}/m$, $m \to \infty$)

$$
\begin{aligned}
\frac{\mathrm{d}S_V}{\mathrm{d}t} &= \mu_V N_H - \alpha S_V - \mu_V S_V \\
\frac{\mathrm{d}E_{1V}}{\mathrm{d}t} &= \alpha p_V \left(f_{H1} \frac{I_{1H1} + I_{2H1}}{N_{H1}} + f_{H2} \frac{I_{1H2} + I_{2H2}}{N_{H2}} \right) (S_V + \varepsilon G_V) - (3\sigma_V + \mu_V) E_{1V} \\
\frac{\mathrm{d}E_{2V}}{\mathrm{d}t} &= 3\sigma_V E_{1V} - (3\sigma_V + \mu_V) E_{2V} \\
\frac{\mathrm{d}E_{3V}}{\mathrm{d}t} &= 3\sigma_V E_{2V} - (3\sigma_V + \mu_V) E_{3V} \\
\frac{\mathrm{d}I_V}{\mathrm{d}t} &= 3\sigma_V E_{3V} - \mu_V I_V \\
\frac{\mathrm{d}G_V}{\mathrm{d}t} &= \alpha \left(1 - p_V \left(f_{H1} \frac{I_{1H1} + I_{2H1}}{N_{H1}} + f_{H2} \frac{I_{1H2} + I_{2H2}}{N_{H2}} \right) \right) S_V \\
&\quad - \alpha p_V \left(f_{H1} \frac{I_{1H1} + I_{2H1}}{N_{H1}} + f_{H2} \frac{I_{1H2} + I_{2H2}}{N_{H2}} \right) \varepsilon G_V - \mu_V G_V
\end{aligned}
$$

Tab 3.2: Parameter notation and values for the stochastic model. We give the parameter notation (Par), a description, the value and the source. In the value column, d = days.

Par	Description	Value	Source
μ_H	Natural human mortality rate	5.4795×10^{-5} d^{-1}	[244]
ω_H	Human recovery rate	0.006 d^{-1}	[117]
σ_H	Human incubation rate	0.0833 d^{-1}	[214]
ϕ_H	Stage 1 to 2 progression rate	0.0019 d^{-1}	[77, 251]
γ_H	Exit rate from Stage 2 by treatment or death	0.006 d^{-1}	[219]
N_H	Human population size	Varies	N/A
μ_V	Tsetse mortality rate	0.03 d^{-1}	[214]
σ_V	Tsetse incubation rate	0.034 d^{-1}	[241, 242]
ε	Reduced non-teneral susceptibility factor	0.05	[219]
α	Tsetse bite rate	0.333 d^{-1}	[240]
m	Relative tsetse density	Varies	N/A
p_H	Probability of human infection per single infective bite	Varies	N/A
m_{eff}	Effective tsetse density	6.56	[219]
p_V	Probability of tsetse infection per single infective bite	0.065	[214]
f_H	Proportion of blood-meals on humans	0.09	[243]
R_1	Proportion of low-risk humans, randomly participating in screening	0.924	[219]
R_2	Proportion of high-risk humans, not participating in screening	0.076	[219]
r	Relative bites taken on high-risk humans compared to low-risk	6.6	[219]
u	Reporting probability for Stage 2 cases	0.26	[219]
δ	Importation of infection rate	3.4×10^{-6} d^{-1}	fitted

3.2.5 Parameters

Model parameters for the compartmental model are taken from Rock *et al.* [219] (Model 4) and applied to our stochastic model. The values of these parameters were taken from literature, where well-defined, and otherwise inferred by model fitting using a Metropolis–Hastings MCMC algorithm, which sought to match the deterministic model to observed cases and screening effort in Yasa-Bonga and Mosango; this is the process that was used in model selection. The values used in this thesis are the median of the distributions inferred using MCMC methodology applied to the aggregate annual data from Yasa-Bonga and Mosango. Therefore, the parameter values are specific to the study region, and the model is well-matched with the incidence data from active and passive surveillance. A full list of parameters is given in Table 3.2.

3.2.6 Active screening

Additional to the epidemiological and demographic processes, we simulate the effect of active screening and passive detection of cases. Passive detection (and disease-induced mortality) is assumed to occur at a fixed (per capita) rate for all Stage 2 infected individuals [129]. Active screening takes place annually; the proportion screened is either replayed from the historic pattern for that settlement or chosen randomly from the set of all screening coverages recorded, allowing a greater range of scenarios to be explored. Since Yasa-Bonga and Mosango are high endemicity regions, we assume that the screening coverage and frequency remains constant over time but note that these quantities are somewhat affected by population size (Figure 3.7).

Larger village populations tend to have more frequent screening, but with a lower proportion of the population tested; small populations tend to have high coverage but infrequent screens (Fig 3.7A and B). The smaller villages were also not typically screened in 2000–2008, while villages with populations greater than 500 were screened over the whole data period 2000–2012 (Fig 3.7C). We do not account for these differences in our analysis but can use these trends to help explain any discrepancies between models and the data.

Individuals in the low-risk group are selected uniformly at random for screening, irrespective of epidemiological status, and those that are found to be infected are moved to the hospitalised class. We assume that screening only applies to low-risk individuals, such that screening coverages greater than 92.4% (the estimated proportion in the low-risk group) [219] are truncated (Fig 3.3B). In the field, the diagnostic process is complex and multi-stage [240]; however, in the model, we collapse this into characteristics for the whole algorithm, which is assumed to be 91% sensitive [252] but 100% specific. False negatives remain undetected in the settlement, but by assuming 100% specificity, there are no false positives since we assume confirmation by microscopy will be carried out due to the low case numbers.

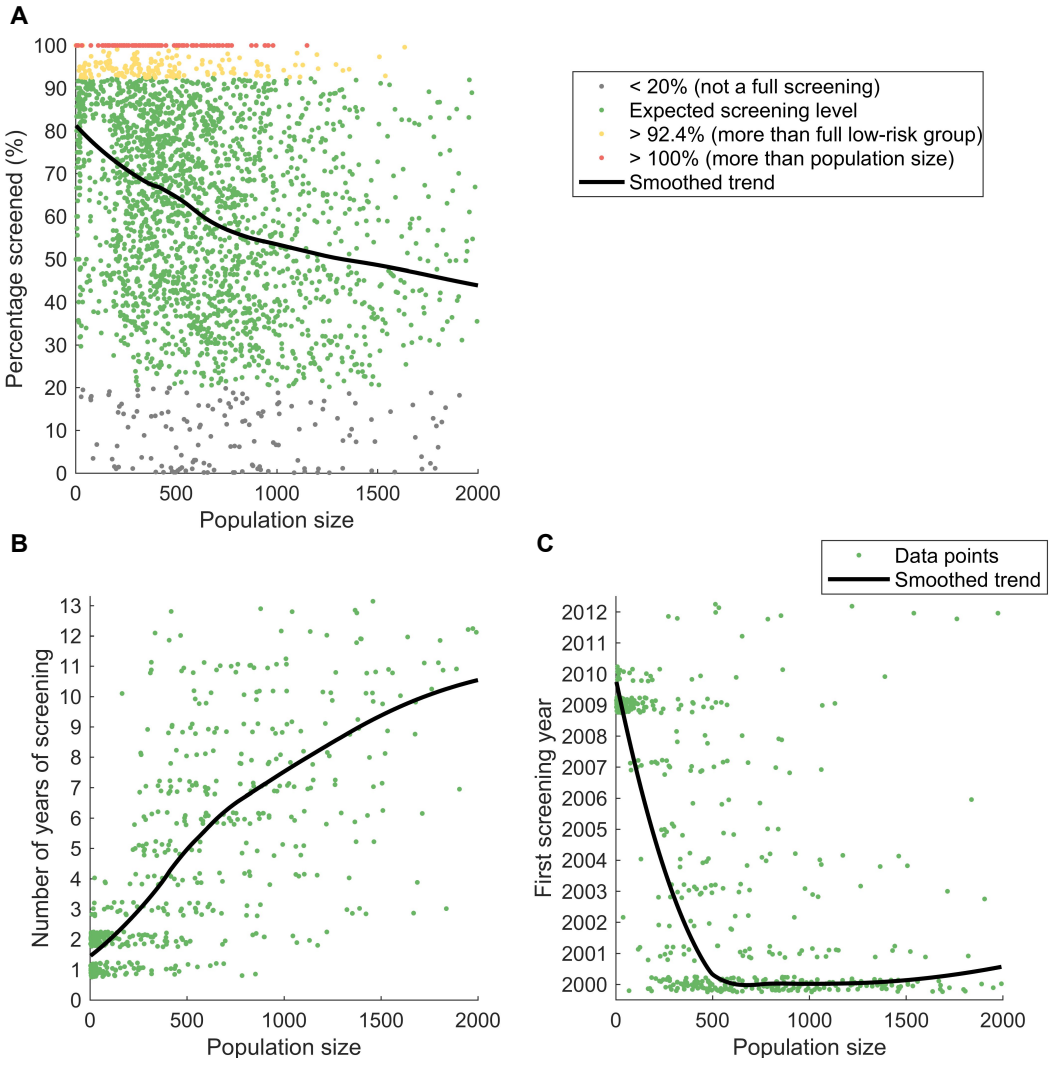


Fig 3.7: Differences in screening between villages. (A) Screening coverage in village populations. (B) Number of screenings at each settlement in the period 2000–2012. Small random perturbations are made to the integer values for visibility. (C) Years in which the first active screenings occur in the villages (within the 2000–2012 period), with small random perturbations for visibility.

3.2.7 Model implementation

To implement this model and generate realisations of the infection process, we employ the tau-leaping algorithm [253]. This algorithm is an approximation for the Gillespie algorithm [205, 254] that can improve the speed of the simulation at some loss of precision. While the Gillespie algorithm simulates the time to the next event, the tau-leaping algorithm 'leaps' forward by some time step and estimates how many of each event occurred in this time interval. To get the best results, we can have an adaptive time interval τ, such that the change in the state of the system is small, however, in this case, we will choose to keep τ fixed and small enough such the value of this parameter does not affect the results. For the full tau-leaping algorithm see Algorithm 3.1.

Algorithm 3.1: The tau-leaping algorithm.

Specify model run time T_{max};
Set initial conditions $t = t_0$ and $\mathbf{x}(t_0) = \mathbf{x_0}$;
Label all events E_1,...,E_n;
while $t < T_{\text{max}}$ **do**
  Compute the event rates $R_i(\mathbf{x}(t))$,...,$R_n(\mathbf{x}(t))$;
  Choose time step τ;
  for $i = 1 : n$ **do**
    Compute occurence of event i $k_i \sim \text{Poisson}(R_i(\mathbf{x}(t))\tau)$;
  end
  Update $\mathbf{x}(t)$;
  Ensure non-negativity of $\mathbf{x}(t)$;
  $t = t + \tau$
end

Our fixed step size for the tau-leaping algorithm was chosen to be 0.1 days. However, the results for persistence of gHAT after 15 years are qualitatively the same for time steps of 0.01 and 1 days, so we are confident that a time step of 0.1 days is not introducing numerical errors (Figure 3.8). 95% confidence intervals are all small, in the range (0.0044, 0.0196) about the plotted mean values.

Using this algorithm, a sample realisation of the infection dynamics in a population with three active screening events, where randomly sampled infected low-risk individuals are moved directly into the hospitalised class, is shown in Figure 3.9.

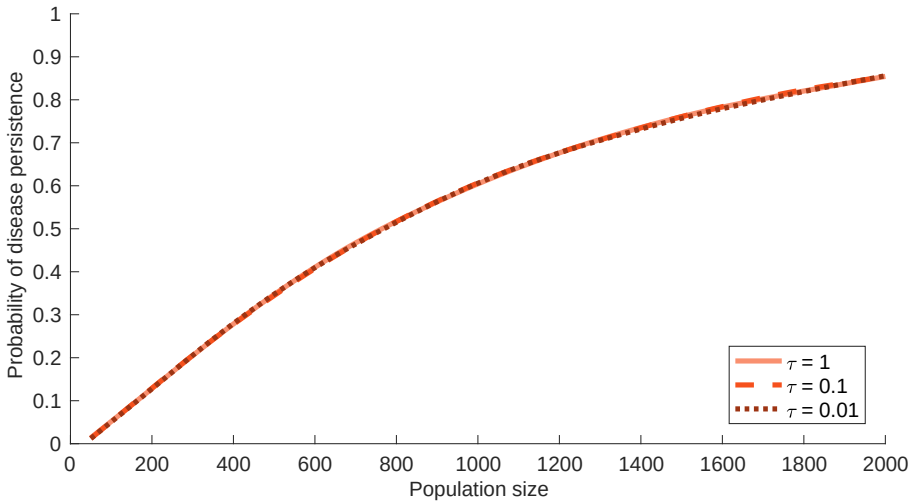


Fig 3.8: Probability of disease persistence under the tau-leaping algorithm for fixed step size. Probability of HAT persistence after 15 years for V:H $\to \infty$, simulated using three different tau-leaping time steps: $\tau = 0.01$, 0.1 and 1. Since the resulting curves are the same, the step size in the algorithm is suitable.

3.2.8 Population dynamics

For the majority of this thesis, we model the dynamics as a closed population, without emigration or immigration, so that once the disease has gone extinct in a population it cannot be re-introduced. This removes a critical dependency in model formulation, and greatly simplifies the presentation of results. In reality, no population is ever completely isolated; however we later show that the expected rate of infectious imports is very low and does not affect our main results or conclusions (see Figure 3.18).

In all model simulations, we have assumed that the population size remains constant, such that when a death occurs, the individual is replaced by a new susceptible individual immediately. We have analysed this assumption by comparing this with different population dynamics (Fig 3.10). We consider four additional population assumptions, all where births and deaths have been decoupled. These are: births and deaths are proportional to the number of individuals of that infection status; deaths are proportional, but births are at the constant rate from the initial equilibrium; and the same two mechanisms, but where the birth rate is increased by the population growth of 2.6%. As the populations change in size, we keep the effective tsetse density constant.

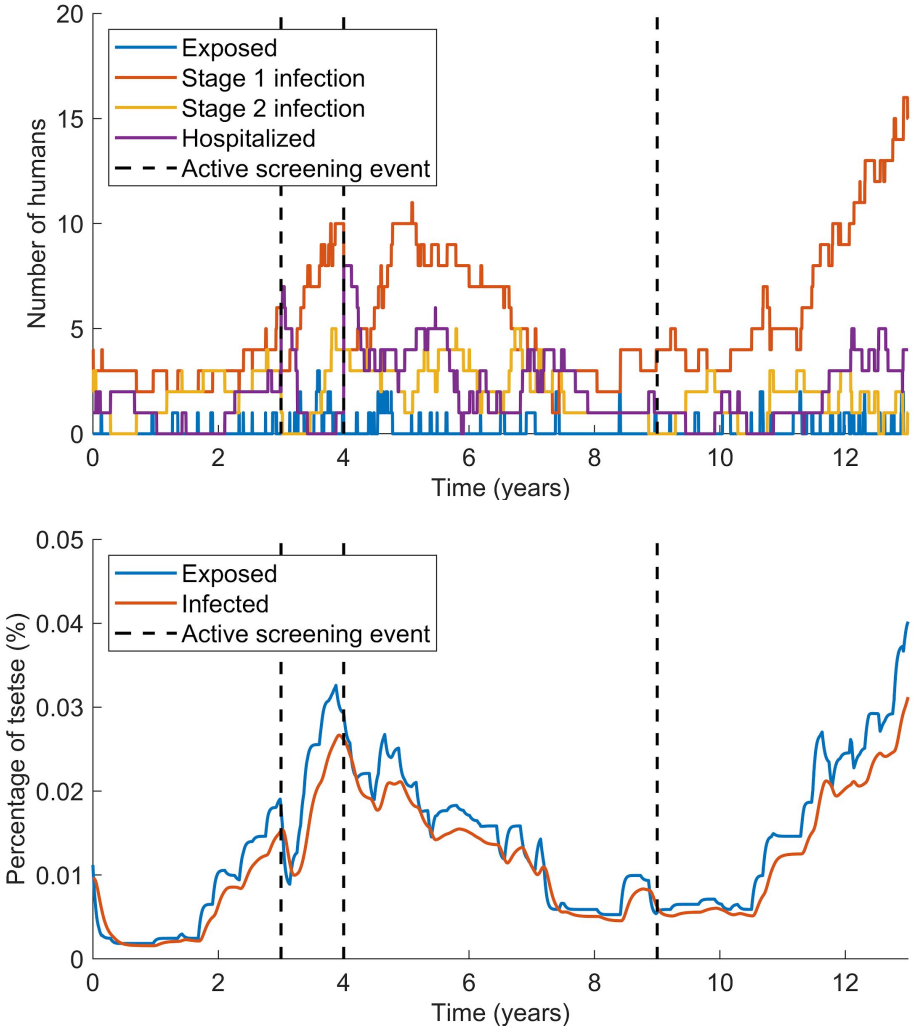


Fig 3.9: A single realisation of the simulated infection dynamics of the model. A population of 514 is used with 404, 430 and 52 of the low-risk population randomly selected for active screening after three, four and nine years, respectively. Note that the number of hospitalised individuals increases at the first two active screening events, but no infected individuals are randomly selected and detected in the third screening event. We only plot those infected (or affected) by HAT, although we note that the vast majority of both populations are susceptible.

We see that the probability of gHAT persistence as time progresses is very similar for all five assumptions, while it does diverge slightly with time (Fig 3.10). For a population size of 2,000, after 15 years, the probability of gHAT persistence using all five assumptions gives results less than 0.01 apart and the confidence intervals all overlap. For a smaller population size, there is a bit more variation and the confidence intervals do not all overlap. The assumptions where the population size grows by 2.6% each year has a higher probability of gHAT persistence, although the maximum difference is still only approximately 0.012 and this represents a significant increase in the village population over the time span. Where there is no population growth the confidence intervals all overlap.

Thus, we can use the assumption of constant population size, since it does not significantly change the results, excluding large increases in population size for small populations, and is significantly easier to implement. Keeping the population size constant also decouples the stochastic extinction of the gHAT infection from any underlying population dynamics. If we were to introduce more elaborate demographics into the model structure, it would be unclear if the extinction of infection was due to the stochastic process of transmission or the fluctuations in population size. Furthermore, we can associate each simulation of a village with a single population size, and can hence show probability of gHAT persistence as a function of population size.

3.3 Comparing the stochastic model to data

While the underlying deterministic model has been fitted to the aggregate data from this region, it is important to assess the behaviour of the stochastic model against village-scale observations.

Unfortunately, local disease extinctions cannot be directly observed; failure to discover any cases does not necessarily mean that the infection is not present, simply that it has not been detected. Thus, to validate our model, we make comparisons between the simulated predictions and WHO HAT Atlas data [106, 107] for the probability of detecting no cases on an active screening (termed zero-detections for brevity), which is

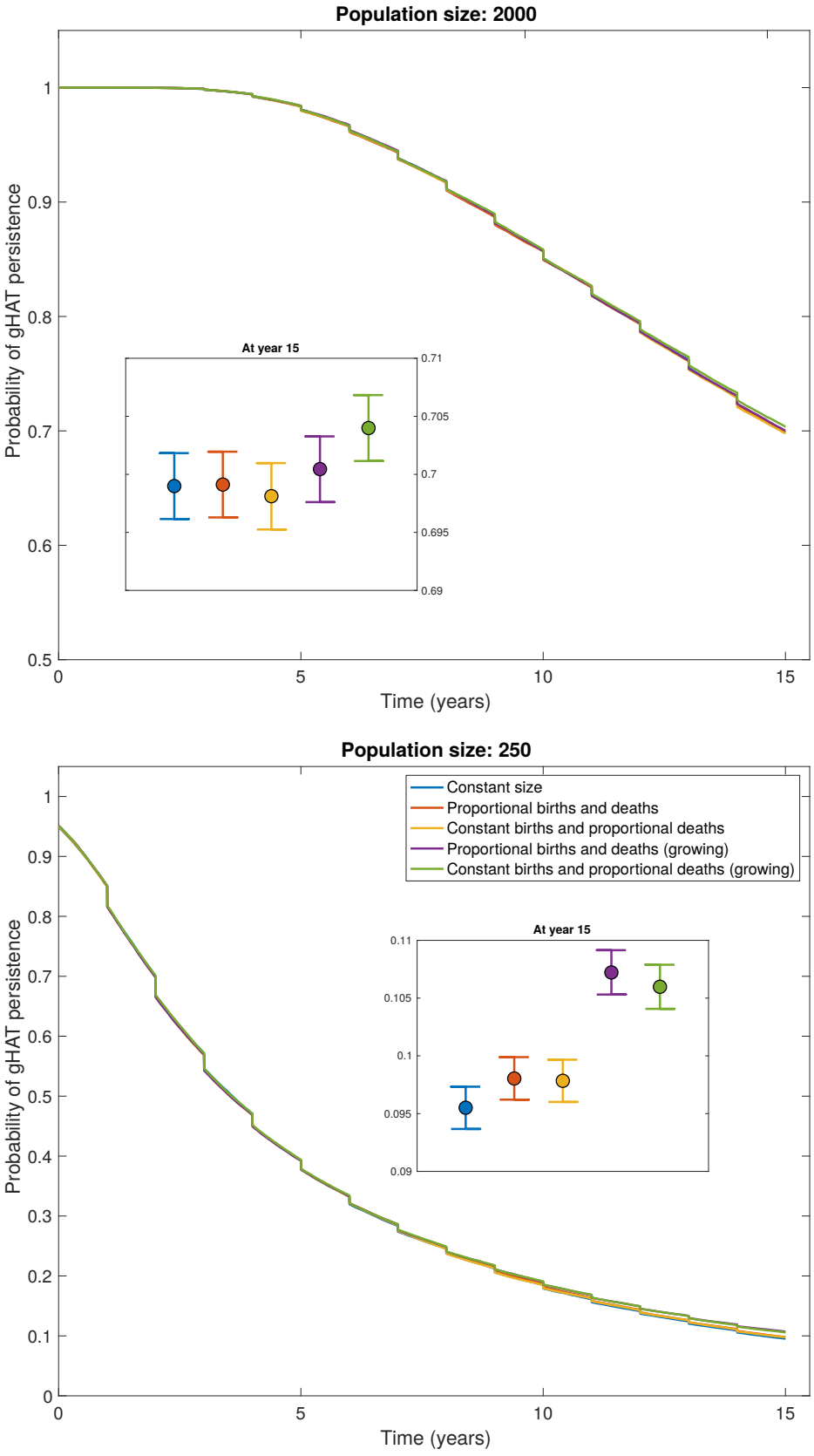


Fig 3.10: Using different population dynamics assumptions. The probability of gHAT persistence is very similar for different assumptions on the population dynamics, diverging only slightly for larger times. Our typical assumption of a constant population, achieved by coupling births with deaths, provides results within the 95% confidence intervals of when births and deaths are decoupled and when births are constant. Results are shown for populations of size 2,000 and 250.

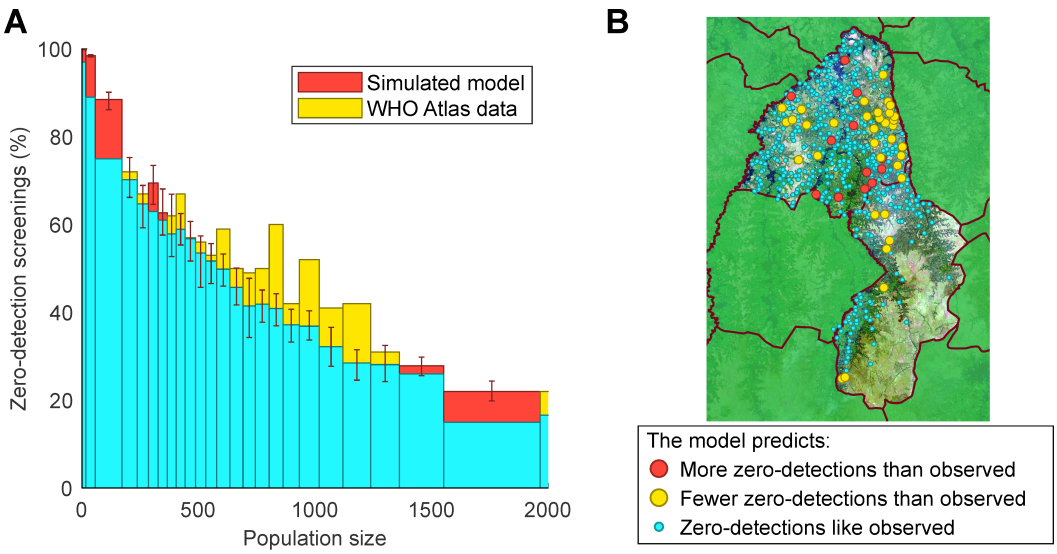


Fig 3.11: Comparison of model predictions and data for active screenings with no detected cases (zero-detections). (A) Histogram by population size of the percentage of active screenings that find no new gHAT cases for both the model and the HAT Atlas data. Each bar represents 100 screenings of simulated results (averaged over 10,000 replicates) from the model that uses the observed pattern of screenings and compares to the data. Values where the model predictions have more zero-detections than the data are in red, while the reverse is shown in yellow. Error bars represent the 95th percentile of model results. (B) Map of populations in Yasa-Bonga and Mosango showing the settlements with significant differences (at the 95% level) in the expected proportion of active screenings with no cases detected. Red circles are where the observed number of active screenings with zero-detections is below the 95th percentile of the model; yellow circles are where the data falls above the 95th percentile; small blue circles are for data that lie within the 95th percentile of predictions and therefore are well described by the model. The satellite image shown is from Landsat-8 accessed through https://earthexplorer.usgs.gov/ from the U.S. Geological Survey.

a combination of failure to detect and local extinction. We compare model predictions to observations by calculating the percentage of zero-detections in aggregations of 100 active screenings with similar village population sizes (Figure 3.11A). We find very strong agreement between model predictions and data, with a pronounced decline in zero-detections for larger populations.

For individual settlements, those where the number of zero-detections lie outside the 95th percentile of model predictions are notably spatially clustered (Figure 3.11B). In 2.1% of settlements (red), there are significantly fewer zero-detections than predicted and hence greater persistence; these villages are generally localised around the main river through the region. In 6.3% of settlements there are significantly more zero-detections than expected (yellow), and these are clustered far from the major rivers

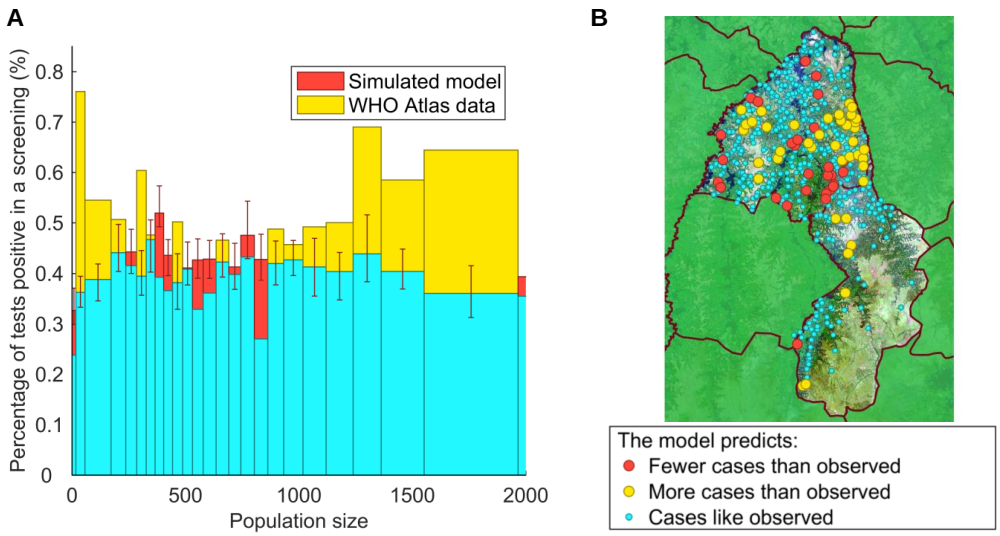


Fig 3.12: Expected proportion tests that are gHAT positive in an active screening. (A) Histogram by population size of the percentage of positive tests in independent active screenings for both the model and the WHO HAT Atlas data. (B) Map of populations in Yasa-Bonga and Mosango showing the settlements with significant differences in the expected proportion of tests positive. The underlying Landsat-8 satellite map is shown courtesy of the U.S. Geological Survey.

and in upland areas. Since tsetse are most densely distributed surrounding riverine areas [81], this spatial clustering may indicate the need for spatially heterogeneous parameters that reflect the suitability of the local environment for tsetse. However, given that more than 90% of villages fall within our prediction intervals, we believe the homogeneous parameters capture the general stochastic behaviour of this region.

Considering a different metric, and comparing the data on the proportion of tests that result in a positive detection of gHAT to the output of the model demonstrates similar results to comparing the screenings with no detected cases, as in Figure 3.11 (Figure 3.12). The difference in coverages of an active screening event mean the results have more variation than simply not detecting any new cases, yet the model is similarly capturing the behaviour seen in the data; there are a few more settlements that stand out as significantly different from the model (69 settlements, 12.3%), but these are similarly clustered where expected.

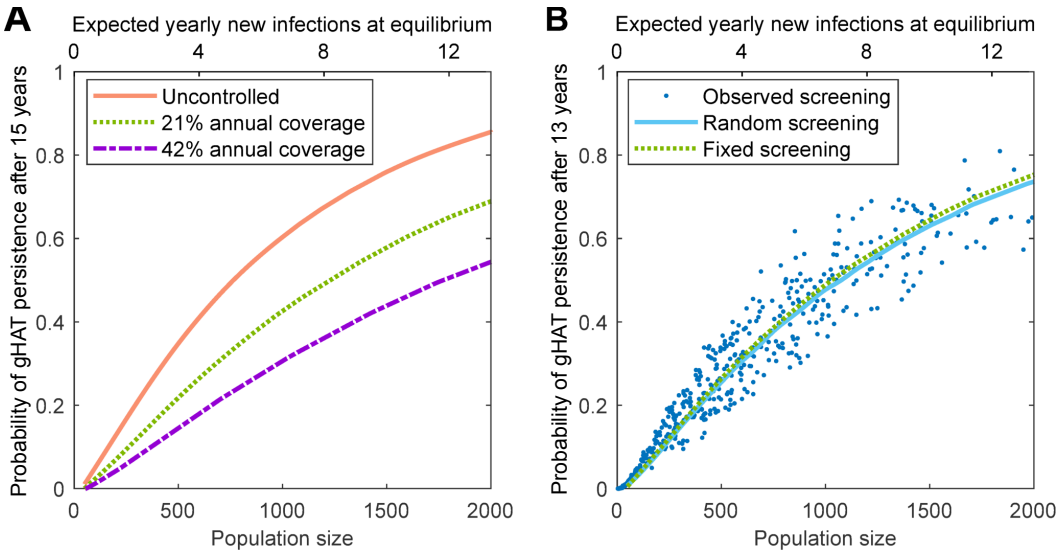


Fig 3.13: Predicted probability of gHAT persistence in isolated settlements. Simulations are started at the endemic (uncontrolled with passive surveillance only) equilibrium from the deterministic model and iterated forwards (without infectious imports), while the persistence of infection is recorded. This is repeated 100,000 times for settlement population sizes between 50 and 2,000 individuals. The expected number of yearly new infections if the system were at deterministic equilibrium is proportional to the population size and is given by the top scale. (A) Impact of active screening on gHAT persistence; annual screening at a fixed coverage per year yields a drop in persistence with increased coverage. (B) Comparison of screening assumptions on the persistence of gHAT. The solid curve shows results where annual screening coverages were randomly sampled from all observed coverages; dots represent the individual settlements recorded in the WHO HAT Atlas for Yasa-Bonga and Mosango health zones[106, 107], where the reported coverage in each year is used. There were sufficient simulations such that confidence intervals are too small to be visible.

3.4 Considering village-scale persistence

Local gHAT persistence, where human infection is maintained in a settlement, is affected by many factors (Figure 3.13), including the population size of the settlement; the vector-to-host (tsetse-to-human) ratio; the exposure to the tsetse; the screening procedure; and any movement of infected individuals between populations. We calculate the probability of persistence by stochastically simulating the epidemic for 16 years from the endemic (uncontrolled, passive surveillance only) disease equilibrium found in the deterministic model. If there is zero gHAT infection in the human population after a given number of years and no further human infection emerges in the following year from infected tsetse, we say in this simulation there is local disease elimination with no immediate threat of re-emergence (Figure 3.15A provides a justification for this criteria). This procedure is repeated for multiple population sizes; the proportion of simulations that retain infection after a given number of years determines the probability of persistence (Figure 3.13).

We focus on settlements with fewer than 2,000 inhabitants, typical of Yasa-Bonga and Mosango (Figure 3.2), and using our regionally specific parameters. In all scenarios investigated, we find that persistence increases with increasing population size (Figure 3.13). This echoes results from other infections [255–258], where small populations with low incidence experience a greater impact of stochasticity and chains of transmission that are more likely to be broken.

The addition of active screening (leading to the treatment of detected cases) decreases the probability of persistence across all population sizes, since removing infected individuals leads to a greater chance of breaking chains of transmission (Figure 3.13A). Increasing levels of screening, beyond the observed 21% average, leads to further reductions for persistence. We compare three assumptions for active screening (Figure 3.13B): that each population is screened annually at a fixed coverage equal to the regional average (21%); that each population experiences screening coverages sampled from the regional pattern including not screening in a given year; and simply replaying the recorded pattern of active screening in each village. Despite the very different distributions of screening effort, all three of these assumptions produce comparable

levels of persistence. As a control strategy, we also find that randomly selecting the screening coverage from all reported coverages slightly out-performs screening annually at the mean coverage (21%), kept fixed across all years, but there is minimal difference (Figure 3.14A). In addition, given that long-term persistence relies on persistence for shorter time intervals, the probability of persistence decreases with time (Figure 3.14B–C). The relatively long persistence times of gHAT, compared to the frequently studied persistence of childhood diseases [255, 256], are attributable to the long time-scale of gHAT infection in the absence of active interventions [245].

3.5 The probability of re-invasion

Following localised elimination of infection, populations remain vulnerable to re-invasion; we investigate the potential for re-establishment of sustained transmission in a settlement for different invasion scenarios. For a village of 1,000 individuals, following the elimination of infection in humans, the level of infection in tsetse falls rapidly, even when starting at the endemic level in these vectors (Figure 3.15A). In approximately 65% of simulations, the initially infected tsetse generate human cases, and the level of infection in vectors rapidly plateaus; otherwise, infection is eliminated from the location within six months due to short vector life expectancies in comparison to the human hosts. This validates the previous simulation assumption (Figure 3.13) that local disease elimination can be considered achieved after a period of one year in which there are no infected humans, as the number of infected tsetse will also become negligible.

The probability that a population of tsetse, infected at the endemic equilibrium (0.02% of tsetse exposed or infected), will lead to re-establishment of infection is predicted to be a function of settlement size (Figure 3.13C). Small populations are unlikely to see any new human infections, and those that are generated fail to persist. This is because tsetse have short lifespans and so are only able to interact with a few humans over their lifetime. However, even for large population sizes of 2,000 individuals, the chance of continued transmission beyond one year is only 55% and is less than 10% over 15 years. For lower levels of infection in the tsetse population, the risk of successful re-establishment is proportionally reduced.

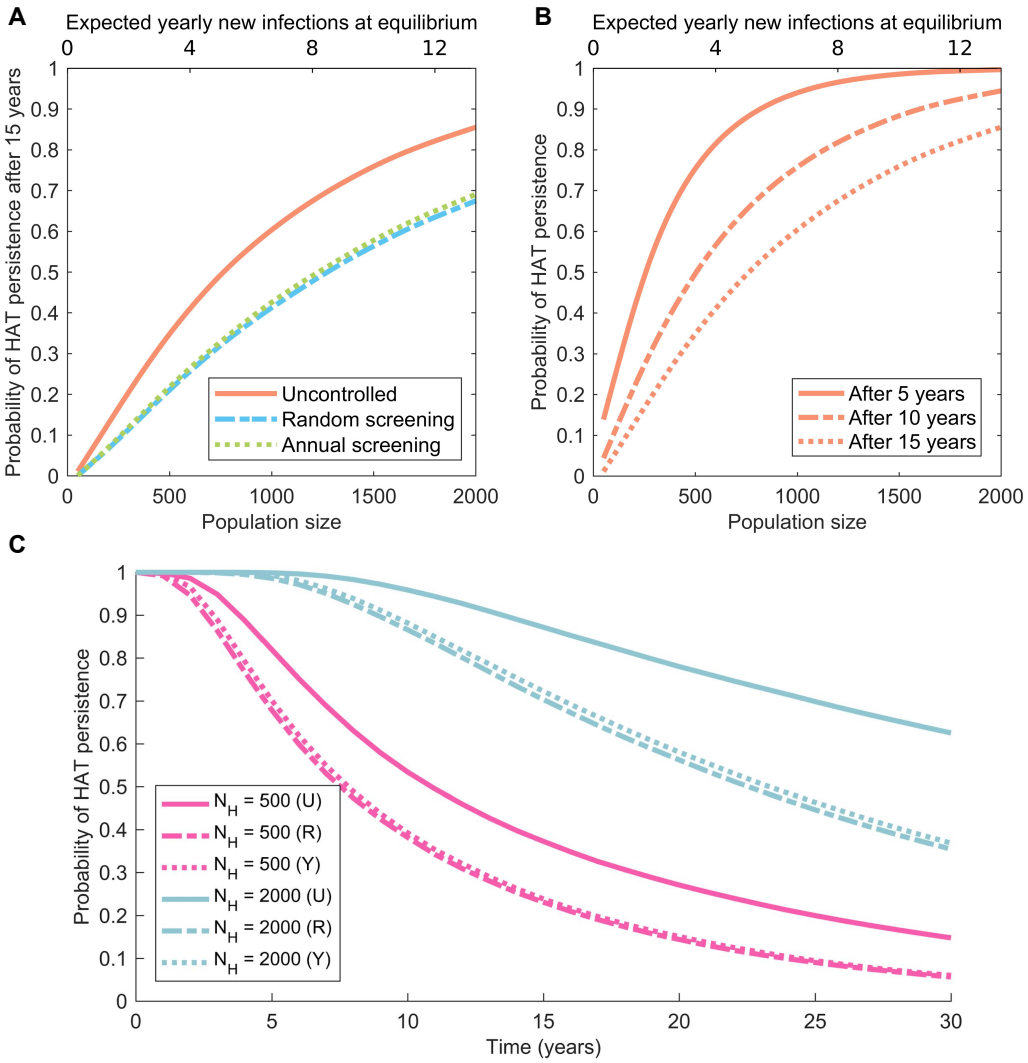


Fig 3.14: Factors affecting predicted probability of HAT persistence in isolated settlements. (A) Impact of active screening on disease persistence: both random screening (sampling from all screening coverages, including zeros) and annual screening at a fixed coverage (the mean annual coverage of 21%) yield a drop in persistence. Choosing the coverage randomly slightly out-performs screening annually at the mean coverage. (B) Probability of HAT persistence 5, 10 and 15 years after starting at the endemic equilibrium. (C) For villages with populations of sizes 500 and 2,000, the probability of HAT persistence in time for the uncontrolled setting (U), random active screening coverage (R) and annual screening at the average coverage of 21% (Y).

In contrast, if re-invasion of an infection-free population is due to the movement of an infected person into the settlement (in the absence of infected tsetse) the probability of re-establishment over different time-frames is largely unaffected by the population size (Figure 3.15B). This is because infection in a human will be present in the population for much longer and so population size will have less impact on whether new transmission events occur. We predict a high probability ($> 70\%$) of short term re-invasion, but a more limited chance ($< 20\%$) that this will generate persistent infection for 15 years or more. This is typical of stochastic dynamics of infection with low basic reproductive ratio (R_0), where, although short chains of transmission are likely, it is difficult for the infection to fully establish. The probability of new transmission events is further increased by the addition of an infected human to endemic levels of tsetse infection (Figure 3.15D).

3.6 Detecting elimination

As discussed above, zero-detections may be an indication that there is no infection in a settlement; however, it may simply be that infected individuals were not screened or were false negatives. Despite this, there is a temptation to associate zero-detection with zero infection; we therefore use our simulations to tease apart this complex relationship. The model replays the observed pattern of screening in each settlement, but we perform multiple simulations to ascertain the probability that infection has been locally eliminated following one, two or three consecutive zero-detection screenings, in which at least 20% of the population are screened (Figure 3.16). We also insist that no passive cases were detected between the screening events. Our standard assumption, in agreement with parameter inference for this region, is that 26% of Stage 2 infections, where people either self-report or likely die of gHAT infection, are detected and reported [219], but we also investigate 0% (no passive reporting) and 100% (all passive cases and deaths are reported) for comparison.

