PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Ledien J, Cucunubá ZM, Parra-Henao G, Rodríguez-Monguí E, Dobson AP, Adamo SB, Castellanos LG, Basáñez M-G, Nouvellet P. 2023 From serological surveys to disease burden: a modelling pipeline for Chagas disease. *Phil. Trans. R. Soc. B* **378**: 20220278. https://doi.org/10.1098/rstb.2022.0278

Received: 20 November 2022
Accepted: 29 June 2023

One contribution of 14 to a theme issue 'Challenges and opportunities in the fight against neglected tropical diseases: a decade from the London Declaration on NTDs'.

Subject Areas:
health and disease and epidemiology

Keywords:
Force-of-Infection, Random Forest, Chagas disease, infectious diseases, modelling, burden of disease

Authors for correspondence:
Julia Ledien
e-mail: j.ledien@sussex.ac.uk
Zulma M. Cucunubá
e-mail: zulma.cucunuba@javeriana.edu.co

†Joint first authors.
‡Joint last authors.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6751758.

THE ROYAL SOCIETY PUBLISHING

From serological surveys to disease burden: a modelling pipeline for Chagas disease

Julia Ledien[1,†], Zulma M. Cucunubá[2,†], Gabriel Parra-Henao[3,4], Eliana Rodríguez-Monguí[2,5], Andrew P. Dobson[6], Susana B. Adamo[7], Luis Gerardo Castellanos[8], María-Gloria Basáñez[9,‡] and Pierre Nouvellet[1,9,‡]

[1]School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9RH, UK
[2]Departamento de Epidemiología Clínica y Bioestadística, Facultad de Medicina, Universidad Pontificia Javeriana, 110231 Bogotá, Colombia
[3]Centro de Investigación en Salud para el Trópico, Universidad Cooperativa de Colombia, 470002, Santa Marta, Colombia
[4]National Institute of Health, 111321 Bogotá, Colombia
[5]Independent consultant to the Neglected, Tropical and Vector Borne Diseases Program, Pan American Health Organization (PAHO), Colombia
[6]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
[7]Center for International Earth Science Information Network (CIESIN), Columbia Climate School, Columbia University, New York, NY 10025, USA
[8]Department of Communicable Diseases and Environmental Determinants of Health, Pan American Health Organization (PAHO), Washington, DC 20037, USA
[9]London Centre for Neglected Tropical Disease Research (LCNTDR) & MRC Centre for Global Infectious Disease Analysis (GIDA), Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London SW7 2AZ, UK

ZMC, 0000-0002-8165-3198; SBA, 0000-0002-9168-7172; M-GB, 0000-0001-5031-3361; PN, 0000-0002-6094-5722

In 2012, the World Health Organization (WHO) set the elimination of Chagas disease intradomiciliary vectorial transmission as a goal by 2020. After a decade, some progress has been made, but the new 2021–2030 WHO roadmap has set even more ambitious targets. Innovative and robust modelling methods are required to monitor progress towards these goals. We present a modelling pipeline using local seroprevalence data to obtain national disease burden estimates by disease stage. Firstly, local seroprevalence information is used to estimate spatio-temporal trends in the Force-of-Infection (FoI). FoI estimates are then used to predict such trends across larger and fine-scale geographical areas. Finally, predicted FoI values are used to estimate disease burden based on a disease progression model. Using Colombia as a case study, we estimated that the number of infected people would reach 506 000 (95% credible interval (CrI) = 395 000–648 000) in 2020 with a 1.0% (95%CrI = 0.8–1.3%) prevalence in the general population and 2400 (95%CrI = 1900–3400) deaths (approx. 0.5% of those infected). The interplay between a decrease in infection exposure (FoI and relative proportion of acute cases) was overcompensated by a large increase in population size and gradual population ageing, leading to an increase in the absolute number of Chagas disease cases over time.

This article is part of the theme issue 'Challenges and opportunities in the fight against neglected tropical diseases: a decade from the London Declaration on NTDs'.

1. Introduction

Chagas disease is a neglected tropical disease (NTD) caused by the protozoan parasite *Trypanosoma cruzi*. Vectorial transmission (by reduviid, triatomine bugs) is the main, but not exclusive, transmission route. While Chagas disease is endemic in 21 Latin American countries, population migration has resulted in its globalization.

Infections can remain asymptomatic for many years, with 20–35% of those infected eventually developing clinical manifestations and requiring medical interventions [1]. Such interventions (including treatment) aim at alleviating symptoms and/or reducing disease progression when possible. Disease control efforts have mainly focused on infection prevention (e.g. through vector control, education and housing improvement) and testing for prompt identification of asymptomatic cases [2].

In 2012, the World Health Organization (WHO) set the elimination of intradomiciliary vectorial transmission in the Americas by 2020 as a goal in its first NTD roadmap [3]. After a decade, progress has been made, but the new 2021–2030 WHO roadmap on NTDs is even more ambitious, proposing that all routes of transmission be interrupted in nearly 40% of endemic countries by 2030 [4]. The application of innovative and robust statistical methods can help to monitor the epidemiological situation and the progress to be made to meet this challenge. To this end, estimating the spatio-temporal variations in disease exposure is critical, but this is hampered by weak surveillance [2]. For example, in Colombia in 2021, 306 chronic and 172 acute cases were reported, with only 170 and 14, respectively, of them being confirmed [5]. By contrast, estimations of the number of cases for the country from WHO, Global Burden Model (GBM) and others ranged between 186 000 and 438 000 for the 2005–2010 period [1,6–8].

Chagas disease is a long-lasting disease. Therefore, current prevalence of infection or disease does not truly reflect the current transmission trends. For instance, high prevalence potentially reflects a high level of past transmission rather than current exposure. Using mathematical modelling, seroprevalence studies can be used to reconstruct temporal trends in the Force-of-Infection (FoI, the per-susceptible rate of parasite acquisition). Therefore, age-stratified seroprevalence studies have the potential to provide a largely untapped resource to predict spatio-temporal trends in Chagas disease incidence, which can, in turn, be used to predict the burden of Chagas disease over time and space at a resolution much finer than that available from current national estimates.

Seroprevalence surveys have been used to estimate past trends in exposure in the context of Chagas disease [9], dengue [10–13], malaria [14], schistosomiasis [15] and yellow fever [16]. Provided enough surveys are available, predictive models can be used to estimate spatio-temporal trends in exposure for Chagas disease [17] and other infections [10,11,16,18]. The crucial next step is linking such trends with models of disease progression so robust estimates of disease burden can be obtained to better target the necessary interventions [7]. However, in some of the applications mentioned above, estimates of the FoI have been assumed to be constant over time [10,11,19], or only average FoI values have been used to fit predictive models [10,13,16,19], substantially neglecting the associated uncertainty. Therefore, appropriately propagating the uncertainty surrounding each step is essential for reliable estimation of disease burden.

In this paper, we present a modelling pipeline to estimate Chagas disease incidence and burden of disease. We collated information from 76 seroprevalence studies in Colombia, from published and unpublished sources between 1990 and 2020. Those studies were used to estimate local temporal trends in the FoI. Spatio-temporal predictive models were used to obtain FoI estimates over the last seven decades at the municipality level across Colombia. Finally, those estimates were used in an age-structured compartmental model linking infection to disease states to estimate the burden of Chagas disease and its spatial and temporal heterogeneities.

Our study highlights the benefit of using currently available but largely under-used seroprevalence studies to inform the burden of Chagas disease. Our modelling pipeline relies on robust statistical modelling that propagates the various uncertainties at each step, providing a more realistic assessment of the past and current epidemiological situation. We also discuss its applications for estimating disease burden across the remaining Chagas disease-endemic countries in the Americas.

2. Models and methods

(a) The DICTUM platform

With support from the Pan American Health Organization (PAHO, the Regional WHO office for the Americas), the 'Decreasing the Impact of Chagas Disease Through Modelling' (DICTUM) platform has been created to collate, standardize and communicate data relevant to Chagas disease epidemiology, including information on serosurveys, vector surveillance and blood-bank screening. A key aim is to use the DICTUM platform to inform public health professionals on crucial aspects of Chagas disease epidemiology; for instance, by obtaining estimates of the number of asymptomatic, chronic, and severe cases by age class, allowing targeting of diagnostics and treatment activities.

The process of estimating the burden of Chagas disease using local serosurveys (figure 1) involves three steps:

(1) Step 1: Local seroprevalence information is used to estimate local trends in temporal exposure (quantified by the FoI).
(2) Step 2: Exposure estimates from various surveys are used to predict spatio-temporal trends across larger geographical areas (using Random Forest (RF) models, as in [20]).
(3) Step 3: Predicted exposure estimates are used at a fine spatial resolution to predict disease burden based on a disease progression model.

Special attention is given to propagating uncertainty between steps and at different spatial and temporal scales. Electronic supplementary material, 'Models and methods', text S1–S3, provides a detailed account of Steps 1–3, with electronic supplementary material, text S2.2 describing the implementation of the RF models and the integration of uncertainty.

(b) Step 1: Force-of-Infection at the serosurvey level (yearly Force-of-Infection catalytic model)

Under the assumption that antibodies are life-long, age-specific seroprevalence is a reflection of the cumulative incidence, and thus, by using catalytic models we can disentangle the rate at which the population became infected, namely the Force-of-Infection (FoI) [8]. For these FoI models, we assumed (a) no age-dependency in transmission, (b) no seroreversion and (c) no specific migration due to Chagas disease infection status. We relied on 76 (unique) serological age-stratified local surveys (serosurveys) conducted

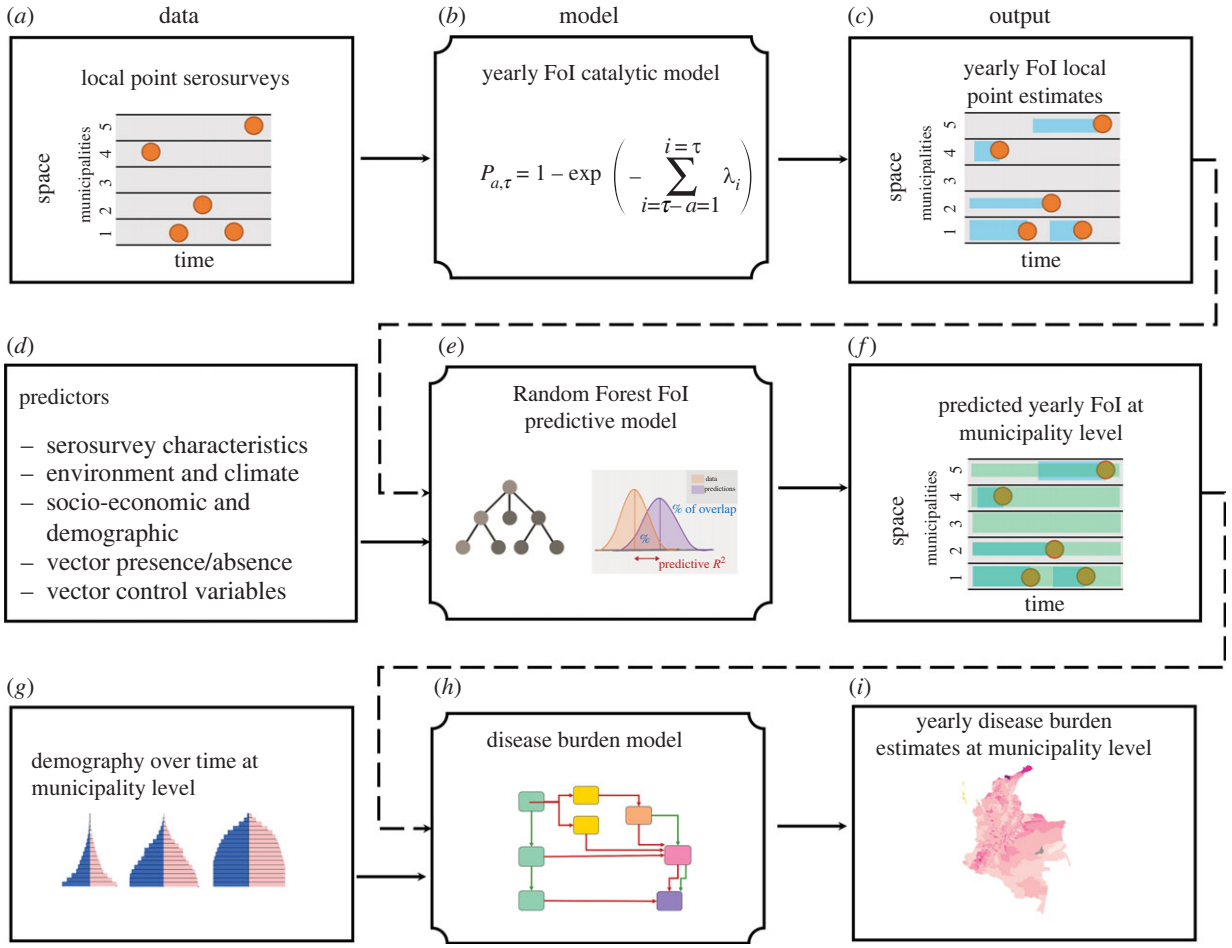

Figure 1. Modelling pipeline, from local serosurveys to sub-national yearly disease burden estimates. Using local point seroprevalence age-stratified data (*a*), the modelling pipeline uses a Force-of-Infection (FoI) catalytic model (*b*) to estimate yearly FoI local point estimates (*c*). Then, a set of covariate predictors (*d*) are used as input for the FoI Random Forest predictive model (*e*) to obtain the predicted yearly FoI values at municipality level (*f*). This information is then combined with detailed demographic information at municipality level over time (*g*) into a detailed disease burden and progression model (*h*) to generate, as final output, the prevalence by year and by age class of different stages of Chagas disease (including mortality) at the municipality level and burden of disease information across the country at the municipality level (*i*) while propagating uncertainty from one model to another.

in Colombia at the municipality level (electronic supplementary material, figure S1 and table S1). These age-stratified serosurveys were used to fit a Bayesian catalytic model and obtain yearly estimates of the local FoI (in the 'catchment areas'), from the birth of the oldest participant to the year of the survey. (We refer to those municipalities where at least one serosurvey was conducted as municipalities 'in catchment areas' [20].) Electronic supplementary material, text S1.1–S1.5 describes the catalytic models used. For those municipalities that have more than one serosurvey conducted at different times (figure 1*a*), the FoI was estimated separately but these estimates were reconciled as described in Step 2 below.

(c) Step 2: Force-of-Infection at the municipality level (Random Forest predictive model)

(i) Predictors included

The second step aimed to predict the FoI in areas where no serosurveys had been conducted (i.e. outside the catchment areas), to obtain yearly FoI predictions across the entire country, at the municipality level (electronic supplementary material, text S2). Informed by previous studies [17,20], and in order to build a pipeline that could be applied in other countries, the predictors selected are available across Latin America and include characteristics of the serosurveys as well as spatio-temporal, climatic, environmental, demographic and socio-economic predictors as described below:

Serosurvey characteristics. The setting where the serosurvey was conducted was defined as urban, rural, indigenous or mixed (composed of urban and rural settings; see electronic supplementary material, S2.1.1 for details on how mixed settings surveys were used). The urban/rural definition follows government guidelines [21].

Spatio-temporal factors. Large urban centres, with a population of >100 000 inhabitants in 1985, were not included in the catchment area of the serosurveys. These represent about 3% of the municipalities (33 out of the 1122 municipalities of Colombia). Therefore, the FoI predictions made relate to small to medium-sized towns/cities. The year when the serosurvey was conducted was included to correct for a selection bias present in the early serosurveys, as those conducted before the year 2000 largely focused on high-risk populations [17], especially in rural settings. We allowed for a temporal trend by including the 'years' of each FoI 'observation' as a covariate, but assume that other

predictors would account for spatial heterogeneities and therefore latitude and longitude were not included [22].

Climatic and environmental predictors. We focused on climatic variables and indicators for presence of triatomine vectors. BioClim data for Colombia were collated between 1979 and 2013 on a 1 km^2 scale. Following the triatomine niche modelling literature [23–34], we focused on relative day–night temperature differences (Bio03), median minimum temperature of the coldest month (Bio06), and seasonality of precipitation (Bio15). We included additional predictors based on available literature, included median municipality elevation [26,30,34,35] and normalized difference vegetation index (NDVI) [26,30,34,36,37]. The year when a municipality was certified free from intradomiciliary transmission was used as a predictor in the model. Our analysis did not use further vector indicators to keep the pipeline flexible in terms of availability of country-specific information. We therefore implicitly assume that the environmental variables included above would encapsulate such information.