For small settlements, given that long-term persistence is unlikely (Figure 3.13), even a single zero-detection screen (Figure 3.16, blue) is frequently associated with local elimination. For larger populations, a single zero-detection has limited predictive power

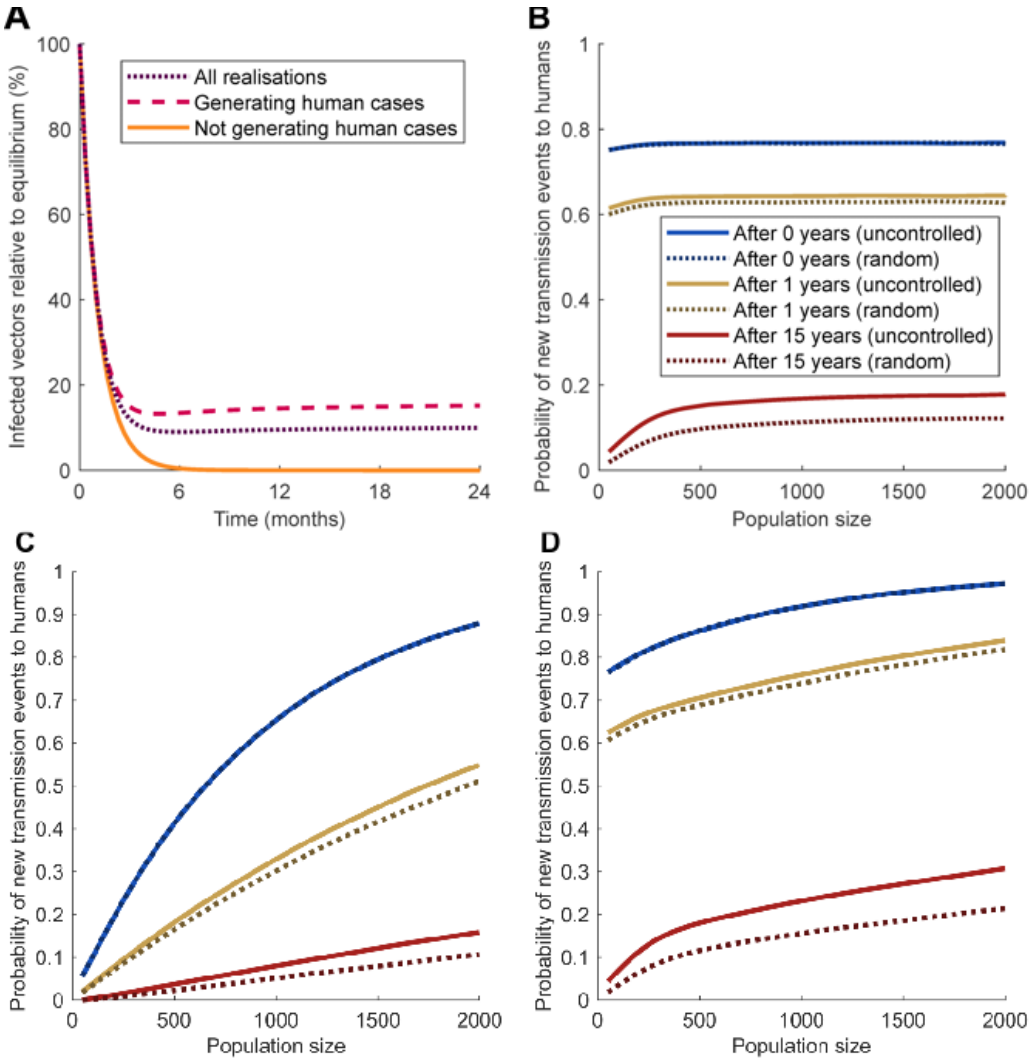


Fig 3.15: Dynamics of extinction and reintroduction. (A) Starting with no infected humans but the tsetse population at its endemic equilibrium and a settlement size of 1,000 individuals, the model predicts a dramatic decline in the infected tsetse population, depending on whether subsequent human cases are generated by the infected tsetse. (B–D) Extending this model further for a range of initial conditions. We examine the probability that at least one human case is generated and the infection persists for a given time, when starting with (B) one infected human and no infected tsetse for both uncontrolled and random screening, (C) with no human infection and endemic-equilibrium levels of infected tsetse, (D) with one infected human and endemic-equilibrium levels of infected tsetse.

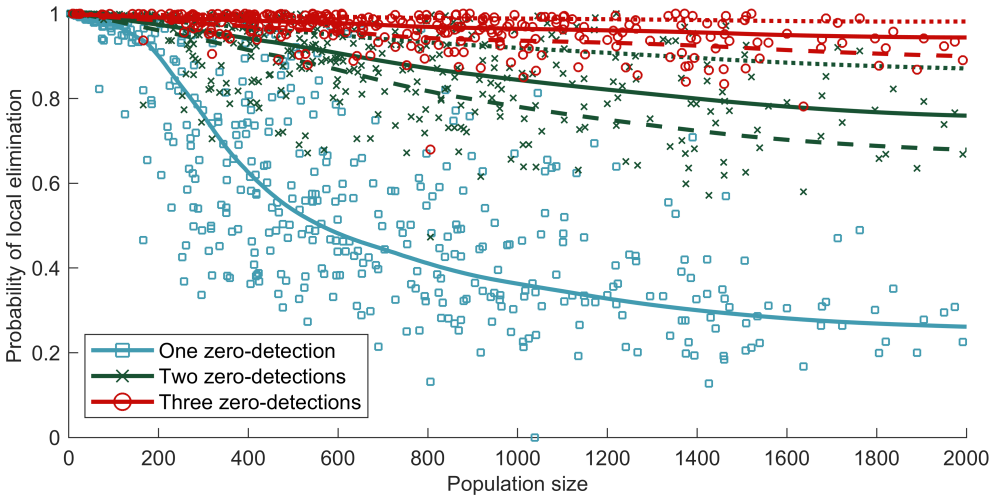


Fig 3.16: Probability of elimination in a settlement, given consecutive zero-detections with no detected passive cases. Consecutive zero-detections means consecutive in the observed years of screenings, not necessarily in consecutive years, and with no passive cases detected in between. Each point represents the average from multiple simulations of individual settlements where the reported pattern of screenings is replayed. The points and solid lines assume a reporting rate of 26% [219], while the dashed and dotted lines show reporting rates of 0% and 100%, respectively. Sufficient simulations are used such that the confidence intervals are small (unobservable on the scale of this graph). Lines represent a weighted local regression fit. Active screenings where fewer than 20% of the population are assessed are excluded from our analysis due to the small sample sizes (alternative cut-offs of less than 10% and less than 50% are presented in Figure 3.17).

and three consecutive zero-detections are needed to have any degree of confidence, in which case the reporting of passive cases plays a noticeable role. There is, however, significant variation between settlements, reflecting very different patterns of reported screening. Moderate population sizes of between 200 and 1,000 individuals show extreme variation in the ability to predict local elimination, while smaller villages have less variation, in part due to rarely being screened before 2009 (Figure 3.7C).

In calculating the probability of infection on detecting no cases, a zero-detection event is different for different screening percentages. We use 20% of the population screened to define a 'full screening'; however, similar results are achieved for the screening regime of Yasa-Bonga and Mosango 2000–2012 when 10% is used. If 50% is used, and so all screenings are of very reasonable quality, there is more confidence that the same number of consecutive zero screenings are a proxy for no infection in the population (Fig 3.17). When screening coverage is higher and no infection is detected, there is more certainty that no infection is present.

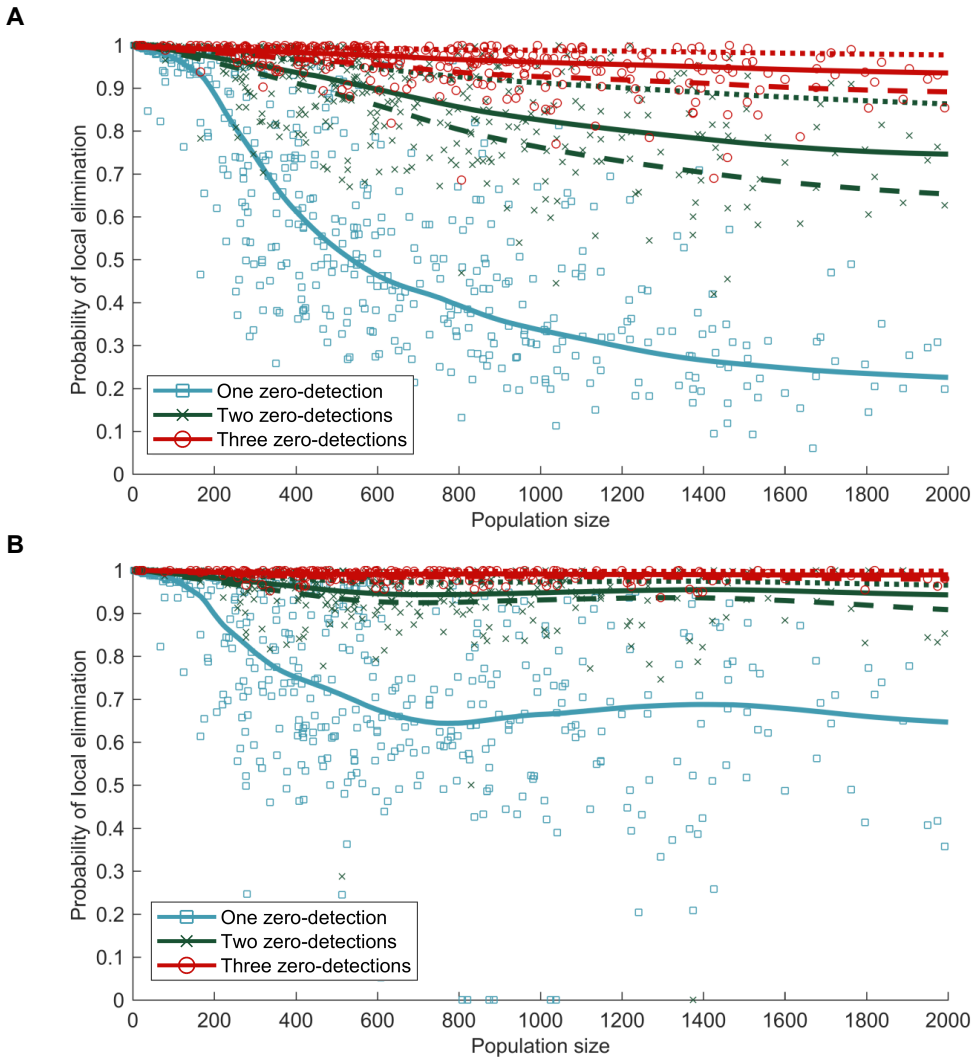


Fig 3.17: The probability of a settlement containing no infection given that a number of consecutive active screenings detecting no cases have occurred. (A) Only active screenings where at least 10% of the population are included are used to calculate the probability. (B) At least 50% of the population must be screened for the screening to be included.

3.7 Estimating the rate of importation

For the majority of this thesis, we have modelled the dynamics as a closed population, without emigration or immigration, so that once the disease has gone extinct in a population it cannot be re-introduced. This removes a critical dependency in model formulation, and greatly simplifies the presentation of results. In many cases, such as community-level persistence, this approach is justified, as we wish to understand the stochastic dynamics in the absence of confounding imports. In reality, no population is ever completely isolated. Regional persistence relies on more than independent persistence in individual settlements. It is likely that the occasional movements of infected people lead to a stochastic meta-population paradigm [259], where rare local extinctions of infection are balanced by external imports. However, the agreement between model and data (Figure 3.11), together with the low prevalence of infection, indicates that imports are likely to be rare.

We make this more quantitative by fitting an importation rate of infection, proportional to the size of the population. Before interventions, the presence or absence of infection in a population reflects the equilibrium balance between extinction and re-colonisation; we can therefore use the presence of infection at the first recorded active screening within a village as measure of the equilibrium state. Fitting the external importation rate to the probability of detecting infection in a village at the first active screen gives the best fit when the rate is small at just 3.4×10^{-6} per person per day using least squares regression (Figure 3.18A). This rate is small and so can be reasonably ignored on the studied time scales, particularly as there is some probability that this importation will not cause any subsequent transmission. Moreover, we can assume that the importation rate declines over time since the overall prevalence in DRC is declining [20]. By fitting the reduction in total cases in the DRC to an exponential decay function, we get a decay rate of 0.1071, which we apply to the import rate (Figure 3.19). In addition, since we are only using villages included in the WHO HAT Atlas [106, 107], we may overestimate importation rate for the whole region as we do not consider the villages not presenting cases.

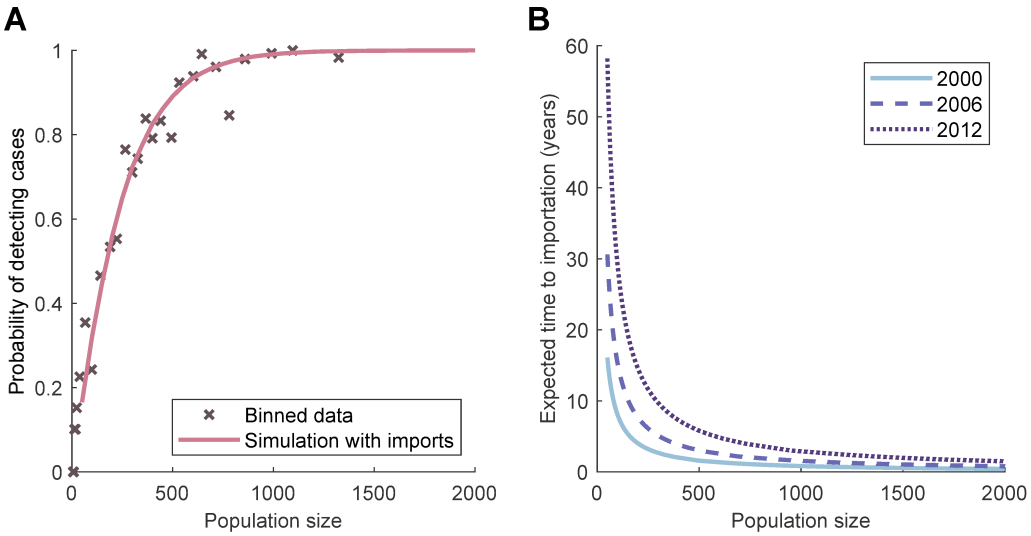


Fig 3.18: Simulating external importations of infection into village populations. (A) By running simulations with different values for the external infection parameter, we find the best fit — to data binned by population size from the WHO HAT Atlas on whether there are any detected cases on the first active screening — is when imports per susceptible individual are equal to 3.4×10^{-6} days^{-1} (solid line). (B) Curves of the expected time for an external importation into a village population using the fitted importation parameter. The importation parameter is assumed to decay at the same rate as the total number of cases in time in the DRC (see Figure 3.19).

For most population sizes in the region, the expected time to importation is therefore relatively long (Figure 3.18B), and in many cases a single importation will not cause further transmission events (Figure 3.15), leading us to conclude that in general the level of importation will not qualitatively change our results. As a validation of our assumption that importations can be safely ignored over the time-scales of this study, we re-examine the match between active-screenings that fail to detect cases and model results including imports (Fig 3.20). This is a counterpoint to Figure 3.11. We find that the inclusion of imports has an extremely small difference on the model predictions: slightly improving the fit between model and data (Fig 3.20A). However, under this new model, there is now an additional village where the model significantly predicts fewer zero-detections than observed, decreasing the percentage of villages within the model prediction intervals to 91.4% (Fig 3.20B). We conclude that this justifies our modelling assumptions that villages act as isolated populations and that importation of infection is unlikely to perturb the dynamics; instead, we are able to separate the processes of local elimination and re-invasion.

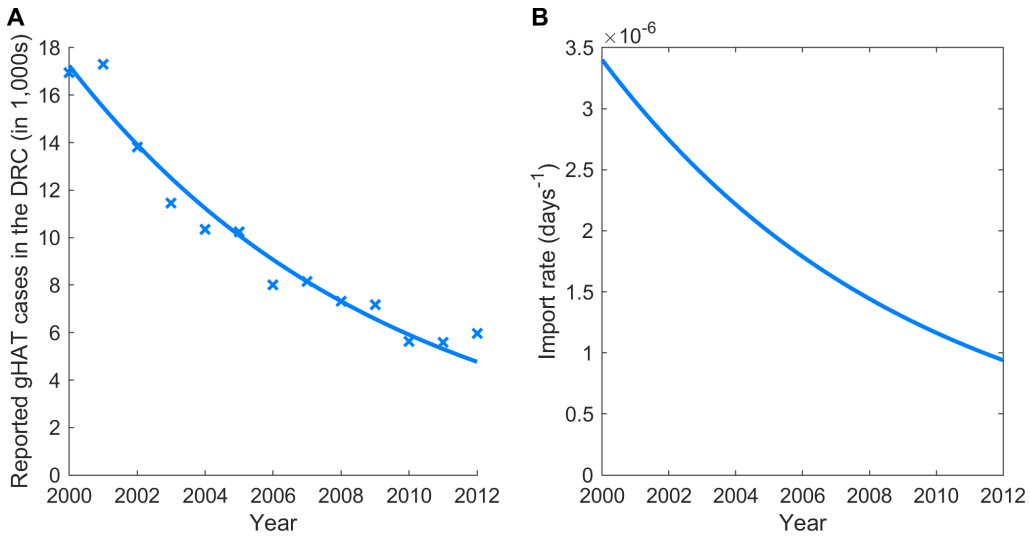


Fig 3.19: Change in the importation of infection parameter in time. (A) New cases of gHAT detected in the DRC with fitted exponential decay [20]. (B) The decay of the rate of importation of infection parameter is matched to the decay rate of cases of HAT in the DRC.

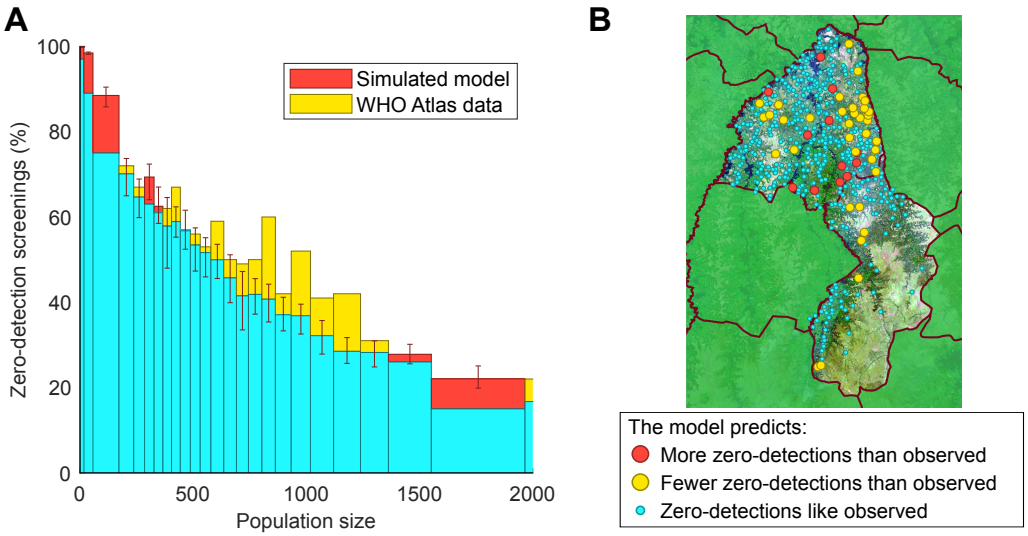


Fig 3.20: Comparison of model predictions and data for active screenings with no detected cases (zero-detections) for the model with importations of infection. (A) Histogram by population size of the percentage of active screenings that find no new HAT cases for both the model and the HAT Atlas data. The differences between this histogram and without importations of infection are very small. (B) Map of populations in Yasa-Bonga and Mosango showing the settlements with significant differences (at the 95% level) in the expected proportion of active screenings with no cases detected. The addition of importations of infection means this model fits the data less well, with an additional village falling outside the prediction intervals. The underlying Landsat-8 satellite map is shown courtesy of the U.S. Geological Survey.

3.8 Discussion

Despite global declines in reported cases over the last decades, *gambiense* gHAT remains a problem in many focal areas [79]. These regions, concentrated primarily in the DRC, represent a significant challenge to achieving the WHO 2020 and 2030 goals of elimination as a public health problem and zero transmission, respectively. Robust models, matched to the available data, are the only viable means of quantitatively assessing future dynamics and the long-term impact of controls [219, 220, 223, 224]. Active screening followed by treatment is one of the main control measures, but this action is deployed at the village level suggesting that village-scale models (which recognise the effects of small population size) may be needed to optimise deployment; these results can then be scaled to an infection focus or national level to measure regional elimination, which is especially important as we approach zero transmission.

We have introduced a dynamic, mechanistic, stochastic gHAT model, which is applied to 559 settlements in the Yasa-Bonga and Mosango health zones within the former Bandundu province of the DRC (Figure 3.2). Using parameters inferred from a deterministic model fitted to aggregated reported cases, our model reliably captures observed detection patterns at the village-scale (Figure 3.11). This comparison highlighted some spatial heterogeneity associated with the local environmental conditions (significantly fewer zero-detections than predicted occurred in regions close to large rivers, where the tsetse density is presumably high); however, 91.6% of settlements fell within the model (95%) prediction intervals, giving us confidence in our predictive ability. The inclusion of such local environmental factors, which modify the underlying parameters, is clearly an area for further research into refining this small-scale model and may help to practically focus localised control measures, in particular for planning tsetse control.

Throughout our simulation experiments, we consistently find that gHAT persists better in larger populations. This is as expected and agrees with theoretical work and analysis of other diseases [255–258, 260, 261]; in small populations, the behaviour of the individual is more important, and hence stochastic effects are magnified. The degree of persistence predicted is, however, surprising; settlements of around 2,000 inhabitants, where yearly incidence is only 13 new infections, frequently persist for 15 years or more

(Figure 3.13). This should be contrasted with frequently studied, highly transmissible diseases, such as measles, where local extinctions are common in population sizes of less than 300,000[255, 256]. We attribute this pronounced difference to the much longer time scales associated with gHAT, meaning single individuals can maintain infection, and the vector-borne nature of gHAT transmission, such that the tsetse act as a short-lived reservoir. We consistently find that incorporating active screening reduces the persistence of infection (Figure 3.13A), although the distribution of screening across years has only a small effect (Figure 3.13B). Increasing the screening coverage beyond the average reported levels (of 21% per year) is predicted to lead to further reductions in persistence, but infection is still predicted to be maintained for over 15 years in many larger populations.

Throughout, we have generally ignored the impact of new infectious individuals entering the population, and indeed have shown that this rate of importation is very low. If re-invasion following local elimination is due to the movement of a single infected individual into the settlement (Figure 3.15B), we predict that the probability of subsequent cases is high (70–80%) and largely independent of population size. However, only a small proportion (10–20%) of such invasions lead to long-term persistence of over 15 years. Re-establishment of infection due to the movement of a limited number of infected tsetse is even less likely. Current uncertainties about the impact of potential reservoirs — either a human reservoir of asymptomatic infections or an animal reservoir — mean there is insufficient knowledge for resurgence to be explicitly modelled by these mechanisms [184], but an animal reservoir that can maintain infection in the absence of humans is likely to represent a worst-case scenario.

A key question, as we approach the 2030 goal of zero transmission, is to ascertain when local elimination has been achieved, allowing policy-makers to scale back control if the infection is no longer present. Due to only a limited proportion of each settlement being screened and the potential for false negatives, a screening can fail to detect any cases even when there is infection in the population. We have shown that, while a single zero-detection screening provides relatively little information of the probability of local gHAT elimination, multiple consecutive zero-detection screenings are a strong indicator of elimination (Figure 3.16). This can be further strengthened if only large screens

($> 50\%$) are included in the analysis (Figure 3.17B), providing valuable public health information. This concurs with WHO guidelines for active screenings, as villages are no longer considered in planning by mobile screening teams after three consecutive years of zero-detections, followed by a further zero-detection after three years [240]; our model would predict local elimination with large probability for this level of surveillance. Consistent with the observed patterns in high endemicity regions, we have assumed that the screening coverage and frequency remains constant over these time scales but note that in other regions it may be important to consider the reduction in active screening as reported gHAT cases decline and elimination is approached.

The ability to capture the stochastic dynamics and persistence of *gambiense* gHAT infection at the village-scale is a major advance in public health modelling, with far-reaching consequences for informing policy decisions. This is particularly pertinent as our models operate at the same spatial scale as controls and can capture the local elimination of infection that is a prerequisite of achieving the 2030 goal of zero transmission globally.

4

Optimising active screening for *gambiense* human African trypanosomiasis

Traditionally, the most effective form of controlling *gambiense* human African trypanosomiasis (gHAT) infection has been active screening and treatment [12, 97, 246]. This is the screening that is carried out by mobile teams that travel to villages in focal disease regions and target the screening of the full village population for gHAT; those determined to have the infection can then be treated at the closest health facility offering treatment.

While active screening has been very effective in reducing case numbers and still has an important role to maintain surveillance and availability of treatments, particularly where access is a problem, it is an expensive intervention in both time and money [123, 154]. As some local elimination of gHAT is seen in focal areas, active screening will be scaled back and gHAT testing will become more integrated into fixed health facilities, as resources can be reallocated and it becomes no longer necessary to screen entire village populations for the infection [127]. Reactive screening can also be implemented, whereby after active screening has stopped, it can return upon a new case being reported due to a diagnosis at a fixed health centre in the area [156].

In this chapter, we weigh up the costs and benefits of different active screening strategies, by considering them in a net monetary benefit (NMB) framework. We explore the modelling results from invoking different active screening intervention strategies and, by applying costs to the outcomes of simulations, consider which strategies prove optimal in both maximising the NMB and having the highest probability of being cost-effective. With a limited number of active screening teams and resources for them to carry out

their duties, it is imperative to optimise their activities whilst sustaining the aim of driving towards elimination.

4.1 Considering the factors in optimising active screening

HAT, including both gHAT and rHAT, caused an estimated 1,364 deaths in 2017 with the global burden approximately 78,990 disability-adjusted life years (DALYS) [262]. The health problems caused by HAT are not just a burden for the individual, but are an obstacle to economic and human development [3]; investments in health are profitable in macroeconomic analysis. Thus, we aim to consider monetary benefits associated with implementing interventions to reduce the health burden, alongside the monetary costs of performing these different interventions.

We have explored mathematical models of gHAT and how they have been used for the prediction of future case numbers and effect of a range of plausible control strategies [146, 179, 186, 219, 220, 223, 224]. However, these have typically considered the infection dynamics and impact of interventions without accounting for the costs of implementing such strategies. We explicitly use an adapted stochastic model of gHAT infection in a village population developed in Chapter 3 of this thesis to simulate different plausible active screening programmes (and passive surveillance) to quantify the relative costs of implementation as well as the effect on averting DALYs compared to a baseline of only considering passive surveillance (the comparator strategy).

The costs of interventions have been previously been evaluated [154, 263–266] and the different strategies for intervention types considered for large populations [142, 182]. We consider the effect of active screening on individual villages in the drive for elimination of transmission (EOT), by determining how active screening can be best implemented to achieve this goal whilst providing value for money.

Tab 4.1: Descriptions of the variables used for defining an active screening strategy.

Variable	Name	Definition	Value range
c	Screening coverage	Proportion of the village population screened in a visit.	0–91.4%
t	Screening interval	Time between active screening visits to a village.	0.25–5 years
z_a	Active zero-detections	Number of consecutive active screenings where no cases are detected for the cessation of active screening.	1–5 screenings
z_r	Reactive zero-detections	Number of consecutive reactive screenings where no cases are detected for the cessation of reactive screening.	1–3 screenings

4.1.1 Active screening strategies

In determining how active screening should be performed, we consider four factors that we construct to formulate a screening strategy for a village: screening coverage c, screening interval t, active zero-detections z_a, and reactive zero-detections z_r (see Table 4.1). Like in the model from Chapter 3, individuals in the human population are classified as either high-risk or low-risk [219], whereby the high-risk population have a higher exposure to tsetse and do not participate in active screening. This means that the screening coverage is assumed to have a maximum of less than 100%, since only the low-risk human population participate (randomly) in active screening. We have estimated in model fitting that the maximum screening coverage is therefore 91.4% (see Section 4.2.2, note that this is a different value from Chapter 3). The screening interval t is the length of time between active screening visits to a village, which we limit to being between four visits per year and one visit every five years. The active zero-detections z_a represent the cessation criterion for the strategy, whereby if a village is screened z_a times consecutively, with no cases detected on each screening, the active screening should terminate. Reactive screening is the resumption of active screening and occurs upon detection of a case in passive surveillance after active screening has been stopped; reactive zero-detections z_r are the number of these screenings that should occur consecutively with no cases detected before this screening is stopped again.

An active screening is generally assumed to aim for as high a screening coverage as possible, depending on the availability and consent of the population to be tested. However, there is limited evidence for how frequently these active screenings should occur and when they should stop if no cases are found. Several active screening strategies have been proposed including a recommendation of three repeated screening rounds with one year intervals [118], or using six month intervals [119]. The cessation criteria also has limited evidence to support its value, although, as we showed in Chapter 3, multiple active zero-detections in a village are required to have confidence of local gHAT elimination. The WHO guidelines currently recommend three years of zero case reporting before stopping active screening in previously endemic villages [267].

To consider how to optimise over these four variables, we return to our stochastic gHAT model to obtain output on the infection dynamics for different active screening strategies.

4.2 Adapting the gHAT model

To capture how implementing different active screening strategies affects the underlying infection dynamics on a village population, we use the stochastic compartmental model from Chapter 3. The stochasticity incorporates the chance events involved in infection transmission into the mechanistic model and is more suited than a deterministic model for the case of optimising active screening for a village population. This is because the population, and hence the number of people infected, is small and so, for the pre-elimination setting, extinction of gHAT can be directly reflected with a stochastic framework. A deterministic formulation, which captures average dynamics, would be less suitable to capture the chance events associated with village level extinction of infection.

The previous stochastic gHAT model from Chapter 3 is updated based on model refinements from Crump *et al.* [268] to both more accurately fit the data from the WHO HAT Atlas [106, 107] (now using data from years 2000–2016) and to better reflect

the underlying biology of the system. This is done by both re-structuring the exact formulation of our model and by re-fitting to the data.

4.2.1 Modifying the model

The tsetse modelling compartments are updated from the model structure used in [219] to include tsetse beginning life in the pupal stage, rather than simply being born susceptible [223, 224]. This introduces four additional parameters, B_V, a constant deposit rate into the pupal stage (a birth rate), ξ_V, the pupal death rate, K_V, pupal density dependence, and $P(\text{survive})$, the probability of pupating. These parameters are not fitted, but taken as fixed values from the literature (see Table 4.2). This new model with an explicit pupal stage, density dependent mortality and target-dependent deaths only occur with feeding, increases the biological relevance of the model by incorporating further aspects of the life cycle explicit to tsetse, while not introducing too many parameters and remaining tractable.

Additionally, we had previously assumed that passive infections were only detected, and hence moved to the hospitalised class, R_H, from Stage 2 infection. We therefore introduce an exit rate from Stage 1 infection η_H. Both the exit rates from Stage 1 and Stage 2, η_H and γ_H, are found in model fitting, but we also assume these values change in time. We assume a sigmoid function, to increase the rate in time, as passive surveillance quality increases. This is parameterised using an additional four parameters that are also found in model fitting. The parameters $\eta_{H\text{amp}}$ and $\gamma_{H\text{amp}}$ specify the maximum increase in the value of their respective parameters, d_{steep} specifies the steepness of the improvement rates, and d_{change} gives the switching year of the improvement rate. We also distinguish between baseline exit rates for pre-1998 and post-1998, whereby there is a step change at 1998 before the sigmoid function is introduced. The pre-1998 rates are used for determining the initial conditions at endemic equilibrium. Thus, $\eta_H(Y) = \eta_H^{\text{post}}\left(1 + \eta_{H\text{amp}}/\left(1 + \exp\left(d_{\text{steep}}\left(t - d_{\text{change}}\right)\right)\right)\right)$, and similarly for $\gamma_H(Y) = \gamma_H^{\text{post}}\left(1 + \gamma_{H\text{amp}}/\left(1 + \exp\left(d_{\text{steep}}\left(t - d_{\text{change}}\right)\right)\right)\right)$.

These changes to the model structure are summarised in Figure 4.1.

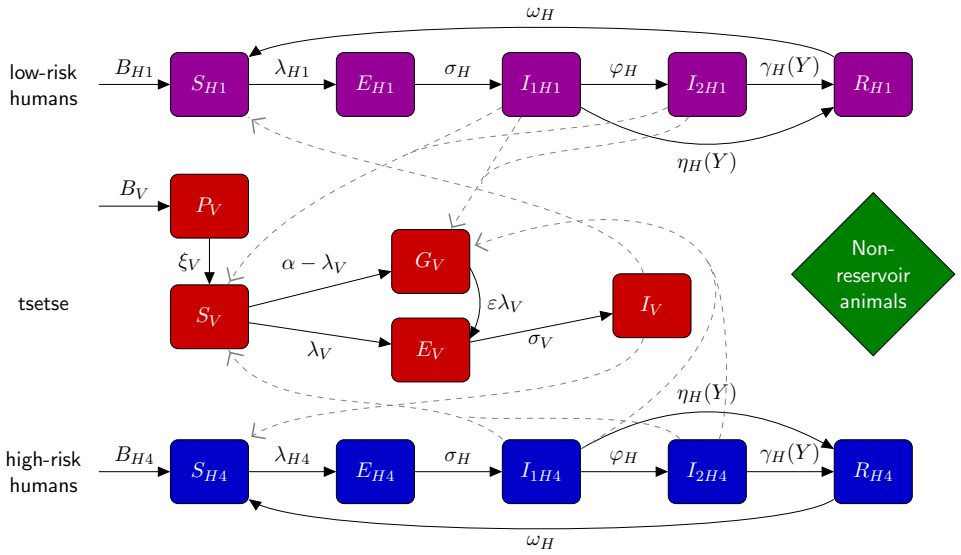


Fig 4.1: An updated model schematic for the gHAT infection dynamics used in this chapter.

4.2.2 Fitting to a new health zone

In this chapter, we choose to use model fitting for a different health zone in the DRC. This health zone, Kwamouth, is located close to Yasa-Bonga and Mosango, but in the Mai-Ndombe province (yet also still in the former Bandundu province) and is highly endemic with gHAT. This health zone was chosen as an example with a high number of gHAT cases and, at the time of choosing, had had no vector control activities that would complicate the modelling dynamics. We will also briefly consider using our methods in this Chapter on the same model, but parameterised for Mosango (rather than Mosango and Yasa-Bonga in Chapter 3) for comparison.

The active screening coverages of these two health zones have similar median values (when ignoring outliers of very high coverages or small coverages, that are likely to be residents of villages that travel to a different active screening, yet are recorded as their home village). We see the median coverage for Kwamouth is 55% and Mosango is 57%. The full distribution is shown in Figure 4.2.

We will primarily focus on Kwamouth in this chapter. A full list of the fitted and fixed parameter values for our model is shown in Table 4.2. These parameter values are

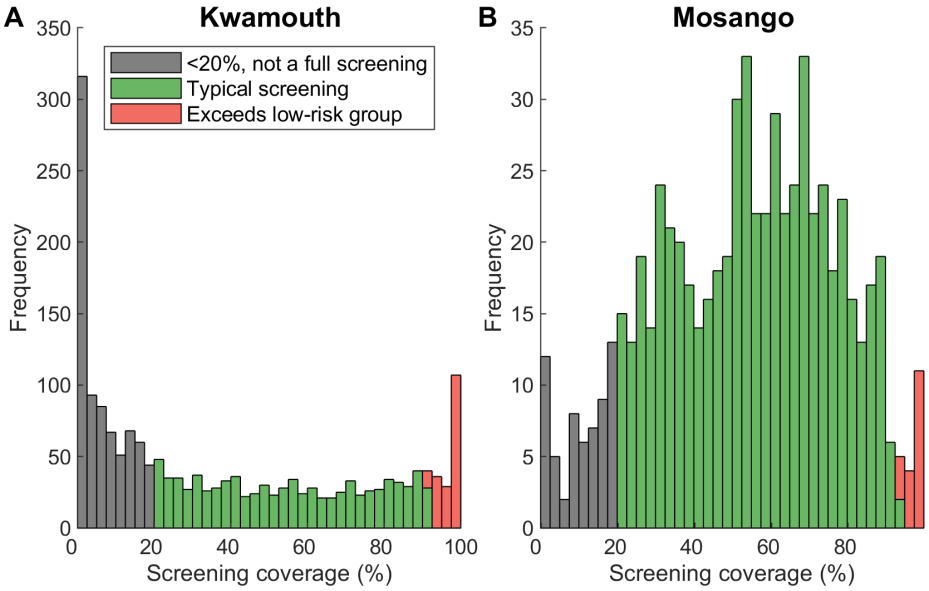


Fig 4.2: Histograms of screening coverages for (A) Kwamouth and (B) Mosango. We show the coverages below 20% in grey and above the full low risk group population screened in red, with typical screening coverages shown in green.

taken from Crump *et al.* [268], which were either sourced from literature, where well-defined, or otherwise (in the case of $m_{\text{eff}}, R_1, R_2, r, u, others$) taken as the median of the distribution obtained by model fitting using a Metropolis–Hastings MCMC algorithm that matched the deterministic version of the model to incidence data from the WHO HAT Atlas [106, 107].

4.2.3 Simulating active screening

In villages targeted for active screening, we consider how the infection dynamics are affected by: the screening coverage c; the screening interval t; active zero-detections z_a; reactive zero-detections z_r (see Table 4.1). This is modelled by taking all combinations of c, t, z_a and z_r for the values $c = 0$, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 91.4%, $t = 0.25$, 0.33, 0.5, 0.67, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.5, 3, 3.5, 4, 4.5, 5 years, $z_a = 1$, 2, 3, 4, 5 screenings, and $z_r = 1$, 2, 3 screenings. The specified screening procedure is then implemented in the model by using the screening coverage to randomly select this proportion of the population from the low-risk sub-population, which move into the hospitalised class if they are correctly identified as being exposed to the infection after each screening interval. This process is stopped after the given number of zero-detections. However, if the infection has not been eliminated after the

Tab 4.2: Parameter notation and values for the gHAT infection model.

Parameter	Description	Value	Source
μ_H	Natural human mortality rate	5.4795×10^{-5} days^{-1}	[269]
ω_H	Human recovery rate	0.006 days^{-1}	[117]
σ_H	Human incubation rate	0.0833 days^{-1}	[214]
ϕ_H	Stage 1 to 2 progression	0.0019 days^{-1}	[77, 245]
η_H^{pre}	Pre-1998 exit rate from Stage 1 by treatment or death	0.00007 days^{-1}	[268]
η_H^{post}	Post-1998 exit rate from Stage 1 by treatment or death	0.00007 days^{-1}	[268]
$\eta_{H\text{amp}}$	Maximum increase in the stage 1 exit rate	0.922	[268]
γ_H^{pre}	Pre-1998 exit rate from Stage 2 by treatment or death	0.006 days^{-1}	[268]
γ_H^{post}	Post-1998 exit rate from Stage 2 by treatment or death	0.006 days^{-1}	[268]
$\gamma_{H\text{amp}}$	Maximum increase in the stage 2 exit rate	0.922	[268]
d_{steep}	Steepness of the improvement rates for stage 1 and stage 2 passive detection	1.106	[268]
d_{change}	Switching year of the improvement rate for stage 1 and stage 2 passive detection	2008	[268]
N_H	Human population size	Varies	N/A
μ_V	Tsetse mortality rate	0.03 days^{-1}	[214]
σ_V	Tsetse incubation rate	0.034 days^{-1}	[242, 270]
ε	Reduced non-teneral susceptibility factor	0.05	[268]
α	Tsetse bite rate	0.333 days^{-1}	[267]
m	Relative tsetse density	Varies	N/A
p_H	Probability of human infection per single infective bite	Varies	N/A
m_{eff}	Effective tsetse density	6.56	[268]
p_V	Probability of tsetse infection per single infective bite	0.065	[214]
f_H	Proportion of blood-meals on humans	0.09	[243]
K_V	Pupal density dependence	111.09	[223]
P (survive)	Probability of pupating	0.75	[223]
ξ_V	Pupal death rate	0.037	[223]
B_V	Total deposit rate	0.0505	[223]
Se	Sensitivity	0.91	[252]
Sp	Specificity	0.99958	[268]
R_1	Proportion of low-risk humans, randomly participating in screening	0.924	[268]
R_2	Proportion of high-risk humans, not participating in screening	0.076	[268]
r	Relative bites taken on high-risk humans compared to low-risk	6.6	[268]
u	Reporting probability for Stage 2 cases	0.26	[268]
δ	Importation of infection rate	3.4×10^{-6} days^{-1}	Chapter 3

active screening has been halted, there is a chance that a new case can be reported by the individual attending a fixed facility to be tested for the infection. The model explicitly incorporates under-reporting of passive case detections, such that only 29% of cases not picked up through active screening will be reported in passive surveillance. This parameter was estimated in model fitting, taken as the median of the distribution inferred using MCMC methodology applied to the aggregate annual data from Kwamouth from the WHO HAT Atlas [106, 107]. Thus, we re-start the screening procedure as reactive screening upon identification of these passive cases, stopping again after the given number of consecutive reactive zero-detections, z_r.

The WHO propose to aim for high screening coverage in active screenings for villages in gHAT-endemic foci, and guidelines suggest these screenings should continue annually until there have been three consecutive years of no new cases, followed by a further screening after three years if there are no detected cases [267]. Our modelling aims to provide evidence to support this strategy or recommend how the strategy could be adapted to make more efficient use of resources.

4.3 Using the net monetary benefit framework

We want to apply costs to our models, since while without costs, we can still provide recommendations on strategies that will achieve elimination or a reduction in cases and the timescales involved, we know there will be monetary constraints on which strategies are feasible to fund. Thus, considering costs and also providing cost-effectiveness analysis provides additional information, important to the potential funders of these programmes.

The cost of an active screening strategy is a function of several component costs: implementing the screening test, confirmation of the infection, carrying out treatments, setting up and maintaining the mobile screening teams, and the change in number of passive tests and treatments caused by the change in active screening activity. We do not consider the additional costs of passive surveillance, such as capital costs, only the costs directly affected by active screening. The active screening strategy costs will vary

depending on the type of screening test and treatment used, and also the type of mobile screening team, like a truck team that can carry more tests and equipment, such as a generator, or a motorbike team, which can reach more remote villages [271].