Demographic and socio-economic predictors. Population size and proportion of the municipal population living in urban settings were also included along with an integrated public use microdata series (IPUMS) indicator characterizing the proportion of houses with unfinished floors [38], a proxy for poverty with relevance to Chagas disease (housing conditions being highly correlated with vectorial intradomiciliary infestation [39]).

A full description of the predictors is given in electronic supplementary material, text S2.1, table S2, and figures S2–S21).

(ii) Model definition

We used the available collated data to predict the spatio-temporal trends in the FoI between 1950 and 2020, at the municipality level across Colombia using an RF regression model [40]. Following previous work [20], a nested resampling was applied for model-tuning with a spatial resampling strategy. In particular, and to partially account for sampling bias, a spatial resampling was used to assess the predictive ability of the model in spatial areas that were not included in the fitting of the model. Cross-validation (CV) was used to assess model performance, as described in electronic supplementary material, text S2.2. The importance of each predictor, i.e. the relative contribution of this variable in the model, was extracted and predictions were calculated (see electronic supplementary material, §S2.2 for details). To propagate the uncertainty inherited from the calculation of the FoI, the fitting and performance evaluation was repeated with 100 bootstrap samples from the posterior distribution of the FoI.

A composite indicator was used to assess model performance, including an estimated mean of the coefficient of determination (R^2) (calculated among the cross-validation (CV) set) and the percentage of overlap between predicted and 'observed' distributions of the FoI (using the function 'overlap' from the R package 'overlapping' [40]), with 'observed' distribution referring to the full posterior distribution of the FoI estimated with the catalytic model, as done in [20]. This performance indicator ensured that predictions reflected the central tendency, while also correctly accounting for the uncertainty in the response variable. The uncertainty was measured using the median absolute deviation coefficient of variation (MAD-CV), which is a coefficient of variation based on the median to account for the asymmetrical distribution of the FoI [41]. Each predictor is then ranked by its importance, the importance reflects the usefulness of the predictor in the model by quantifying how often the predictor has been used and how much variance it helps to explain. More detail on the modelling process is available in electronic supplementary material, text S2.2 and figure S21.

As previously mentioned, serosurveys were not available for large cities; therefore, our predictions of 'urban' exposure were representative of small to medium cities, and not of large cities. For cities with a population size >100 000 in 1985 (based on National Administrative Department of Statistics (DANE) estimates [21]), the prevalence observed in blood banks (voluntary donors between 18 and 65 years of age) was used to estimate a time-constant FoI.

(d) Step 3: from Force-of-Infection to disease burden (Chagas Disease Burden Model)

To estimate the spatio-temporal trends in the burden of Chagas disease in Colombia, we developed a model of disease progression (figure 2). The disease progression model is an age-specific compartmental model that estimates the prevalence of each stage of the disease for each age class, using parameters that describe disease progression and mortality (detailed in electronic supplementary material, text S3 and tables S3–S5).

In the progression model, individuals may acquire the parasite at a rate specified by the predicted, municipality-specific FoI for a given year. Following infection, some individuals may present no or mild symptoms, while a given proportion may develop acute symptoms (i.e. acute phase of Chagas disease), with symptoms including cardiomyopathy. Cases with no or mild symptoms transition to the indeterminate phase, during which they will remain largely asymptomatic. Cases in the indeterminate and severe acute phase can progress to the mild and thereafter severe chronic phases (electronic supplementary material, figure S22).

Individuals in the acute, chronic mild and chronic severe phases contribute most to the mortality associated with Chagas disease [42]. Progression to mild and severe chronic phases (e.g. with mild or severe cardiomyopathy) may be associated with co-morbidities rather than *T. cruzi* infection itself. We, therefore, allowed all individuals, infected and non-infected, to transition to those phases regardless of *T. cruzi* infection status. Digestive forms of Chagas disease (e.g. megaoesophagus or megacolon) were not included in this model as they are uncommon in Colombia [7] (but would have to be taken into account for other endemic countries).

From this progression model, we tracked, for each yearly cohort, the proportion of individuals in each stage, as well as the yearly proportion of the cohort deaths that were directly attributable to *T. cruzi* infection. Cohort size over time was informed by census data, while yearly mortality per cohort was informed by death record line-listing officially estimated [21]. The uncertainty associated with the estimated prevalence and the numbers of cases for each disease stage and/or death was characterized by the interquartile range divided

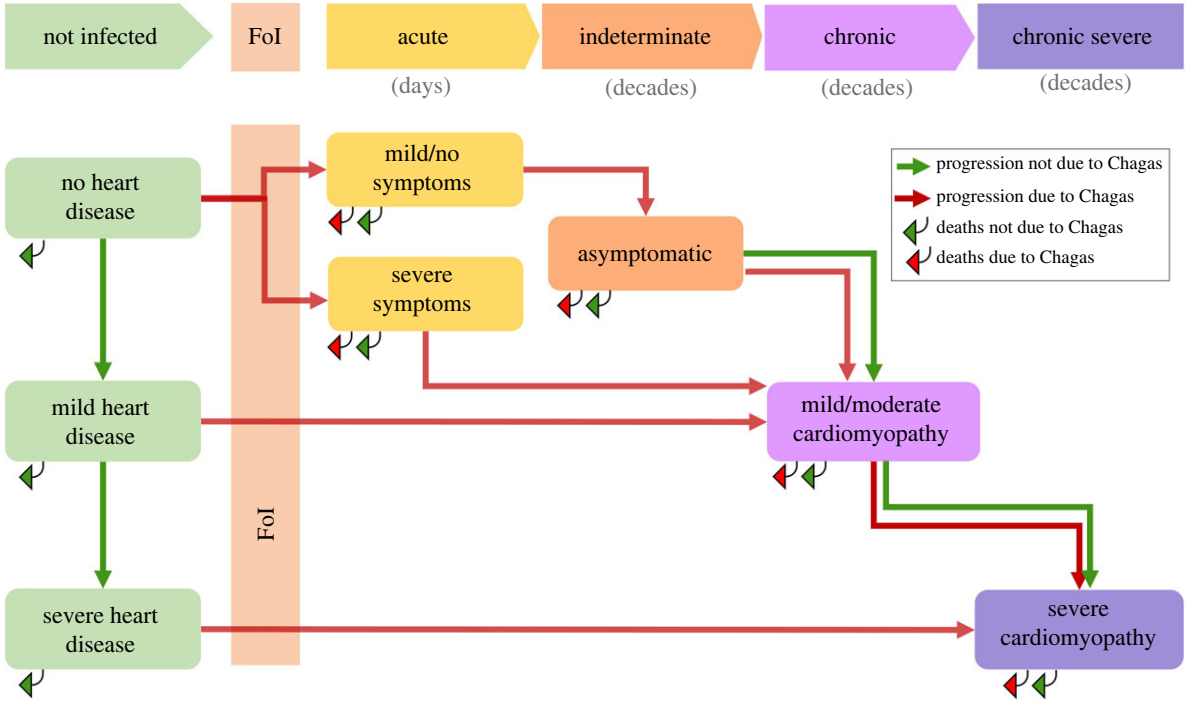

Figure 2. Chagas Disease Burden Model. Schematic representation of the compartmental model used to calculate the burden of disease from spatio-temporal predictions of the Force-of-Infection (FoI). The model takes into account the occurrence of co-morbidities, such that individuals in each compartment can have heart disease before becoming infected with *Trypanosoma cruzi*. For each compartment in the model, the age class-specific prevalence of each stage was calculated along with the mortality caused or not caused by Chagas disease depending on whether the progression was or was not due to *T. cruzi* infection. Details on the progression rates and compartments used are provided in electronic supplementary material, text S3 and tables S3–S5.

by the median (i.e. a measure of uncertainty relative to the central estimate).

All the analyses were conducted using R and its environment RStudio (https://www.posit.co/); Bayesian FoI models were fitted using RStan (https://mc-stan.org/), and the maps were prepared using QGIS 3.16.3-Hannover (https://www.qgis.org/es/site/).

3. Results

(a) Force-of-Infection at the serosurvey level

The age-stratified seroprevalence data were used to back-calculate the FoI using a time-varying FoI catalytic model [9]. Figure 3 presents the fitting for 76 surveys. Serosurvey-specific estimated FoI, convergence parameters and residual plots are provided in electronic supplementary material, text S4 and figures S24–S26. Note that some serosurveys were targeted at younger age classes, leading to larger sample sizes among them and to larger uncertainty of the predicted seroprevalence for older age classes.

(b) Predicted yearly Force-of-Infection at the municipal level

The RF predictive model of the FoI showed good performance, with a coefficient of determination (R^2) on the CV set of 64% in urban and 71% in rural areas. Model uncertainty was well propagated, with predicted and observed FoI distributions showing an overlap of 59% (electronic supplementary material, text S5, table S6 and figures S26 and S27).

To accurately predict FoI values, inclusion, as a predictor, of the year when the serosurvey was conducted was essential. The importance of this predictor in the model (which represents how helpful the predictor has been to the models) reached 99.2 (table 1). Accounting for setting type was also crucial, especially distinguishing between indigenous versus non-indigenous settings (importance of 87.7). Further accounting for differences between urban and rural settings had more marginal importance (around 11). All environmental (excluding NDVI), demographic and socio-economic predictors substantially improved the fit, with importance ranging from 19.0 to 43.7. Both year of the FoI and NDVI were associated with more limited improvement (approx. 10).

Predicted FoI across Colombia showed similar patterns in urban and rural settings (figure 4), with strong spatial heterogeneity and a substantial decrease in infection exposure over time, especially in the Andean and northeast regions. In urban settings, the median of the municipal FoI ranged from 2.0×10^{-4} to 2.8×10^{-3} new infections per susceptible per year in 1995, and from 1.6×10^{-4} to 2.7×10^{-3} in 2020 (a 4 to 20% reduction). In rural settings the values ranged from 2.1×10^{-4} to 2.8×10^{-3} in 1995 and from $1.6–1.8 \times 10^{-4}$ to 2.8×10^{-3} in 2020 (a 14 to 24% reduction). Uncertainty decreased slightly over time (with high uncertainty observed for 9% of the municipalities in 1995 and for 8% in 2020 across urban and rural settings), and was larger in the south of the country (where fewer surveys were available). In this context, high uncertainty was defined as a MAD-CV above 2.

(c) Burden model over 1985–2020

Chagas disease prevalence showed spatial heterogeneity, with the northern part of the country generally indicating higher FoI values (figure 4). The prevalence of infection was higher

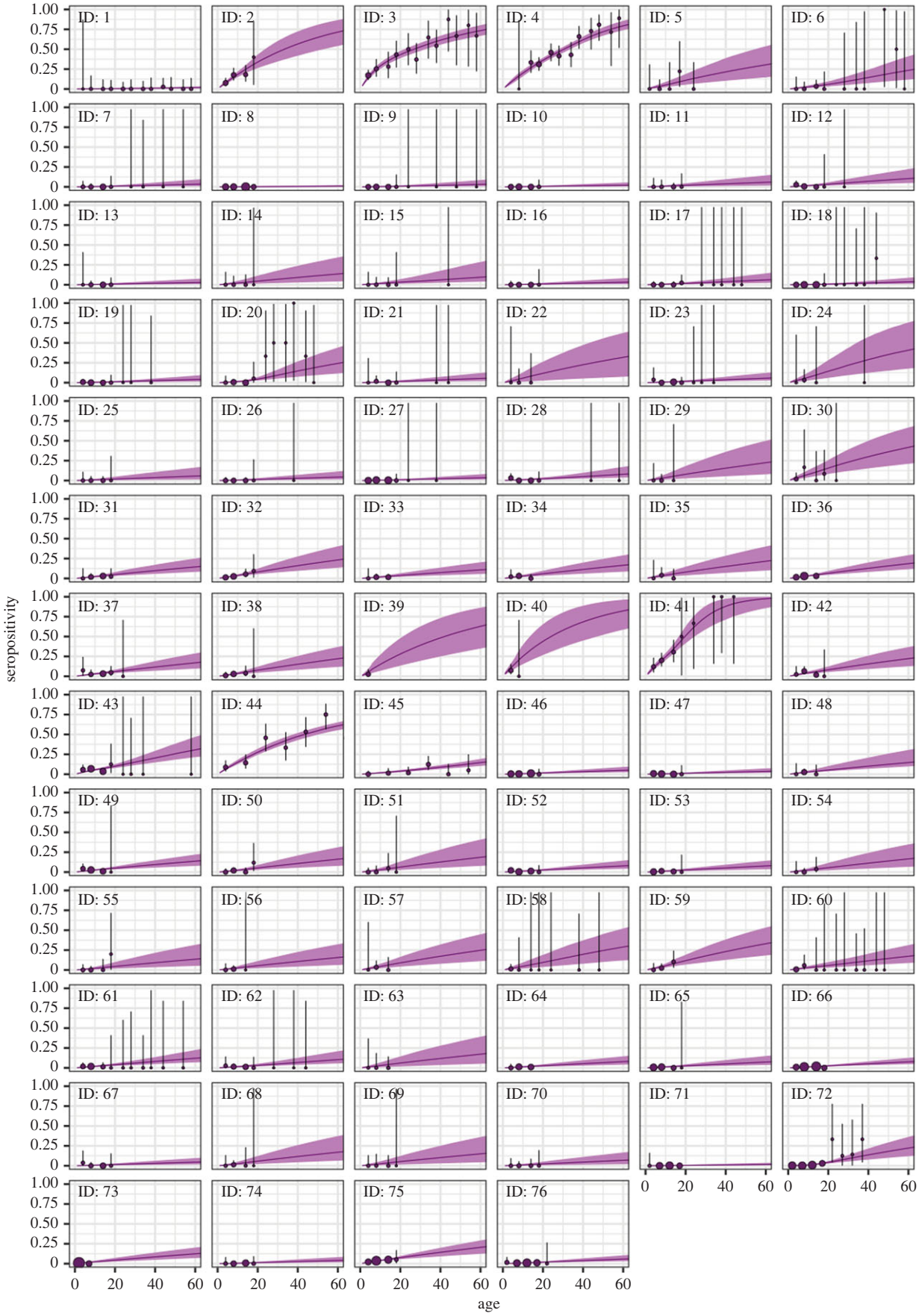

Figure 3. Observed and predicted seroprevalence of *Trypanosoma cruzi* infection. Predicted medians (purple solid lines) and 95% credible intervals (shaded areas) of the posterior distribution of age seroprevalence profiles are compared with median observed data (purple circles, size representing sample size for each age class in each serosurvey; vertical lines representing the 95% binomial confidence intervals). The graph numbering corresponds to the serosurvey ID in electronic supplementary material, table S1, where sample sizes can be found.

in rural areas, with a municipal median and interquartile range (IQR) prevalence in 2020 of 1.36% [1.15–1.57%] in rural settings, and 1.26% [1.10–1.45%] in urban areas irrespective of the population size of each municipality. However, given the higher population size in urban settings (76% of the total population in 2020), 61% of the predicted cases belonged to urban

Table 1. Ranking of the importance of the predictors used in the Random Forest model for predicting the Force-of-Infection of Chagas disease in Colombia at the municipal level.

predictor	importance
serosurvey characteristics:	
year when the serosurvey was conducted	99.2
setting type:	
urban	11.1
rural	10.9
indigenous	87.7
temporal coordinates:	
year	11.7
environmental predictors:	
year of certification for intradomiciliated vector elimination	34.7
isothermality (Bio03)	43.7
minimum temperature of the coldest month (Bio06)	26.8
seasonality of precipitation (Bio15)	36.4
normalized difference vegetation index (NDVI)	9.9
elevation	29.3
demographic predictors:	
population size	36.5
proportion of urban population	29
proportion of household with unfinished floor material	19

areas in 2020 (electronic supplementary material, text S5 and figures S28–S31).