As well as considering the changes in monetary costs, we want to consider the change in the health benefit of implementing different active screening programmes. Therefore, we consider the number of DALYs averted [272]. The number of DALYs are the discounted sum of the number of years of life lost (YLL) and years lived with disability (YLD), where YLL is the number of years of life lost due to premature death and YLD is the number of years of healthy years lost with a weighting for the severity of the condition [273]. We calculate the number of DALYs averted by a particular screening programme by comparing with a baseline of doing only passive surveillance and no active screening (our comparator strategy).

We evaluate the net monetary benefit (NMB) to assess the cost-effectiveness. For each active screening strategy, the NMB was calculated as:

$$\begin{aligned}\text{NMB} = {} & \text{WTP} \times \text{DALYs averted} \\ & - \text{Cost of active screening strategy compared to passive surveillance only.}\end{aligned} \tag{4.1}$$

The willingness to pay (WTP) is the maximum amount of money that the funder is prepared to pay to gain the health benefit of averting one DALY. The number of DALYs averted is the negative change in DALYs from the comparator strategy (specifically, implementing the active screening programme from only passive surveillance and no active screening). The change in costs is the cost of implementing the active screening strategy, including the consequential change in cost of passive surveillance, minus the cost of baseline passive surveillance. A particular strategy is more beneficial than the comparator strategy if the NMB is positive. The NMB will be highly dependent on the value of the WTP, hence, we consider a range of fixed WTP thresholds, such that an optimal strategy can be determined for a given WTP. Typically, the WTP is taken as the product of the gross domestic product (GDP) per capita of a country and a multiplying factor. This factor is traditionally given as 3 [274], but this is often considered too high

for low-income countries [275] and so 0.5 is also used [276, 277]. Additional to the WTP, the NMB will be affected by the population size of a village, N_H, the proportion of infections, excluding those detected in active screening, that are detected and treated passively p_t, and the initial level of infection in the population.

To evaluate the NMB, we consider all component costs and the associated health benefits of implementing an active screening strategy. However, we do not consider out-of-pocket expenses, and so frame the model as from the perspective of the funder of the programme. Since we are considering the NMB of an active screening strategy, we only consider costs impacted by this and so do not consider costs such as maintaining a fixed health centre, which would exist regardless, given our baseline strategy of continued passive surveillance. We do note that passive treatment costs will depend on the quantity of active screening carried out and additionally on the proportion of infections that are detected passively. Since we assume 29% of infections that are not found in active screening are reported, we can make the worst case assumption that while all these reported infections are treated passively, none of the unreported infections will be treated. Alternatively, we use the parameter p_t for the proportion of infections, excluding those detected in active screening, that are treated passively. This parameter takes values in the interval 29–100%, such that the proportion of infections being treated is between only those reported being treated and all infections being treated.

Using the formulation of the NMB given by 4.1, we then define the constituent parts of the equation as follows:

$$\text{WTP} = WTP_c \times G, \tag{4.2}$$

DALYs averted

$$= D_1 (1 - p_t) \sum_t (\text{Change exits from Stage 2 infection }) (t) \zeta (t) \tag{4.3}$$

$$+ D_2 \int (\text{Change in Stage 1 infections}) (t) \zeta (t) \, \mathrm{d}t \tag{4.4}$$

$$+ D_3 \int (\text{Change in Stage 2 infections}) (t) \zeta (t) \, \mathrm{d}t, \tag{4.5}$$

Change in costs

$$= C_1 N_H (\text{Years of active screening}) \zeta (t) \tag{4.6}$$

$$+ C_2 N_H (\text{Active screening visits}) \zeta (t) \tag{4.7}$$

$$+ C_3 \sum_t (\text{Number of people in active screening}) (t) \zeta (t) \tag{4.8}$$

$$+ (C_4 + C_5) \sum_t (\text{Active screening true positives}) (t) \zeta (t) \tag{4.9}$$

$$+ C_4 \sum_t (\text{Active screening false positives}) (t) \zeta (t) \tag{4.10}$$

$$+ C_6 \sum_t \Big((\text{Change in Stage 1 passive infections}) (t) \zeta (t)$$

$$+ p_t (\text{Change in Stage 2 passive infections}) (t) \zeta (t) \Big) \tag{4.11}$$

$$+ C_7 \sum_t (\text{Treatment of Stage 1 cases}) (t) \zeta (t) \tag{4.12}$$

$$+ C_8 \sum_t (\text{Treatment of Stage 2 cases}) (t) \zeta (t) . \tag{4.13}$$

The willingness-to-pay, which is in units of 2018 US dollars per DALY, is split into the GDP per capita of the DRC, G, and a multiplying factor, WTP_c as is common in the literature (Equation 4.2) [274]. For the health benefit of the intervention, the DALYs averted is the discounted sum of the number of years of life lost (YLL) and years lived with disability (YLD). The number of years of life lost is given by the sum of the change in number of people that exit the Stage 2 infection class multiplied by both the proportion of these infections not treated $(1 - p_t)$ and the discounted average years of life lost per death (Equation 4.3). The number of years lived with disability is given by the total time

spent in each infection class multiplied by the associated disability weight (Equations 4.4 and 4.5).

The change in costs is simply the costs incurred by implementing the active screening strategy and the effect on costs of operating passive surveillance. This is given as a capital cost of active screening (Equation 4.6) and the recurrent cost of operating each visit (Equation 4.7), the number of screening tests carried out (Equation 4.8), confirmation of the infection and stage determination for the true positive (Equation 4.9) and negative confirmation for the false positives (Equation 4.10), the change in testing, confirmation and stage determination for passive infections (Equation 4.11), and the treatment of detected cases (Equations 4.12 and 4.13).

For calculating both the health benefit costs and active screening implementation costs, we use a discounting rate of 3% annually [278], denoted in the equations by $\zeta(t) = \exp(-\delta t)$. The time horizon used is 30 years, which is sufficient to capture almost all the costs of the active screening programme (see Figure 4.12). The cost parameters are taken as fixed values and are a function of the cost of the test or treatment, the cost of implementation and the cost of hospitalisation [142]. These costs are also dependent on the type of treatment; we primarily consider active screening to use the CATT algorithm, while Stage 1 and Stage 2 treatments use pentamidine and NECT respectively, but later consider the additional treatments of fexinidazole and acoziborale. The former we expect to become the standard in 2020 for many gHAT patients [279], and the latter could replace it as a single dose cure a few years later if it passes phase III clinical trials and receives an appropriate recommendation [280] (see Figure 4.13).

4.4 Determining cost and health parameters

In developing a parameterised cost function for active screening strategies, we have ensured transferability of costs across time by updating all values to 2018 US dollars (USD). All costs are converted to local currency units in the year of the study, inflated with the consumer price index (CPI) to find the 2018 values, and then converted to 2018 USD with the exchange rate. We also ensure transferability of costs across settings,

by following the 2003 WHO Guide to Cost-effectiveness recommendations [278]. For non-traded items (nurse and doctor time) we turn the USD or local currency unit (LCU) prices into PPP (international dollars) in the year of the cost study and then turn the value in international dollars to local currency (still in the year of the study) of the country where a cost estimate is needed. Then, we use the CPI to inflate costs to 2018 levels and then use the exchange rate with USD to get 2018 USD values.

We provide a list of all the component parameters we used to determine our model cost parameters (Table 4.3). We then convert these parameters into values that can be used directly in our NMB function (Table 4.4). All NMB parameters are taken as fixed values in this study.

All parameters were obtained by an extensive literature search, which were then converted, as described above into 2018 USD. Alternative costings are available in Table 4.3 and explored in Figure 4.13. We are aware that costs change over time and people can have different opinions on the value they should take [7]. We therefore provide a full choice for the values of our cost parameters in generating our results. These can be selected using the developed companion webapp at `https://christopherdavis.shinyapps.io/optimising-ghat-active-screening/`, where a selection of results from this chapter can be viewed for any chosen parameter value.

4.5 A breakdown of the costs of active screening

We now use the stochastic compartmental model from Chapter 3 with updated model structure and a new parameterisation for the Kwamouth health zone in the DRC to simulate our different strategies for active screening. We simulate each strategy by changing the value of our variables: screening coverage c, screening interval t, active zero-detections z_a, and reactive zero-detections z_r (see Table 4.1). In particular we note that in the model, individuals in the human population are classified as either high-risk or low-risk [219], whereby the high-risk population (a small minority, which has been previously estimated to be 8.6% the study health zone of Kwamouth [268]) have a higher exposure to tsetse and do not participate in active screening. This means that the

Tab 4.3: Net monetary benefit component parameters that determine our full NMB parameters. All values are assumed fixed. All costs are in 2018 USD.

Parameter	Description	Value	Source
D_{death}	Age of death from infection	27 years	[281]
D_{life}	Life expectancy	60.21 years	[244]
D_{stage1}	Disability weight – Stage 1	0.133	[282]
D_{stage2}	Disability weight – Stage 2	0.5432	[282]
N_{cap}	Active screening capacity per year	60,000	[154]
C_{cap}	Capital costs of a traditional active control team	\$11,406	[266]
C_{rec}	Recurrent costs of a traditional active control team	\$40,000	[266]
C_{catt}	CATT algorithm cost	\$0.57	[154, 263]
M_{wasa}	CATT wastage in active screening mark up	1.08	[266]
M_{wastep}	CATT wastage in passive surveillance mark up	1.25	[265]
C_{rdt}	RDT algorithm cost	\$1.496	[154]
C_{mic1}	Microscopy (blood sample, LNA, mAECT)cost	\$9.53	[265]
C_{mic2}	Microscopy (blood sample, LNA, CTC, mAECT) cost	\$11.30	[265]
C_{lumb}	Lumbar puncture and laboratory exam cost	\$23.02	[283]
C_{out}	Outpatient consultation cost	\$0.43	[284, 285]
C_{in}	Hospital day cost	\$1.40	[284, 285]
C_{pent}	Course of pentamidine cost	\$54	[264]
L_{pent}	Length of hospital stay for pentamidine	7 days	[286]
C_{nect}	Course of NECT cost	\$460	[178]
L_{nect}	Length of hospital stay for NECT	10 days	[178]
C_{fex}	Course of fexinidazole cost	\$50	[142]
L_{fex}	Length of hospital stay for fexinidazole	10 days	[178]
M_{del}	Drug delivery mark up	1.45	[284, 285]
M_{prog}	National programme mark up	1.15	[266]

Tab 4.4: Parameters for calculating the NMB.

Parameter	Name	Value	Formulation
WTP_c	Willingness-to-pay per DALY	Varies	WTP_c
G	GDP per capita for the DRC[1]	\$457.85	G
D_1	Discounted average years of life lost per death	21.03 years	$D_{\text{life}} - D_{\text{death}}$
D_2	Stage 1 disability weight	0.1330	D_{stage1}
D_3	Stage 2 disability weight	0.5432	D_{stage2}
C_1	Active screening capital cost per person	\$0.22	$C_{\text{cap}} M_{\text{prog}} / N_{\text{cap}}$
C_2	Active screening recurrent cost per person	\$0.77	$C_{\text{rec}} M_{\text{prog}} / N_{\text{cap}}$
C_3	Active screening test per person [2]	\$1.03	$C_{\text{catt}} M_{\text{del}} M_{\text{wastea}} M_{\text{prog}}$
C_4	Confirmation per person [3]	\$10.96	$(C_{\text{mic1}} + C_{\text{lumb}})\, M_{\text{prog}}$
C_5	Stage determination per person [3]	\$26.46	$C_{\text{mic1}} M_{\text{prog}}$
C_6	Passive screening person [3,4]	\$39.12	$(C_{\text{catt}} M_{\text{del}} M_{\text{wasp}} + C_{\text{out}} + C_{\text{mic1}} + C_{\text{lumb}}) M_{\text{prog}}$
C_7	Stage treatment per person [5]	\$101.32	$(C_{\text{pent}} M_{\text{del}} + C_{\text{in}} L_{\text{pent}})\, M_{\text{prog}}$
C_8	Stage 2 treatment per person [6]	\$783.15	$(C_{\text{nect}} M_{\text{del}} + C_{\text{in}} L_{\text{nect}})\, M_{\text{prog}}$
N_H	Population size	Varies	N_H

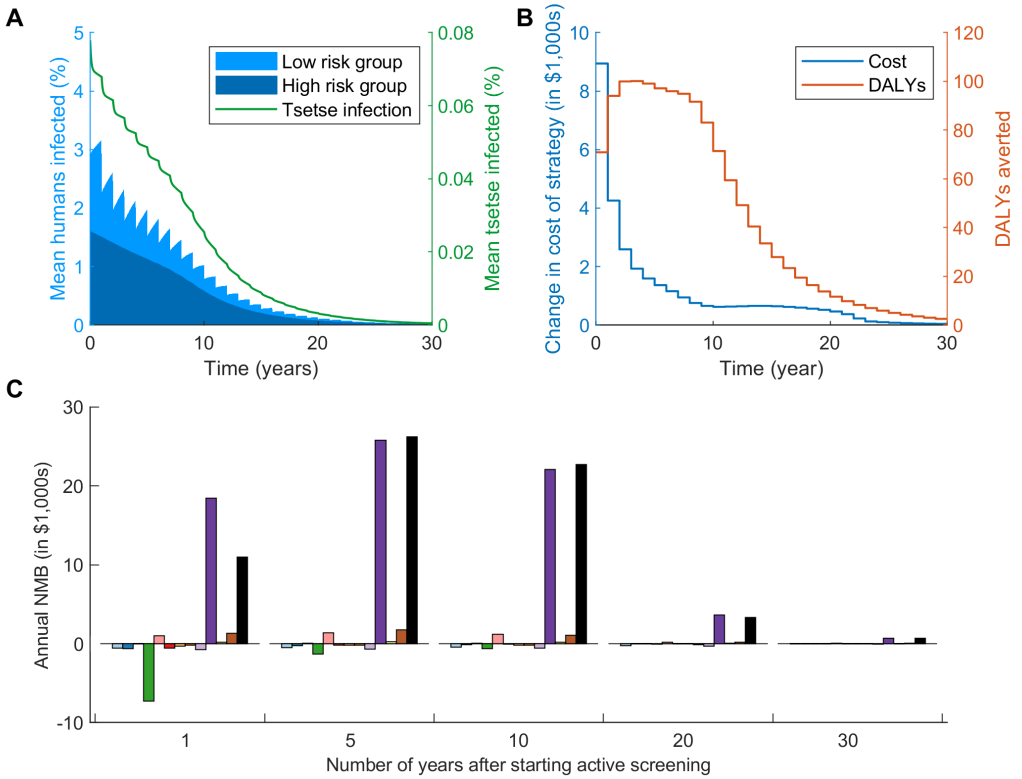
[1]2018 value used. [2]We assume the use of the card agglutination test for trypanosomiasis (CATT) test. [3]We assume confirmation is by microscopy using a blood sample, lymph node aspiration (LNA) and mini Anion Exchange Centrifugation Technique (mAECT). [4]Passive screening includes the screening test (CATT), outpatient consultation, confirmation and stage determination. [5]We assume the use of pentamidine. [6]We assume the use of nifurtimox-eflornithine combination therapy (NECT).

screening coverage is assumed to have a maximum of 91.4%, since only the low-risk human population participate (randomly) in active screening.

We first consider an active screening strategy with a typical screening coverage of 55% (median coverage for Kwamouth), carried out annually, and with three active zero-detections and one reactive zero-detection required for the cessation of screening ($c = 55\%$, $t = 1$ year, $z_a = 3$, and $z_r = 1$), for a village population starting from endemic equilibrium of size $N_H = 1,000$, and with 29% of infections not detected in active screening treated passively ($p_t = 29\%$). We simulate the model for one million realisations and calculate mean values of the simulations. For the chosen strategy,

the infection in both the human and tsetse populations rapidly decays towards zero (Figure 4.3A). The annual cost of implementing this strategy also decreases with time (Figure 4.3B); this is in part due to 3% discounting, the method of adjusting future costs to present day value (which is applied to both costs and DALYs averted) [278], but also because decreasing the infection in the population reduces the required number of treatments. Since even with no active screening, the infection may die out in the village due to stochastic fade out, there is a small annual increase in costs after ten years since the difference in the number of infections treated in passive surveillance is smaller in later years, however, this decays towards zero as the probability of gHAT extinction increases with time. While some recurrent costs will remain, the number of treatments will decrease in time with the corresponding reduction in infections. In addition, the cost also decreases since if the consecutive zero detection threshold is reached, the cost of the active screening is completely removed. With the assumption of $p_t = 29\%$, the number of DALYs averted will initially increase each year, due to the effect of fewer people becoming infected and more being treated, before decreasing with the discounting (Figure 4.3B).

A breakdown of the components for the NMB of implementing this active screening strategy shows that the biggest costs are the treatment of Stage 2 individuals, the recurrent costs of the active screening and screening populations with the CATT test (Figure 4.3C). However, assuming a WTP of 50% of the GDP per capita of the DRC ($WTP_c = 0.5$ giving a total WTP of \$280.89 [287]) the monetary benefit in reducing the years of life lost is dominant and the biggest factor in maximising the NMB. The total NMB (black bars) shows the full benefit of this active screening strategy is always positive and so, on average, this strategy is better than the comparator of no active screening. Further into the future, the NMB moves closer to zero, both because of discounting and since there is a higher probability the infection will be locally extinct, and so active screening not required. The table in Figure 4.3C shows the NMB breakdown in full.



		Annual net monetary benefit in year n ($)				
Colour	Benefit/cost description	$n = 1$	$n = 5$	$n = 10$	$n = 20$	$n = 30$
	Active screening tests	-559.45	-496.19	-426.52	-252.74	-28.74
	Active Stage 1 treatment	-642.69	-278.26	-143.37	-13.58	-1.36
	Passive Stage 1 treatment	$+25.10$	$+38.16$	$+73.69$	$+17.21$	$+3.31$
	Active Stage 2 treatment	$-7,259.56$	$-1,304.34$	-614.10	-54.08	-5.42
	Passive Stage 2 treatment	$+1,006.08$	$+1,407.29$	$+1,203.59$	$+199.86$	$+36.48$
	Active gHAT confirmation	-579.10	-106.30	-78.14	-5.12	-0.76
	Passive gHAT confirmation	-308.39	-216.94	-193.73	-27.83	-4.26
	Active capital costs	-215.37	-191.01	-164.20	-97.30	-11.06
	Active recurrent costs	-766.67	-679.97	-583.27	-324.41	-34.81
	Averted years of life lost	$+18,448.30$	$+25,805.15$	$+22,069.92$	$+3,664.74$	$+669.01$
	Averted years lost due to Stage 1 disability	$+174.77$	$+251.10$	$+187.28$	$+36.36$	$+6.98$
	Averted years lost due to Stage 2 disability	$+1,286.63$	$+1,770.72$	$+1,077.10$	$+162.83$	$+29.76$
	Total	$+10,967.30$	$+26,237.66$	$+22,682.17$	$+3,345.44$	$+666.50$

Fig 4.3: The cost of active screening for a coverage of 55%, a screening interval of 1 year, stopping active screening after 3 screenings when no cases are detected and stopping reactive screening after 1 screening with no cases. (A) The number of infected people dramatically decreases with time for this coverage (total shaded blue area, with left axis) with the majority of these infections being in the high-risk group (darker blue fraction). The proportion of tsetse that are also infective is reduced with time (green line, with right axis). (B) The mean total change in costs of implementing a particular screening strategy (left axis) and the mean number of DALYs averted from the baseline of only passive surveillance (right axis). (C) The contribution to the cost from each component of the cost function for years 1, 5, 10, 20 and 30 after starting an active screening program. Full costs are given in the table below. A population size of $N_H = 1,000$ is used. All costs in 2018 US dollars.

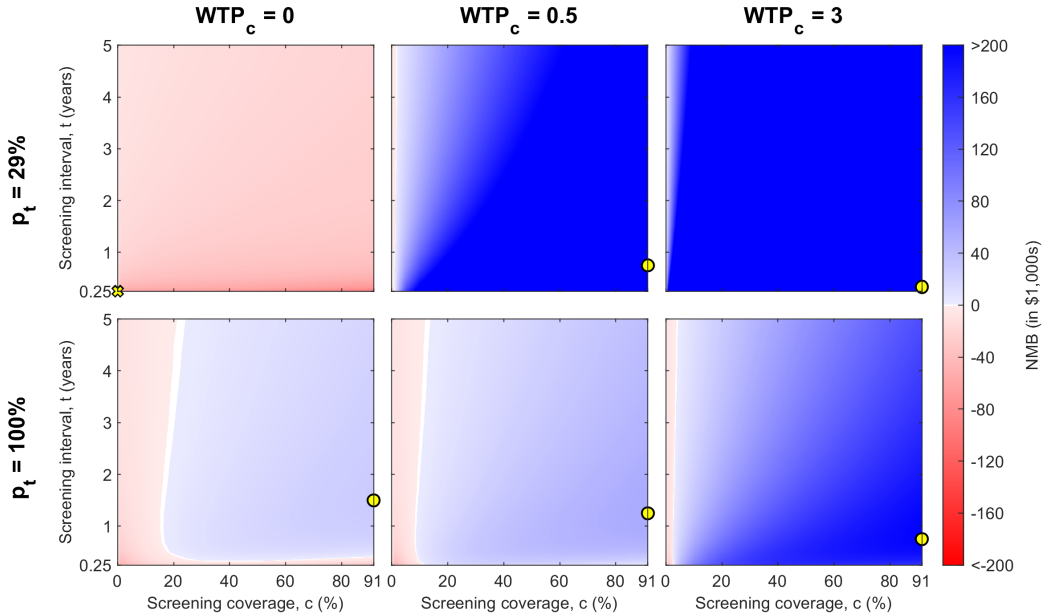


Fig 4.4: The NMB of different active screening strategies for given WTP per DALY averted (given as multiplication factor WTP_c of GDP per capita from the DRC) and the proportion of passive infections that are treated, p_t. We show the mean NMB value over one million simulations. The red areas show a negative NMB, while blue areas are positive NMB, with white at the boundary of no change. The maximum NMB for each WTP_c and p_t combination is marked by a yellow circle on each heatmap, with a cross if the maximum is for no active screening (only observed here for $WTP_c = 0$ and $p_t = 29\%$). A population size of $N_H = 1,000$ is used and we fix $z_a = 3$.

4.6 Variations of costs between strategies

The NMB for some strategies is highly dependent on both the WTP and the proportion of passive infections treated p_t, with $(1 - p_t)$ suffering disease-induced mortality. Thus, for different values of WTP and p_t, we directly consider the value of the NMB across different screening strategies, by varying the screening coverage c and screening interval t (Figure 4.4). Increasing the WTP is particularly effective in increasing the NMB for active screening strategies where lots of people are screened who otherwise would not have been; more infections are likely to be detected and so more DALYs will be averted, which are increasingly high valued.

When $p_t = 100\%$, the assumption is that all infections are eventually treated; this means that there is no loss of life, which would otherwise be a large component of the change in NMB. Thus, for high p_t, active screening strategies with low coverage will have a

lower NMB than for when p_t is smaller, since the implementation costs will be high, but the effect will be minimal as baseline passive surveillance is already very effective. Fewer strategies will achieve a positive NMB for higher p_t, however, if the coverage is high enough, these strategies can still have a positive NMB (Figure 4.4).

For $WTP_c = 0.5$ and $p_t = 29\%$ (our standard assumption), the maximum NMB is found when the screening coverage is at the maximum of 91.4% and the screening interval is at 0.75 years (Figure 4.4; yellow dot in top centre panel). This is true for all values of the number of zero-detections, with the minimum found at $z_a = 1$ and $z_r = 1$.

4.7 Maximising the net monetary benefit

We present the active screening strategy that gives the maximum increase in NMB over our comparator strategy over a range of values of WTP, since the actual value of WTP is entirely at the discretion of the funder. To do this, we apply costs (see Table 4.4) to the mean simulation outputs to see which strategy, on average, provides the largest NMB. Each line in Figure 4.5A–C shows the value of c, t and z_a that together give the optimum strategy, for given p_t values at each value of WTP. For $WTP_c = 0.5$ and $p_t = 29\%$, we maximum NMB is obtained for $c = 91.4\%$, $t = 0.75$ years, and $z_a = 1$ (Figure 4.5A–C). Figure 4.5D–F considers the same results for different fixed N_H and initial conditions of the simulation, all with $p_t = 29\%$. We fix $z_r = 1$ for all simulations.

Screening coverage, c, is optimal when either 0% (not doing any screening) or at a very high coverage (the maximum of 91.4%). For $p_t = 29\%$, active screening at the maximum coverage becomes optimal for $WTP_c > 0.03$ (Figure 4.5A). This means if there is no WTP to avert DALYs, it is best not to incur any screening costs, since the NMB will be negative, however, if the WTP is above this threshold it becomes worth screening entire village populations to reduce the number of infected people and prevent further transmission. The threshold WTP where maximal screening coverage becomes optimal decreases for larger values of p_t. This is because more infections are already being treated and so the additional cost of active screening is smaller and it becomes beneficial to treat all cases faster if they will eventually be treated regardless.

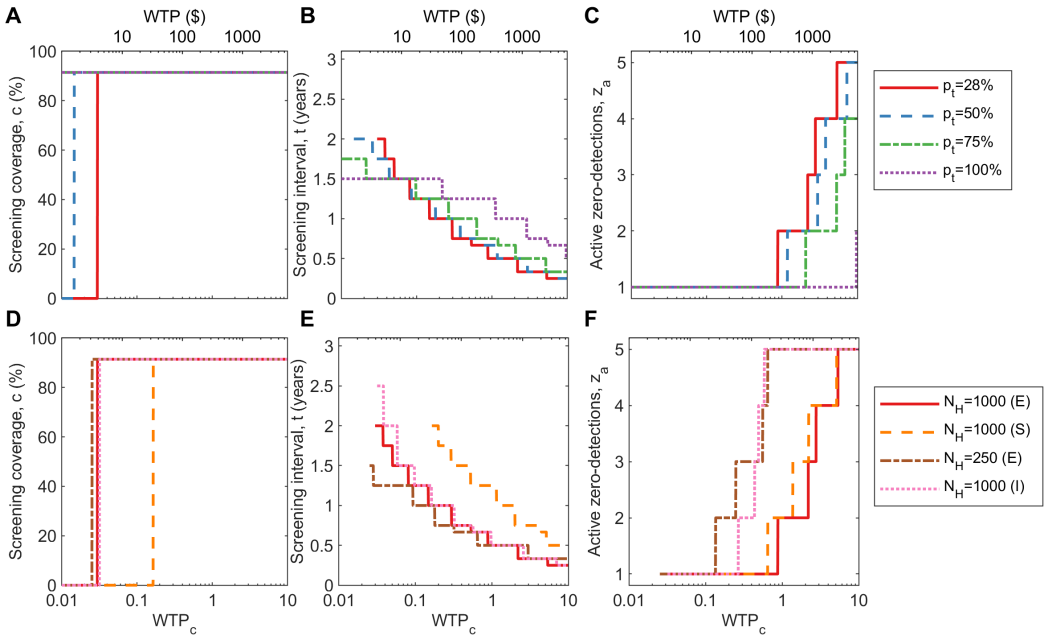


Fig 4.5: Theoretical optimum strategy for mean simulation of infection dynamics given WTP. (A-C) The screening coverage, screening interval and number of zero-detections required to stop screening (respectively) that achieves the highest NMB for given WTP assuming a village population of 1,000, where the disease is endemic. (D-F) Optimum screening coverage, screening interval and number of zero-detections required to stop screening additionally for a village population of 250, a population of 1,000, where only one person is initially infected (S) as opposed to endemic (E), and an endemic population of 1,000 with a small probability of infectious importations being introduced (I).

The screening interval, t, decreases with increasing WTP, since screening more frequently will reduce transmission more quickly, which is more advantageous when DALYs are valued more highly. We also show that in this scenario active screening can only be optimal when it is implemented every two years or more frequently, typically approximately annually (Figure 4.5B). For most WTP values, a single zero-detection is enough to justify cessation of active screening, however as the funder becomes willing to pay larger amounts to avert DALYs caused by gHAT, there is benefit in waiting longer before stopping active screening to ensure the infection is completely removed from the population and to prevent any DALYs being incurred (Figure 4.5C).

Additionally, we have considered three other scenarios: a smaller village population $N_H = 250$; the population starting with a single infection as opposed to the disease being endemic, mimicking a local post-elimination reintroduction of the infection; and when a small chance an imported infection is present (Figure 4.5D–F). For scenarios where the population size $N_H = 1,000$, the results are qualitatively similar, with less active screening optimal for when there are fewer infections (due to the single reintroduction), and more active screening optimal when there are more infections (due to a low level of imported infections). For a smaller population $N_H = 250$, it is more effective to have a shorter screening interval and more zero-detections to be sure of local elimination, since the cost of active screening is smaller when there are fewer people needing to be screened.

The results in Figure 4.5 calculate the optimal strategy for the mean infection dynamics for specific p_t values. However, since p_t is a continuous variable that can take any value in the interval between the values presented in Figure 4.4, we additionally show heatmaps for the optimal strategy across all p_t values (Figure 4.6A–C). The associated NMB of the optimal strategy and broken down into the cost of the strategy and the number of DALYs averted is also displayed in Figure 4.6D–F).

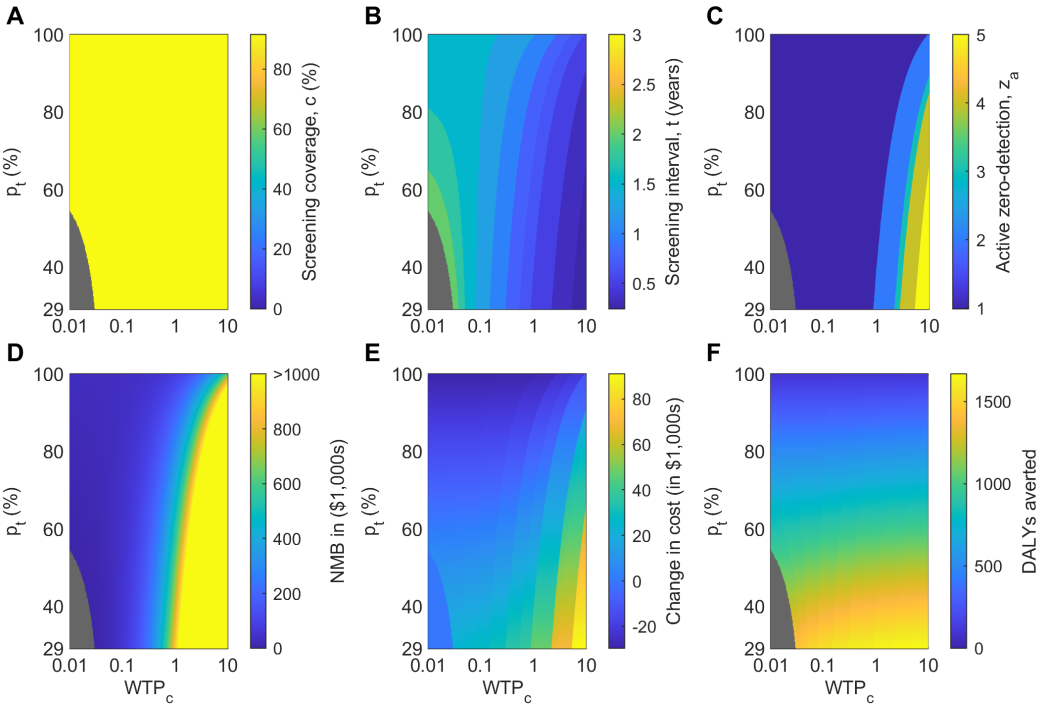


Fig 4.6: Theoretical optimum strategy for mean simulation of infection dynamics given WTP and p_t. (A-C) The screening coverage, screening interval and number of zero-detections required to stop screening (respectively) that achieves the highest NMB. (D) The NMB of that strategy. (E) The change in cost. (F) The number of DALYs averted. We use a village population of 1,000, where the disease is endemic. The plots are grey where no active screening is the optimal strategy.

4.8 Limiting analysis to practical strategies

A high screening coverage on an active screening is desirable for strategy planners in order that the maximum number of people are screened for a minimum number of visits [115]. We have also shown that this is the most beneficial way of implementing active screening (Figure 4.5). However, unlike the screening interval and active zero-detections, which are determined by when active screening teams choose to visit, it is not realistic to choose a level of screening coverage. The screening coverage will depend on availability and consent of the population to be tested. The village population may be reluctant to submit to screening or simply not be present in the village when the screening teams arrive [117]; attendance at active screening is often low [288]. In Kwamouth, for 2000–2016, a median screening coverage of 55% was achieved for all village-level active screenings taken from the WHO HAT Atlas [106, 107].

Therefore, since a high screening coverage cannot be guaranteed, we consider the optimum active screening strategy over mean infection dynamics when we have imposed a maximum on the screening coverage. This achieves an understanding of what the best strategy could be for a given level of obtainable screening coverage. A higher maximum level of screening coverage means the optimum active screening strategy will have a large screening interval and a small number of zero-detections required to justify the cessation of active screening. For the very low maximum screening coverage of 5%, the optimum is a screening four times a year $t = 0.25$ and five zero-detections to stop $z_a = 5$, while for high coverages, we see the expected result of $t \approx 1$ and $z_a = 1$ (Figure 4.7). For the median screening coverage in Kwamouth of 55%, the optimal strategy is an active screening every six months, with two active zero-detection required for cessation.

4.9 Considering cost-effectiveness

While we have determined which strategy, on average, maximises the NMB and so can be considered optimal (and for achievable levels of screening coverage), we now consider cost-effectiveness. Therefore, we now restrict the number of strategies we consider to a smaller number of options. For this process, we have selected seven options: doing no

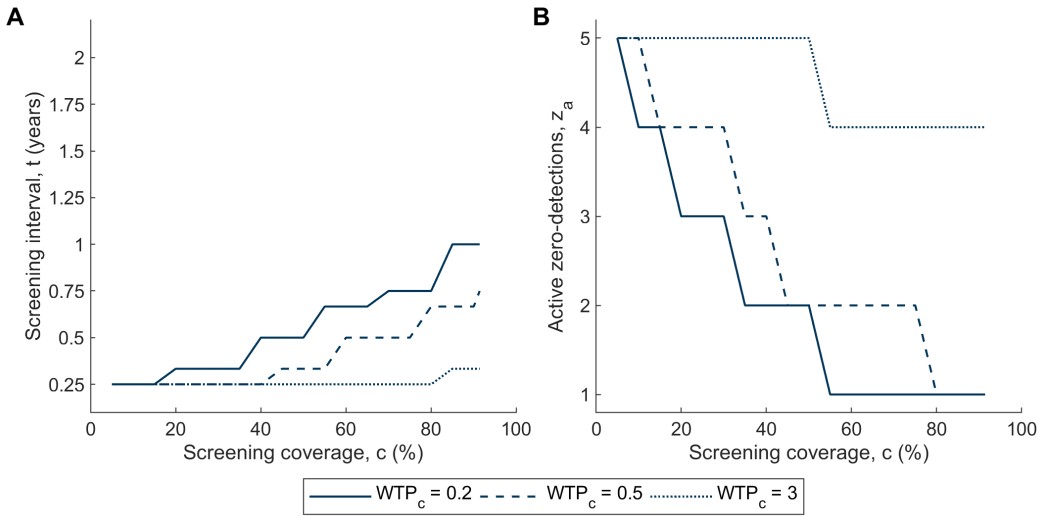


Fig 4.7: Optimal strategy given a maximum screening coverage. Results are shown for a $WTP_c = 0.2, 0.5, 3$. $z_r = 1$, $p_t = 29\%$, $N_H = 1,000$ are fixed and the optimum t and z_a is found simultaneously. (A) Optimal screening interval t. (B) Optimal number of zero-detections to stop screening z_a.

active screening and six realistic proposal schemes for active screening including biennial and annual screening with different cessation criteria. These active screening strategies are shown in Table 4.5 and we keep $z_r = 1$, $p_t = 29\%$ and $N_H = 1,000$ constant, with simulations started from the endemic equilibrium initial conditions.

For our comparator strategy, passive surveillance only, the total cost of implementing this as a strategy for active screening, is the cost of testing and treating self-presenting individuals infected with gHAT. We still do not include the fixed costs of continually operating passive surveillance, as we assume that implementing a strategy does not change this cost. Thus, Table 4.5 shows the average cost of only treating self-presenting patients is $36,255.05 and since we are considering this our baseline strategy, zero DALYs are averted from this process.

By implementing some active screening, the total costs increase, and the strategies are able to avert DALYs. Annual screening costs more than biennial screening, since the screening is occurring more regularly, but a correspondingly larger number of DALYs are averted. Increasing the number of active zero-detections, increases costs, but on average, only a very small number of additional DALYs are averted for annual screening (and indeed the 95% confidence intervals for total DALYs averted slightly overlap).

Tab 4.5: Active screening strategies considered in the probability of cost-effectiveness calculations. Costs are in 2018 US dollars.

Strategy	Name	Screening coverage, c (%)	Screening interval, t (years)	Active zero-detections, z_a	Total cost ($)	Total DALYs averted	ICER ($/DALY)
1	Passive surveillance only[1]	0	N/A	N/A	36,255.05 (± 17.97)	0	Minimum cost
2	Biennial screening with one zero for cessation	55	2	1	58,041.25 (± 24.65)	1,073.28 (± 0.68)	20.30[2]
3	Biennial screening with two zeros for cessation	55	2	2	58,563.48 (± 24.36)	1,073.73 (± 0.68)	1,145.74[3]
4	Biennial screening with three zeros for cessation	55	2	3	59,001.25 (± 24.12)	1,074.16 (± 0.68)	1,029.39[4]
5	Annual screening with one zero for cessation	55	1	1	64,822.23 (± 27.22)	1,340.66 (± 0.54)	25.36[3]
6	Annual screening with two zeros for cessation	55	1	2	65,524.62 (± 26.72)	1,341.14 (± 0.54)	1,483.83[5]
7	Annual screening with three zeros for cessation	55	1	3	66,165.95 (± 26.52)	1,341.44 (± 0.54)	2,152.04[6]

[1]The comparator strategy. [2]Relative to Strategy 1. [3]Relative to Strategy 2. [4]Relative to Strategy 3. [5]Relative to Strategy 5. [6]Relative to Strategy 6.

The ICER is the ratio of the change in cost to change in DALYs averted relative to the next best option. These values are shown in Table 4.5. The values determine the WTP thresholds for a screening strategy to be chosen over the next best option in terms of DALYs averted. For biennial active screening to occur the funder needs to be willing to pay $17.91 per DALY, and for annual screening to occur over biennial active screening, $24.60 per DALY is required. Significantly larger WTP values are needed for multiple active zero-detections, but if the WTP was large enough these could be implemented, although not instead of a shorter time interval. However, in considering the high WTP needed for multiple active zero-detections, we have not included any additional benefits for achieving local gHAT elimination that may be highly valued enough to include more active zero-detections in the screening strategy at the additional cost.

Not every village will experience gHAT infection the same as the mean infection profile. Hence, we aim to account for uncertainty and consider each individual realisation of the stochastic epidemic process to measure the probability a strategy is cost-effective. Using the full range of possibilities for the infection dynamics is particularly important for

gHAT infection in a village as we know there is potential for large differences between seemingly identical villages, due to the focal nature of the infection. Therefore, we have simulated the infection dynamics of each strategy one million times to compare how the costs and number of DALYs averted vary.

Evidently, even when there is no active screening, there will be variations in the cost and the number of DALYs incurred. Here, this is particularly well correlated as the cost is due solely to passive detection and treatment. The more infections there are in the village, the more treatments that will be performed at increased cost and the more DALYs that will be incurred, decreasing the total number of DALYs averted from the expected value. When there are few infections, the total number of DALYs for the village will be lower (and so more DALYs averted) and the cost of detection and treatment will also be lower. There is a large variation in the possible outcome in terms of both cost and DALYs averted when there is no active screening, which is entirely dependent on the number of cases of infection that occur (Figure 4.8A).

There is a similar pattern for when active screening exists at some coverage, but the range of introduced costs and the number of occurring treatments increases, expanding the number of possibilities for the overall infection dynamics, and so the correlation is smaller. However, reducing the screening interval, both increases the costs and reduces the number of DALYs incurred. Results for varying the number of zero-detections are still marginal when considering all realisations of the process, as compared with the mean values (Figure 4.8A).

We have calculated the probability a strategy is cost-effective for a given WTP value by taking the proportion of our one million simulations that each strategy has the largest NMB (Figure 4.8B). These are the cost-effectiveness acceptability curves (CEACs). The strategy with no active screening has the highest probability of being cost-effective for low WTP, however, the probability is approximately 50% for the lowest WTP, meaning that the probability of doing some active screening being cost effective, is still relatively high. No active screening has negligible probability of being the most cost-effective for approximately $WTP_c > 0.3$. For the largest WTP values, Strategy 7 (annual screen-

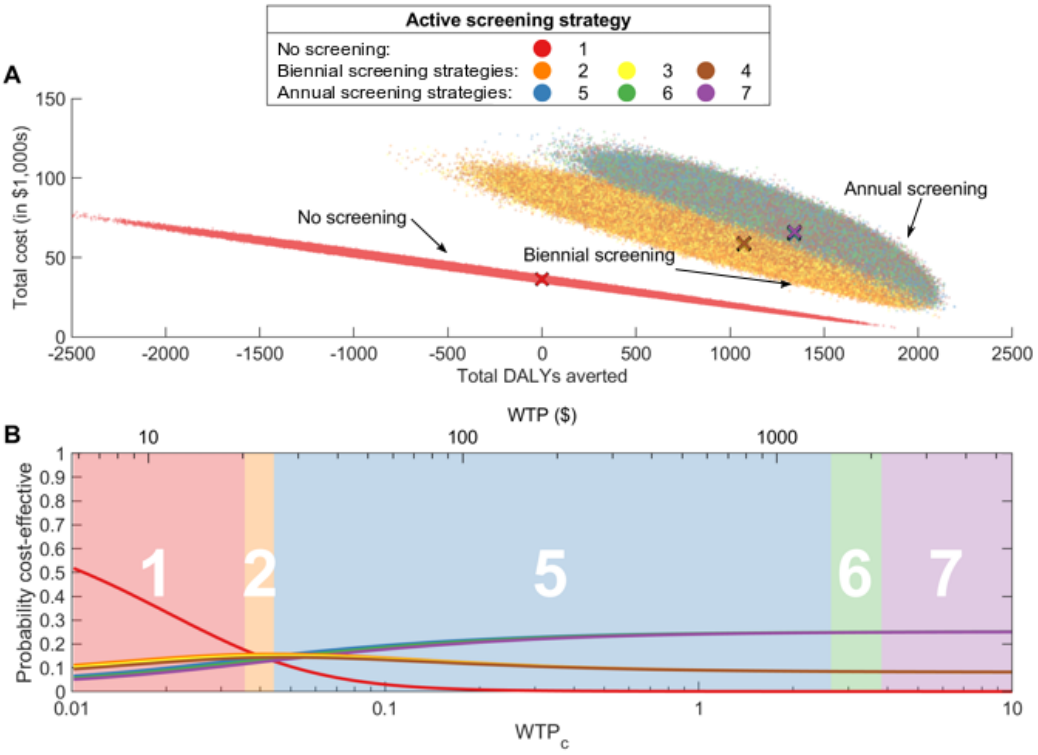


Fig 4.8: The cost-effectiveness of active screening strategies. See Table 4.5 for full description of strategies. (A) Cost-effectiveness plane showing the total cost of a strategy and the associated total number of DALYs averted from the mean value of the comparator strategy. Mean values for each strategy are shown by the coloured crosses. (B) Cost-effectiveness acceptability curves (CEACs) for each strategy are shown by lines, with the cost-effectiveness acceptability frontier (CEAF) shown by the numbered background colour, which demonstrated the values for the ICER. WTP is shown in 2018 USD on the top and as the WTP_c coefficient on the bottom, where the coefficient is the multiplier of the GDP per capita of the DRC.

ing with three zero-detections), marginally has the largest probability, with similar probabilities for the other strategies with annual screening.

It is notable that for values of WTP, the one with the highest probability of being cost effective does not directly correspond with the outcome from the ICER calculations. These calculations are demonstrated by the cost-effectiveness acceptability frontier (CEAF), the optimal strategy in maximising the expected NMB, which is shown as the shaded and numbered background (Figure 4.8B). It is the CEAF which is used to provide the recommended strategy for a given WTP threshold.

4.10 The effects of reactive screening

So far we have only considered when only one reactive screening with no cases detected is required to stop screening again $z_r = 1$. However, to be more certain that there is no infection, more may be required. By considering when $c = 91.4\%$, $t = 1$, and $z_a = 3$, we calculate the probability of being most cost-effective for the four strategies of no active screening and $z_r = 1, 2, 3$ (Figure 4.9). Similarly to Figure 4.8B, a low WTP of less than 3.5% of the GDP per capita of the DRC means no active screening has the highest probability of being cost-effective, and for higher WTP, active screening is beneficial. The effect of z_r is, however, very minimal, with the chosen three values of z_r giving almost equal probability, although $z_r = 1$ has the highest probability and $z_r = 3$ the lowest.

This is explained by there only being a small probability that a reactive screening occurs at all (Figure 4.10A). When $z_a = 3$ and three consecutive zero-detections have already occurred, the probability of continuing to find cases to restart screening is small, roughly 5%. Secondly, if reactive screening does occur, it will occur a long time in the future, where discounting has more of an effect and hence the differences in z_r will be negligible.

We observe when $z_a = 1$ the expected number of screenings will vary a little depending on z_r, but much less for larger values (Figure 4.10B). Indeed, when looking at the

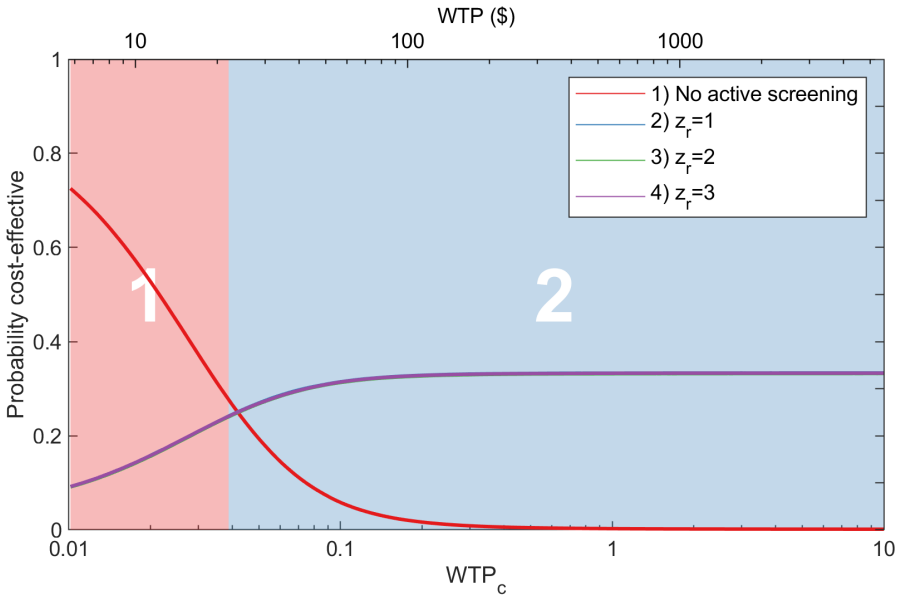


Fig 4.9: The effect of the number of zero-detections in reactive screening. Probabilities that a reactive screening strategy is the most cost effective (CEACs) and the cost-effectiveness acceptability frontier shown in the coloured background. The CEACs for different z_r values show very similar results and cannot be distinguished in the figure. We assume $c = 91.4\%$, $t = 1$, and $z_a = 3$ and the population is of size 1,000 starting from endemic equilibrium with $p_t = 29\%$.

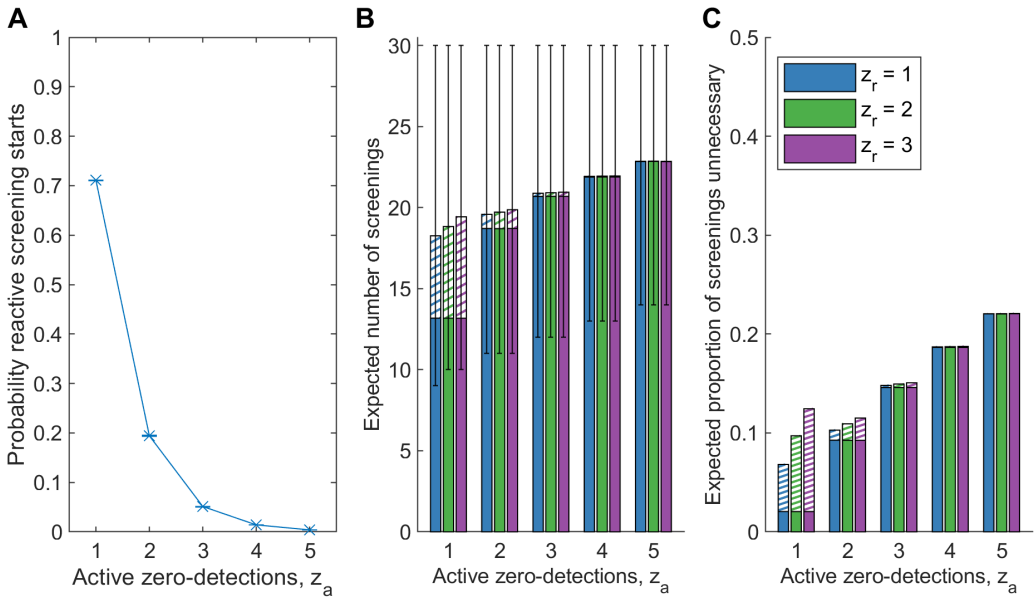


Fig 4.10: Active and reactive screenings. (A) The probability that reactive screening starts for different z_a values. (B) The expected number of screenings. Active screening are solid bars, while striped bars are reactive screenings. The colour of the bar indicates the value of the reactive zero-detections. 95% prediction intervals are shown. (C) The expected proportion of screenings where there is no infection. Sufficient simulations have been run such that confidence intervals are too small to plot.

number of screenings that were unnecessary, with full information of the infection (since there was actually no infection), there are more unnecessary screenings when there are more active zero-detections, but more reactive zero-detections only has significant impact when z_a =1 and 2 (Figure 4.10C).

Hence, we cannot say much about the optimum number of reactive screenings when considering the NMB. The model does not directly consider the benefit of extensively screening a village to achieve local elimination in a village and there is little differentiation in the values for z_r.

Also important to determine reactive screening strategy, is the time between the detection of a passive case when there is no active screening in the village and the starting of reactive screening to that detection. We have primarily assumed that reactive screening restarts immediately, although we recognise there would be some delay. However, we show that there is limited effect of this time interval, except for when the number of active zero-detections is small (Figure 4.11). When $z_a = 1$, there is only one active zero-detection required to stop active screening, and so very possible there is still infection despite when no cases are observed on a particular screening. This means that when the time interval for reactive screening is small, reactive screening can resume very quickly upon a false zero detection. We see the effect of a delay in reactive screening reduces the total NMB slightly for $z_a = 1$ and very negligibly for $z_a = 2$, with any effect of higher z_a values completely outweighed by the stochasticity of the epidemic process. We note in practice, there would be some delay before reactive screening began, but this difference is small, less than $1,000 difference for a delay between 0 and 1 years for the $z_a = 1$ value with the largest effect.

4.11 Time horizons

In all simulations in this chapter, we have used a time horizon for 30 years, since this is sufficient to capture the majority of the costs of an active screening strategy on a village until cessation and similarly the number of DALYs averted. Figure 4.12 shows the distribution of the change in costs and the number of DALYs averted when the

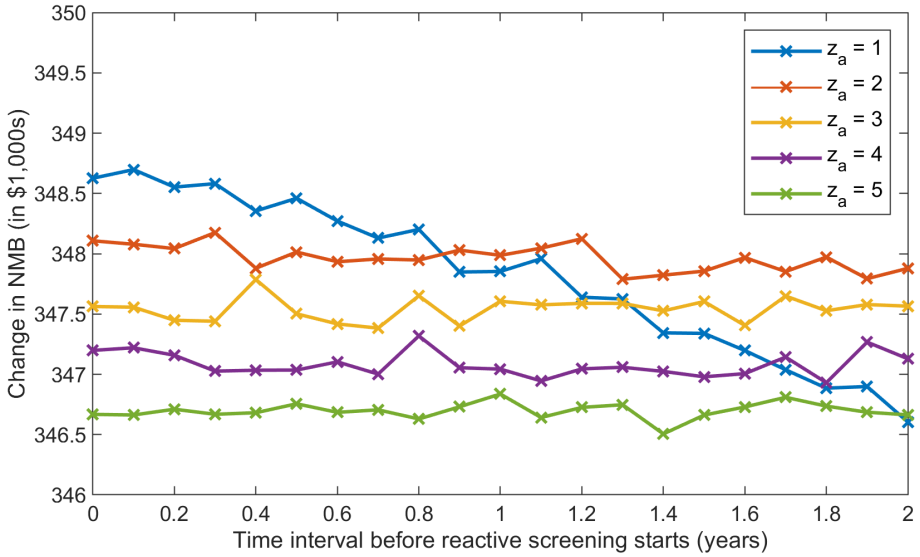


Fig 4.11: The time interval between observing a passive detection and resuming reactive screening for different values of active zero-detections z_a.

simulation is run for both 30 and 100 years using a 55% screening coverage annually with 3 zero-detections in a population of 1,000 starting from endemic equilibrium with 29% of passive infections treated ($c = 55\%$, $t = 1$ year, $z_a = 3$, $N_H = 1,000$, $p_t = 29\%$). The distribution is broadly identical since only a minimal amount of cost is included additionally after the 30 years. Indeed, we see that in 94.6% of the simulations no additional cost is included or DALYs incurred by expanding the time horizon to 100 years. The mean cost for 30 years is 99.2% of the mean cost for 100 years, with the mean DALYs incurred in 30 years being 99.8% of those incurred over 100 years. Hence, we are satisfied using a time horizon of 30 years.

4.12 Alternative diagnostics, treatments are active screening methods

We have used fixed cost and benefit parameters for all simulations. These costs are liable to change over time and have the potential to change dramatically if new practices are implemented, such as a change in treatment. Thus, we have a supplementary R Shiny [289] web app at `https://christopherdavis.shinyapps.io/`

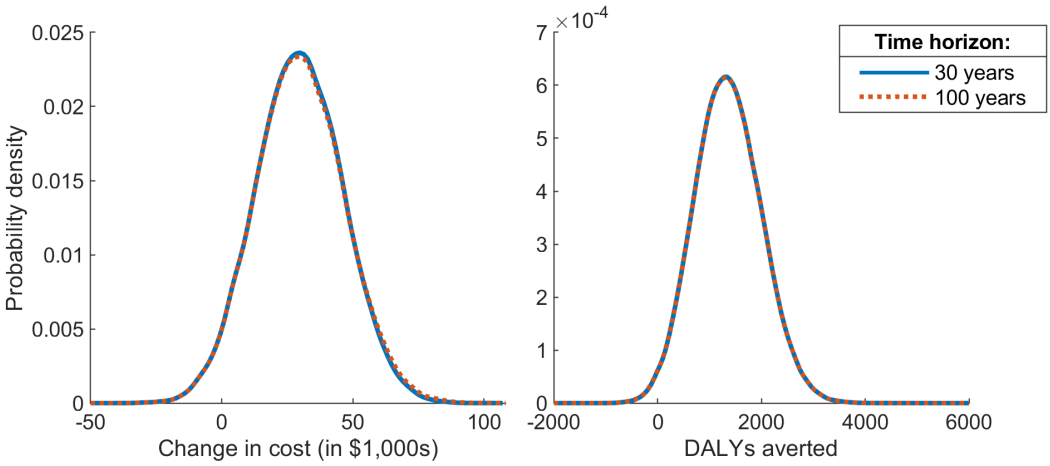


Fig 4.12: Distributions of (A) the change in cost of an active screening strategy and (B) number of DALYs averted for 30 and 100 year time horizons. The simulation is for 55% screening coverage annually with 3 zero-detections in a population of 1,000 starting from endemic equilibrium with 29% of passive infections treated.

`optimising-ghat-active-screening/`. A selection of figures can then be generated for any updated costings for the active screening strategies.

In particular, we focus here on how the change in costs would change in four different scenarios: CATT screening tests were replaced by RDT; capillary tube centrifugation (CTC) was additionally used in the microscopy procedure; NECT for Stage 2 treatment was replaced by fexinidazole; all treatments replaced by fexinidazole (Figure 4.13). The assumed increased cost of RDT compared to CATT increases the total change in cost most significantly as it would be applied to all those undergoing active screening, while the reduced cost of fexinidazole treatment, reduces costs slightly. The effect of CTC in microscopy is negligible when considering a whole village for 30 years.

4.13 Differences between health zones

We have only considered the health zone Kwamouth in the analyses in this chapter. To consider different health zone parameterisations show vastly different results, we use our gHAT model parameterised to incidence data from the Mosango health zone and compare the probability of a strategy being cost-effective the same strategies as Kwamouth, but using the median screening coverage of the typical screenings from Mosango (57%), along with the model parameters for this health zone.

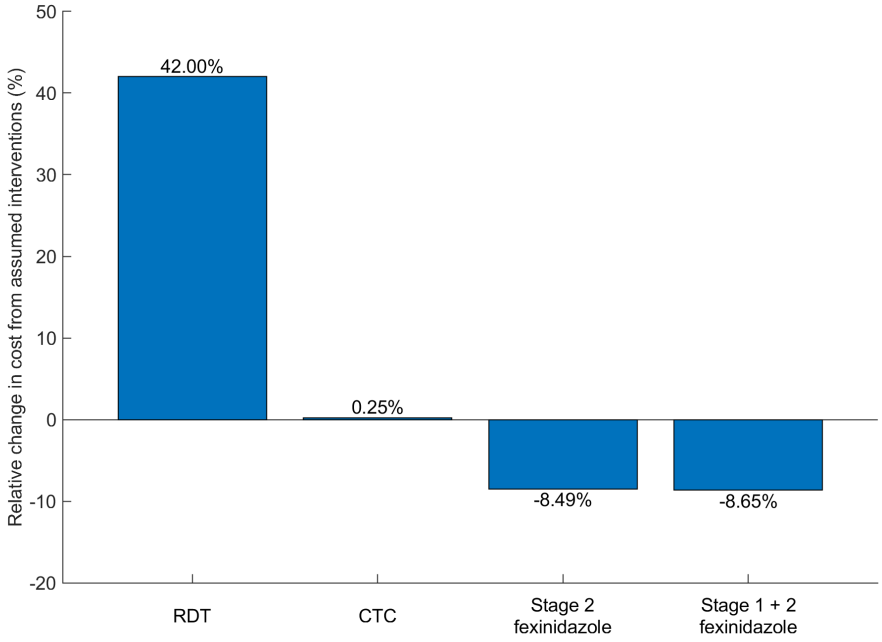


Fig 4.13: Mean change in costs relative to the assumed interventions for a village with 55% screening coverage annually with 3 zero-detections. We consider the replacement of CATT with RDT, the addition of CTC in microscopy and the use of fexinidazole in Stage 2 patients and all patients. We use a population of 1,000 starting from endemic equilibrium with 29% of passive infections treated.

Qualitatively, we see that the cost-effectiveness plane for Mosango looks similar to Kwamouth, however, since there are fewer reported cases in Mosango, it is not possible for so many DALYs to be averted. Hence the cost of implementing the active screening strategies is less, as they are terminated earlier than Kwamouth (Figure 4.8; Figure 4.14). The WTP required to carry out active screening is also less for Mosango, shown by the cost-effective acceptability frontier (Figure 4.14), but for $WTP_C > 0.03$, annual active screening is the most cost-effective strategy in terms of both the cost-effective acceptability frontier and cost-effective acceptability curves.

4.14 Discussion

To achieve the goal of eliminating gHAT it is essential to have robust models that can inform policy makers about the potential of different intervention strategies [201]. Furthermore, the addition of economic analysis will further develop the use of this work,

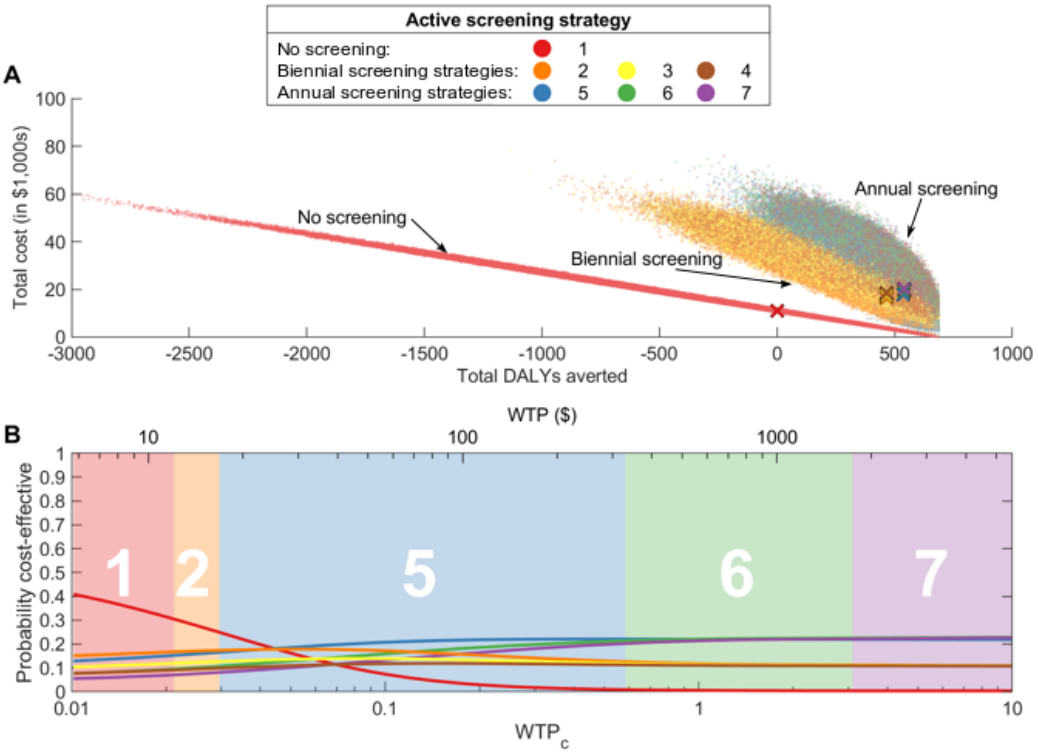


Fig 4.14: The cost-effectiveness of active screening strategies for the health zone Mosango. (A) Cost-effectiveness plane showing the total cost of a strategy and the associated total number of DALYs averted from the mean value of the comparator strategy. Mean values for each strategy are shown by the coloured crosses and ellipses are fitted, which contain 95% of all data points from the one million simulated realisations of the infection process for each strategy. (B) Cost-effectiveness acceptability curves (CEACs) for each strategy are shown by lines, with the cost-effectiveness acceptability frontier (CEAF) shown by the numbered background colour, which demonstrated the values for the ICER. WTP is shown in 2018 USD on the top and as the WTP_c coefficient on the bottom, where the coefficient is the multiplier of the GDP per capita of the DRC.

as to not only evaluate which strategies are able to decrease infection in the population, but which are additionally cost-effective.

We have presented a stochastic model for individual villages that demonstrates how active screening should be considered, by determining costs of implementing the screening for different screening coverages c, screening intervals t, and the number of zero-detections observed to stop screening z_a and z_r. Unsurprisingly, we find that a big factor in choosing a strategy to implement is how much the programme funder, ministry of health or external donor is willing to pay to avert a DALY. Thus, we present our results across a range of WTP values. We have typically used a WTP value of 0.5 of the GDP per capita of the DRC ($WTP_c = 0.5$), which is commonly used in the literature [276, 277]; and we find on average that ideally screening would be done approximately yearly with maximal screening coverage and ceased when no infection is found in a single screening ($c = 91.4\%$, $t = 0.75$ years, $z_a = 1$) (Figure 4.5). While the optimum for the screening interval is found to be 0.75 years, practically this might be difficult to implement when it is close to yearly. It is noted that screening coverages will rarely be able to be achieved this high and so multiple visits where no infection is observed may be necessary to optimise control. This is in line with WHO guidelines of annual active screenings until there have been three consecutive years of no new cases, followed by a further screening with no cases after three years [267].

In particular, we note that the modelling assumption that after stopping screening, the reactive screening should be immediately resumed on identification of a passive infection will not actually occur, but has little effect on the results (see Figure 4.10). In fact, this can lead to an optimal solution with a low z_a, such that it is possible to restart screening many times after quickly reaching the cessation criteria and so induce a higher frequency of screening. However, the time horizon of 30 years is sufficient in the village context to capture the dynamics with roughly 99.2% of costs and 99.8% of DALYs attributable to the first 30 years (Figure 4.12).

As new treatments and active screening modes are introduced, the costs of the model will change, however, the biggest effect is that of the number of DALYs averted, assuming

the WTP threshold is set to a reasonable level. Details of the variation in NMB for different screening diagnostics and medical treatments can be found in Figure 4.13.

We also note that from the perspective of cost-effectiveness of a single village, we do not put any weighting on the benefit of extensive active screening in a village to achieve local elimination, so that this village does not have to be re-visited in the future. This means we often observe the result that the optimum screening strategy should be to terminate the programme after a single zero-detection ($z_a = 1$), rather than continuing to screen to aim for local elimination. However, there will only be high confidence that local gHAT elimination has been achieved with at least three zero-detections for villages of this size ($N_H = 1,000$) (see Chapter 3), but it is unclear how much monetary value we should attribute to meeting EOT targets.

The result of a single zero-detection being optimal is particularly notable in conjunction with when the screening coverage c is at the maximum. In this case, there is higher confidence that local gHAT elimination is achieved, as almost all the population is screened, above what is typically seen, and there are no cases detected. Even if infection temporarily persists after this, there is a large probability it will soon die out due to stochastic fade out (see Chapter 3). More realistic screening coverages will require more zero-detections are required to terminate active screening (Figure 4.7).

Additionally, we note that there will also be additional costs in restarting active screening as reactive screening, if regional cessation has led to trained mobile teams being disbanded. We have not accounted for this in our model and it may increase the optimal number of active zero-detections. In particular we note there is a high probability of reactive screening being required in a village where the number of active zero-detection required to stop screening is only one (Figure 4.10A).

When considering different gHAT interventions, active screening is known to be effective in reducing case numbers and hence the infection in the population [12, 97, 290]. Thus, with a limited number of active screening teams and resources for them to carry out their duties, it is important to optimise their activities with the aim of driving towards elimination.

5

Exploring simplified dynamics of *gambiense* human African trypanosomiasis

We have so far used simulation-based methods to consider the dynamics of gHAT infection, primarily using the tau-leaping algorithm. However, through the use of master equations (forward Kolmogorov equations), we can consider the full probability distribution of being in all possible infection states. This method enables us to explore the dynamics further, including giving exact methods to calculate expected disease extinction times.

In this chapter, we develop a simplified model of gHAT infection that reduces the possible state space, to analytically investigate our model for probabilities of gHAT persistence and expected times until elimination of transmission (EOT). Furthermore, we explore the interaction between villages using this method, by considering importations of infection into each individual village population as proportional to the global level of infection across all villages. This provides a link between all villages in a region, whereby we can scale up the village-level model to consider elimination on a larger scale. This method retains both the stochastic properties of the model and allows us to continue to consider interventions at the level at which they occur, the village-level. We use an adapted model structure and parameterisation from Chapter 3.

5.1 Considering the master equations

The purpose of considering the master equations of our system in this chapter is due to the disadvantages of using the event-driven simulation approaches in analysing our models.

Both the tau-leaping algorithm and the Gillespie direct algorithm are integer-based event-driven simulations, which capture the dynamics of random events, allowing for deviations from the deterministic equilibrium solution to be observed. However, each realisation of the process is just one simulation and so rather than capturing something fundamental of the model, they can be considered as the result of an *in silico* experiment, and thus should be treated in the same way as any other statistical observation [291]. This means a large number of replicates are required to have confidence in the results and this is still only an approximation of the true probability distribution of the potential trajectories for the infection dynamics in time. Furthermore, if we are particularly interested in an incredibly rare event, to ensure we capture the frequency at which this event occurs, or even to observe it at all, a vast number of simulations may be required, and these will take time to compute.

An alternative is to consider the full probability distribution of all combinations of individuals being in each infection state. This complete ensemble, known as the master equations (forward Kolmogorov equations) [292], will comprise of a very large number of ordinary differential equations (ODEs). The system can be written in a matrix formulation, since it is linear in terms of the probability of being in each state [291, 293]. The system is thus easy to solve and provides a complete description of the dynamics. Furthermore, using this approach we are able to directly calculate quantities of interest, such as probabilities of persistence or expected extinction times using exact methods.

The drawback of such methods, and the reason for using the tau-leaping algorithm to approximate the dynamics in previous work (Chapter 3 and Chapter 4), is that the method becomes infeasible for very large population sizes and more biologically realistic models [291]. In the case of our gHAT model, we have two populations of humans in five infection states and a population of tsetse in six infection states. Since the master

equation approach requires the full state space of all possible combinations of this would be a practically impossible number of equations, particularly considering we do not know the total number of tsetse in our populations (and are already taking the infinite limit and approximating their dynamics as ordinary differential equations in previous chapters). Simulation methods take a lot of computation time, master equation methods take a lot of memory storage capacity [294].

To calculate the number of equations required for the full system, we note that there are

$$\binom{n+k-1}{n} \tag{5.1}$$

ways of putting n indistinguishable objects in k distinguishable classes [295]. Thus if we have N_{H1} low-risk individuals and N_{H2} high-risk individuals, the total number of equations will be $\binom{N_{H1}+4}{N_{H1}}\binom{N_{H2}+4}{N_{H2}}$, as there are five classes for each risk group. If we only considered the humans in a typical village population of 1,000 people, using the parameterisation from Chapter 3, with 924 low-risk people and 76 high-risk, this is a very large number of equations. For this population, there are $\binom{924+4}{924} = 30,702,178,200$ possible combination for the low-risk group and $\binom{76+4}{76} = 1,581,580$ for the high-risk group, leading to 48,557,950,997,556,000 ($\sim 4.8 \times 10^{16}$) ODEs for the master equations, even before vectors are considered. For a population of 10,000, this number grows to $\sim 1.7 \times 10^{22}$. Therefore, we need to simplify our model if we are going to take this approach.

5.2 Simplifying the gHAT model

To simplify our model, we need to reduce the potential state space, so that the model can be studied at larger population sizes. As such, one immediate simplification of the model is to observe that the vector dynamics occur at a faster rate than the human dynamics, due to the short life span of the tsetse (approximately one month [230]), compared to the long infections times of humans (typically three years if untreated [77]). Thus, we can assume that the vector dynamics are at quasi-equilibrium and solve the model ODE equations for the vectors (i.e. $\frac{\mathrm{d}I_V}{\mathrm{d}t} = 0$) to find I_V^Q, which we can then substitute into our expressions in the human dynamics. This is common method in the literature for

removing vector dynamics [6, 210, 211, 296, 297], alternative to simply modelling with a force of infection term [298].

Our full vector equations at steady state (for full mode equations and parameters descriptions see Table 3.1 and Table 3.2 respectively):

$$\begin{aligned}
\frac{\mathrm{d}S_V}{\mathrm{d}t} &= \mu_V N_H - \alpha S_V - \mu_V S_V && = 0, && (5.2)\\
\frac{\mathrm{d}E_{1V}}{\mathrm{d}t} &= \alpha p_V \left(\frac{f_{H1} I_{H1}}{N_{H1}} + \frac{f_{H2} I_{H2}}{N_{H2}}\right)(S_V + \epsilon G_V) \\
&\quad - (3\sigma_V + \mu_V) E_{1V} && = 0, && (5.3)\\
\frac{\mathrm{d}E_{2V}}{\mathrm{d}t} &= 3\sigma_V E_{1V} - (3\sigma_V + \mu_V) E_{2V} && = 0, && (5.4)\\
\frac{\mathrm{d}E_{3V}}{\mathrm{d}t} &= 3\sigma_V E_{2V} - (3\sigma_V + \mu_V) E_{3V} && = 0, && (5.5)\\
\frac{\mathrm{d}I_V}{\mathrm{d}t} &= 3\sigma_V E_{3V} - \mu_V I_V && = 0, && (5.6)\\
\frac{\mathrm{d}G_V}{\mathrm{d}t} &= \alpha \left(1 - \left(p_V \left(\frac{f_{H1} I_{H1}}{N_{H1}} + \frac{f_{H2} I_{H2}}{N_{H2}}\right)\right)\right) S_V \\
&\quad - \alpha p_V \left(\frac{f_{H1} I_{H1}}{N_{H1}} + \frac{f_{H2} I_{H2}}{N_{H2}}\right) \epsilon G_V - \mu_V G_V && = 0. && (5.7)
\end{aligned}$$

Solving these equations gives:

$$I_V^Q(I_{H1}, I_{H2}) = \frac{27\sigma_V^3 \alpha p_V \left(\frac{f_{H1} I_{H1}}{N_{H1}} + \frac{f_{H2} I_{H2}}{N_{H2}}\right) N_H (\mu_V + \epsilon\alpha)}{(\alpha + \mu_V)\left(\alpha p_V \left(\frac{f_{H1} I_{H1}}{N_{H1}} + \frac{f_{H2} I_{H2}}{N_{H2}}\right)\epsilon + \mu_V\right)(3\sigma + \mu_V)^3}. \quad (5.8)$$

This expression is a function of just I_{H1} and I_{H2}, since the vectors becomes infected upon taking a blood-meal on an infected human, and so we can replace the occurrence of I_V terms in our description of the human dynamics with this term I_V^Q.

If we were considering the model in the deterministic ODE framework, this simplification reduces the number of ordinary differential equations (ODEs) from sixteen to ten. We further note that as we are using a constant population size for each risk class, we don't need to consider one class explicitly, since, for example, $R_{H1} = N_{H1} - (S_{H1} + E_{H1} +$

$I_{1H1} + I_{2H1}$). However, this is still an eight dimensional system for the state space of the master equations, and, as calculated above, already requires a huge number of ODEs to present the master equations for the system. We need to simplify the system further.

In all our model simplifications, we want to retain the risk structure in the human population as different behaviour is observed from the risk groups, with individuals in the high-risk group having a higher risk of exposure to tsetse and not participating in active screening [117, 219]. Therefore, we simplify the infection profile of humans from passing through exposed, Stage 1 infection and Stage 2 infection, to just a single infected state that includes all these stages. We are reducing the model from a two population SEIIR model to a two population SIR model. Directly adapting the existing model this would now be described in the ODE framework as:

$$\frac{\mathrm{d}S_{Hi}}{\mathrm{d}t} = \mu_H N_{Hi} + \omega_H R_{Hi} - \alpha m_{\text{eff}} \frac{f_i S_{Hi}}{N_{Hi}} I_V^Q - \mu_H S_{Hi} \tag{5.9}$$

$$\frac{\mathrm{d}I_{Hi}}{\mathrm{d}t} = \alpha m_{\text{eff}} \frac{f_i S_{Hi}}{N_{Hi}} I_V^Q - (\mu_H + \psi_H) I_{Hi}, \tag{5.10}$$

$$\frac{\mathrm{d}R_{Hi}}{\mathrm{d}t} = \psi_H I_{Hi} - (\omega_H + \mu_H) R_{Hi}, \tag{5.11}$$

for $i = 1, 2$ for both low- and high-risk humans. We also introduce new parameter $\psi_H = 1/(1/\sigma_H + 1/\phi_H + 1/\gamma_H)$, as the mean expected time to hospitalisation from exposure. We also scale I_{Hi} terms in I_V^Q by $(1/\phi_H + 1/\gamma_H)/(1/\sigma_H + 1/\phi_H + 1/\gamma_H)$ to account for the lack of infectiousness when only recently exposed.

To achieve minimal dimensionality for our model, while keeping the risk structure, we could further reduce the number of ODEs from six to four, by removing the recovered (hospitalised) class. Since the total number of people in the recovered R_H class is small (typically less than 0.4% of the population, unless a large active screening occurs, when using the ODE framework; Figure 5.1), we choose to simply omit this class in our simplified model approximation and move individuals straight from infected (I) to susceptible (S) following treatment.

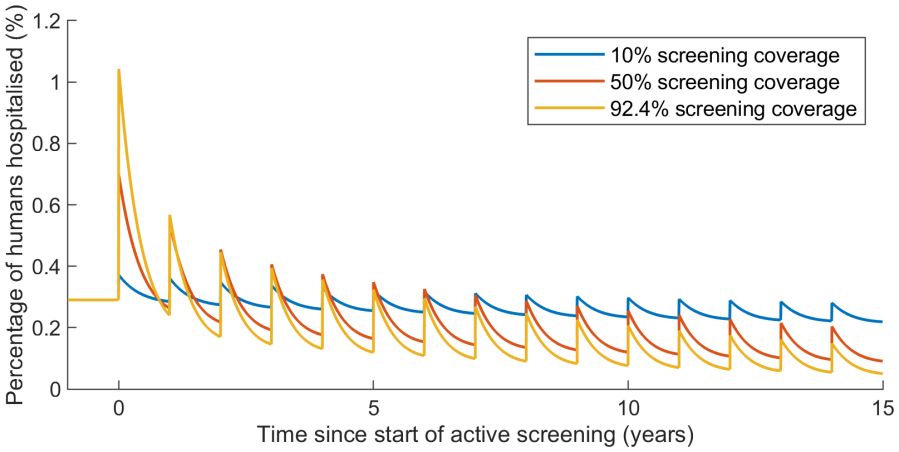


Fig 5.1: Dynamics of gHAT infection in the recovered class using the full deterministic model for 15 years with active screening.

This means the model structure is given by the following four ODES, two for each risk class:

$$\frac{\mathrm{d}S_{Hi}}{\mathrm{d}t} = (\mu_H + \psi_H)\, I_{Hi} - \alpha m_{\text{eff}} f_i \frac{S_{Hi}}{N_{Hi}} I_V^Q, \tag{5.12}$$

$$\frac{\mathrm{d}I_{Hi}}{\mathrm{d}t} = \alpha m_{\text{eff}} f_i \frac{S_{Hi}}{N_{Hi}} I_V^Q - (\mu_H + \psi_H)\, I_{Hi}, \tag{5.13}$$

for $i = 1, 2$.

We note that $\mu_H I_{Hi} = \mu_H N_{Hi} - \mu_H S_{Hi}$, since $N_{Hi} = S_{Hi} + I_{Hi}$. These terms ensure the population size remains constant in time. The constant population size also means that the system is now two dimensional, and we only need to consider the number of infected individuals in each risk class.

To understand the effects of changing the model structure, while retaining the same parameterisation of Chapter 3, we plot the infection dynamics of the simple ODE model with different levels of active screening (Figure 5.2).

All of the infection dynamics are qualitatively the same, with minor differences in the exact proportions for the different models (Figure 5.2). Removal of the explicit vector dynamics means that the vector dynamics happens instantaneously, dependent on the human population, rather than with the delay of infection decline due to a reduced pool of human infection due to active screening. Hence, with active screening, the infection

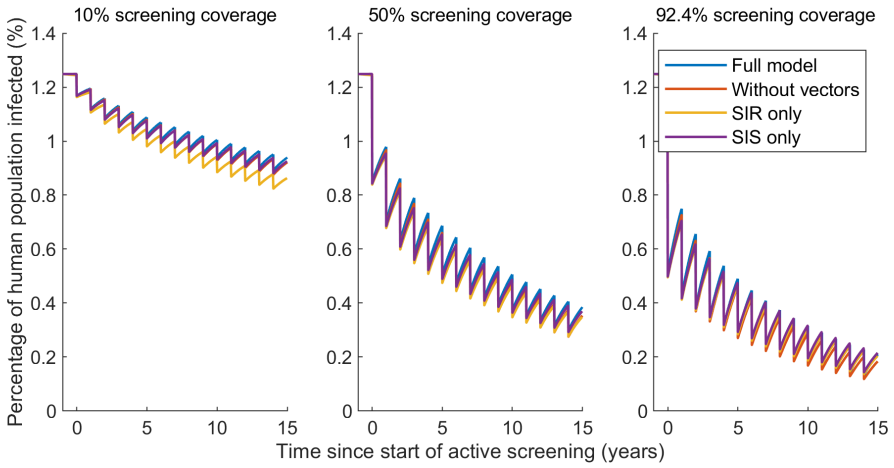


Fig 5.2: Number of infected individuals for 15 years with active screening using different model structures. We consider 10% (low), 50% (medium) and 92.4% (maximum, full low-risk population).

decays faster than it otherwise would. However, because of the short lifespan of the tsetse, meaning that the vector dynamics are much faster than the human infection dynamics, the change is very minimal (<0.05%). This means removing the vector dynamics is still a very good approximation.

The amalgamation of the exposed class and the two infected classes reduces the number of compartments in our model to a simple SIRS model. This means that instead of assuming separate exponentially distributed time periods in each class, we assume one exponentially distributed infected period with a mean time in the class equal to the sum of the mean times in the original classes. This is achieved by the definition of our parameter $\psi_H = 1/(1/\sigma_H + 1/\phi_H + 1/\gamma_H)$. We expect that by removing the latent period, the infections will increase slightly [250], however this effect is counteracted by scaling the infected compartment by the expected proportion of those that are infectious (not exposed), $(1/\phi_H + 1/\gamma_H)/(1/\sigma_H + 1/\phi_H + 1/\gamma_H)$. Indeed, there are fewer infected humans in time with this model, particularly when the active screening coverage is small and so these effects are at a maximum. When the screening coverage is high, this greatly disrupts the dynamics and so relatively dampens the effect of the model structure (Figure 5.2).

Our final reduction of our model to the simplest form removes the recovered (or hospitalised) class and so reduces the time for potential reinfection in the model. This will have the effect of increasing the number of infected humans, as more will be

susceptible to be infected (as opposed to hospitalised and hence immune). Since the number of recovered humans is small compared to the susceptible individuals (Figure 5.1), the effect is however minimal and furthermore counteracts the other effects of other model reductions. This means that we observe the SIS dynamics to be very similar to the dynamics of the full model (Figure 5.2).

This evidence means that the SIS model is a good approximation to the full dynamics and greatly decreases the complexity of the model to allow for a master equation approach to the analysis, without an infeasible number of equations. However, while this parameterisation gives a very good approximation to our full model, we note that future work would be to actually obtain a new parameterisation for this reduced model by fitting it to the underlying case data used to fit the full model. For the rest of this chapter, we will however only consider this SIS model with the current parameterisation.

5.2.1 The master equation

From the ODE equations for the SIS model (Equations 5.12 and 5.13), we can construct the master equation by considering the possible events that occur. There are four possible events that can change the probability of the low-risk population being in that state with S_{H1} susceptible people and I_{H1} infected people:

- an infected human can recover or die at rate $(\mu_H + \psi_H)I_{H1}$ to decrease the infected population from I_{H1} to $I_{H1} - 1$,
- an infected human can recover or die at rate $(\mu_H + \psi_H)(I_{H1} + 1)$ to decrease the infected population from $I_{H1} + 1$ to I_{H1},
- a susceptible human can get infected at rate $\alpha m_{\text{eff}} f_{H1} \frac{S_{H1}}{N_H} I_V^Q$ to increase the infected population from I_{H1} to $I_{H1} + 1$,
- a susceptible human can get infected at rate $\alpha m_{\text{eff}} f_{H1} \frac{(S_{H1}+1)}{N_H} I_V^Q$ to increase the infected population from $I_{H1} - 1$ to I_{H1}.