While the FoI showed an overall decreasing trend between 1995 and 2020 (figure 4), the prevalence of infection was relatively stable, with the national median prevalence and IQR across urban and rural settings being, respectively, 1.0% [0.8–1.2%] in 1995 and 1.0% [0.8–1.3%] in 2020. A 6% decline in the prevalence of cases in the acute stage was predicted between 1995 and 2020, but this was compensated by a 13% increase, during the same period, in the predicted prevalence of cases in the chronic severe stage (figure 5).

As the relative prevalence of infection remained largely stable but the population increased by 39% between 1995 and 2020 (electronic supplementary material, figures S28 and S29), the overall burden of Chagas disease was predicted to have increased significantly (table 2). We estimated that the total number of infections across Colombia had increased by 43% between 1995 and 2020, reaching half a million cases. Between 1995 and 2020, even larger increases, of 57 and 79%, applied to cases with severe cardiomyopathy and deaths attributable to Chagas disease, respectively (purple line and shading in figure 5*c*). These were driven by an increase in both population size and the gradual ageing of the population (electronic supplementary material, figures S28 and S29).

Large spatial heterogeneity and clustering in the burden were observed, with three departments having the largest number of deaths attributable to Chagas disease in 1995, namely, Bogotá Distrito Capital (DC), Cundinamarca and Santander, which accounted for 31% of the deaths but with only 25% of the total population.

4. Discussion

We have developed a modelling pipeline that uses local seroprevalence data to obtain national disease burden estimates at the municipality level for Chagas disease and have used Colombia as a case study. From the unique 76 serosurveys conducted in Colombia, we estimated that the number of people infected with *T. cruzi* would have reached 506 000 (95% credible interval (CrI) = 395 000–648 000) in 2020, with a 1.0% (95%CrI = 0.8%–1.3%) infection prevalence in the general population and 2400 (95%CrI = 1900–3400) deaths due to Chagas disease in the same year. Temporally, the interplay between a slight decrease in exposure, measured as lower values of FoI and lower prevalence of new acute cases, was overcompensated by a large increase in population size and the gradual ageing of the population over the same period, both of which led to a substantial increase in the estimated burden of Chagas disease over time. The burden of disease dynamics reflects the protracted nature of disease progression. The substantial spatial heterogeneities in predicted disease burden suggest that spatial targeting of interventions could improve the cost-effectiveness of resource allocation (although we did not conduct an economic analysis in the present study). Our results could inform such spatial targeting, as our approach can help to predict locations where reducing exposure through vector control, for instance, would be most impactful. Other interventions, such as improving diagnosis and access to treatment, can also be considered, and this may lead to targeting different locations.

The performances of the FoI predictive models were good, with cross-validation performances that would suggest limited overfitting (electronic supplementary material, table S6). However, a formal validation of the disease burden model with external data would be far more complex, as sources to achieve this are limited. In fact, the surveillance system for Chagas disease currently in place cannot be used to validate our disease burden estimates, as the reporting is extremely sparse, although, we observe that cases are detected in almost all departments of Colombia (electronic supplementary material, table S7). The four departments having reported confirmed chronic cases in 2019 (Arauca, Santander, Cesar and Boyacá) represent only 14% of the severe cases obtained according to our results. These departments are endemic areas that likely have adequate infrastructure to conduct diagnosis, as well as trained staff and population with greater awareness of the disease than in other areas of the country. In Colombia, in the general population, the prevalence of heart conditions has been estimated at around 11% [43]. In 2019, a total of 41 848 deaths associated with cardiovascular disease and conditions were reported across the country [21]. If, as suggested by our models, 2400 of these deaths had been caused by Chagas disease, then Chagas disease would have accounted for 6% of the heart disease-related deaths in 2019 in Colombia, in line with the situation in Brazil, where Chagas disease was estimated to be responsible for between 1 and 21% of in-patient cases with heart failure [44]. Our model also estimates a total of 88 226 deaths among people suffering from heart disease, and 23 158 among people suffering from severe heart conditions, in 2020.

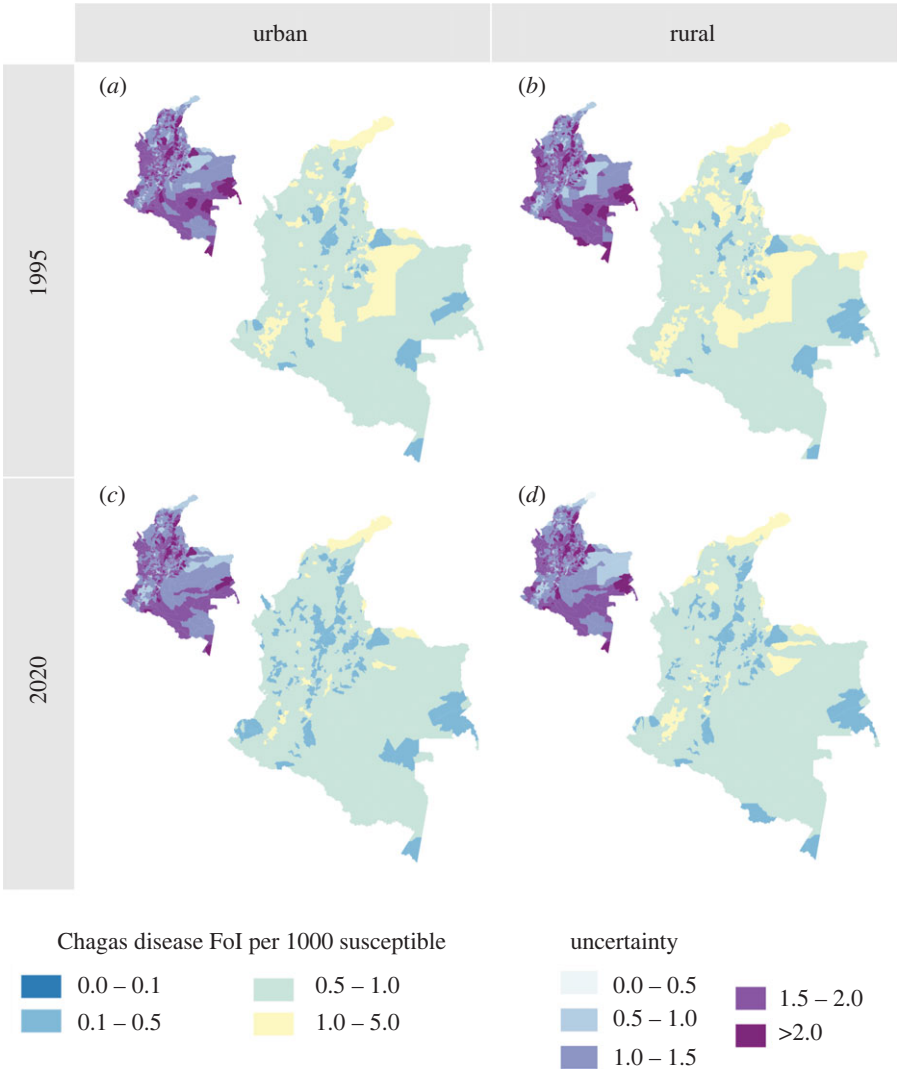

Figure 4. Spatial distribution of the Force-of-Infection of Chagas disease and the associated uncertainty (per year and per 1000 susceptible individuals), in Colombia at the municipality level in 1995 and 2020. (*a*) Urban areas, 1995; (*b*) rural areas, 1995; (*c*) urban areas, 2020; (*d*) rural areas, 2020. The predicted distribution of the FoI was generated using a Random Forest (RF) model (main maps); the associated uncertainty (small map insets) presents the median absolute deviation coefficient of variation (MAD-CV).

Our disease burden estimates are broadly comparable to those provided by WHO (with 438 000 cases of *T. cruzi* infection estimated in 2010 and a 1.0% prevalence) [6], as well as by Moncayo *et al*. (with 436 000 cases of infection estimated in 2005) [1]. Disconcertingly, our estimated increase in disease burden contrasts with the sharp decrease predicted by WHO and by Moncayo and colleagues. Estimates from the Global Burden of Disease Study (GBD) [8] showed a similar temporal increase in burden in terms of deaths (143% increase between 1995 and 2019, compared with our estimate of 71% increase in deaths), but a much lower absolute burden, with 170 (95% uncertainty interval = 74–283) deaths predicted in Colombia, compared with the 2400 deaths provided by our median estimate for 2019. In terms of infection prevalence and number of cases, the GBD predicted a stable number of cases between 1995 and 2019, with 123 000 (95% UI = 106 000–144 000) cases and a decreasing infection prevalence, from 0.34% (0.29–0.39%) to 0.26% (0.22–0.31%) between 1995 and 2019, which are again below our estimates. The estimates from the GBD, as well as ours, rely on demographic data, and therefore the underlying population data used will influence the results. The demography in Colombia has shown dramatic changes over the 1995–2019 period, experiencing a sharp increase in population size as well as transitioning, through changes in birth and death rates, towards an older population (electronic supplementary material, figure S29). The GBD, at the global scale, estimated a 3% decrease in Chagas disease-related deaths for the same period, with a decrease in Brazil and Argentina, but not in the remaining endemic countries [8]. The lower estimation reported by the GBD might be explained by the methodology used but also by differences in the data used, as our study is the first to our knowledge to comprehensively incorporate published and unpublished serological data.

Our modelling approach is a first attempt at estimating the burden of Chagas disease in Colombia over several years at the municipality level. Serosurveys that were conducted at a geographical scale larger than the municipality level were excluded from our analyses to be able to account for finer granularity in Chagas disease transmission heterogeneity and provide estimates at the most operational implementation and evaluation unit.

(a) Limitations

Given the complexity of the modelling tasks required, major simplifications were made which could influence our results. For instance, while we used information from serological

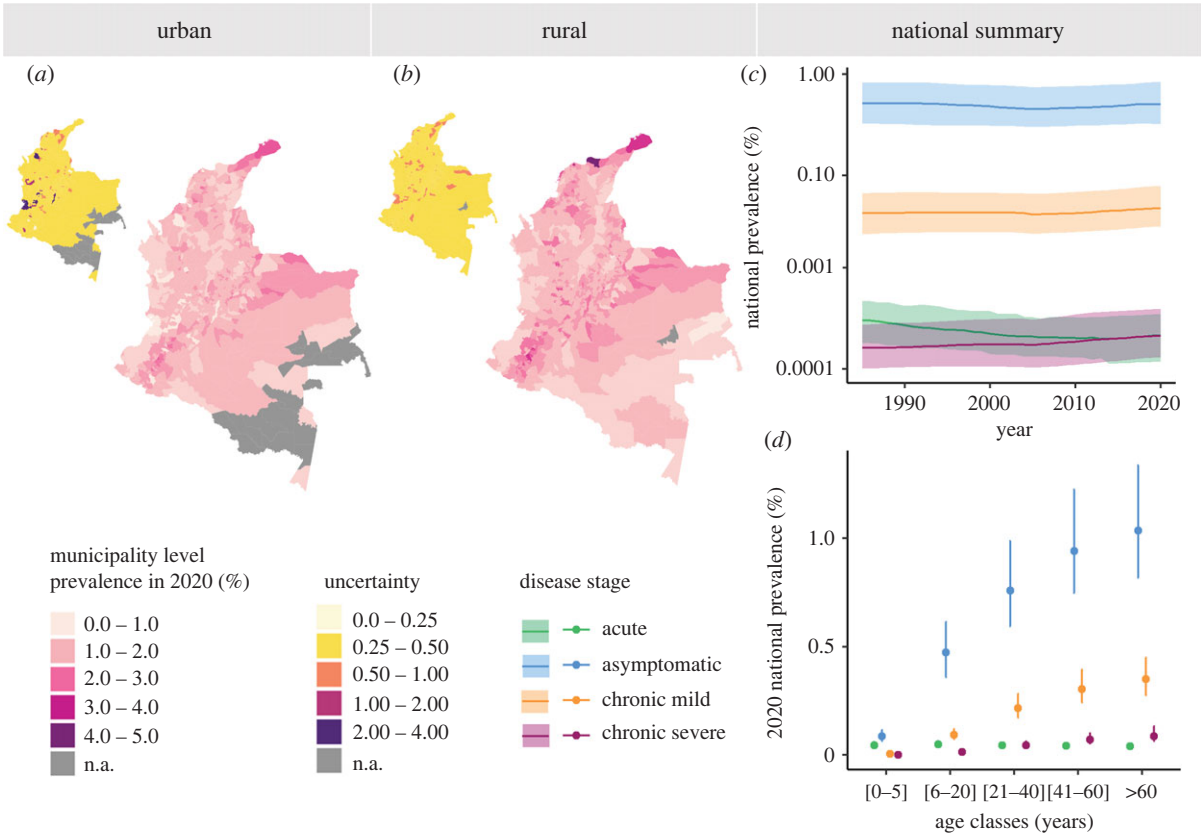

Figure 5. Spatial, temporal and age-class distribution of the prevalence of Chagas disease in Colombia. Municipal prevalence of *Trypanosoma cruzi* in 2020 in urban (*a*) and rural (*b*) areas (main maps); the associated uncertainty (small map insets) represents the interquartile range divided by the median. (*c*) Yearly national prevalence (median, solid line and interquartile, ribbon) from 1985 to 2020 of cases in the acute stage (green), asymptomatic (blue), chronic mild stage (orange) and chronic severe stage (purple). (*d*) National prevalence of Chagas disease stages by age class, in years (median, point and interquartile range, error bar) in 2020; each colour corresponds to a different disease stage as described for (*c*). The figure presents all prevalence values as percentages. n.a., data not available.

surveys conducted in indigenous populations to inform spatio-temporal trends in exposure, we did not include indigenous populations in our burden estimates. From our datasets only three serosurveys were conducted in indigenous settings at the municipality level, and we felt this was wholly insufficient to map exposure in specific indigenous settings across the country, although they may be broadly contained within rural estimates. We believe that our overall estimates of burden would remain largely unaffected, as only 0.6% of the population is estimated to live in indigenous settings, characterized, among other features, by the construction and use of traditional houses mostly in rural areas [21]. However, improving our understanding of the spatial distribution of the disease burden in these settings would be critical to improve interventions, as a prevalence of infection of 48.7% (95%CI = 42.6–51.6%) had been estimated in indigenous communities in 2012 [9].

We acknowledge the use of secondary data that are not necessarily representative of the total population. Indeed, the initial purpose of data collection for each serosurvey varied in time and space, and different sampling biases might have been introduced. Detecting and accounting for those biases is challenging, and we have attempted to address the main ones. First, the spatial target of the survey on endemic areas could limit our ability to extrapolate across Colombia. To overcome this, we used a spatial resampling strategy to assess the predictive ability of the model on spatial units of data that have not been included in the fitting of the model (see electronic supplementary material, text S2.2).

Second, we highlighted the importance of including the year when the serosurvey was conducted, as sampling tended to be directed at high-risk populations for serosurveys conducted before 2000. This effect of year may contribute to underestimating the FoI in the past, making it difficult to disentangle an actual decline in FoI from a reduction in sampling bias over time. Therefore, our results, and in particular temporal trends in older estimates of Chagas burden, should be treated with caution. We believe that our most recent estimates should be relevant, but that currently available data are insufficient to demonstrate a decline in Chagas disease burden.

Third, some serosurveys targeted a particular age class, as in the case of serosurveys sampling only or primarily children (i.e. a surveillance sentinel population for Chagas disease); such a sampling strategy might have limited our ability to better infer past transmission patterns beyond the age classes that were included in the surveys.

Other limitations are related to the structure and methodological choices made to build the modelling pipeline. We developed our pipeline to be flexible, robust and transparent, with most variables used to predict the FoI and estimate the disease burden being available across Latin America. However, our model currently does not account for the contribution of digestive morbidity and mortality, which may be less prominent in Colombia but of greater importance in other endemic regions such as the Southern Cone of Latin America, where megacolon and megaoesophagus can be present in 6% of the cases [1,45].