The same four events can also occur in the high-risk population, meaning a total of eight possible events exist.

Since $S_{Hi} + I_{Hi} = N_{Hi}$, for $i = 1, 2$, with N_{Hi} fixed, we can also write $S_{Hi} = N_{Hi} - I_{Hi}$ and only consider the equations in terms of I_{Hi} henceforth. Thus, we introduce the notation $P_{I_{H1},I_{H2}}$ for the probability of being in the state with I_{H1} low-risk infected people and I_{H2} high-risk infected people, which totally describes the state of the system. We also write $I_V^Q(I_{H1}, I_{H2})$ as a function of the number of infected people in each risk class. By using this notation in conjunction with the eight possible events, we derive the master equation for our system as:

$$\begin{aligned}
\frac{\mathrm{d}P_{I_{H1},I_{H2}}}{\mathrm{d}t} = & -P_{I_{H1},I_{H2}} \alpha m_{\text{eff}} I_V^Q(I_{H1}, I_{H2}) \left(\frac{f_{H1}\,(N_{H1} - I_{H1})}{N_{H1}} + \frac{f_{H2}\,(N_{H2} - I_{H2})}{N_{H2}} \right) \\
& - P_{I_{H1},I_{H2}} \left(\mu_H + \psi_H\right) \left(I_{H1} + I_{H2}\right) \\
& + P_{I_{H1}-1,I_{H2}} \left(\alpha m_{\text{eff}} I_V^Q(I_{H1} - 1, I_{H2}) \frac{f_{H1}\,(N_{H1} - I_{H1} + 1)}{N_{H1}} \right) \\
& + P_{I_{H1},I_{H2}-1} \left(\alpha m_{\text{eff}} I_V^Q(I_{H1}, I_{H2} - 1) \frac{f_{H2}\,(N_{H2} - I_{H2} + 1)}{N_{H2}} \right) \\
& + P_{I_{H1}+1,I_{H2}} \left(\mu_H + \psi_H\right) \left(I_{H1} + 1\right) \\
& + P_{I_{H1},I_{H2}+1} \left(\mu_H + \psi_H\right) \left(I_{H2} + 1\right), \qquad (5.14)
\end{aligned}$$

where $I_{H1} = 0, ..., N_{H1}$ and $I_{H2} = 0, ..., N_{H2}$.

5.2.2 Calculating the rate matrix

The master equations (Equation 5.14) comprise of $(N_{H1} + 1)(N_{H2} + 1)$ equations (since the low-risk infected population can take any value between 0 and N_{H1} and similarly for the high-risk, or alternatively using Equation 5.1). We simplify the notation by writing the equations in matrix form. This is possible since all the equations are linear in terms of the probabilities. Thus, by defining the probability vector:

$$\mathbf{p} = (P_{0,0}, P_{1,0}, ..., P_{N_{H1},0}, P_{0,1}, ..., P_{N_{H1},N_{H2}}), \qquad (5.15)$$

we obtain the master equations in matrix form as:

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{pQ}, \tag{5.16}$$

where **Q** is the rate matrix of all transition rates. **Q** will be a sparse matrix, with elements only on the tridiagonal (for exiting the current state and transitioning in the low-risk population) and two further diagonals (for transitioning in the high-risk population).

Since **Q**, is still a very large matrix for large populations and using this matrix may be infeasible when computation time and memory is considered, we further simplify Equation 5.16. Since the prevalence of gHAT is typically very low, even in the most highly endemic areas [102], and we know that at steady state in our model the infected population is only $\sim 1.25\%$, the probability of being in highly infected states such as $P_{N_{H1},N_{H2}}$ will be very small. Therefore, we can impose maxima M_{H1} and M_{H2} on the values that I_{H1} and I_{H2} can take. If we choose the maxima large enough, there will be a negligible impact on the dynamics. When considering the dynamics with a high number of cases, we typically take the approach that if $I_{Hi} = M_{Hi}$, the probability of further infection is zero. However, if we are interested in how often our approximation fails, we can additionally add a fail class, P_F, whereby if $I_{Hi} = M_{Hi}$ and an infection occurs, the state moves to P_F and does not return. This means there are now $(M_{H1}+1)(M_{H2}+1)$ equations, or $(M_{H1}+1)(M_{H2}+1)+1$ if the fails are explicitly tracked.

5.2.3 Solving the master equations

Once we have formulated the master equations in the form of Equation 5.16, this differential equation is easy to solve, as

$$\mathbf{p}(t) = \mathbf{p}(0)\exp(\mathbf{Q}t). \tag{5.17}$$

This provides a full analytic form for the solution of the master equations, which is easy to compute, by just taking a matrix exponential [299]. Alternatively, we can also solve the system by taking an eigenvalue approach [291]. If we define the eigenvalues of **Q**

as $(\lambda_1, ..., \lambda_{(M_{H1}+1)(M_{H2}+1)})$ and the left eigenvectors as $(\mathbf{l}_1, ..., \mathbf{l}_{(M_{H1}+1)(M_{H2}+1)})$, we can specify the dynamics by the equation

$$\mathbf{p}(t) = \sum_{n=1}^{(M_{H1}+1)(M_{H2}+1)} q_n \exp(\lambda_n t)\mathbf{l}_n, \tag{5.18}$$

where $q_n = \mathbf{r}_n \cdot \mathbf{p}(0)$, with $\mathbf{r}_n$ the normalised right eigenvectors. This formulation of the solution does not rely on computing exponential of large matrices, which can take a long time and we can also exploit some useful properties due to the system we are studying. We will order the eigenvalues such that $0 = \lambda_1 > Re(\lambda_2) > ... > Re(\lambda_{(M_{H1}+1)(M_{H2}+1)})$.

5.2.4 Finding the endemic equilibrium

Without the inclusion of infectious imports in the model, there are two long term steady-state solutions of our equations. There is probability 1 of being in the state with no infected individuals or of failing, since these states are absorbing. This can be seen by considering the eigenvectors, or taking the matrix exponential at large times. However, by considering the steady-state of the deterministic model, we can construct a state vector using the product of two binomial distributions, $\mathrm{B}(M_{Hi}, I_{Hi_0})$ for $i = 1, 2$, where I_{Hi_0} is the expected proportion of infected in risk class i. Considering a population of size 1,000, this distribution is centered around approximately eight low-risk infected people and four high-risk infected people (Figure 5.3). By using the rate matrix $\mathbf{Q}$, we can also show how the dynamics are likely to change from a particular initial condition. The probability distribution is very tight with less than 0.001 probability of $I_{H1} > 20$ or $I_{H2} > 13$. Alternatively, and more accurately, we could calculate the limiting conditional distribution of the model to obtain the endemic quasi-equilibrium.

5.2.5 Adding active screening

With the infection dynamics described and passive detection assumed to be included in the rate of returning from the infected class to the susceptible class, we model active screening by the removal of a subset to individuals in the low-risk infected class into the

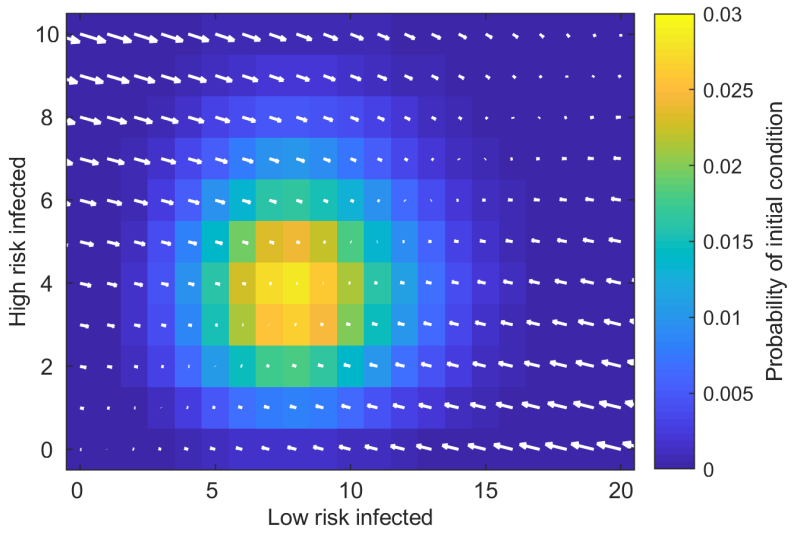


Fig 5.3: Distribution of initial conditions starting from endemic equilibrium with a population $N_H = 1,000$ and no infectious imports. We show percentiles of highest cumulative probabilities with a quiver plot overlaid to show the dynamics.

corresponding susceptible class. In our master equation framework, we do not consider this action as part of the rate matrix $\mathbf{Q}$, but as a separate action we can simulate when considering the time evolution of our master equations. However, this is not as simple as the ODE or tau-leaping framework, as we still need to consider the action of active screening in the probabilistic sense. (Alternatively, we could represent active screening by a lower-triangular transition matrix A and embed this as a discrete time Markov chain.)

We do this by considering the hypergeometric distribution $\text{HG}(N_{H1}, c, k)$ to get the number of real infected individuals chosen for screening. We consider the number of infected people detected c from the population N_{H1} of low-risk people available to be screened, when a total of k people are screened. Subsequently, if c infected people are screened, we use a binomial distribution $\text{B}(c, \text{Se})$, where the test sensitivity is $\text{Se} = 91\%$ [252] to get the number of infected people screened that are identified. We assume specificity is 100%. Taking the product of these probabilities and applying them to every possible state in terms of number of infected people in the low-risk class, we formulate a probability vector for the probability of detecting each different number of people, given the number screened and the number infected. Hence, given the probability of being in each state at time t, we can apply these probability vectors to determine a new probability of being in each state after an active screening of given coverage.

Using this method and calculating the expected number of infected people with time, we see that active screening is working correctly in our simulations (Figure 5.4). Larger population sizes have more infected people and a lower probability of extinction of infection, as expected, with active screening reducing the number of infected with time.

5.2.6 Choosing the maximum number of infected individuals

Since the number of ODEs in the master equations can be very large for villages with large population sizes and so induce memory problems, as mentioned in Section 5.2.2, we can limit the state space to only realistic numbers of infected people. However, this needs to be done such that the probability of being in the state space excluded by this artificial cut off is negligible.

We first consider the probability of failure (the probability of being in the excluded state space) after 15 years for different population sizes given fixed thresholds for the maximum number infected, where M_L is the maximum number of low-risk and M_H is the maximum number of high-risk. We exclude active screening here, since this will reduce the infection level and we want to consider the worst case scenario.

Using fixed thresholds, it is immediate that increasing the maximum threshold for the number of infected people, decreases the probability of the state space being too small (Figure 5.5). Using typical village population sizes of less than 2,000 people, using a maximum size of 200 in each risk group achieves effectively no error, with a maximum of 150 achieving less than 10^{-10} probability of failure. For larger population sizes, such as $N_H = 100,000$, these approximations are certain to fail however, as the expected number of infected people is larger than the imposed maxima. For a maximum of 300, reasonable results are obtained up to approximately $N_H = 10,000$, however, we may want to consider larger populations.

It is evident that the maximum should be chosen as a function of population size. Thus, we start using $M_L = M_H = 50$, and increase either maxima in increments of 50 if either the expected number of infected individuals is larger than the maximum, or if

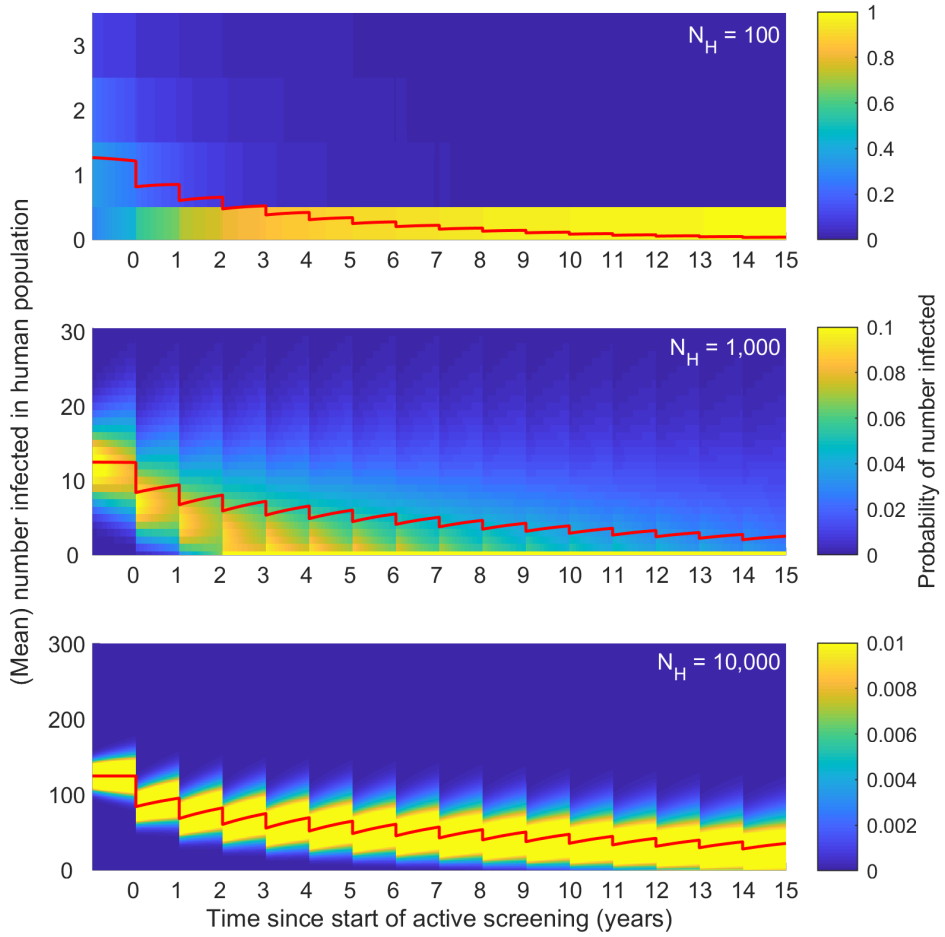


Fig 5.4: Probability of being in any given infection state in time for 15 years with an active screening coverage of 50%. Marked in red is the expected number of infected people. Populations size of $N_H = 100$, $1,000$ and $10,000$ are used. For larger populations, there are a greater number of possible states and so the scale on the colour bar is adjusted for the smaller probabilities of being in individual states. We assume no imports of infection into these villages.

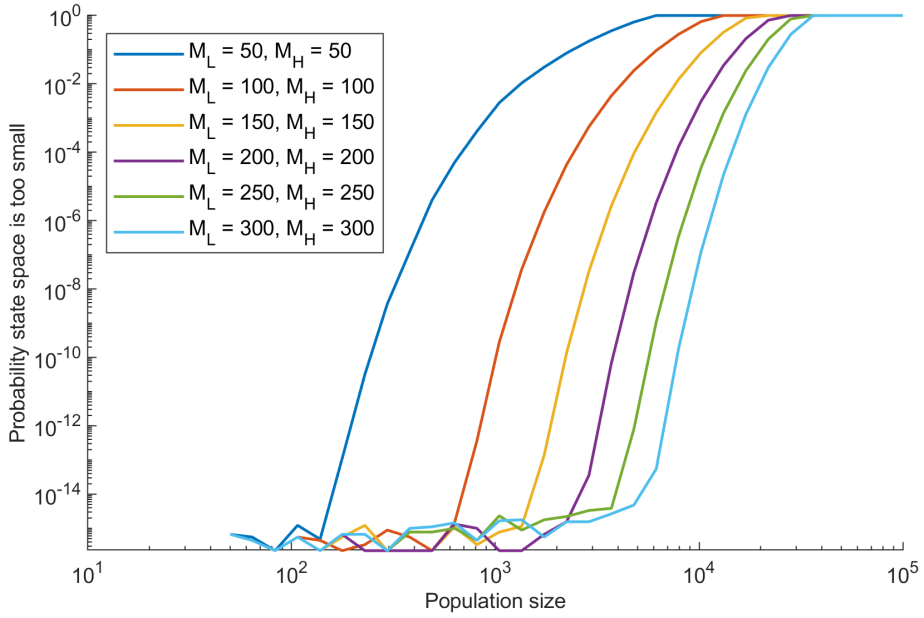


Fig 5.5: The probability that the state space for the number of infected individuals is not sufficiently large for a given imposed maximum. Probabilities of approximately 10^{-16} are due to using floating point numbers, which are only accurate to 2^{-52} ($\sim 2.22 \times 10^{-16}$), and so these values can be assumed approximately zero.

the probability of failure is determined to be greater than 10^{-8} when considering the dynamics after 15 years with no active screening (Figure 5.6).

By increasing the maximum threshold number of infected people if the population size means the error is too large, it is possible to model larger populations with a reasonable state space size and negligible probability that the state space is too small. This ensures the full probability distribution is captured (Figure 5.6). Even for very large villages of size $N_H \sim 20,000$, using $M_L = 550$ and $M_H = 300$ captures all the dynamics (with a probability less than 10^{-8} of the maximum infected exceeding either limit). The values of maxima shown in Figure 5.6 do not always fall on 50 unit increments, since, for the high-risk individuals, this can exceed the number in the population, which is a small minority of the total population.

Using this knowledge, for the rest of this thesis, we can use the correct limits to ensure this condition and exclude the probability of failure from calculations as it will always be small. We however will always keep $M_L > 150$ and $M_H > 150$, since there is little issue with memory or computation speed for matrices of this size.

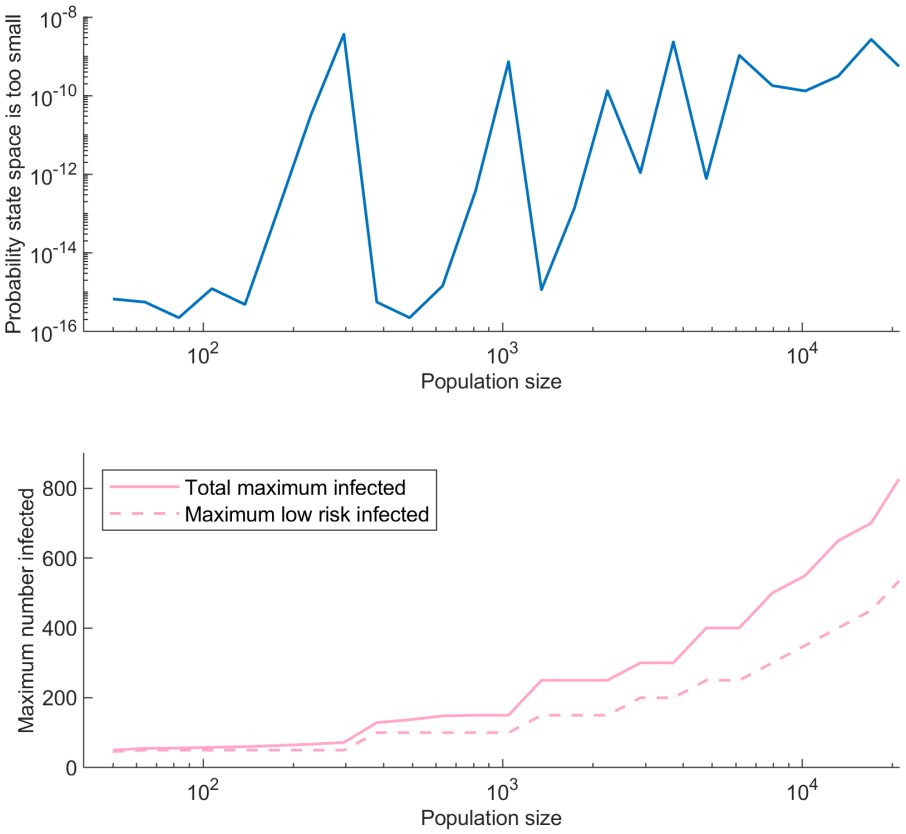


Fig 5.6: The probability that the state space is too small when increasing the maximum number low- and high-risk infected people in intervals of 50, to ensure there is negligible probability that the state space is too small (less than 10^{-8}). The probability and resulting maxima in state space are shown in the two panels.

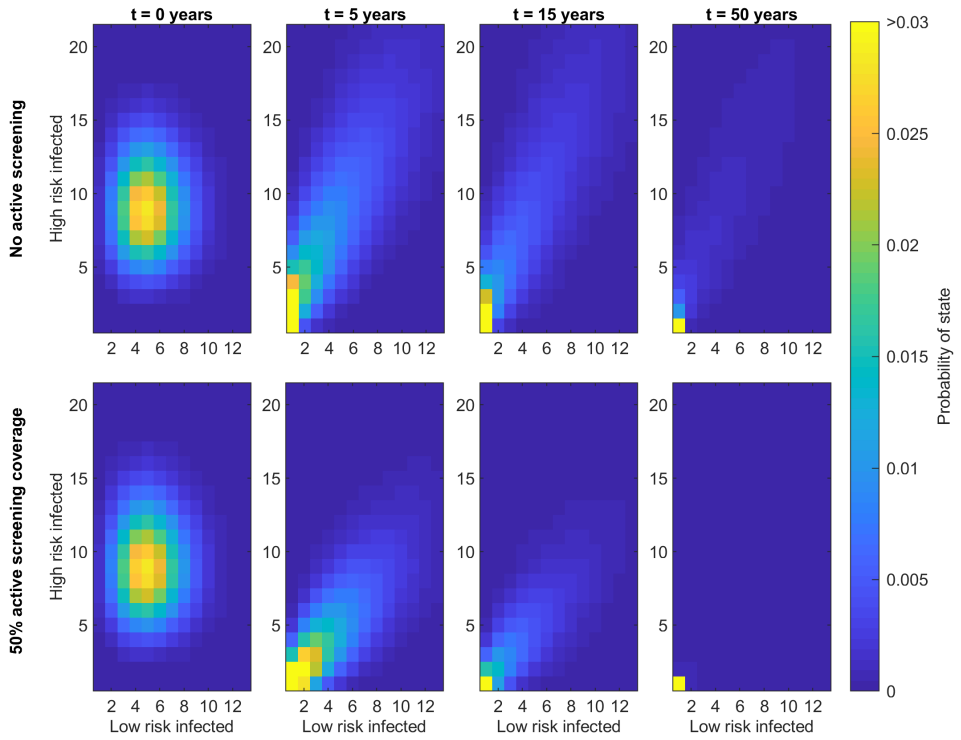


Fig 5.7: Time evolution of the infection distribution in the absence of imports from the deterministic endemic equilibrium both with and without active screening.

5.3 Considering a single village

Similarly to Chapter 3, we can consider the dynamics of gHAT using our master equation approach on a single village, isolated from other populations. Without importations of infection, the first row of the rate matrix will consist entirely of zeros, meaning the largest eigenvalue will be zero, with the corresponding eigenvector being a one in the first entry, indicating a steady state of disease extinction.

5.3.1 The distribution of infection

The dynamics of the infection with time will always move towards a disease-free population, since this state is absorbing. Beginning at the endemic equilibrium found in Section 5.2.4, and considering the time evolution of the system for an example population of size 1,000, we see that infection levels fall, particularly in the low-risk population, although there is some probability that the number of infections would increase in

individual simulations (Figure 5.7). The addition of active screening accelerates the process of disease extinction within the population, such that at large times, almost all the probability is given by no infection, whereas without the screening, there is some probability the disease will persist.

5.3.2 Elimination of gHAT in the absence of importations

Probability of persistence

In Chapter 3, we considered the probability of local gHAT persistence and how this was linked to the population size of the village. We found that persistence increases with increasing population size, since transmission chains are more likely to be broken when there are fewer people to infect and there is greater impact of the stochasticity (Figure 3.13). While we are using a different model structure here and the method of simulation is also different, we confirm that we observe very similar results to this original model when using the master equation approach (Figure 5.8). In this figure, the dashed lines correspond to the results found in Figure 3.13, with the new results plotted as solid lines. Hence, we are confident of achieving qualitatively meaningful results in this chapter, even if the model needs to be fully parameterised.

We also note that in this model, when comparing the results for different population sizes, the effect of rounding on the fixed number of people in each risk class is large, due to the small population size of villages. For example, increasing the population size from 19 people to 20 people will mean the rounded number of high risk people increases from 1 to 2, greatly changing the dynamics. Hence, in Figure 5.8, we consider the number of high-risk people to be both integer values either side of the proportion $R_2 \times N_H$ and take a weighted average of the resulting two values for the probability of persistence to achieve a smooth curve for this probability plotted against population size. We do not need to take account of this effect when considering single large populations.

Similarly, considering the dynamics in time and comparing to Figure 3.14, the thinner darker lines in Figure 5.9 show the local gHAT persistence for the master equation model.

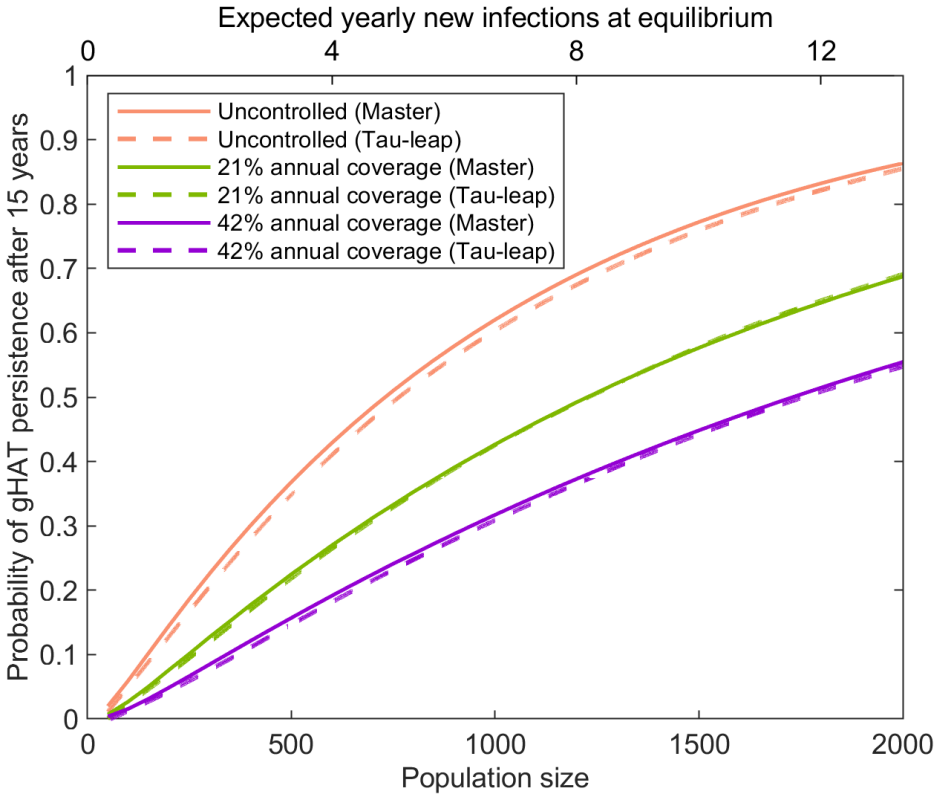


Fig 5.8: Probability of gHAT persistence in isolated settlements for the two different models. As in Figure 3.13, the tau-leaping algorithm was used with 100,000 realisation, such that the confidence interval for the mean persistence was too small to be plotted in this figure. In both models annual screening at a fixed coverage per year yields a drop in persistence with increased coverage. Uncontrolled means passive surveillance only.

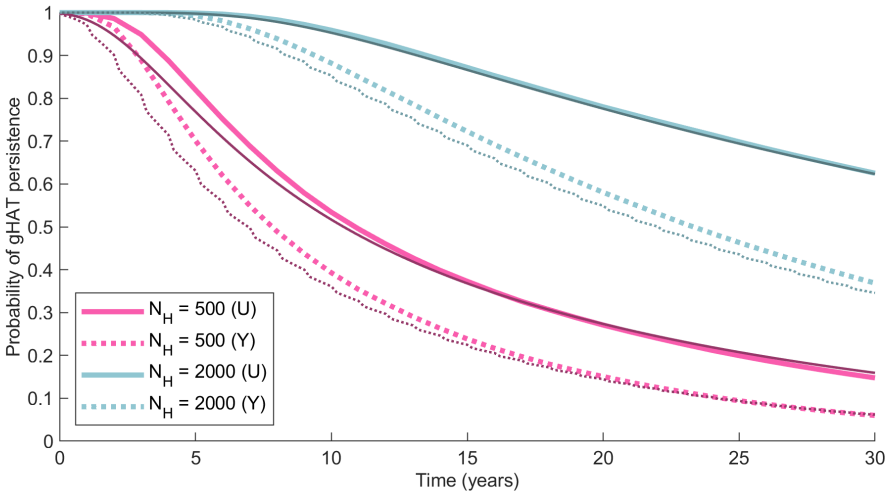


Fig 5.9: The probability of gHAT persistence in time for the uncontrolled setting (U), and annual screening at the average coverage of 21% (Y) for villages with populations of sizes 500 and 2,000. The thick lighter colours show the tau-leap results, and the thin dark colours show the master equation results.

For large times, the results are very similar; there is some discrepancy for smaller times, likely due to the potential for a faster recovery, compared to the full model with multiple infection classes [250]. However, qualitatively the same dynamics are observed, in both different population sizes and active screening statuses (Figure 5.9). Note that as in Figure 3.14 we plotted the time series from one data point per year, the active screening is not visible, which is retained in Figure 5.9. Whereas for the master equation approach we plot additional time points, such that the active screening events are visible.

Expected number of infecteds

In addition to the probability of local gHAT persistence, the expected number of infected people is another quantity easy to calculate from this model. It is simply obtained by taking the dot product of the probability vector and the number of infected people in each state.

The expected number of infections increases linearly for large population sizes, with smaller populations having fewer expected infections due to the effects of an increased number of disease extinction events (Figure 5.10A). Larger active screening coverage, as expected, also reduces the expected number of infections. Without infectious imports to

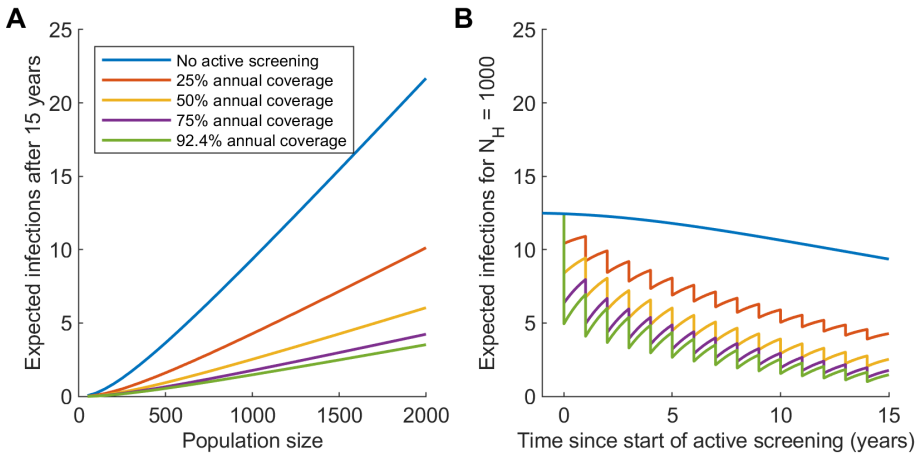


Fig 5.10: The expected number of gHAT infections for different annual screening coverages for (A) population sizes, and (B) time points, using the master equations.

balance the population of infected individuals, the expected number will also decrease in time, even without active screening, due to local gHAT elimination, with further reductions, when active screening is also added (Figure 5.10B).

Since the increase in the expected number of infections is almost linear in population size (Figure 5.10A), there is a similar relationship between the expected number of infections and the probability of gHAT persistence, as there is between the population size and the probability of gHAT persistence. Larger populations, with lower active screening coverage and small times have the highest persistence, as is intuitive. However, if the expected number of infections is known the biggest factor in determining the probability of gHAT persistence is the time since the simulations began, irrespective of the active screening coverage or population size (Figure 5.11).

Time to extinction

Since the disease-free state is absorbing, an additional quantity of interest is the expected time to reach this state. This time to extinction can be calculated very simply using the master equation approach. By taking the rate matrix $\mathbf{Q}$ and removing the rows and columns corresponding to the absorbing state, we obtain the matrix $\mathbf{Q}_0$. The expected time to extinction from each state $\mathbf{T}$ is now given by the solution to the equation $\mathbf{Q}_0\mathbf{T} = -\mathbf{1}$ [300].

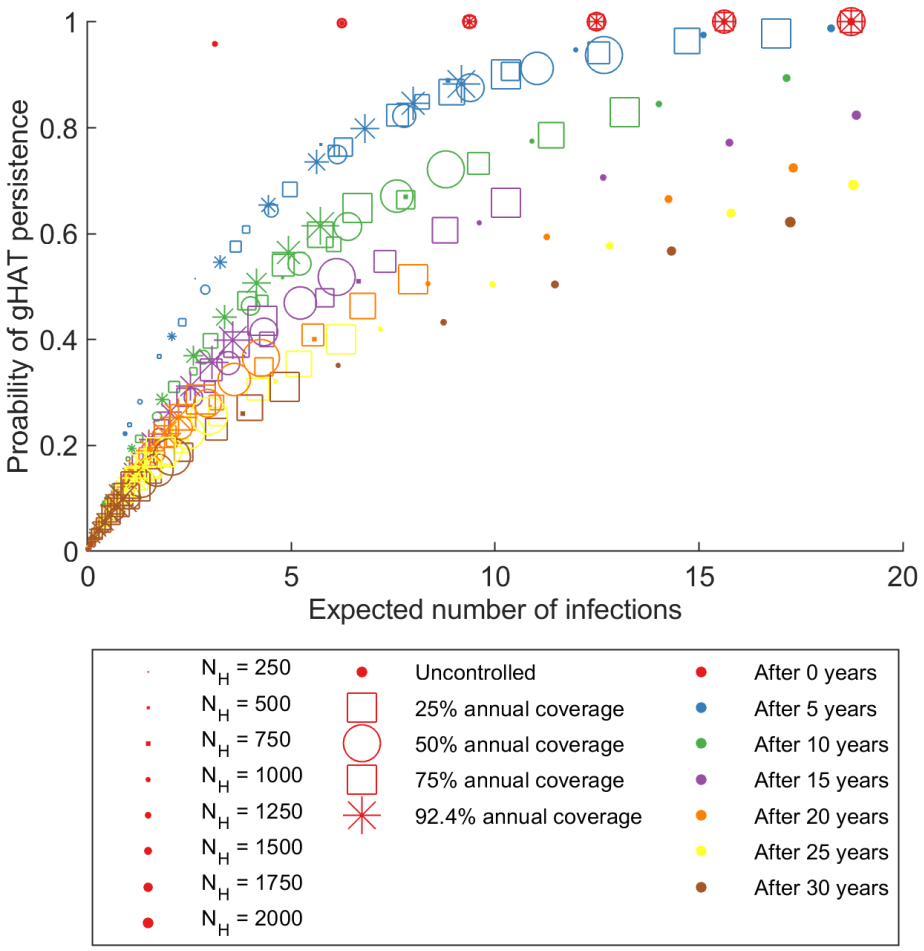


Fig 5.11: Scatter plot of the probability of gHAT persistence and the expected number of gHAT infections for different population sizes, annual screening coverages, and time points.

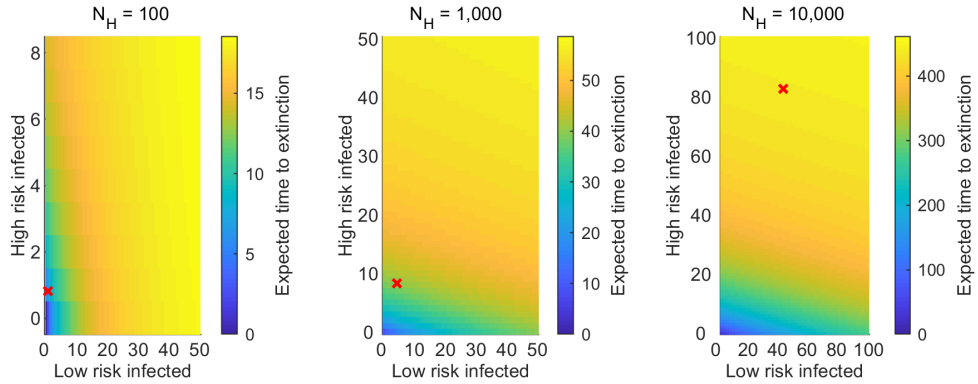


Fig 5.12: The expected time to extinction for different population sizes, given the initial number of high- and low-risk infected people and no active screening. The red cross marks the expected number of high- and low-risk individuals for the population size, given the endemic equilibrium.

The expected time to extinction is small if initial states are close to the absorbing state, however, the time is very long for large populations with a large number of infections in the absence of interventions (Figure 5.12). Corresponding with the previous results on persistence, large populations have longer expected times to extinction. When a population size of $N_H = 10,000$ is considered, and is started from endemic equilibrium, the expected time to extinction is 400 years.

When considering active screening, the same method cannot be used with this formulation to calculate the time to extinction since active screening is not part of the rate matrix formulation. We would need to implement an additional discrete time Markov chain. However, the result can be obtained by calculating the state space up to large time intervals. By considering 100 years, we can obtain the probability that the infection went extinct within each of these years, and the probability the infection remains, to get the full probability distribution of the time to extinction for different annual active screening coverages, starting from endemic equilibrium. Considering $N_H = 1,000$, there is a small probability of not achieving local gHAT elimination within 100 years without active screening, but with increased active screening coverages, this becomes negligible, with a high probability of local elimination, even within ten years (Figure 5.13).

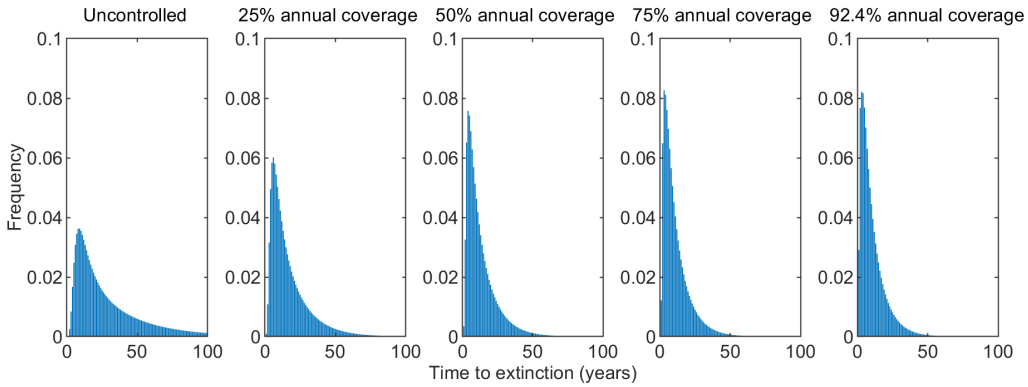


Fig 5.13: Distribution of the expected time to extinction for a population of N_H =1,000, for different levels of annual active screening coverage.

The extinction rate

The rate of extinction is the rate that the population moves from one high-risk infected and no low-risk infected to none infected at all, or one low-risk infected and no high-risk infected to none infected at all. In our model, this will occur at rate $(\mu_H+\psi_H)(P_{0,1}+P_{1,0})$. The extinction rate increases as the expected infection decreases and so there is a higher probability the populations are approaching extinction. This rate will then plateau and decay as more populations actually go extinct (Figure 5.14A). Similar behaviour is observed for the probability that an active screening causes an extinction event. However, for very high coverages, all the infected low-risk population will be detected and treated (except false negative test results), and so future active screening will have a lower probability of causing gHAT extinction, since all infection is in the high-risk population that are not screened (Figure 5.14B).

5.3.3 The effects of active screening

So far we have always assumed that all active screening occurs at the very start of each year, however, this will not actually be the case and active screening will occur throughout [115]. For a single village, however, it is only likely to occur once, rather than multiple times in a year. We show that there is minimal effect on whether the active screening occurs at the start or end of the year, other than a year delay in starting the screening (Figure 5.15). However, if the same amount of active screening is done

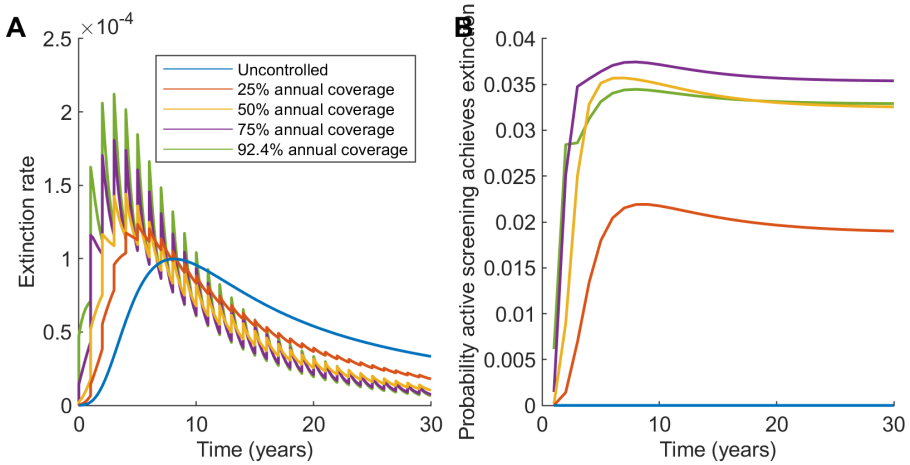


Fig 5.14: (A) Extinction rate of gHAT in time for a population of N_H =1,000. (B) The probability that an active screening in each year causes a gHAT extinction event.