Table 2. Burden of Chagas disease cases in 1995 and 2020.

	median (95% credible interval)	
	1995	2020
estimated number of cases:		
total	355 000 (278 000–451 000)	506 000 (395 000–648 000)
chronic mild	70 000 (55 000–88 000)	102 000 (82 000–133 000)
chronic severe	14 000 (11 000–19 000)	22 000 (17 000–31 000)
estimated prevalence of *T. cruzi* infection (%):		
total	1.0 (0.8–1.2)	1.0 (0.8–1.3)
urban	0.7 (0.6–0.9)	0.8 (0.6–1.1)
rural	1.6 (1.1–2.1)	1.6 (1.2–2.2)
children (0–5 years)	0.1 (0.1–0.2)	0.1 (0.1–0.2)
older (>60 years)	1.8 (1.4–2.3)	1.5 (1.2–2.0)
estimated annual number of deaths due to Chagas disease:		
total	1400 (1100–1900)	2400 (1900–3400)
Bogotá DC	160 (64–443)	238 (93–700)
Cundinamarca	142 (101–200)	240 (164–341)
Santander	129 (87–173)	210 (142–248)

Currently, our model only considers the 'horizontal' transmission route (i.e. not accounting for mother-to-child—or vertical—transmission), and is best suited to model vectorial transmission. In contexts where vectorial transmission has been interrupted, the model will need to integrate other transmission routes, which include mother-to-child transmission during pregnancy (congenital transmission) and through organ transplant or blood transfusion (transplant and transfusion transmission). While a substantial reduction in the transmission risk related to blood transfusion and organ transplantation has been observed, it remains the main transmission route in non-endemic countries [46]. Crucially, in the case of congenital transmission, even if the diagnosis is made during pregnancy, treatment should be delayed until after the delivery as it is contraindicated during pregnancy [47]. Systematic screening of pregnant women has not yet been implemented and remains an important challenge in the fight against Chagas disease. Finally, we have not included the oral transmission route because although this is of great importance these events are less frequent, with a different transmission dynamic that leads to outbreaks with a high proportion of acute cases and high mortality [48–50], whereas we have focused on the endemic epidemiological situation.

Given the current low access to treatment, the model did not consider medical improvements that might help reduce mortality due to cardiomyopathy. Finally, there were important population movements in Colombia during the period of our study that are not accounted for in our model, and thus people who tested seropositive in a municipality could have acquired the infection in another municipality. This likely contributes to explaining why our model showed similar FoI patterns in urban and rural areas, while vectorial transmission is expected to have greater prominence in rural areas.

(b) Recommendations for future work

Collecting additional seroprevalence data would greatly assist with the validation of the modelling pipeline presented here. However, any sampling bias that may be introduced by the sampling strategy chosen has to be considered, minimized or at least well documented. Our results also strongly demonstrate that (1) given the chronic nature of Chagas disease, any recent reduction in incidence would have little measurable impact on short-term burden, and (2) understanding demographic trends is essential to estimate the burden of Chagas disease. Therefore, a stable FoI is predicted to lead to constant prevalence and burden in a stable population, but if the population is both increasing and ageing, burden is predicted to increase. Better understanding of the temporal trends in Chagas disease burden would therefore require more representative serosurveys (across locations, settings and age classes).

Using our framework, sampling could be made in areas where the model suggests higher uncertainty, leading to a cycle of model fit and improvement. While rural areas are currently well represented, this is not the case for urban ones. Indeed, serosurveys conducted in areas defined as urban are unlikely to be representative of densely populated cities. Conducting serosurveys in large cities would help to improve the estimates of disease burden and is crucial as population migration towards cities increases; these large cities already represent 27% of the population [21] (electronic supplementary material, figure S27). Currently, the modelling pipeline uses seroprevalence information from blood banks for large cities, with these cities accounting for 20% of the cases in 2020 and for 17% of the cases in 1995. Our study also highlights the importance of sampling across all age-groups to reconstruct the history of exposure and extrapolate burden, which will be relevant across other chronic NTDs.

In this paper, we have used Colombia as a case study. However, the DICTUM platform is increasingly hosting data from a larger number of endemic countries in Latin America in collaboration with national Chagas disease control programmes, and the modelling process is ready to be implemented in other contexts. Although such implementation will consider study quality and sampling design by assigning, for instance, appropriate weights, the criteria should initially not be too stringent, as this would limit the number of serosurveys included and it would be preferable to cover a large proportion of the population and capture as much exposure variation as possible during preliminary, exploratory analyses. A process of refinement, estimation of uncertainty, and identification of data gaps will ensue with national and regional programme managers and stakeholders. It is our hope to expand the work presented here to other endemic countries, so FoI estimates and their spatio-temporal trends can be updated, together with burden of disease estimates, to inform optimal targeting of interventions and help realize the WHO roadmap goals for Chagas disease by 2030 and beyond.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. The data and analytics routines are publicly available from the Github repositories: github.com/zmcucunuba/chfoi-col and mcucunuba/ChagasBurdenModel.

Authors' contributions. J.L.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, writing—original draft, writing—review and editing; Z.M.C.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—review and editing; G.P.-H.: data curation, investigation, methodology, validation, writing—review and editing; E.R.-M.: data curation, investigation, validation, writing—review and editing; A.P.D.: conceptualization, funding acquisition, methodology, supervision, validation, writing—review and editing; S.B.A.: investigation, methodology, writing—review and editing; L.G.C.: conceptualization, funding acquisition, resources, writing—review and editing; M.-G.B.: conceptualization, investigation, methodology, project administration, resources, supervision, validation, writing—review and editing; P.N.: conceptualization, formal analysis, investigation, methodology, project administration, resources, software, supervision, validation, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed herein.

Conflict of interest declaration. All authors declare that they have no competing interests.

Funding. J.L. acknowledges funding from a University of Sussex PhD Studentship. Z.M.C. was funded by MR/R024855/1. Z.M.C, M.-G.B. and P.N. acknowledge funding from the Medical Research Council (MRC) Centre for Global Infectious Disease Analysis (grant no. MR/R015600/1), jointly funded by the UK MRC and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the European and Developing Countries Clinical Trials Partnership (EDCTP2) programme supported by the European Union. E.R.-M. was funded by the Pan American Health Organization.

Acknowledgements. We would like to thank the Pan American Health Organization (PAHO) for their support and encouragement over the years and throughout all the stages of this work. Also, we would like to thank the large number of people from governmental, non-governmental and academic organizations who originally conducted the serosurveys and made their results available for us to use in this modelling effort.