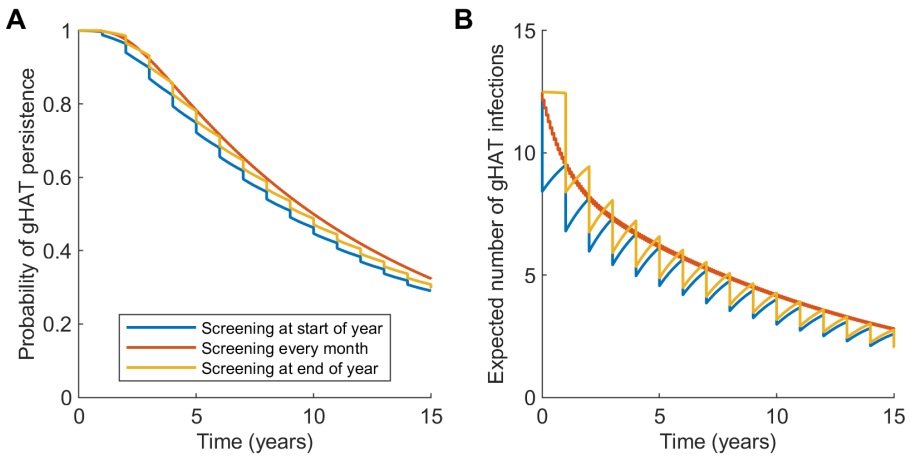


Fig 5.15: Timing of active screening. Active screening at the start of each year, throughout or at the end of the year affects (A) the probability of gHAT persistence, and (B) the expected number of infections.

in the village, but throughout the year, there will be more infection in the populations with more frequent screening with a higher probability of persistence. This is dual-fold; the largest factor being that since we consider a single annual screening and screening throughout the year to have the same number of people screened, but at different time points, there is the opportunity for individuals to be screened multiple times, effectively reducing the screening coverage. In addition, there is a smaller effect of the village missing out on the benefit of mass screening all at once, and so removing infected people from transmitting the infection throughout the year. A similar effect to this was seen in Figure 3.13B, where a random screening coverage level, which could be large, was more effective in reducing the infection level than regular screening at a lower coverage. We consider these effects in more detail in Section 5.5.3.

5.4 Importation of infection into a village

While considering villages as separate populations provides information on the duration of persistence (or time to extinction) of the infection, we know that the populations are not actually isolated and the movement of people can both sustain the infection in villages and re-introduce the disease where it is previously been locally eliminated (see Section 3.5). We have previously estimated a value for the rate of importation into a village population of 3.4×10^{-6} days^{-1} due some background level of infection (Figure 3.18), and so can introduce this into our SIS model using the master equations.

5.4.1 Considering invasion

The introduction of a single infected person into a village with no infection may have no effect on the population if this person is treated (or dies) before infecting other people. However, as shown in Figure 3.15, there is a very significant probability that further transmission will occur after an infectious import.

Using our master equation approach, with the initial condition of full probability of starting with one infected individual (in the high-risk class), the probability of the village returning to elimination grows quickly with time, although there is a significant probability of further transmission causing more infections (Figure 5.16). The expected number of infections increases above one in the first ten years, since there is a small probability that a single infectious import can cause a very large number of new infections, despite the overwhelming probability that there is a return to local elimination.

With only one initial infection and no subsequent invasion, the time to extinction is not greatly impacted by the population size of the village, with very similar results for populations varying between $N_H = 100$ and $10,000$ (Figure 5.17). The probability of extinction within one year is just less than 0.3, with the infection otherwise persisting. This probability of persistence on this reintroduction means we now consider the effect of regular imports into a population.

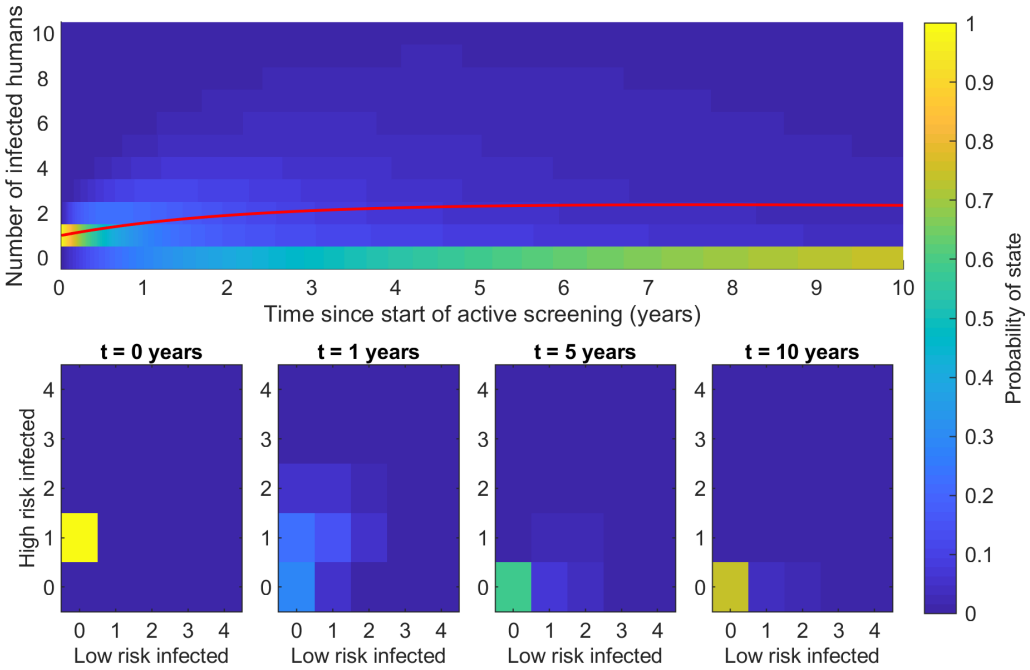


Fig 5.16: The distribution of states of a village upon the reintroduction of a single infected high-risk individual and no subsequent invasion. The red line in the top panel shows the expected number of infected people. The distribution between the risk classes is shown in panels below at times $t = 0, 1, 5$, and 10 years. A population size of $N_H = 1,000$ is used.

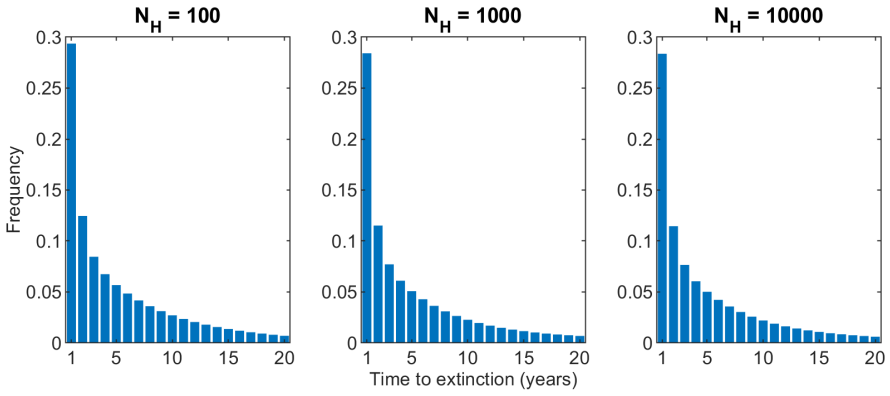


Fig 5.17: The distribution of time to extinction of gHAT infection without any active screening for populations of different sizes from the initial conditions of a single infected person from the high-risk class.

5.4.2 Explicit inclusion of infectious imports

For the explicit and continuous inclusion of imports, we can expand our rate matrix by the addition of a new component that describes the infectious importation events into the village **E**. As such, our full rate matrix will now be **Q**+**E**, where **Q** describes the disease events of infection and recovery and **E** describes the movement events of importation of infection. We assume the import can occur by a susceptible person travelling to another village, being infected there, and returning to their home village. To calculate the matrix **E**, we consider the two possible changes of state for each risk class:

- a susceptible human can become infected by movement and introduce the infection into the village at rate δS_{H1} to increase the infected population from I_{H1} to $I_{H1}+1$,

- a susceptible human can become infected by movement and introduce the infection into the village at rate $\delta(S_{H1}+1)$ to increase the infected population from $I_{H1}-1$ to I_{H1}.

By introducing importations of infection into the equations, while still considering just a single village, long term local elimination is no longer possible, since there is now always a chance of recrudescence due to a person visiting another village with the infection and bringing it back to their home village. Thus, starting with the endemic equilibrium distribution, including the imports, and simulating forwards in time with annual active screening with 50% coverage, the level of infection decays, however it will reach an equilibrium due to the number of imports entering the village (Figure 5.18).

We can also measure the extinction rate as in Figure 5.14 for a village with imports (Figure 5.19). When there is no active screening from endemic equilibrium, the extinction rate remains constant, but it will increase in time if active screening is causing additional removals of infected people due to hospitalisation and treatment.

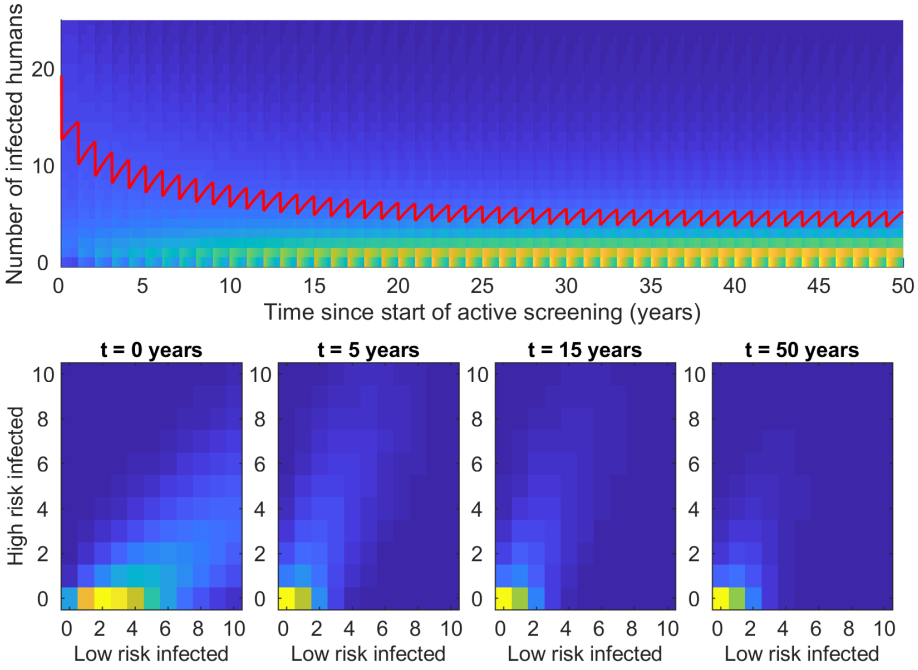


Fig 5.18: The probability of infection in a village of size $N_H = 1,000$ with infectious imports given the rate 3.4×10^{-6} per day); the red line in the top panel shows the expected number of infected people. The distribution between the risk classes is shown in panels below at times $t = 0, 5, 15$, and 50 years.

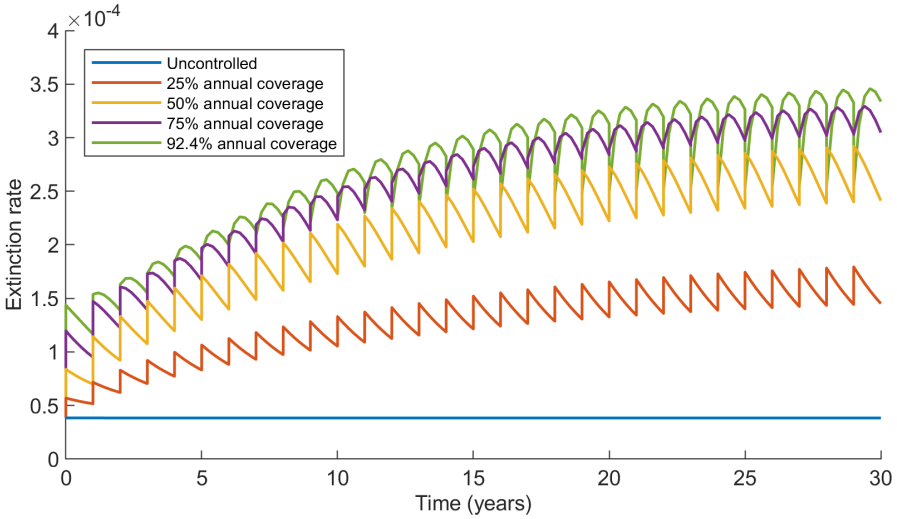


Fig 5.19: The extinction rate for a single village with a population size of $N_H = 1,000$ with infectious imports for different levels of active screening.

While the addition of infectious imports adds to the realism of the model, the assumption of a constant rate of importation depends on other villages not undergoing changes in their infection dynamics due to active screening and it is assumed they remain at endemic equilibrium. For a better understanding of the infection dynamics of gHAT around local elimination, we need to consider the interactions between villages, such that the importation rate into a village decays as the global infection rate falls.

5.5 Considering multiple villages

When considering elimination of gHAT, elimination in a whole region will depend on both the persistence in individual settlements and the interaction between these settlements. Local extinctions of infection will be balanced by external imports until global elimination is achieved.

5.5.1 The model with multiple villages

To construct a model where multiple villages can interact, we consider the villages for the most part as independent, as we have done throughout. We construct rate matrices **Q** for each population size and solve the equations, as before, where each rate matrix represents an infinite number of realisations of a single village of a given population size. For a group of villages, we can calculate the total expected number of infected people by summing the expected number from each population size, counting some population sizes multiple times, if there are multiple villages of the same size within the group. Similarly, the probability of local gHAT extinction for the group is calculated by the product of extinction probabilities of the single populations, noting again that the rate matrix will not need to be calculated multiple times for the same population size, but will be repeated in this calculation.

However, when importations of infection are considered in each village of a given population size, rather than giving the rate of importation by $\delta(N_{Hi} - I_{Hi})$, where δ is the rate of importations, we multiply this whole term by $I_{\text{Total}}(t)/I_{\text{Total}}(0)$, where I_{Total} is the sum of all expected infections across all the villages we are considering. This

gives the term $\delta(I_{\text{Total}}(t)/I_{\text{Total}}(0))(N_{Hi} - I_{Hi})$. This means that our fitted importation of infection rate in any given village will decay with time as the total number of infections in all villages falls. We are making the simplification that $I_{\text{Total}}(t)$, the total expected number of infections across all the villages, is not dependent on the current state of a particular village, but formed simply from the total expected number across all villages. We do not want to compute the probabilities of all possible combinations of village states, as this would greatly increase the size of our system. This is assumption will be a good approximation for a large number of villages, although we note in making it we are neglecting second order effects. However, with this method, we have now linked the villages so that when there is no infection I_{Total} will be zero and so there will be no importations of infection and elimination of transmission will have occurred in the region.

Since our model is fitted to data from the WHO HAT Atlas [106, 107], we explore this data to get reasonable village parameters for our model. As in Figure 3.3A, we have census estimates of the villages sizes (Figure 5.20A). We can also obtain reasonable parameters for the expected screening behaviours by considering the mean active screening coverage across all villages than have been screened and the probability that any particular village listed in the WHO HAT Atlas is screened in a given year (Figure 5.20B–C). The mean of these values from 2000–2012 are given by 62.5% and 36.8% respectively (dashed red lines). We can then incorporate this into the model by a new parameter for the probability of a screen. Thus, when a screening can occur we take the full probability distribution and take the linear combination of the probability of no screening multiplied by the current distribution plus, the probability of a screening multiplied by the distribution after an active screening of given coverage.

5.5.2 An example of villages interacting

To show how the dynamics are affected by the interactions of villages, we take two hypothetical areas with 10 villages. In the first each village has a population size of $N_H = 1,000$; in the second, each has $N_H = 10,000$.

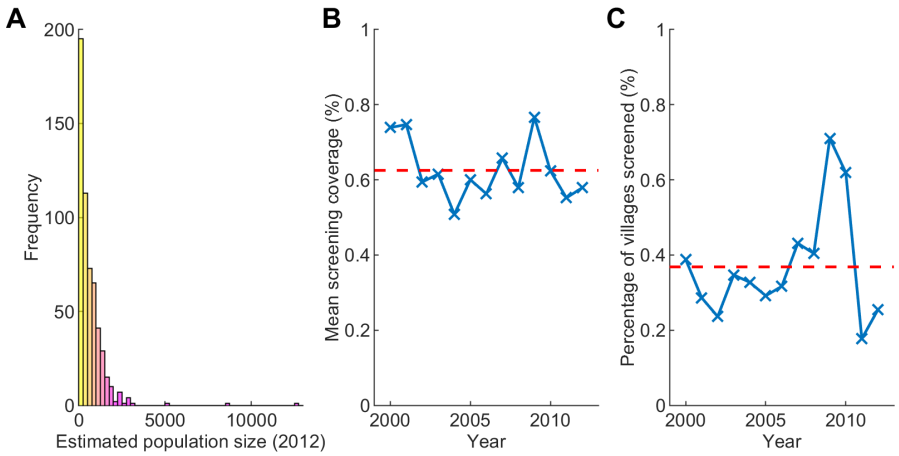


Fig 5.20: The villages of Yasa-Bonga and Mosango health zones, DRC. (A) Histogram of the estimated population sizes of the villages from 2012, based on census data from the WHO HAT Atlas [106, 107]. (B) The annual mean screening coverage for a village when it is screened in Yasa-Bonga and Mosagno (2000–2012). (C) The proportion of the villages of Yasa-Bonga and Mosagno listed in the WHO HAT Atlas that are screened each year (2000–2012).

For $N_H = 1,000$, there is a strong increase in the probability of gHAT becoming extinct within the first three years, although there infection also has the potential to grow, shown by the expected number of infections decaying much slower (Figure 5.21). For $N_H = 10,000$, because the populations are so much larger, the probability of extinction is very low (Figure 5.22). The effect of some villages being screened and others not, is very clear in the risk distribution plots, that is not apparent for the smaller population sizes due to the low numbers of infected individuals. After one active screening, the number of low-risk can decrease or remain constant, to give a bimodal distribution across the villages. This slowly merges at later time points.

In considering gHAT persistence in time, we know that larger populations will have a higher probability of persistence (see Chapter 3). It is also intuitive that a collection of populations of the same size together will have a higher level of persistence, since there are more people total. We additionally show that one population will have a higher probability of persistence than ten populations, each a tenth of the size (Figure 5.23). For example, after 50 years, the probability of persistence of one population of size $N_H = 1,000$ is approximately 0.35, while the total probability of persistence of ten populations of size $N_H = 100$ is less than 0.05. This is because each village is more likely to observe a local extinction with fewer susceptible people to infect and less mixing of the whole population than in a single village.

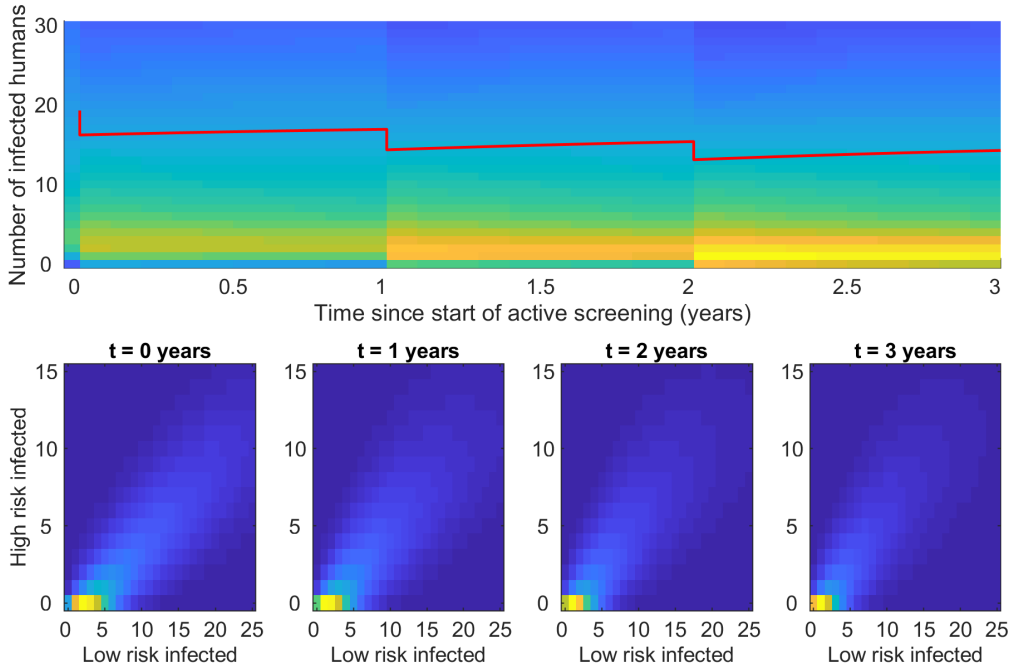


Fig 5.21: Example early infection dynamics for ten village populations of size N_H =1,000. Risk distributions are shown in the panels below for one of these villages.

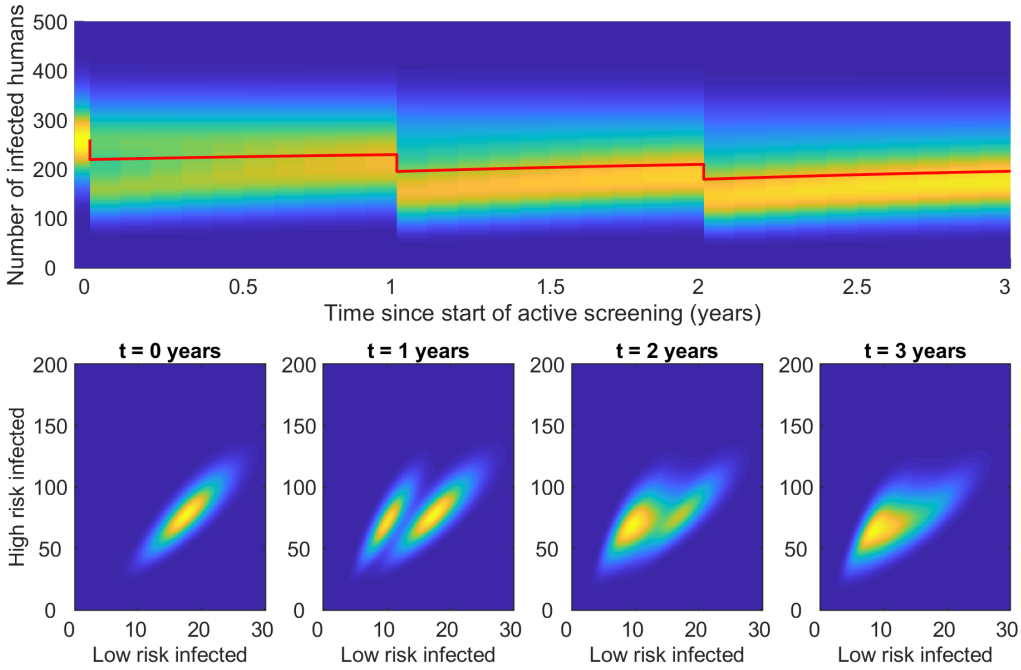


Fig 5.22: Example early infection dynamics for ten village populations of size N_H =10,000. Risk distributions are shown in the panels below for one of these villages.

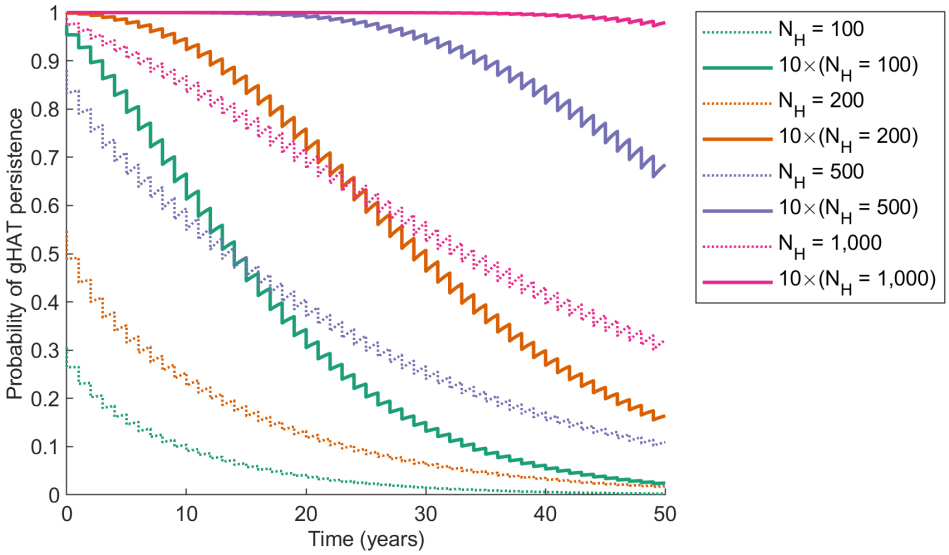


Fig 5.23: Probability of gHAT persistence in time for different populations. The paler dotted lines show the persistence for a single village within the larger cluster of ten, and the darker solid lines show the persistence across all ten villages.

This effect is pronounced when comparing back to Section 5.4, with a single population with imports. Single village populations with imports have much higher probabilities of gHAT persistence, than when multiple villages with a connecting importation rate (Figure 5.24).

5.5.3 Varying active screening procedures

As previously mentioned, in most model simulations in this thesis, we have considered that active screening occurs on the first day of the year. However, this is not truly representative of the real system, since screening occurs throughout the year. With a model with multiple villages we can test the effect of this simplification by considering a hypothetical region of 12 villages, each with a population of 100 people. If all 12 villages are screened on the first day of the year we observe an immediate large drop in the expected number of infected individuals, however slow resurgence in infections will occur throughout the year.

Alternatively, if one village was screened on each first day of the month across all years, the decrease in the expected number of infections will be more staggered. However, we show that at the end of each year, when the same total amount of active screening

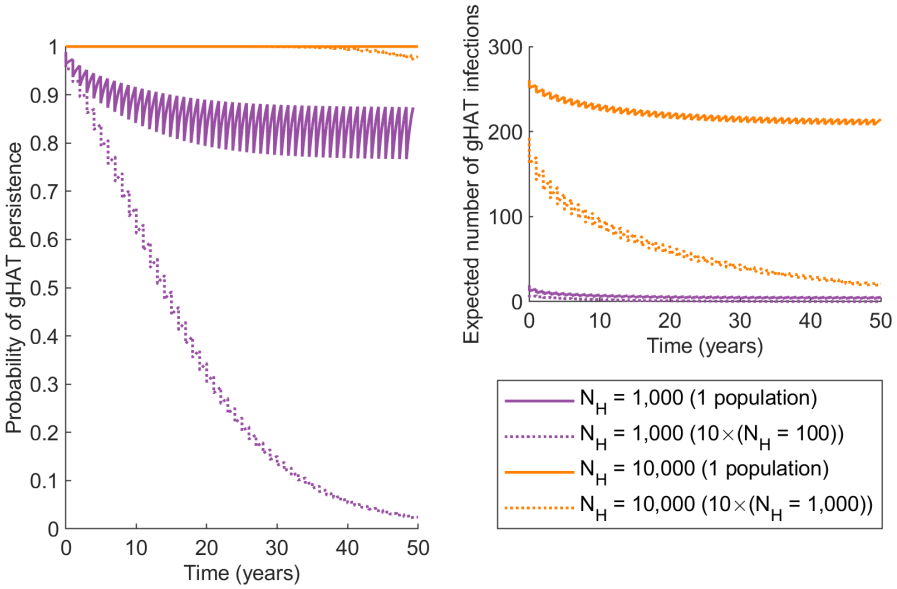


Fig 5.24: Probability of gHAT persistence in time for different model structures. Either a single village with importations of infection, or ten villages where imports are proportional to the total infection.

will have been done, there is very little difference in the expected number of infections (Figure 5.25). There are very marginally more infections when the screening is spread across the year, as there is benefit in reducing the infection numbers as much as possible at a given time to reduce transmission within the year, but the difference is so negligible, it is a reasonable approximation to assume all screening occurs on the first day of the year (or indeed any other day of the year). Therefore, in the modelling, assuming that all screening is done on the first day of the year is a useful simplification, since the available data will often not indicate exactly when active screenings occurred, simply providing a year; this lack of information will have minimal impact on predictions.

Furthermore, we can also explore other active screening strategy factors, such as the coverage, frequency and probability of villages being screened (Figure 5.26). If the probability of a village being screened is reduced, while the coverage of the proposed active screening increased, such that the expected number of people screened remains constant, there is almost no difference in the expected number of infections over time. Higher coverage reduces the infection marginally (as shown in the inset panel of Figure 5.26A). Similarly, this provides evidence of the validity of using higher-aggregation models when only the total number of people screened is known.

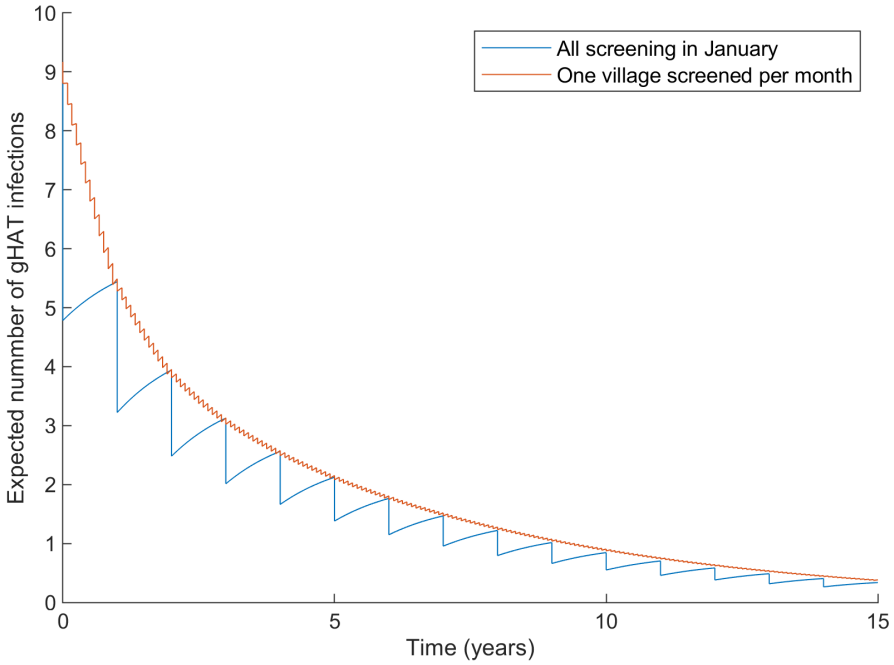


Fig 5.25: The effect of the timing of active screening interventions on the expected number of infections. Dynamics are shown for all villages screened annually at the same time and for different villages screened each month.

The same message is however not true of trading screening coverage for increased frequency of active screenings. Higher coverages are better than frequent screenings, as demonstrated in Chapter 4. In this case, more frequent screenings at lower coverages may involve screening the same people multiple times and so the effective coverage will be larger for a single large screen that ensures a larger proportion of the population have been screened (Figure 5.26B).

5.5.4 The interactions of villages in Yasa-Bonga and Mosango health zones, DRC

Finally, using all this information, we can carry out a full simulation of the infection dynamics for all the villages of the two health zones of Yasa-Bonga and Mosango in the DRC. We consider all villages using the estimated population sizes given in Figure 5.20A and also use the mean screening coverage and village active screening probability, also from this figure.

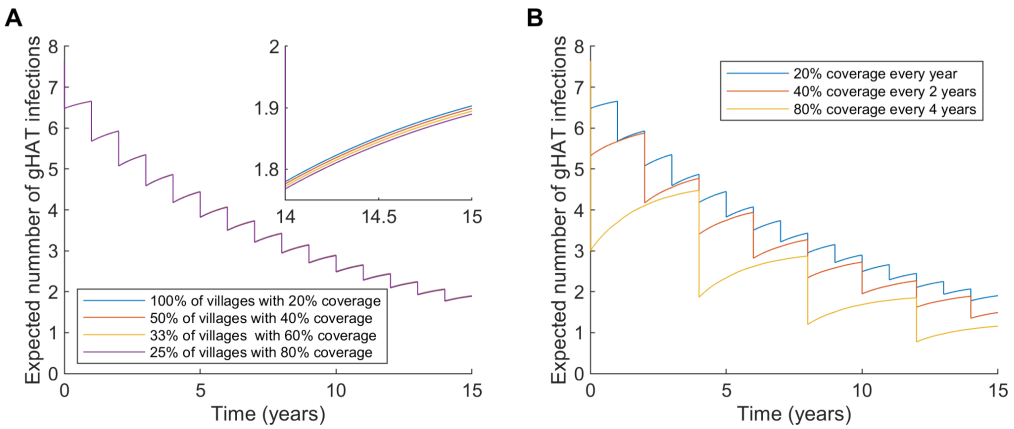


Fig 5.26: Active screening probability, coverage and frequency. (A) The effect of changing the screening coverage and the probability a village is screened. The inset panel shows a close up for the last year plotted. (B) The effect of changing the screening coverage and screening frequency.

In simulating this system, we obtain a full probability distribution in time for each village, of which we present the probability distribution of a selection for certain time points (Figure 5.27). For the smallest villages, there is a large probability of local gHAT elimination, as given the probability distribution of the endemic equilibrium, there is likely to be no infection in a given village.

However, for the largest villages, such as the one with a population of $N_H = 12,645$ (displayed in the rightmost column of Figure 5.27), there is a high probability of a large number of cases. This reduces in time, particularly as active screening begins to remove some of the low-risk infections; the panel after a year ($t = 1$), clearly shows the effect active screening can have in reducing infection, with the distribution reflecting peaks of a village, where screening has and has not occurred. While after 50 years ($t = 50$), the number of infected humans has decreased greatly, there is still a very high probability of a small number of cases persisting. This is shown more clearly in Figure 5.28.

Each village will have a probability of persistence that depends on the population size, with many small villages being very likely to achieve gHAT elimination, and largest villages being unlikely to obtain elimination, particularly with the high levels of imports, unless active screening is further intensified (Figure 5.28). The largest villages correspond to the greener lines than are plotted at the top of the left hand panel, showing the probability of persistence as close to one across the fifty years. The mean probability of gHAT persistence (the probability of persistence in a randomly

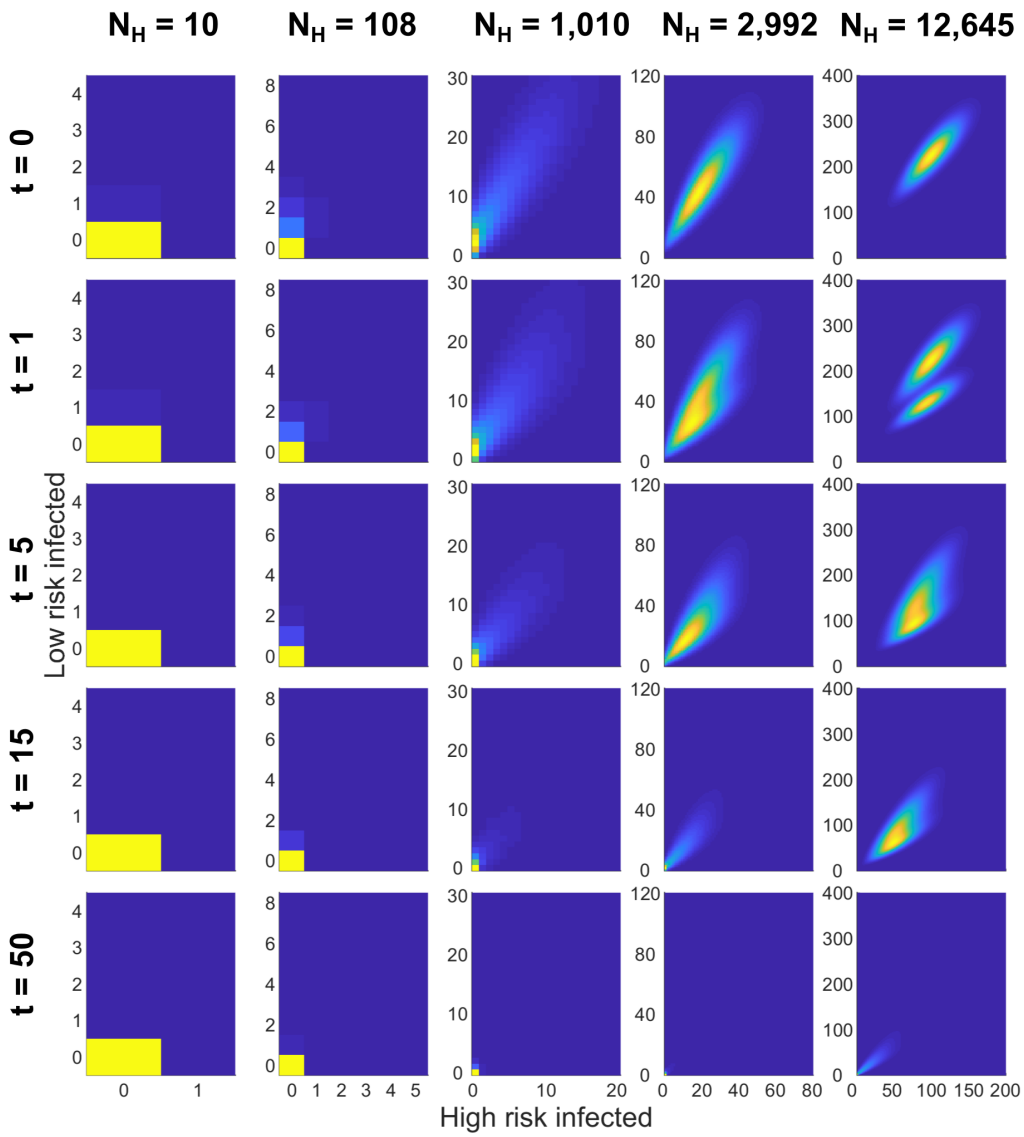


Fig 5.27: The risk distribution of infection in a subset of the villages of Yasa-Bonga and Mosango and selected time points. There are 559 villages with populations ranging between 3 and 12,645.

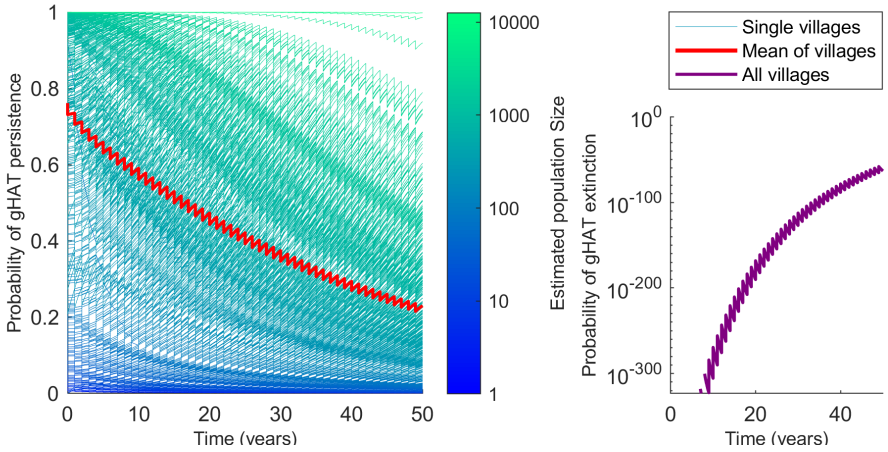


Fig 5.28: The probability of gHAT persistence in the villages of Yasa-Bonga and Mosango in time with mean active screening. Each blue/green line shows a single village, with the red line giving the average over all villages. The probability of simultaneous extinction across all villages is shown in the right panel.

selected village), decreases from approximately 0.8 to 0.2 in the simulated 50 years, reflecting the effects of active screening and stochastic extinction when the imports start to decrease. However, the continued persistence of infection in the largest villages means that the total probability of disease extinction across the whole health zones is exceedingly unlikely, even in the long time frames simulated here.

5.6 Discussion

By simplifying the gHAT model from a Ross–MacDonald–type model that includes vector dynamics (humans SEIIR, vectors SEEEI) to a simple Susceptible–Infected—Susceptible (SIS) model of infection, has allowed us to utilise a powerful master equation approach to our modelling, while maintaining a good correspondence to the original model. The SIS master equation model is now very fast to evaluate, and the nature of the implementation means that various interesting properties can easily be explored, with exact methods for calculating extinction times and expected dynamics [291]. This approach has also allowed the model to be easily extended to consider the interaction of multiple villages and even to consider the dynamics of persistence at the health zone level by linking the total number of infected individuals to the rate of infectious imports into the villages.

Using this model, we conclude that based upon the strategy of active screening at the mean level, gHAT will continue to persist in Yasa-Bonga and Mosango for more than 50 years. This is in line with deterministic predictions using this parameterisation; Rock *et al.* [219] predicted that the year of elimination as a public health problem would be 2140 (95% confidence interval (2103, 2199)) using the most similar model (Model 4 in that study) and the mean coverage of active screening.

This simplified model would benefit from many of the recent improvements in the model structure and model fitting of a more recent version of the full deterministic model [268], many of which are detailed in Section 4.2.2. This updated model makes use of more recent data from the WHO HAT Atlas [99, 106, 107] and considers improvements in the interventions for gHAT. For example, in the model of this chapter, we do not account for any improvement in the passive surveillance, which is known to have occurred [156], and would contribute to a decline in infection.

We are also aware that we have only used village populations that are listed in the WHO HAT Atlas for these predictions and there are known to be villages within these health zones that are not listed as they have never been screened. As shown in Figure 3.7, we also know that screening coverage and frequency is correlated with population size of a settlement, and so adapting our model to account for this would also likely result in more accurate predictions. The proportion of people in each risk class is also constant for all population sizes in the model, whereas we could speculate that the larger populations are more town-like and perhaps have fewer high-risk members. This potential over-estimation of high-risk people in the large populations could explain some of the very high gHAT persistence probabilities seen in results for these populations (Figure 5.28).

In addition, we consider a probability of active screening every year, despite the fact that this is unlikely to be necessary for small villages, where the disease is almost certain to be locally eliminated. To improve the plausibility of the model, we would like to add a cessation criterion, similar to the work in Chapter 4. This is less straightforward to implement in this probabilistic framework than in the tau-leaping scheme, as we do not consider specific realisations of the model where the infection is either detected or

not, but have a full probability distribution of all possible infection states. One potential solution could be to link the probability of being screened to the probability of observing a case in active screening.

As considered in Chapter 3, there is also no data on the movement of people between villages, and so the rate of importation was estimated by matching to the probability of detecting infection on the first active screening in a village. However, since in this model, we initially simulate forward in time with a given level of importations until the balance between extinction events and importations means we have found the endemic equilibrium, we could estimate this parameter more accurately for a given health zone, by matching the rate of importation to achieve the expected number of infected people at the endemic equilibrium; the import parameter currently gives an over-estimate of the endemic equilibrium found in the ODE framework.

Overall in this chapter, we have shown that a simplified model can achieve very similar results to a more biologically realistic version, while introducing a method of obtaining analytic results for extinction times and expected number of infected individuals. The results would benefit from updating the model parameterisation to fit to more recent data, however the predictions provided are inline with previous deterministic and stochastic results. The model implementation provides a framework to scale between modelling at a health zone or national level and at the village level. This is an important development, since the data obtained and the actual interventions (active screening) are provided at village level.

The current model from this chapter suggests that with mean active screening and limited passive surveillance, and with no additional interventions, such as vector control, the infection is almost certain to persist for long periods. This indicates that additional controls are required to achieve elimination of transmission, such as tsetse control or improved passive surveillance. Recent known improvements in the detection rates in passive surveillance could be added to this model however, and would decrease the probability of gHAT persistence. Indeed, this model structure with further refinements and fitting to multiple health zones could expand the range of analytical projections it is possible to generate.

6 Conclusions and discussion

6.1 Aims and previous work

The aim of this thesis was to take the existing modelling literature on *gambiense* human African trypanosomiasis (gHAT) and extend the work in terms of modelling at different spatial scales and incorporating stochasticity into the models. There was already a substantial amount of literature regarding the modelling of gHAT, considering its status as a neglected tropical disease, of which research started appearing from the 1980s [214, 215]. Many of the underlying epidemiological features of the disease were well-understood; some literature also gave insights into the effectiveness of control strategies and their potential to achieve elimination [14, 186, 218, 220]. However, since the beginning of this PhD thesis in 2016, there has been a large increase in the number of mathematical modelling publications for gHAT (for example the following studies [146, 179, 190, 221, 223–226]).

The work within this thesis has contributed significantly to this literature in many ways, having successfully formulated a stochastic infection model for gHAT that has been applied to village-level analyses, and with the ability to scale this model to a health zone, while maintaining village interactions. I detail the developments made in the work in the discussion below.

6.2 Discussion of the thesis and future work

In Chapter 2, I have described HAT with an introduction to its biology, transmission and disease symptoms. The history is then examined to put the current status of the disease into context. Subsequently, I aimed to build an understanding of the previous

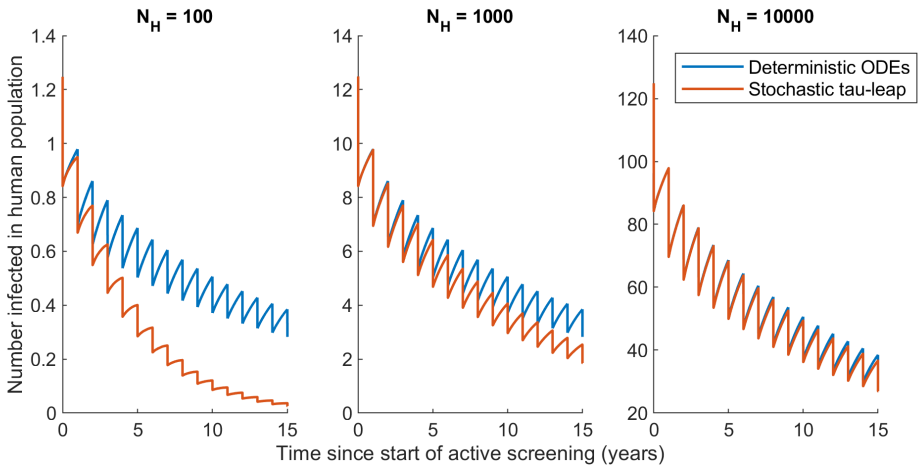


Fig 6.1: The expected number of infected individuals in time using deterministic ODEs and the stochastic tau-leaping algorithm for the same model structure of the full vector-host model for villages of different population sizes, N_H.

modelling literature for HAT, by reviewing the types of models and methods used for HAT and some other vector-borne diseases, before the implications of these models were considered.

In Chapter 3, I introduced a stochastic model for gHAT infection dynamics. At the time, this was one of the first models to use this methodology, although others have introduced stochastic gHAT models to the literature subsequently [225]. In particular, the use of the stochastic model was most appropriate in this thesis, since the focus is on small scale dynamics, modelling village populations with population sizes of typically less than 2,000 people (Figure 3.3).

These small population sizes mean that extinction events are more common, since the chain of transmission can be broken more readily with fewer susceptible people available to be infected. These extinction events are particularly common for a disease such as gHAT, where the incidence is so low. This means that the expected number of infected people will be less in stochastic models than deterministic models using the same model structure. It is evident that this effect is negligible for large populations, for example at the health zone scale with approximately 100,000 people, however, there is a large effect for smaller populations of size 100 or 1,000 people (Figure 6.1).

Throughout this chapter, I demonstrated that by using parameters fitted to the deterministic version of the model at the health zone level, and applied to an event-driven

stochastic simulation at the village level, provided a good correspondence with expected trends, seen from data (Figure 6.1). I also show that gHAT infection is able to persist for surprisingly long times when compared to other infectious diseases, even in these small populations and without the influence of infectious imports (Figure 3.13). An important result for policy is considering the distinction between the observation of detecting no cases of gHAT by an active screening and the actuality of that village having achieved elimination of transmission (EOT). I show that we would typically need at least three observations of no cases detected on active screenings and no passive detection in between them to have high confidence that local elimination has been met (Figure 3.16).

This work was completed with parameters inferred for a region combined of the two health zones, Yasa-Bonga and Mosango, and in particular I noted that there was some spatial variation in the applicability of the model to different villages, seemingly dependent on the proximity to a large river, where we know that the areas surrounding rivers typically have the highest tsetse density [81]. I would therefore like to extend this work, such that an updated model is parameterised based on the data from a single health zone, and so consider the probability of persistence of gHAT in multiple different health zones; by doing this we may better understand where the disease is more likely to persist and the learn more about the impact of the underlying parameters on gHAT persistence. In addition, this work assumed that all villages started with the infection levels associated with the calculated endemic equilibrium for the year 2000. It would be beneficial to consider the probability of persistence starting from different initial conditions, given a variety of interventions, such as tsetse control, improved passive surveillance and different coverages of active surveillance, have now been used in different health zones. Furthermore, when considering the probability of elimination of transmission in a village, we would like to relate the effects of the time of active screenings in the history of the village to the probability of local elimination.

In Chapter 4, I used an updated stochastic model from Chapter 3, to consider the potential efficiency of a range of plausible active screening strategies, which were defined in terms of the screening coverage, screening frequency and cessation of screening criteria. While, higher screening coverages, with increased frequency and no cessation

would be the best strategy in terms of achieving and maintaining local gHAT elimination, this is exceedingly unlikely to be an efficient use of resources, and completely infeasible when scaled up a national level, or across every village located in the gHAT-endemic regions.

Thus, I explore the potential of using the net monetary benefit (NMB) framework for determining prospective optimal strategies. This involved running the stochastic model simulations for a large range of strategies and applying an array of relevant costs to the resulting output. The NMB of each strategy is considered to see where the maximum was achieved for a given willingness to pay (WTP) of the funder (Figure 4.4). A high screening coverage is particularly important when considering cost-effectiveness of a strategy, however, this may be difficult to achieve practically, due to individuals being not present or unwilling to be screened for gHAT. Therefore, I also limited the strategies under consideration to a range of more practical strategies, to determine the probability each was cost-effective (Figure 4.8). The results show that active screening is typically cost-effective for low to moderate WTP thresholds, when screening is carried out (approximately) annually with a high coverage.

The results from this chapter provide a quantitative backing to the current WHO guidelines for active screening [267] and indicates that annual screening should continue to be a priority in the absence of additional interventions, such as tsetse control. The work provides an important confirmation of the benefit of active screening programmes within a cost-effectiveness framework.

This work used a model that was matched to the health zone Kwamouth, while a comparison was also made for the health zone of Mosango. Future analyses could therefore focus on extending the work to additional health zones to observe any local variation in the results. Additionally, each village was considered independent of all other villages in this modelling work, while in practice, the costs of maintaining active surveillance teams in a region will depend on the interaction with all villages in that region. For example, it may not be cost-effective to have a fully-staffed mobile team conducting active screenings, if this was only required for a single village in the health zone.

Therefore, this model could be extended to consider a model based upon all the villages visited by a single mobile screening team, rather than a single village. The availability of active screening is dependent on the number and location of the mobile active screening teams [97]. The whole system would therefore be captured and modelled, since the basis of this chapter is optimising the use of a mobile team. This would also allow for a better understanding of the costs involved and clarify the use of reactive screening in a village, whereby if cases are presented in a village previously thought to have locally eliminated gHAT, reactive screening could quickly occur in a region where there already exists a mobile screening team, however, there may be greater costs involved if such a team has already been disbanded from the whole region. In addition, additional strategies could be modelled such as the reactive use of the oral acoziborale drug in mass administration for single villages.

In Chapter 5, I considered a different methodology for exploring gHAT dynamics — the use of the master equations. Through the use of a master equation model I obtained numerically exact results, rather than relying on a large number of realisations from an event-driven simulation model. Furthermore, rather than examining the summary statistics of these many replications, the full probability distribution for the state space of possible infection states in time could be analysed. Exact methods to calculate expected disease extinction times were also used.

To use the master equation approach, I simplified the model dynamics, using the quasi-equilibrium assumption for the tsetse infection dynamics and reducing the number of human infection classes to simply susceptible infected, although maintaining the risk class structure. This reduction in state space allowed for the rate matrix of the system to be formulated within the boundary of reasonable memory allowance.

The examination of the reduced model using the three different methodologies discussed in this thesis — deterministic ordinary differential equations, the stochastic tau-leaping approximation and the use of the master equations — shows that we observe expected results. Using sufficient tau-leaping realisations will provide identical expected infection dynamics to the master equations, however due to extinction events in the two

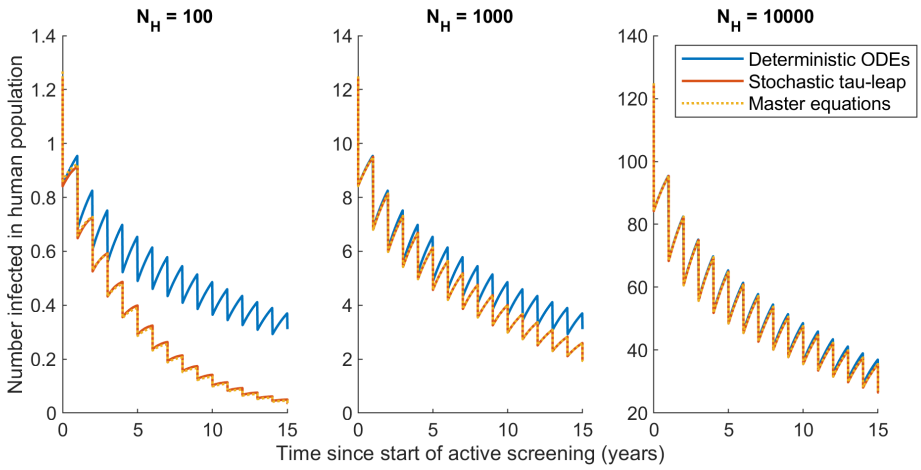


Fig 6.2: The expected number of infected individuals in time using deterministic ODEs, the stochastic tau-leaping algorithm and the master equations for the same model structure of the simplified SIS model.

stochastic methods, the expected number of infected individuals will be lower than the deterministic model for small populations (Figure 6.2).

The use of the master equations provided the necessary computational efficiency to evaluate the effect of the infection dynamics in a group of villages, where movement between them can introduce importations of infection. This link between all villages in a region, means that results at the village-level can be scaled up to consider regional infection dynamics, probabilities of persistence, and the effects of different intervention strategies.

In particular, by using the model parameterisation and calculated rate of importations of infection (3.4×10^{-6} per day, proportional to the number of susceptible people in the village), we can simulate the gHAT dynamics for a full region comprised on individual villages, using village and screening parameters directly from the WHO HAT Atlas [106, 107] (Figure 5.20). Furthermore, we can also compare these results to a deterministic model of a single population the size of the entire health zones (Yasa-Bonga and Mosango), using the same total annual active screening coverage for the region (Figure 5.27).

We see that for the same total population size that with the level of infectious imports calculated in Section 3.7 gives a higher number of initial infected than the deterministic equivalent with one population for the whole region. This overestimation of the impor-

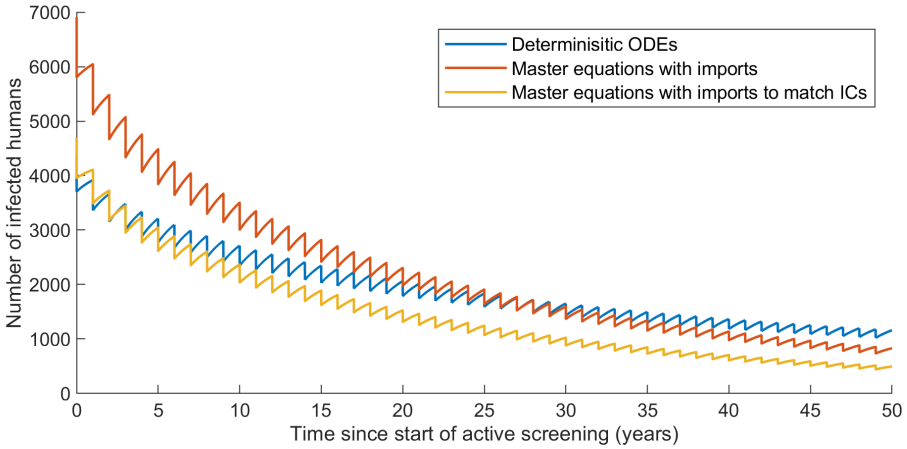


Fig 6.3: The expected number of infected individuals in time using deterministic ODEs for the whole region of Yasa-Bonga and Mosango, and the master equations for connected villages, using the SIS model structure. The rate of importations is given by 3.4×10^{-6} per day (blue) and 1.7×10^{-6} per day (yellow), both proportional to the number of susceptible people in the village.

tation rate may be that the rate was obtained using only data from villages included in the WHO HAT Atlas [106, 107], and thus we have not accounted for the villages not thought to be presenting cases, which would reduce this rate for the whole region. By matching the rate of importation to the initial conditions from the deterministic model, we can achieve a better fit (Figure 6.3). As we expect, the expected number of infected people will reduce faster in the master equation method than the deterministic model in time because of the local extinction events and because the population is subdivided into villages, in which people will primarily only mix with other people within their own village, limiting the potential infection spread.

Future work for this model would be to update the model structure and model parameters to more recent model fits and to focus more upon the calculation of the rate of infectious imports between the villages. We could also simulate a large number of villages with spatial interaction if there were sufficient data to support this. However, this methodology is an important step in the scaling between models at the village level and more broadly considering gHAT across larger regions.

6.3 Conclusions

Overall, this thesis has thoroughly investigated the dynamic, system behaviour of gHAT infection in individual village populations with some implications for modelling gHAT at different spatial scales. The development of stochastic infection models for gHAT and the ability to capture processes such as village persistence and to be able to make recommendations using these models in cost-effectiveness frameworks is a major advance in gHAT modelling for public health.

To achieve the elimination target set by WHO for 2030 [11], the infection needs to be understood at the scale of the village, since it is at this level the controls are implemented. I hope this work can therefore provide some additional insights for achieving this ambitious goal.

Bibliography

1. Roth, G. A., Abate, D., Abate, K. H., *et al.* Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392,** 1736–1788 (2018) (cit. on p. 1).

2. Global Health Estimates. *Deaths by Cause, Age, Sex, by Country and by Region, 2000–2016* 2018 (cit. on p. 1).

3. Boutayeb, A. The burden of communicable and non-communicable diseases in developing countries. *Handbook of disease burdens and quality of life measures,* 531–546 (2010) (cit. on pp. 1, 74).

4. Ross, R. An application of the theory of probabilities to the study of a priori pathometry.—Part I. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character* **92,** 204–230 (1916) (cit. on pp. 2, 27, 31).

5. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* **115,** 700–721 (1927) (cit. on pp. 2, 27).

6. Keeling, M. J. & Rohani, P. *Modeling infectious diseases in humans and animals* (Princeton University Press, 2011) (cit. on pp. 2, 27, 28, 32, 113).

7. Whitty, C. J. *What makes an academic paper useful for health policy?* (BioMed Central, 2015) (cit. on pp. 2, 86).

8. Behrend, M. R., Basáñez, M.-G., Hamley, J. I., *et al.* Modelling for policy: The five principles of the Neglected Tropical Diseases Modelling Consortium. *PLoS neglected tropical diseases* **14,** e0008033 (2020) (cit. on p. 2).

9. Irvine, M. A. & Hollingsworth, T. D. Making transmission models accessible to end-users: the example of TRANSFIL. *PLoS neglected tropical diseases* **11,** e0005206 (2017) (cit. on p. 2).

10. Davis, C. N., Hollingsworth, T. D., Caudron, Q. & Irvine, M. A. The use of mixture density networks in the emulation of complex epidemiological individual-based models. *PLoS computational biology* **16,** e1006869 (2020) (cit. on p. 2).

11. World Health Organization. *Proposed goals and milestones for each NTD* tech. rep. (World Health Organization, 2019) (cit. on pp. 2, 16, 158).

12. Büscher, P., Cecchi, G., Jamonneau, V. & Priotto, G. Human african trypanosomiasis. *The Lancet* **390,** 2397–2409 (2017) (cit. on pp. 4, 5, 9, 15, 21–23, 25, 43, 73, 109).

13. Aksoy, S., Buscher, P., Lehane, M., Solano, P. & Van Den Abbeele, J. Human African trypanosomiasis control: achievements and challenges. *PLoS neglected tropical diseases* **11,** e0005454 (2017) (cit. on pp. 4, 15, 19, 24).

14. Rock, K. S., Stone, C. M., Hastings, I. M., *et al.* in *Advances in parasitology* 53–133 (Elsevier, 2015) (cit. on pp. 4, 31, 45, 151).

15. Jamonneau, V., Ilboudo, H., Kaboré, J., *et al.* Untreated human infections by Trypanosoma brucei gambiense are not 100% fatal. *PLoS neglected tropical diseases* **6,** e1691 (2012) (cit. on pp. 5, 9, 24).

16. Franco, J. R., Cecchi, G., Priotto, G., *et al.* Monitoring the elimination of human African trypanosomiasis: Update to 2014. *PLoS neglected tropical diseases* **11,** e0005585 (2017) (cit. on pp. 5, 16).

17. Smith, D. H., Pepin, J. & Stich, A. H. Human African trypanosomiasis: an emerging public health crisis. *British Medical Bulletin* **54,** 341–355 (1998) (cit. on p. 5).

18. Brun, R., Blum, J., Chappuis, F. & Burri, C. Human african trypanosomiasis. *The Lancet* **375,** 148–159 (2010) (cit. on pp. 5, 38).

19. Simarro, P. P., Franco, J. R., Cecchi, G., *et al.* Human African Trypanosomiasis in Non-Endemic Countries (2000–2010). *Journal of travel medicine* **19,** 44–53 (2011) (cit. on p. 5).

20. World Health Organization. *Global health observatory data repository* tech. rep. (World Health Organization, 2018) (cit. on pp. 5, 6, 15, 16, 21, 25, 67, 69).

21. Losos, G. J. & Ikede, B. Review of pathology of diseases in domestic and laboratory animals caused by Trypanosoma congolense, T. vivax, T. brucei, T. rhodesiense and T. gambiense. *Veterinary Pathology* **9,** 1–79 (1972) (cit. on p. 6).

22. Molyneux, D. *et al.* Animal reservoirs and Gambian trypanosomiasis. *Ann Soc Belg Med Trop* **53,** 605–618 (1973) (cit. on p. 6).

23. Njiokou, F., Nimpaye, H., Simo, G., *et al.* Domestic animals as potential reservoir hosts of Trypanosoma brucei gambiense in sleeping sickness foci in Cameroon. *Parasite* **17,** 61–66 (2010) (cit. on p. 6).

24. Berman, J. J. *Taxonomic guide to infectious diseases: understanding the biologic classes of pathogenic organisms* (Academic Press, 2019) (cit. on p. 6).

25. Bastin, P. The trypanosome journey in the tsetse fly. *Comptes Rendus Biologies* **342,** 273–275 (2019) (cit. on p. 6).

26. Stephens, N. A., Kieft, R., MacLeod, A. & Hajduk, S. L. Trypanosome resistance to human innate immunity: targeting Achilles' heel. *Trends in parasitology* **28,** 539–545 (2012) (cit. on p. 6).

27. Pays, E., Vanhollebeke, B., Uzureau, P., Lecordier, L. & Pérez-Morga, D. The molecular arms race between African trypanosomes and humans. *Nature Reviews Microbiology* **12,** 575 (2014) (cit. on p. 6).

28. Vanhollebeke, B. & Pays, E. The trypanolytic factor of human serum: many ways to enter the parasite, a single way to kill. *Molecular microbiology* **76,** 806–814 (2010) (cit. on p. 6).

29.Thomson, R. & Finkelstein, A. Human trypanolytic factor APOL1 forms pH-gated cation-selective channels in planar lipid bilayers: relevance to trypanosome lysis. *Proceedings of the National Academy of Sciences* **112,** 2894–2899 (2015) (cit. on p. 6).

30.Lugli, E. B., Pouliot, M., Portela, M. d. P. M., Loomis, M. R. & Raper, J. Characterization of primate trypanosome lytic factors. *Molecular and biochemical parasitology* **138,** 9–20 (2004) (cit. on p. 6).

31.Nakayima, J., Nakao, R., Alhassan, A., *et al.* Molecular epidemiological studies on animal trypanosomiases in Ghana. *Parasites & vectors* **5,** 217 (2012) (cit. on p. 6).

32.Hoare, C. A. & Wallace, F. G. Developmental stages of trypanosomatid flagellates: a new terminology. *Nature* **212,** 1385–1386 (1966) (cit. on p. 7).

33.Cecchi, G., Mattioli, R. C., Slingenbergh, J. & De La Rocque, S. Land cover and tsetse fly distributions in sub-Saharan Africa. *Medical and veterinary entomology* **22,** 364–373 (2008) (cit. on p. 7).

34.Grébaut, P., Melachio, T., Nyangmang, S., *et al.* Xenomonitoring of sleeping sickness transmission in Campo (Cameroon). *Parasites & vectors* **9,** 201 (2016) (cit. on p. 7).

35.Shereni, W., Anderson, N. E., Nyakupinda, L. & Cecchi, G. Spatial distribution and trypanosome infection of tsetse flies in the sleeping sickness focus of Zimbabwe in Hurungwe District. *Parasites & vectors* **9,** 605 (2016) (cit. on p. 7).

36.Leak, S. G. *Tsetse biology and ecology: their role in the epidemiology and control of trypanosomosis* (ILRI (aka ILCA and ILRAD), 1999) (cit. on p. 7).

37.Hargrove, J., English, S., Torr, S. J., *et al.* Wing length and host location in tsetse (Glossina spp.): implications for control using stationary baits. *Parasites & vectors* **12,** 24 (2019) (cit. on p. 7).

38.Moloo, S., Sabwa, C. & Kabata, J. Vector competence of Glossina pallidipes and G. morsitans centralis for Trypanosoma vivax, T. congolense and T. b. brucei. *Acta tropica* **51,** 271–280 (1992) (cit. on pp. 7, 9).

39.Dale, C., Welburn, S., Maudlin, I. & Milligan, P. The kinetics of maturation of trypanosome infections in tsetse. *Parasitology* **111,** 187–191 (1995) (cit. on p. 7).

40.Akoda, K., Van den Bossche, P., Marcotty, T., *et al.* Nutritional stress affects the tsetse fly's immune gene expression. *Medical and veterinary entomology* **23,** 195–201 (2009) (cit. on p. 7).

41.Vreysen, M. J., Seck, M. T., Sall, B. & Bouyer, J. Tsetse flies: their biology and control using area-wide integrated pest management approaches. *Journal of invertebrate pathology* **112,** S15–S25 (2013) (cit. on p. 7).

42.Hagan, H. R. Embryology of the viviparous insects. *Transactions of the New York Academy of Sciences* **12,** 112–112 (1950) (cit. on p. 7).

43.Lewis, D. The behaviour of the larvae of tsetse-flies before pupation. *Bulletin of Entomological Research* **25,** 195–200 (1934) (cit. on p. 7).

44.Hargrove, J. Tsetse: the limits to population growth. *Medical and veterinary Entomology* **2,** 203–217 (1988) (cit. on p. 8).

45.Pollock, J. Male accessory secretions, their use and replenishment in Glossina (Diptera, Glossinidae). *Bulletin of Entomological Research* **64,** 533–539 (1974) (cit. on p. 8).

46. Glasgow, J. P. *et al.* The distribution and abundance of tsetse. *The Distribution and Abundance of Tsetse.* (1963) (cit. on p. 8).

47. Rogers, D. J. & Randolph, S. E. Estimation of rates of predation on tsetse. *Medical and veterinary entomology* **4,** 195–204 (1990) (cit. on p. 8).

48. Rogers, D. J. & Randolph, S. E. Distribution and abundance of tsetse flies (Glossina spp.) *The Journal of Animal Ecology,* 1007–1025 (1986) (cit. on p. 8).

49. Hargrove, J. The flight performance of tsetse flies. *Journal of Insect Physiology* **21,** 1385–1395 (1975) (cit. on p. 8).

50. Vale, G. & Torr, S. User-friendly models of the costs and efficacy of tsetse control: application to sterilizing and insecticidal techniques. *Medical and veterinary entomology* **19,** 293–305 (2005) (cit. on pp. 8, 35).

51. Hargrove, J. & Williams, B. Optimized simulation as an aid to modelling, with an application to the study of a population of tsetse flies, Glossina morsitans morsitans (Diptera: Glossinidae). *Bulletin of entomological research* **88,** 425–435 (1998) (cit. on pp. 8, 35).

52. Hargrove, J. A theoretical study of the invasion of cleared areas by tsetse flies (Diptera: Glossinidae). *Bulletin of Entomological Research* **90,** 201–209 (2000) (cit. on p. 8).

53. Van Den Abbeele, J., Claes, Y., Van Bockstaele, D., Le Ray, D. & Coosemans, M. Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology* **118,** 469–478 (1999) (cit. on p. 8).

54. Kristensson, K., Masocha, W. & Bentivoglio, M. in *Handbook of clinical neurology* 11–22 (Elsevier, 2013) (cit. on p. 8).

55. Ashcroft, M., Burtt, E. & Fairbairn, H. The experimental infection of some African wild animals with Trypanosoma rhodesiense, T. brucei and T. congolense. *Annals of Tropical Medicine & Parasitology* **53,** 147–161 (1959) (cit. on p. 8).

56. Horn, D. Antigenic variation in African trypanosomes. *Molecular and biochemical parasitology* **195,** 123–129 (2014) (cit. on p. 9).

57. Mugnier, M. R., Stebbins, C. E. & Papavasiliou, F. N. Masters of disguise: antigenic variation and the VSG coat in Trypanosoma brucei. *PLoS pathogens* **12,** e1005784 (2016) (cit. on p. 9).

58. Rotureau, B. & Van Den Abbeele, J. Through the dark continent: African trypanosome development in the tsetse fly. *Frontiers in cellular and infection microbiology* **3,** 53 (2013) (cit. on p. 9).

59. Sharma, R., Gluenz, E., Peacock, L., *et al.* The heart of darkness: growth and form of Trypanosoma brucei in the tsetse fly. *Trends in parasitology* **25,** 517–524 (2009) (cit. on p. 9).

60. Matetovici, I., Caljon, G. & Van Den Abbeele, J. Tsetse fly tolerance to T. brucei infection: transcriptome analysis of trypanosome-associated changes in the tsetse fly salivary gland. *BMC genomics* **17,** 971 (2016) (cit. on p. 9).

61. Hao, Z., Kasumba, I. & Aksoy, S. Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: Glossinidiae). *Insect biochemistry and molecular biology* **33,** 1155–1164 (2003) (cit. on p. 9).

62.Walshe, D. P., Lehane, M. J. & Haines, L. R. Post eclosion age predicts the prevalence of midgut trypanosome infections in Glossina. *PloS one* **6** (2011) (cit. on p. 9).

63.Haines, L. R. Examining the tsetse teneral phenomenon and permissiveness to trypanosome infection. *Frontiers in cellular and infection microbiology* **3,** 84 (2013) (cit. on pp. 9, 34, 45).

64.Lindner, A. K. & Priotto, G. The unknown risk of vertical transmission in sleeping sickness—a literature review. *PLoS neglected tropical diseases* **4,** e783 (2010) (cit. on p. 10).

65.De Kyvon, M.-A. L.-C., Maakaroun-Vermesse, Z., Lanotte, P., *et al.* Congenital trypanosomiasis in child born in France to African mother. *Emerging infectious diseases* **22,** 935 (2016) (cit. on p. 10).

66.Rocha, G., Martins, A., Gama, G., Brandao, F. & Atouguia, J. Possible cases of sexual and congenital transmission of sleeping sickness. *The Lancet* **363,** 247 (2004) (cit. on p. 10).

67.Hira, P. & Husein, S. Some transfusion-induced parasitic infections in Zambia. *Journal of hygiene, epidemiology, microbiology, and immunology* **23,** 436–444 (1979) (cit. on p. 10).

68.Herwaldt, B. L. Laboratory-acquired parasitic infections from accidental exposures. *Clinical microbiology reviews* **14,** 659–688 (2001) (cit. on p. 10).

69.Kennedy, P. G. Human African trypanosomiasis of the CNS: current issues and challenges. *The Journal of clinical investigation* **113,** 496–504 (2004) (cit. on p. 10).

70.Kennedy, P. G. Human African trypanosomiasis–neurological aspects. *Journal of neurology* **253,** 411–416 (2006) (cit. on p. 10).

71.Kuepfer, I., Hhary, E. P., Allan, M., *et al.* Clinical presentation of Tb rhodesiense sleeping sickness in second stage patients from Tanzania and Uganda. *PLoS neglected tropical diseases* **5,** e968 (2011) (cit. on p. 10).

72.Buguet, A., Bert, J., Tapie, P., *et al.* Sleep-wake cycle in human African trypanosomiasis. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society* **10,** 190–196 (1993) (cit. on p. 10).

73.Mpandzou, G., Cespuglio, R., Ngampo, S., *et al.* Polysomnography as a diagnosis and post-treatment follow-up tool in human African trypanosomiasis: a case study in an infant. *Journal of the neurological sciences* **305,** 112–115 (2011) (cit. on p. 10).

74.Njamnshi, A. K., Gettinby, G. & Kennedy, P. G. The challenging problem of disease staging in human African trypanosomiasis (sleeping sickness): a new approach to a circular question. *Transactions of The Royal Society of Tropical Medicine and Hygiene* **111,** 199–203 (2017) (cit. on p. 10).

75.Buguet, A., Bourdon, L., Bouteille, B., *et al.* The duality of sleeping sickness: focusing on sleep. *Sleep medicine reviews* **5,** 139–153 (2001) (cit. on p. 10).

76.Blum, J., Schmid, C. & Burri, C. Clinical aspects of 2541 patients with second stage human African trypanosomiasis. *Acta tropica* **97,** 55–64 (2006) (cit. on p. 10).

77.Checchi, F., Funk, S., Chandramohan, D., Haydon, D. T. & Chappuis, F. Updated estimate of the duration of the meningo-encephalitic stage in gambiense human African trypanosomiasis. *BMC research notes* **8,** 292 (2015) (cit. on pp. 11, 36, 49, 80, 112).

78.Odiit, M., Kansiime, F. & Enyaru, J. Duration of symptoms and case fatality of sleeping sickness caused by Trypanosoma brucei rhodesiense in Tororo, Uganda. *East African medical journal* **74,** 792–795 (1997) (cit. on p. 11).

79.Steverding, D. The history of African trypanosomiasis. *Parasites & vectors* **1,** 3 (2008) (cit. on pp. 11–13, 70).

80.Lambrecht, F. L. Trypanosomes and hominid evolution. *Bioscience* **35,** 640–646 (1985) (cit. on p. 11).

81.Franco, J. R., Simarro, P. P., Diarra, A. & Jannin, J. G. Epidemiology of human African trypanosomiasis. *Clinical epidemiology* **6,** 257 (2014) (cit. on pp. 11, 58, 153).

82.Griffith, F. L. *The Petrie Papyri: Hieratic Papyri from Kahun and Gurob (principally of the Middle Kingdom)*. (B. Quaritch, 1898) (cit. on p. 11).

83.Cox, F. E. History of sleeping sickness (African trypanosomiasis). *Infectious Disease Clinics* **18,** 231–245 (2004) (cit. on pp. 11, 12).

84.De Raadt, P. The history of sleeping sickness. *Fourth International Cours on African Trypanosomoses* (2005) (cit. on pp. 11–13).

85.Bruce, D. Preliminary report on the tsetse fly disease or nagana. *Zululand (London: Durban, Bennett & Davis)* (1895) (cit. on p. 12).

86.Forde, R. M. Some clinical notes on a European patient in whose blood a trypanosome was observed. *J Trop Med* **5,** 261–263 (1902) (cit. on p. 12).

87.Dutton, J. E. Preliminary note upon a trypanosome occurring in the blood of man (1902) (cit. on p. 12).

88.Castellani, A. On the discovery of a species of Trypanosoma in the cerebrospinal fluid of cases of sleeping sickness. *Proceedings of the Royal Society of London* **71,** 501–508 (1903) (cit. on p. 12).

89.Kleine, F. Weitere wissenschaftliche Beobachtungen über die Entwicklung von Trypanosomen in Glossinen. *Dtsch Med Wochenschr* **35,** 924–925 (1909) (cit. on p. 12).

90.Stephens, J. W. W. & Fantham, H. B. On the peculiar morphology of a trypanosome from a case of sleeping sickness and the possibility of its being a new species (T. rhodesiense). *Annals of Tropical Medicine & Parasitology* **4,** 343–350 (1910) (cit. on p. 12).

91.Louis, W. R. & Gifford, P. *France and Britain in Africa: Imperial rivalry and colonial rule* (Yale University Press, 1971) (cit. on p. 12).

92.Fèvre, E. M., Coleman, P. G., Welburn, S. C. & Maudlin, I. Reanalyzing the 1900–1920 sleeping sickness epidemic in Uganda. *Emerg Infect Dis* **10,** 567–573 (2004) (cit. on p. 13).

93.Fage, J. D., Crowder, M. & Oliver, R. A. *The Cambridge History of Africa* (Cambridge University Press, 1975) (cit. on p. 13).

94.Mbopi-Keou, F.-X., Belec, L., Milleliri, J.-M. & Teo, C.-G. The Legacies of Eugene Jamot and La Jamotique. *PLoS neglected tropical diseases* **8** (2014) (cit. on p. 13).

95.World Health Organization. *WHO report on global surveillance of epidemic-prone infectious diseases* tech. rep. (World Health Organization, 2000) (cit. on pp. 14, 15).

96. Hasker, E., Lutumba, P., Chappuis, F., *et al.* Human African trypanosomiasis in the Democratic Republic of the Congo: a looming emergency? *PLoS neglected tropical diseases* **6,** e1950 (2012) (cit. on pp. 14, 16).

97. Robays, J., Bilengue, M. M. C., Stuyft, P. V. d. & Boelaert, M. The effectiveness of active population screening and treatment for sleeping sickness control in the Democratic Republic of Congo. *Tropical Medicine & International Health* **9,** 542–550 (2004) (cit. on pp. 15, 18, 73, 109, 155).

98. World Health Organization. Trypanosomiasis, human African (sleeping sickness). *Fact sheet* **259** (2012) (cit. on p. 16).

99. Franco, J. R., Cecchi, G., Priotto, G., *et al.* Monitoring the elimination of human African trypanosomiasis at continental and country level: Update to 2018. *PLOS Neglected Tropical Diseases* **14,** e0008261 (2020) (cit. on pp. 16, 17, 149).

100. Organization, W. H. *Accelerating work to overcome the global impact of neglected tropical diseases: a roadmap for implementation: executive summary* tech. rep. (World Health Organization, 2012) (cit. on pp. 16, 17).

101. Organization, W. H. *Sustaining the drive to overcome the global impact of neglected tropical diseases: second WHO report on neglected diseases* (World Health Organization, 2013) (cit. on pp. 16, 17).

102. Franco, J., Cecchi, G., Priotto, G., *et al.* Monitoring the elimination of human African trypanosomiasis: Update to 2016. *PLoS neglected tropical diseases* **12,** e0006890 (2018) (cit. on pp. 17, 24, 26, 119).

103. Jannin, J., Louis, F., Lucas, P. & Simarro, P. Control of human African trypanosomiasis: back to square one. *Medecine tropicale: revue du Corps de sante colonial* **61,** 437–440 (2001) (cit. on p. 17).

104. Bardosh, K. L. Towards a science of global health delivery: A socio-anthropological framework to improve the effectiveness of neglected tropical disease interventions. *PLoS neglected tropical diseases* **12** (2018) (cit. on p. 17).

105. Cattand, P., Jannin, J. & Lucas, P. Sleeping sickness surveillance: an essential step towards elimination. *Tropical medicine & international health* **6,** 348–361 (2001) (cit. on p. 17).

106. Simarro, P. P., Cecchi, G., Paone, M., *et al.* The Atlas of human African trypanosomiasis: a contribution to global mapping of neglected tropical diseases. *International journal of health geographics* **9,** 57 (2010) (cit. on pp. 17, 32, 34, 39–42, 47, 55, 59, 67, 76, 79, 81, 96, 140, 141, 149, 156, 157).

107. Simarro, P. P., Cecchi, G., Franco, J. R., *et al.* Monitoring the progress towards the elimination of gambiense human African trypanosomiasis. *PLoS neglected tropical diseases* **9** (2015) (cit. on pp. 17, 32, 34, 39–42, 47, 55, 59, 67, 76, 79, 81, 96, 140, 141, 149, 156, 157).

108. Cecchi, G., Paone, M., Franco, J. R., *et al.* Towards the Atlas of human African trypanosomiasis. *International Journal of Health Geographics* **8,** 15 (2009) (cit. on p. 17).

109. Solano, P., Torr, S. J., *et al.* Is vector control needed to eliminate gambiense human African trypanosomiasis? *Frontiers in cellular and infection microbiology* **3,** 33 (2013) (cit. on pp. 17, 20).

110. Organization, W. H. *Control and surveillance of human African trypanosomiasis: report of a WHO expert committee* (World Health Organization, 2013) (cit. on pp. 18, 22).

111. Jamot, E. La maladie du sommeil au Cameroun. *Bulletin de la Société de Pathologie Exotique* **18,** 762–69 (1925) (cit. on p. 18).

112. Headrick, D. R. Sleeping sickness epidemics and colonial responses in East and Central Africa, 1900–1940. *PLoS neglected tropical diseases* **8** (2014) (cit. on p. 18).

113. Robays, J., Lefevre, P., Lutumba, P., *et al.* Drug toxicity and cost as barriers to community participation in HAT control in the Democratic Republic of Congo. *Tropical Medicine & International Health* **12,** 290–298 (2007) (cit. on p. 18).

114. Courtin, F., Camara, O., Camara, M., *et al.* Sleeping sickness in the historical focus of forested Guinea: update using a geographically based method. *Parasite* **26** (2019) (cit. on p. 18).

115. Louis, F., Kohagne, L. T., Ebo'O, V. E. & Simarro, P. Organizing an active screening campaign for human African trypanosomiasis due to Trypanosoma brucei gambiense. *Médecine tropicale: revue du Corps de santé colonial* **68,** 11–16 (2008) (cit. on pp. 18, 96, 133).

116. Mulenga, P., Boelaert, M., Lutumba, P., *et al.* Integration of Human African trypanosomiasis control activities into primary health services in the democratic republic of the Congo: A qualitative study of stakeholder perceptions. *The American journal of tropical medicine and hygiene* **100,** 899–906 (2019) (cit. on p. 18).

117. Mpanya, A., Hendrickx, D., Vuna, M., *et al.* Should I get screened for sleeping sickness? A qualitative study in Kasai province, Democratic Republic of Congo. *PLoS neglected tropical diseases* **6,** e1467 (2012) (cit. on pp. 18, 19, 24, 36, 45, 49, 80, 96, 114).

118. Van Nieuwenhove, S. in *Progress in human African trypanosomiasis, sleeping sickness* 253–280 (Springer, 1999) (cit. on pp. 18, 76).

119. Simarro, P., Franco, J. & Asumu, P. N. Has the focus of human African trypanosomiasis in Luba, Equatorial Guinea been eradicated? *Medecine tropicale: revue du Corps de sante colonial* **61,** 441–444 (2001) (cit. on pp. 18, 76).

120. Molyneux, D., Ndung'u, J. & Maudlin, I. Controlling sleeping sickness—"when will they ever learn?" *PLoS neglected tropical diseases* **4** (2010) (cit. on p. 18).

121. Barrett, M. P. The elimination of human African trypanosomiasis is in sight: Report from the third WHO stakeholders meeting on elimination of gambiense human African trypanosomiasis. *PLoS neglected tropical diseases* **12,** e0006925 (2018) (cit. on p. 18).

122. Simarro, P. P., Jannin, J. & Cattand, P. Eliminating human African trypanosomiasis: where do we stand and what comes next? *PLoS medicine* **5,** e55 (2008) (cit. on p. 18).

123. Koffi, M., N'Djetchi, M., Ilboudo, H., *et al.* A targeted door-to-door strategy for sleeping sickness detection in low-prevalence settings in Côte d'Ivoire. *Parasite* **23** (2016) (cit. on pp. 19, 73).

124. Hasker, E., Lumbala, C., Mpanya, A., *et al.* Alternative strategies for case finding in human African trypanosomiasis in the Democratic Republic of the Congo: PS2. 093. *Tropical Medicine & International Health* **20** (2015) (cit. on p. 19).

125. Pepin, J., Guern, C., Milord, F. & Bokelo, M. Integration of African human trypanosomiasis control in a network of multipurpose health centers. *Bulletin of the World Health Organization* **67,** 301–308 (1989) (cit. on p. 19).

126. Franco, J., Simarro, P., Diarra, A., Ruiz-Postigo, J. & Jannin, J. The journey towards elimination of gambiense human African trypanosomiasis: not far, nor easy. *Parasitology* **141,** 748–760 (2014) (cit. on p. 19).

127. Mitashi, P., Hasker, E., Mbo, F., *et al.* Integration of diagnosis and treatment of sleeping sickness in primary healthcare facilities in the Democratic Republic of the Congo. *Tropical Medicine & International Health* **20,** 98–105 (2015) (cit. on pp. 19, 73).

128. Mudji, J., Blum, A., Grize, L., *et al.* Gambiense Human African Trypanosomiasis Sequelae after Treatment: A Follow-Up Study 12 Years after Treatment. *Tropical Medicine and Infectious Disease* **5,** 10 (2020) (cit. on p. 19).

129. Mpanya, A., Hendrickx, D., Baloji, S., *et al.* From health advice to taboo: community perspectives on the treatment of sleeping sickness in the Democratic Republic of Congo, a qualitative study. *PLoS neglected tropical diseases* **9,** e0003686 (2015) (cit. on pp. 19, 50).

130. Checchi, F., Funk, S., Chandramohan, D., Chappuis, F. & Haydon, D. T. The impact of passive case detection on the transmission dynamics of gambiense Human African Trypanosomiasis. *PLoS neglected tropical diseases* **12,** e0006276 (2018) (cit. on p. 19).

131. Lehane, M., Alfaroukh, I., Bucheton, B., *et al.* Tsetse control and the elimination of Gambian sleeping sickness. *PLoS neglected tropical diseases* **10,** e0004437 (2016) (cit. on pp. 19, 20).

132. World Health Organization. Guidelines for malaria vector control (2019) (cit. on p. 20).

133. World Health Organization. Global strategy for dengue prevention and control 2012-2020 (2012) (cit. on p. 20).

134. Vreysen, M. J., Seck, M. T., Sall, B. & Bouyer, J. Tsetse flies: their biology and control using area-wide integrated pest management approaches. *Journal of invertebrate pathology* **112,** S15–S25 (2013) (cit. on p. 20).

135. Rayaisse, J. B., Esterhuizen, J., Tirados, I., *et al.* Towards an optimal design of target for tsetse control: comparisons of novel targets for the control of Palpalis group tsetse in West Africa. *PLoS Neglected Tropical Diseases* **5,** e1332 (2011) (cit. on p. 20).

136. Esterhuizen, J., Njiru, B., Vale, G. A., Lehane, M. J. & Torr, S. J. Vegetation and the importance of insecticide-treated target siting for control of Glossina fuscipes fuscipes. *PLoS neglected tropical diseases* **5** (2011) (cit. on p. 20).

137. Esterhuizen, J., Rayaisse, J. B., Tirados, I., *et al.* Improving the cost-effectiveness of visual devices for the control of riverine tsetse flies, the major vectors of human African trypanosomiasis. *PLoS neglected tropical diseases* **5** (2011) (cit. on p. 20).

138. Tirados, I., Esterhuizen, J., Rayaisse, J. B., *et al.* How do tsetse recognise their hosts? The role of shape in the responses of tsetse (Glossina fuscipes and G. palpalis) to artificial hosts. *PLoS neglected tropical diseases* **5** (2011) (cit. on p. 20).

139. Lindh, J. M., Torr, S. J., Vale, G. A. & Lehane, M. J. Improving the cost-effectiveness of artificial visual baits for controlling the tsetse fly Glossina fuscipes fuscipes. *PLoS Neglected Tropical Diseases* **3,** e474 (2009) (cit. on p. 20).

140. Shaw, A., Tirados, I., CTN, M., *et al.* Costs Of Using "Tiny Targets" to Control Glossina fuscipes fuscipes, a Vector of Gambiense Sleeping Sickness in Arua District of Uganda. *PLoS Neglected Tropical Diseases* **9,** e0003624 (2015) (cit. on p. 20).

141. Rayaisse, J., Tirados, I., Kaba, D., *et al.* Prospects for the development of odour baits to control the tsetse flies Glossina tachinoides and G. palpalis sl. *PLoS neglected tropical diseases* **4** (2010) (cit. on p. 20).

142. Sutherland, C. S., Stone, C. M., Steinmann, P., Tanner, M. & Tediosi, F. Seeing beyond 2020: an economic evaluation of contemporary and emerging strategies for elimination of Trypanosoma brucei gambiense. *The Lancet Global Health* **5,** e69–e79 (2017) (cit. on pp. 20, 74, 85, 87).

143. Bouyer, J., Carter, N. H., Batavia, C. & Nelson, M. P. The ethics of eliminating harmful species: The case of the tsetse fly. *Bioscience* **69,** 125–135 (2019) (cit. on p. 20).

144. Tirados, I., Esterhuizen, J., Kovacic, V., *et al.* Tsetse control and Gambian sleeping sickness; implications for control strategy. *PLoS neglected tropical diseases* **9,** e0003822 (2015) (cit. on p. 20).

145. Courtin, F., Camara, M., Rayaisse, J.-B., *et al.* Reducing human-tsetse contact significantly enhances the efficacy of sleeping sickness active screening campaigns: a promising result in the context of elimination. *PLoS neglected tropical diseases* **9** (2015) (cit. on p. 20).

146. Mahamat, M. H., Peka, M., Rayaisse, J.-b., *et al.* Adding tsetse control to medical activities contributes to decreasing transmission of sleeping sickness in the Mandoul focus (Chad). *PLoS Neglected Tropical Diseases* **11,** e0005792 (2017) (cit. on pp. 20, 32, 34, 43, 74, 151).

147. Selby, R., Wamboga, C., Erpha, s. O., *et al.* Gambian human African trypanosomiasis in North West Uganda. Are we on course for the 2020 target? *PLoS neglected tropical diseases* **13,** e0007550 (2019) (cit. on p. 20).

148. Camara, M., Ouattara, E., Duvignaud, A., *et al.* Impact of the Ebola outbreak on Trypanosoma brucei gambiense infection medical activities in coastal Guinea, 2014-2015: A retrospective analysis from the Guinean national Human African Trypanosomiasis control program. *PLoS neglected tropical diseases* **11,** e0006060 (2017) (cit. on p. 21).

149. Miaka, E. M., Hasker, E., Verlé, P., Torr, S. J. & Boelaert, M. Sleeping sickness in the Democratic Republic of the Congo. *The Lancet Neurology* **18,** 988–989 (2019) (cit. on pp. 21, 23).

150. Lejon, V., Hasker, E. & Büscher, P. Rapid Diagnostic Tests for Human African Trypanosomiasis. *Revolutionizing Tropical Medicine: Point-of-Care Tests, New Imaging Technologies and Digital Health,* 159–169 (2019) (cit. on p. 21).

151. Magnus, E., Van Meirvenne, N., Vervoort, T., Le Ray, D., Wéry, M., *et al.* Use of freeze-dried trypanosomes in the indirect fluorescent antibody test for the serodiagnosis of sleeping sickness. *Ann Soc Belg Méd Trop* **58,** 103–109 (1978) (cit. on p. 21).

152. Jamonneau, V. & Bucheton, B. The challenge of serodiagnosis of sleeping sickness in the context of elimination. *The Lancet Global Health* **2,** e306–e307 (2014) (cit. on p. 21).

153. Kagbadouno, M. S., Camara, M., Rouamba, J., *et al.* Epidemiology of sleeping sickness in Boffa (Guinea): where are the trypanosomes? *PLoS neglected tropical diseases* **6,** e1949 (2012) (cit. on p. 21).

154.Bessell, P. R., Lumbala, C., Lutumba, P., *et al.* Cost-effectiveness of using a rapid diagnostic test to screen for human African trypanosomiasis in the Democratic Republic of the Congo. *PloS one* **13** (2018) (cit. on pp. 21, 73, 74, 87).

155.Lee, S. J. & Palmer, J. J. Integrating innovations: a qualitative analysis of referral non-completion among rapid diagnostic test-positive patients in Uganda's human African trypanosomiasis elimination programme. *Infectious diseases of poverty* **7,** 1–16 (2018) (cit. on pp. 21, 22).

156.Wamboga, C., Matovu, E., Bessell, P. R., *et al.* Enhanced passive screening and diagnosis for gambiense human African trypanosomiasis in north-western Uganda–moving towards elimination. *PLoS One* **12** (2017) (cit. on pp. 21, 73, 149).

157.Büscher, P., Gilleman, Q. & Lejon, V. Rapid diagnostic test for sleeping sickness. *New England Journal of Medicine* **368,** 1069–1070 (2013) (cit. on p. 21).

158.Büscher, P., Mertens, P., Leclipteux, T., *et al.* Sensitivity and specificity of HAT Sero-K-SeT, a rapid diagnostic test for serodiagnosis of sleeping sickness caused by Trypanosoma brucei gambiense: a case-control study. *The Lancet Global Health* **2,** e359–e363 (2014) (cit. on p. 21).

159.Boelaert, M., Mukendi, D., Bottieau, E., *et al.* A phase III diagnostic accuracy study of a rapid diagnostic test for diagnosis of second-stage human African trypanosomiasis in the Democratic Republic of the Congo. *EBioMedicine* **27,** 11–17 (2018) (cit. on p. 21).

160.Bisser, S., Lumbala, C., Nguertoum, E., *et al.* Sensitivity and specificity of a prototype rapid diagnostic test for the detection of Trypanosoma brucei gambiense infection: a multi-centric prospective study. *PLoS neglected tropical diseases* **10** (2016) (cit. on p. 21).

161.Lumbala, C., Bessell, P. R., Lutumba, P., *et al.* 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** (2017) (cit. on p. 21).

162.Lumbala, C., Biéler, S., Kayembe, S., *et al.* Prospective evaluation of a rapid diagnostic test for Trypanosoma brucei gambiense infection developed using recombinant antigens. *PLoS neglected tropical diseases* **12,** e0006386 (2018) (cit. on p. 21).

163.Chappuis, F., Loutan, L., Simarro, P., Lejon, V. & Büscher, P. Options for field diagnosis of human African trypanosomiasis. *Clinical microbiology reviews* **18,** 133–146 (2005) (cit. on p. 21).

164.Jamonneau, V., Bucheton, B., Kaboré, J., *et al.* Revisiting the immune trypanolysis test to optimise epidemiological surveillance and control of sleeping sickness in West Africa. *PLoS neglected tropical diseases* **4** (2010) (cit. on p. 22).

165.Dama, E., Camara, O., Kaba, D., *et al.* Immune trypanolysis test as a promising bioassay to monitor the elimination of gambiense human African trypanosomiasis. *Parasite* **26** (2019) (cit. on p. 22).

166.Mitashi, P., Hasker, E., Lejon, V., *et al.* Human African trypanosomiasis diagnosis in first-line health services of endemic countries, a systematic review. *PLoS neglected tropical diseases* **6** (2012) (cit. on p. 22).

167.Büscher, P. & Deborggraeve, S. How can molecular diagnostics contribute to the elimination of human African trypanosomiasis? *Expert review of molecular diagnostics* **15,** 607–615 (2015) (cit. on p. 22).

168. Deborggraeve, S. & Büscher, P. Molecular diagnostics for sleeping sickness: what is the benefit for the patient? *The Lancet infectious diseases* **10,** 433–439 (2010) (cit. on p. 22).

169. Kennedy, P. G. Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *The Lancet Neurology* **12,** 186–194 (2013) (cit. on p. 22).

170. Mumba Ngoyi, D., Menten, J., Pyana, P. P., Büscher, P. & Lejon, V. Stage determination in sleeping sickness: comparison of two cell counting and two parasite detection techniques. *Tropical Medicine & International Health* **18,** 778–782 (2013) (cit. on p. 22).

171. Simarro, P., Franco, J., Diarra, A., Postigo, J. R. & Jannin, J. Update on field use of the available drugs for the chemotherapy of human African trypanosomiasis. *Parasitology* **139,** 842–846 (2012) (cit. on p. 22).

172. Priotto, G., Kasparian, S., Mutombo, W., *et al.* Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. *The Lancet* **374,** 56–64 (2009) (cit. on p. 23).

173. Yun, O., Priotto, G., Tong, J., Flevaud, L. & Chappuis, F. NECT is next: implementing the new drug combination therapy for Trypanosoma brucei gambiense sleeping sickness. *PLoS neglected tropical diseases* **4** (2010) (cit. on p. 23).

174. Kuepfer, I., Schmid, C., Allan, M., *et al.* Safety and efficacy of the 10-day melarsoprol schedule for the treatment of second stage Rhodesiense sleeping sickness. *PLoS neglected tropical diseases* **6** (2012) (cit. on p. 23).

175. Tarral, A., Blesson, S., Mordt, O. V., *et al.* Determination of an optimal dosing regimen for fexinidazole, a novel oral drug for the treatment of human African trypanosomiasis: first-in-human studies. *Clinical pharmacokinetics* **53,** 565–580 (2014) (cit. on p. 23).

176. Organization, W. H. *WHO interim guidelines for the treatment of gambiense human African trypanosomiasis* tech. rep. (World Health Organization, 2019) (cit. on p. 23).

177. Lindner, A. K., Lejon, V., Chappuis, F., *et al.* New WHO guidelines for treatment of gambiense human African trypanosomiasis including fexinidazole: substantial changes for clinical practice. *The Lancet Infectious Diseases* (2019) (cit. on p. 23).

178. Mesu, V. K. B. K., Kalonji, W. M., Bardonneau, C., *et al.* Oral fexinidazole for late-stage African Trypanosoma brucei gambiense trypanosomiasis: a pivotal multicentre, randomised, non-inferiority trial. *The Lancet* **391,** 144–154 (2018) (cit. on pp. 23, 87).

179. Castaño, M. S., Ndeffo-Mbah, M. L., Rock, K. S., *et al.* Assessing the impact of aggregating disease stage data in model predictions of human African trypanosomiasis transmission and control activities in Bandundu province (DRC). *PLOS Neglected Tropical Diseases* **14,** e0007976 (2020) (cit. on pp. 23, 32, 74, 151).

180. Burton, A. Sleeping sickness in West and Central Africa: is eradication just skin deep? *The Lancet. Neurology* **18,** 332 (2019) (cit. on pp. 23, 24).

181. Jacobs, R. T., Nare, B., Wring, S. A., *et al.* SCYX-7158, an orally-active benzoxaborole for the treatment of stage 2 human African trypanosomiasis. *PLoS neglected tropical diseases* **5** (2011) (cit. on p. 23).

182. Sutherland, C. S. & Tediosi, F. Is the elimination of 'sleeping sickness' affordable? Who will pay the price? Assessing the financial burden for the elimination of human African trypanosomiasis Trypanosoma brucei gambiense in sub-Saharan Africa. *BMJ global health* **4,** e001173 (2019) (cit. on pp. 24, 25, 74).

183. Kim, Y. E., Sicuri, E. & Tediosi, F. Financial and economic costs of the elimination and eradication of onchocerciasis (River Blindness) in Africa. *PLoS neglected tropical diseases* **9** (2015) (cit. on p. 24).

184. Büscher, P., Bart, J.-M., Boelaert, M., *et al.* Do cryptic reservoirs threaten gambiense-sleeping sickness elimination? *Trends in parasitology* **34,** 197–207 (2018) (cit. on pp. 24–26, 32, 71).

185. Viana, M., Mancy, R., Biek, R., *et al.* Assembling evidence for identifying reservoirs of infection. *Trends in ecology & evolution* **29,** 270–279 (2014) (cit. on p. 24).

186. Pandey, A., Atkins, K. E., Bucheton, B., *et al.* Evaluating long-term effectiveness of sleeping sickness control measures in Guinea. *Parasites & vectors* **8,** 550 (2015) (cit. on pp. 24, 32–34, 38, 43, 74, 151).

187. Capewell, P., Cren-Travaillé, C., Marchesi, F., *et al.* The skin is a significant but overlooked anatomical reservoir for vector-borne African trypanosomes. *Elife* **5,** e17716 (2016) (cit. on p. 24).

188. Caljon, G., Van Reet, N., De Trez, C., *et al.* The dermis as a delivery site of Trypanosoma brucei for tsetse flies. *PLoS pathogens* **12** (2016) (cit. on p. 24).

189. Capewell, P., Atkins, K., Weir, W., *et al.* Resolving the apparent transmission paradox of African sleeping sickness. *PLoS biology* **17,** e3000105 (2019) (cit. on p. 25).

190. Pandey, A. & Galvani, A. Strategies for Trypanosoma brucei gambiense elimination. *The Lancet Global Health* **5,** e10–e11 (2017) (cit. on pp. 25, 151).

191. Coffeng, L., Le Rutte, E. A., Muñoz, J., *et al.* Impact of Changes in Detection Effort on Control of Visceral Leishmaniasis in the Indian Subcontinent. *Journal of Infectious Disease* (2019) (cit. on p. 25).

192. NTD Modelling Consortium Discussion Group on Gambiense Human African Trypanosomiasis. Insights from quantitative and mathematical modelling on the proposed 2030 goal for gambiense human African trypanosomiasis (gHAT) [version 1; peer review: 1 approved, 1 approved with reservations]. *Gates Open Research* **3** (2019) (cit. on p. 25).

193. Picado, A. & Ndung'u, J. Elimination of sleeping sickness in Uganda could be jeopardised by conflict in South Sudan. *The Lancet Global Health* **5,** e28–e29 (2017) (cit. on p. 25).

194. Cecchi, G., Paone, M., Feldmann, U., *et al.* Assembling a geospatial database of tsetse-transmitted animal trypanosomosis for Africa. *Parasites & vectors* **7,** 39 (2014) (cit. on p. 25).

195. N'Djetchi, M. K., Ilboudo, H., Koffi, M., *et al.* The study of trypanosome species circulating in domestic animals in two human African trypanosomiasis foci of Cote d'Ivoire identifies pigs and cattle as potential reservoirs of Trypanosoma brucei gambiense. *PLoS neglected tropical diseases* **11,** e0005993 (2017) (cit. on p. 25).

196. Simarro, P., Franco, J., Ndongo, P., *et al.* The elimination of Trypanosoma brucei gambiense sleeping sickness in the focus of Luba, Bioko Island, Equatorial Guinea. *Tropical Medicine & International Health* **11,** 636–646 (2006) (cit. on p. 25).

197. Cordon-Obras, C., Rodriguez, Y., Fernandez-Martinez, A., *et al.* Molecular evidence of a Trypanosoma brucei gambiense sylvatic cycle in the human african trypanosomiasis foci of Equatorial Guinea. *Frontiers in microbiology* **6,** 765 (2015) (cit. on p. 25).

198. Holmes, P. On the road to elimination of rhodesiense human African trypanosomiasis: first WHO meeting of stakeholders. *PLoS neglected tropical diseases* **9** (2015) (cit. on p. 25).

199. Lenk, E., Redekop, W., Luyendijk, M., *et al.* Socioeconomic benefit to individuals of achieving 2020 targets for four neglected tropical diseases controlled/eliminated by innovative and intensified disease management: Human African trypanosomiasis, leprosy, visceral leishmaniasis, Chagas disease. *PLoS neglected tropical diseases* **12,** e0006250 (2018) (cit. on p. 25).

200. Camara, M., Ouattara, E., Duvignaud, A., *et al.* Impact of the Ebola outbreak on Trypanosoma brucei gambiense infection medical activities in coastal Guinea, 2014-2015: A retrospective analysis from the Guinean national Human African Trypanosomiasis control program. *PLoS neglected tropical diseases* **11,** e0006060 (2017) (cit. on p. 25).

201. Hollingsworth, T. D., Adams, E. R., Anderson, R. M., *et al.* Quantitative analyses and modelling to support achievement of the 2020 goals for nine neglected tropical diseases. *Parasites & Vectors* **8,** 630 (2015) (cit. on pp. 26, 38, 106).

202. Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control* (Oxford university press, 1992) (cit. on pp. 27, 31).

203. Diekmann, O. & Heesterbeek, J. A. P. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation* (John Wiley & Sons, 2000) (cit. on p. 27).

204. Andersson, H. & Britton, T. *Stochastic epidemic models and their statistical analysis* (Springer Science & Business Media, 2012) (cit. on p. 27).

205. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry* **81,** 2340–2361 (1977) (cit. on pp. 29, 52).

206. Ross, R. & Hudson, H. P. An application of the theory of probabilities to the study of a priori pathometry.—Part II. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **93,** 212–225 (1917) (cit. on p. 31).

207. Macdonald, G. *et al.* The analysis of infection rates in diseases in which super infection occurs. *Tropical diseases bulletin* **47,** 907–915 (1950) (cit. on p. 31).

208. MACDONALD, G. *et al.* The analysis of the sporozoite rate. *Tropical diseases bulletin* **49** (1952) (cit. on p. 31).

209. Alderton, S., Macleod, E. T., Anderson, N. E., *et al.* Exploring the effect of human and animal population growth on vector-borne disease transmission with an agent-based model of Rhodesian human African trypanosomiasis in eastern province, Zambia. *PLoS neglected tropical diseases* **12** (2018) (cit. on p. 31).

210. Styer, L. M., Minnick, S. L., Sun, A. K. & Scott, T. W. Mortality and reproductive dynamics of Aedes aegypti (Diptera: Culicidae) fed human blood. *Vector-borne and zoonotic diseases* **7,** 86–98 (2007) (cit. on pp. 31, 113).

211. Bellan, S. E. The importance of age dependent mortality and the extrinsic incubation period in models of mosquito-borne disease transmission and control. *PLoS One* **5** (2010) (cit. on pp. 31, 113).

212. Kilpatrick, A. M. Facilitating the evolution of resistance to avian malaria in Hawaiian birds. *Biological Conservation* **128,** 475–485 (2006) (cit. on p. 31).

213. Aron, J. L. & May, R. M. in *The population dynamics of infectious diseases: theory and applications* 139–179 (Springer, 1982) (cit. on p. 31).

214. Rogers, D. A general model for the African trypanosomiases. *Parasitology* **97,** 193–212 (1988) (cit. on pp. 31, 36, 49, 80, 151).

215. Milligan, P. & Baker, R. A model of tsetse-transmitted animal trypanosomiasis. *Parasitology* **96,** 211–239 (1988) (cit. on pp. 31, 151).

216. Artzrouni, M. & Gouteux, J.-P. Control strategies for sleeping sickness in Central Africa: a model-based approach. *Tropical Medicine & International Health* **1,** 753–764 (1996) (cit. on p. 31).

217. Baker, R., Maudlin, I., Milligan, P., Molyneux, D. & Welburn, S. The possible role of Rickettsia-like organisms in trypanosomiasis epidemiology. *Parasitology* **100,** 209–217 (1990) (cit. on p. 31).

218. Funk, S., Nishiura, H., Heesterbeek, H., Edmunds, W. J. & Checchi, F. Identifying transmission cycles at the human-animal interface: the role of animal reservoirs in maintaining gambiense human african trypanosomiasis. *PLoS computational biology* **9** (2013) (cit. on pp. 32, 43, 151).

219. Rock, K. S., Torr, S. J., Lumbala, C. & Keeling, M. J. Quantitative evaluation of the strategy to eliminate human African trypanosomiasis in the Democratic Republic of Congo. *Parasites & vectors* **8,** 532 (2015) (cit. on pp. 32, 34, 38, 42–47, 49, 50, 63, 65, 70, 74, 75, 77, 86, 114, 149).

220. Stone, C. M. & Chitnis, N. Implications of heterogeneous biting exposure and animal hosts on Trypanosomiasis brucei gambiense transmission and control. *PLoS computational biology* **11,** e1004514 (2015) (cit. on pp. 32, 33, 38, 45, 70, 74, 151).

221. Rock, K. S., Ndeffo-Mbah, M. L., Castaño, S., *et al.* Assessing strategies against Gambiense sleeping sickness through mathematical modeling. *Clinical infectious diseases* **66,** S286–S292 (2018) (cit. on pp. 32, 35, 151).

222. Meisner, J., Barnabas, R. V. & Rabinowitz, P. M. A mathematical model for evaluating the role of trypanocide treatment of cattle in the epidemiology and control of Trypanosoma brucei rhodesiense and T. b. gambiense sleeping sickness in Uganda. *Parasite epidemiology and control* **5,** e00106 (2019) (cit. on p. 32).

223. Rock, K. S., Torr, S. J., Lumbala, C. & Keeling, M. J. Predicting the impact of intervention strategies for sleeping sickness in two high-endemicity health zones of the Democratic Republic of Congo. *PLoS neglected tropical diseases* **11,** e0005162 (2017) (cit. on pp. 32, 34, 70, 74, 77, 80, 151).

224. Rock, K., Pandey, A., Ndeffo-Mbah, M., *et al.* Data-driven models to predict the elimination of sleeping sickness in former Equateur province of DRC. *Epidemics* **18,** 101–112 (2017) (cit. on pp. 32, 34, 70, 74, 77, 151).

225. Castaño, M. S., Aliee, M., Mwamba Miaka, E., *et al.* Screening Strategies for a Sustainable Endpoint for Gambiense Sleeping Sickness. *The Journal of Infectious Diseases* (2019) (cit. on pp. 32, 151, 152).

226. Ndeffo-Mbah, M. L., Pandey, A., Atkins, K. E., Aksoy, S. & Galvani, A. P. The impact of vector migration on the effectiveness of strategies to control gambiense human African trypanosomiasis. *PLoS Neglected Tropical Diseases* **13** (2019) (cit. on pp. 32, 34, 151).

227. Grébaut, P., Girardin, K., Fédérico, V. & Bousquet, F. Simulating the elimination of sleeping sickness with an agent-based model. *Parasite* **23** (2016) (cit. on p. 35).

228. Alderton, S., Macleod, E. T., Anderson, N. E., *et al.* An agent-based model of tsetse fly response to seasonal climatic drivers: Assessing the impact on sleeping sickness transmission rates. *PLoS neglected tropical diseases* **12,** e0006188 (2018) (cit. on p. 35).

229. De Vries, H., Wagelmans, A. P., Hasker, E., *et al.* Forecasting human African trypanosomiasis prevalences from population screening data using continuous time models. *PLoS computational biology* **12,** e1005103 (2016) (cit. on pp. 35, 38).

230. Hargrove, J. W. *et al.* Tsetse population dynamics. *The trypanosomiases* **2004,** 139–180 (2004) (cit. on pp. 35, 112).

231. Hargrove, J., Ouifki, R. & Ameh, J. A general model for mortality in adult tsetse (Glossina spp.) *Medical and veterinary entomology* **25,** 385–394 (2011) (cit. on p. 35).

232. Vale, G. A., Hargrove, J. W., Lehane, M. J., Solano, P. & Torr, S. J. Optimal strategies for controlling riverine tsetse flies using targets: a modelling study. *PLoS neglected tropical diseases* **9,** e0003615 (2015) (cit. on p. 35).

233. Torr, S. J. & Vale, G. A. Is the even distribution of insecticide-treated cattle essential for tsetse control? Modelling the impact of baits in heterogeneous environments. *PLoS neglected tropical diseases* **5,** e1360 (2011) (cit. on p. 35).

234. Hargrove, J. W., Ouifki, R., Kajunguri, D., Vale, G. A. & Torr, S. J. Modeling the control of trypanosomiasis using trypanocides or insecticide-treated livestock. *PLoS neglected tropical diseases* **6,** e1615 (2012) (cit. on p. 35).

235. Lord, J. S., Mthombothi, Z., Lagat, V. K., Atuhaire, F. & Hargrove, J. W. Host-seeking efficiency can explain population dynamics of the tsetse fly Glossina morsitans morsitans in response to host density decline. *PLoS neglected tropical diseases* **11,** e0005730 (2017) (cit. on p. 35).

236. Stanton, M. C., Esterhuizen, J., Tirados, I., Betts, H. & Torr, S. J. The development of high resolution maps of tsetse abundance to guide interventions against human African trypanosomiasis in northern Uganda. *Parasites & Vectors* **11,** 340 (2018) (cit. on p. 35).

237. Courtin, F., Jamonneau, V., Oké, E., *et al.* Towards understanding the presence/absence of Human African Trypanosomosis in a focus of Côte d'Ivoire: a spatial analysis of the pathogenic system. *International Journal of Health Geographics* **4,** 27 (2005) (cit. on p. 35).

238. Lord, J. S., Hargrove, J. W., Torr, S. J. & Vale, G. A. Climate change and African trypanosomiasis vector populations in Zimbabwe's Zambezi Valley: a mathematical modelling study. *PLoS medicine* **15,** e1002675 (2018) (cit. on p. 35).

239.Moore, S., Shrestha, S., Tomlinson, K. W. & Vuong, H. Predicting the effect of climate change on African trypanosomiasis: integrating epidemiology with parasite and vector biology. *Journal of the Royal Society Interface* **9,** 817–830 (2011) (cit. on p. 35).

240.WHO. Control and surveillance of human African trypanosomiasis. *World Health Organization technical report series* (2013) (cit. on pp. 36, 49, 50, 72).

241.Davis, S., Aksoy, S. & Galvani, A. A global sensitivity analysis for African sleeping sickness. *Parasitology* **138,** 516–526 (2011) (cit. on pp. 36, 49).

242.Ravel, S., Grébaut, P., Cuisance, D. & Cuny, G. Monitoring the developmental status of Trypanosoma brucei gambiense in the tsetse fly by means of PCR analysis of anal and saliva drops. *Acta tropica* **88,** 161–165 (2003) (cit. on pp. 36, 49, 80).

243.Clausen, P., Adeyemi, I., Bauer, B., *et al.* Host preferences of tsetse (Diptera: Glossinidae) based on bloodmeal identifications. *Medical and veterinary entomology* **12,** 169–180 (1998) (cit. on pp. 36, 49, 80).

244.Indicators, W. D. *World Bank* tech. rep. (2015) (cit. on pp. 36, 49, 87).

245.Checchi, F., Filipe, J. A., Haydon, D. T., Chandramohan, D. & Chappuis, F. Estimates of the duration of the early and late stage of gambiense sleeping sickness. *BMC infectious diseases* **8,** 16 (2008) (cit. on pp. 36, 61, 80).

246.Lumbala, C., Simarro, P. P., Cecchi, G., *et al.* Human African trypanosomiasis in the Democratic Republic of the Congo: disease distribution and risk. *International journal of health geographics* **14,** 20 (2015) (cit. on pp. 39, 73).

247.Ducomble, T. & Gignoux, E. Learning from a massive epidemic: measles in DRC. *The Lancet Infectious Diseases* **20,** 542 (2020) (cit. on p. 39).

248.UN. World statistics pocketbook (2016) (cit. on p. 40).

249.Gelman, A., Carlin, J., Stern, H. & Rubin, D. Texts in statistical science. *Bayesian data analysis* (2004) (cit. on p. 42).

250.Wearing, H. J., Rohani, P. & Keeling, M. J. Appropriate models for the management of infectious diseases. *PLoS medicine* **2,** e174 (2005) (cit. on pp. 45, 116, 129).

251.Checchi, F., Filipe, J. A., Barrett, M. P. & Chandramohan, D. The natural progression of Gambiense sleeping sickness: what is the evidence? *PLoS neglected tropical diseases* **2,** e303 (2008) (cit. on p. 49).

252.Checchi, F., Chappuis, F., Karunakara, U., Priotto, G. & Chandramohan, D. Accuracy of five algorithms to diagnose gambiense human African trypanosomiasis. *PLoS neglected tropical diseases* **5,** e1233 (2011) (cit. on pp. 50, 80, 121).

253.Cao, Y., Gillespie, D. T. & Petzold, L. R. Avoiding negative populations in explicit Poisson tau-leaping. *The Journal of chemical physics* **123,** 054104 (2005) (cit. on p. 52).

254.Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics* **115,** 1716–1733 (2001) (cit. on p. 52).

255.Bartlett, M. S. Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (General)* **120,** 48–70 (1957) (cit. on pp. 60, 61, 70, 71).

256.Keeling, M. J. & Grenfell, B. T. Disease extinction and community size: modeling the persistence of measles. *Science* **275,** 65–67 (1997) (cit. on pp. 60, 61, 70, 71).

257. Hagenaars, T., Donnelly, C. & Ferguson, N. Spatial heterogeneity and the persistence of infectious diseases. *Journal of theoretical biology* **229,** 349–359 (2004) (cit. on pp. 60, 70).

258. Brooks-Pollock, E. & Keeling, M. Herd size and bovine tuberculosis persistence in cattle farms in Great Britain. *Preventive veterinary medicine* **92,** 360–365 (2009) (cit. on pp. 60, 70).

259. Grenfell, B. & Harwood, J. (Meta) population dynamics of infectious diseases. *Trends in ecology & evolution* **12,** 395–399 (1997) (cit. on p. 67).

260. Van Herwaarden, O. A. & Grasman, J. Stochastic epidemics: major outbreaks and the duration of the endemic period. *Journal of mathematical biology* **33,** 581–601 (1995) (cit. on p. 70).

261. Grasman, J. Stochastic epidemics: the expected duration of the endemic period in higher dimensional models. *Mathematical biosciences* **152,** 13–27 (1998) (cit. on p. 70).

262. Institute for Health Metrics and Evaluation. Global Burden of Disease (GBD) Compare (2019) (cit. on p. 74).

263. Lutumba, P., Meheus, F., Robays, J., *et al.* Cost-effectiveness of algorithms for confirmation test of human African trypanosomiasis. *Emerging infectious diseases* **13,** 1484 (2007) (cit. on pp. 74, 87).

264. Keating, J., Yukich, J. O., Sutherland, C. S., Woods, G. & Tediosi, F. Human African trypanosomiasis prevention, treatment and control costs: a systematic review. *Acta tropica* **150,** 4–13 (2015) (cit. on pp. 74, 87).

265. Snijders, R., Fukinsia, A., Mulenga, P., *et al.* Evaluation and cost of an integrated sleeping screening system at district level in Democratic Republic of Congo. *Submitted, EID* (2020) (cit. on pp. 74, 87).

266. Snijders, R., Fukinsia, A., Claeys, Y., *et al.* Cost of a new method of active screening for human African trypanosomiasis in the Democratic Republic of the Congo. *Submitted, PLOS NTD* (2020) (cit. on pp. 74, 87).

267. World Health Organization. Control and surveillance of human African trypanosomiasis (2013) (cit. on pp. 76, 80, 81, 108, 154).

268. Crump, R., Huang, C., Knock, E., *et al.* Quantifying epidemiological drivers of gambiense human African Trypanosomiasis across the Democratic Republic of Congo. *In prep* (2020) (cit. on pp. 76, 79, 80, 86, 149).

269. The World Bank. Data: Democratic Republic of Congo (2015) (cit. on p. 80).

270. Davis, S., Aksoy, S. & Galvani, A. A global sensitivity analysis for African sleeping sickness. *Parasitology* **138,** 516–526 (2011) (cit. on p. 80).

271. Steinmann, P., Stone, C. M., Sutherland, C. S., Tanner, M. & Tediosi, F. Contemporary and emerging strategies for eliminating human African trypanosomiasis due to Trypanosoma brucei gambiense: Review. **20,** 707–718 (2015) (cit. on p. 82).

272. Lea, R. A. World Development Report 1993: 'Investing in Health'. *Forum for Development Studies* **20,** 114–117 (1993) (cit. on p. 82).

273. Murray, C. Quantifying the burden of disease : the technical basis for disability-adjusted life years. *Bulletin of the World Health Organization* **72,** 429–445 (1994) (cit. on p. 82).

274. Hutubessy, R., Chisholm, D. & Edejer, T. T.-T. Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. *Cost effectiveness and resource allocation* **1,** 8 (2003) (cit. on pp. 82, 84).

275. Leech, A. A., Kim, D. D., Cohen, J. T. & Neumann, P. J. Use and misuse of cost-effectiveness analysis thresholds in low-and middle-income countries: trends in cost-per-DALY studies. *Value in Health* **21,** 759–761 (2018) (cit. on p. 83).

276. Ochalek, J. M., Lomas, J. & Claxton, K. P. Cost per DALY averted thresholds for low-and middle-income countries: evidence from cross country data, 1–50 (2015) (cit. on pp. 83, 108).

277. Woods, B., Revill, P., Sculpher, M. & Claxton, K. Country-level cost-effectiveness thresholds: Initial estimates and the need for further research. *Value in Health* **19,** 929–935 (2016) (cit. on pp. 83, 108).

278. Baltussen, R. M., Adam, T., Tan-Torres Edejer, T., *et al.* Making choices in health: WHO guide to cost-effectiveness analysis (2003) (cit. on pp. 85, 86, 89).

279. World Health Organization. WHO interim guidelines for the treatment of gambiense human African trypanosomiasis (2019) (cit. on p. 85).

280. Wall, R. J., Rico, E., Lukac, I., *et al.* Clinical and veterinary trypanocidal benzoxaboroles target CPSF3. *Proceedings of the National Academy of Sciences of the United States of America* **115,** 9616–9621 (2018) (cit. on p. 85).

281. Lutumba, P., Makieya, E., Shaw, A., Meheus, F. & Boelaert, M. Human African trypanosomiasis in a rural community, Democratic Republic of Congo. *Emerging infectious diseases* **13,** 248 (2007) (cit. on p. 87).

282. Kyu, H. H., Abate, D., Abate, K. H., *et al.* Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392,** 1859–1922 (2018) (cit. on p. 87).

283. Irurzun-Lopez, M., Erondu, N. A., Djibo, A., *et al.* The actual and potential costs of meningitis surveillance in the African meningitis belt: Results from Chad and Niger. *Vaccine* **34,** 1133–1138 (2016) (cit. on p. 87).

284. Bertram, M. Y., Stenberg, K., Brindley, C., *et al.* Disease control programme support costs: an update of WHO-CHOICE methodology, price databases and quantity assumptions. *Cost Effectiveness and Resource Allocation* **15,** 21 (2017) (cit. on p. 87).

285. Stenberg, K., Lauer, J. A., Gkountouras, G., Fitzpatrick, C. & Stanciole, A. Econometric estimation of WHO-CHOICE country-specific costs for inpatient and outpatient health service delivery. *Cost Effectiveness and Resource Allocation* **16,** 11 (2018) (cit. on p. 87).

286. Shaw, A. & Cattand, P. Analytical tools for planning cost-effective surveillance in Gambiense sleeping sickness. *Medecine tropicale: revue du Corps de sante colonial* **61,** 412–421 (2001) (cit. on p. 87).

287. The World Bank. Data: GDP per capita (current US$)Congo, Dem. Rep. (2018) (cit. on p. 89).

288. Welburn, S., Maudlin, I. & Simarro, P. Controlling sleeping sickness–a review. *Parasitology* **136,** 1943–1949 (2009) (cit. on p. 96).

289. Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. shiny: Web Application Framework for R v1.4.0 (2019) (cit. on p. 104).

290. Lumbala, C., Simarro, P. P., Cecchi, G., *et al.* Human African trypanosomiasis in the Democratic Republic of the Congo: disease distribution and risk. *International journal of health geographics* **14,** 20 (2015) (cit. on p. 109).

291. Keeling, M. J. & Ross, J. V. On methods for studying stochastic disease dynamics. *Journal of the Royal Society Interface* **5,** 171–181 (2007) (cit. on pp. 111, 119, 148).

292. Gardiner, C. W. *et al. Handbook of stochastic methods* (springer Berlin, 1985) (cit. on p. 111).

293. Keeling, M. & Ross, J. Efficient methods for studying stochastic disease and population dynamics. *Theoretical population biology* **75,** 133–141 (2009) (cit. on p. 111).

294. Albert, J. A hybrid of the chemical master equation and the Gillespie algorithm for efficient stochastic simulations of sub-networks. *PloS one* **11** (2016) (cit. on p. 112).

295. Feller, W. *An introduction to probability theory and its applications* (John Wiley & Sons, 1950) (cit. on p. 112).

296. Dietz, K., Molineaux, L. & Thomas, A. A malaria model tested in the African savannah. *Bulletin of the World Health Organization* **50,** 347 (1974) (cit. on p. 113).

297. Rocha, F., Aguiar, M., Souza, M. & Stollenwerk, N. Time-scale separation and centre manifold analysis describing vector-borne disease dynamics. *International Journal of Computer Mathematics* **90,** 2105–2125 (2013) (cit. on p. 113).

298. Smith, T., Maire, N., Dietz, K., *et al.* Relationship between the entomologic inoculation rate and the force of infection for Plasmodium falciparum malaria. *The American journal of tropical medicine and hygiene* **75,** 11–18 (2006) (cit. on p. 113).

299. Golub, G. H. & Van Loan, C. F. *Matrix computations* (JHU press, 2012) (cit. on p. 119).

300. Mangel, M. & Tier, C. Four facts every conservation biologists should know about persistence. *Ecology* **75,** 607–614 (1994) (cit. on p. 130).