References

1. Moncayo Á, Silveira AC. 2017 Current epidemiological trends of Chagas disease in Latin America and future challenges: epidemiology, surveillance, and health policies. In *American trypanosomiasis, Chagas disease: one hundred years of research*, 2nd edn (eds J Telleria, M Tibayrenc), pp. 59–88. Cambridge, MA: Elsevier.
2. Stanaway JD, Roth G. 2015 The burden of Chagas disease: estimates and challenges. *Glob. Heart* **10**, 139–144. (doi:10.1016/j.gheart.2015.06.001)
3. World Health Organization. 2015 *Investing to overcome the global impact of neglected tropical diseases: Third WHO Report on Neglected Tropical Diseases 2015*. Geneva, Switzerland: World Health Organization. See https://apps.who.int/iris/handle/10665/152781.
4. Ntuli MM. 2021 *Ending the neglect to attain the sustainable development goals: a road map for neglected tropical diseases 2021–2030*. Geneva, Switzerland: World Health Organization. See https://www.who.int/publications/i/item/9789240010352.
5. Instituto Nacional de Salud. 2021 *Informe de evento. Chagas crónico, periodo epidemiológico XII. Colombia 2021*. See http://www.ins.gov.co/buscador-eventos/Informesdeevento/CHAGAS%20CR%C3%93NICO%20PE%20XII%202021.pdf.
6. World Health Organization. 2015 Chagas disease in Latin America: an epidemiological update based on 2010 estimates. *Wkly Epidemiol. Rec.* **90**, 33–43.
7. Cucunubá ZM. 2017 Modelling the epidemiology and healthcare burden of Chagas disease in Colombia. PhD thesis, Imperial College London. (doi:10.25560/69755)
8. GBD. 2015 Disease and Injury Incidence and Prevalence Collaborators. 2016 Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1545–1602. (doi:10.1016/S0140-6736(16)31678-6)
9. Cucunubá ZM, Nouvellet P, Conteh L, Vera MJ, Angulo VM, Dib JC, Parra-Henao GJ, Basáñez MG. 2017 Modelling historical changes in the force-of-infection of Chagas disease to inform control and elimination programmes: application in Colombia. *BMJ Glob. Health* **2**, e000345. (doi:10.1136/bmjgh-2017-000345)
10. Cattarino L, Rodriguez-Barraquer I, Imai N, Cummings DAT, Ferguson NM. 2020 Mapping global variation in dengue transmission intensity. *Sci. Transl. Med.* **12**, eaax4144. (doi:10.1126/scitranslmed.aax4144)
11. O'Driscoll M, Imai N, Ferguson NM, Hadinegoro SR, Satari HI, Tam CC, Dorigatti I. 2020 Spatiotemporal variability in dengue transmission intensity in Jakarta, Indonesia. *PLoS Negl. Trop. Dis.* **14**, e0008102. (doi:10.1371/journal.pntd.0008102)
12. Biggs JR *et al*. 2021 Estimating the annual dengue force of infection from the age of reporting primary infections across urban centres in endemic countries. *BMC Med.* **19**, 217. (doi:10.1186/s12916-021-02101-6)
13. Lim JK *et al*. 2021 Estimating the force of infection for dengue virus using repeated serosurveys, Ouagadougou, Burkina Faso. *Emerg. Infect. Dis.* **27**, 130–139. (doi:10.3201/eid2701.191650)
14. Aleshnick M, Ganusov VV, Nasir G, Yenokyan G, Sinnis P. 2020 Experimental determination of the force of malaria infection reveals a non-linear relationship to mosquito sporozoite loads. *PLoS Pathog.* **16**, e1008181. (doi:10.1371/journal.ppat.1008181)
15. Arnold BF, Kanyi H, Njenga SM, Rawago FO, Priest JW, Secor WE, Lammie PJ, Won KY, Odiere MR. 2020 Fine-scale heterogeneity in *Schistosoma mansoni* force of infection measured through antibody response. *Proc. Natl Acad. Sci. USA* **117**, 23 174–23 181. (doi:10.1073/pnas.2008951117)
16. Garske T, Van Kerkhove MD, Yactayo S, Ronveaux O, Lewis RF, Staples JE, Perea W, Ferguson NM, for the Yellow Fever Expert Committee. 2014 Yellow fever in Africa:estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLoS Med.* **11**, e1001638. (doi:10.1371/journal.pmed.1001638)
17. Ledien J, Cucunubá ZM, Parra-Henao G, Rodríguez-Monguí E, Dobson AP, Basáñez M-G, Nouvellet P. 2022 Spatiotemporal variations in exposure: Chagas disease in Colombia as a case study. *BMC Med. Res. Methodol.* **22**, 13. (doi:10.1186/s12874-021-01477-6)
18. Moore SM. 2021 The current burden of Japanese encephalitis and the estimated impacts of vaccination: combining estimates of the spatial distribution and transmission intensity of a zoonotic pathogen. *PLoS Negl. Trop. Dis.* **15**, e0009385. (doi:10.1371/journal.pntd.0009385)
19. Alleman MM *et al*. 2016 Estimating the burden of rubella virus infection and congenital rubella syndrome through a rubella immunity assessment among pregnant women in the Democratic Republic of the Congo: potential impact on vaccination policy. *Vaccine* **34**, 6502–6511. (doi:10.1016/j.vaccine.2016.10.059)
20. Ledien J, Cucunubá ZM, Parra-Henao G, Rodríguez-Monguí E, Dobson AP, Adamo SB, Basáñez M-G, Nouvellet P. 2022 Linear and machine learning modelling for spatiotemporal disease predictions: force-of-Infection of Chagas disease. *PLoS Negl.*

Trop. Dis. **16**, e0010594. (doi:10.1371/journal.pntd.0010594)

21. Departamento Administrativo Nacional de Estadística (DANE). *Estadísticas de demografía y población de Colombia* [*Statistics for the demography of the population of Colombia*]. See https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion (accessed on 17 January 2022). [In Spanish.]
22. Meyer H, Reudenbach C, Wöllauer S, Nauss T. 2019 Importance of spatial predictor variable selection in machine learning applications – moving from data reproduction to spatial prediction. *Ecol. Modell*. **411**, 108815. (doi:10.1016/j.ecolmodel.2019.108815)
23. Carmona-Castro O, Moo-Llanes DA, Ramsey JM. 2018 Impact of climate change on vector transmission of *Trypanosoma cruzi* (Chagas, 1909) in North America. *Med. Vet. Entomol.* **32**, 84–101. (doi:10.1111/mve.12269)
24. Cavallo MJ, Amelotti I, Gorla DE. 2016 Invasion of rural houses by wild Triatominae in the arid Chaco. *J. Vector Ecol.* **41**, 97–102. (doi:10.1111/jvec.12199)
25. de La Vega GJ, Schilman PE. 2018 Ecological and physiological thermal niches to understand distribution of Chagas disease vectors in Latin America. *Med. Vet. Entomol.* **32**, 1–13. (doi:10.1111/mve.12262)
26. Parra-Henao G, Angulo VM, Osorio L, Jaramillo-O N. 2016 Geographic distribution and ecology of *Triatoma dimidiata* (Hemiptera: Reduviidae) in Colombia. *J. Med. Entomol.* **53**, 122–129. (doi:10.1093/jme/tjv163)
27. Espinoza Echeverria J, Rodriguez AN, Cortez MR, Diotaiuti LG, Gorla DE. 2017 Spatial and temporal distribution of house infestation by *Triatoma infestans* in the Toro Toro municipality, Potosi, Bolivia. *Parasit. Vectors* **10**, 58. (doi:10.1186/s13071-017-1984-0)
28. Ferro e Silva AM *et al.* 2018 Spatial prediction of risk areas for vector transmission of Trypanosoma cruzi in the State of Paraná, southern Brazil. *PLoS Negl. Trop. Dis.* **12**, e0006907. (doi:10.1371/journal.pntd.0006907)
29. Gurgel-Gonçalves R, Cuba CAC. 2009 Predicting the potential geographical distribution of *Rhodnius neglectus* (Hemiptera, Reduviidae) based on ecological niche modeling. *J. Med. Entomol.* **46**, 952–960. (doi:10.1603/033.046.0430)
30. Hernández J, Núñez I, Bacigalupo A, Cattan PE. 2013 Modeling the spatial distribution of Chagas disease vectors using environmental variables and people's knowledge. *Int. J. Health Geogr.* **12**, 29. (doi:10.1186/1476-072X-12-29)
31. Tapia-Garay V *et al.* 2018 Assessing the risk zones of Chagas' disease in Chile, in a world marked by global climatic change. *Mem. Inst. Oswaldo Cruz* **113**, 24. (doi:10.1590/0074-02760170172)
32. Medone P, Ceccarelli S, Parham PE, Figuera A, Rabinovich JE. 2015 The impact of climate change on the geographical distribution of two vectors of Chagas disease: implications for the force of infection. *Phil. Trans. R. Soc. B* **370**, 20130560. (doi:10.1098/rstb.2013.0560)
33. Pereira JM, de Almeida P, de Sousa A, de Paula A, Machado RB, Gurgel-Gonçalves R. 2013 Climatic factors influencing triatomine occurrence in central-west Brazil. *Mem. Inst. Oswaldo Cruz* **108**, 335–341. (doi:10.1590/S0074-02762013000300012)
34. Mischler P, Kearney M, McCarroll JC, Scholte RGC, Vounatsou P, Malone JB. 2012 Environmental and socio-economic risk modelling for Chagas disease in Bolivia. *Geospat. Health* **6**, S59–S66. (doi:10.4081/gh.2012.123)
35. Carbajal de la Fuente AL, Porcasi X, Noireau F, Diotaiutie L, Gorla DE. 2009 The association between the geographic distribution of *Triatoma pseudomaculata* and *Triatoma wygodzinskyi* (Hemiptera: Reduviidae) with environmental variables recorded by remote sensors. *Infect. Genet. Evol.* **9**, 54–61. (doi:10.1016/j.meegid.2008.09.008)
36. Dias JVL, Queiroz DRM, Martins HR, Gorla DE, Pires HHR, Diotaiuti L. 2016 Spatial distribution of triatomines in domiciles of an urban area of the Brazilian southeast region. *Mem. Inst. Oswaldo Cruz* **111**, 43–50. (doi:10.1590/0074-02760150352)
37. Gottdenker NL, Chaves LF, Calzada JE, Saldaña A, Carroll CR. 2012 Host life history strategy, species diversity, and habitat influence *Trypanosoma cruzi* vector infection in changing landscapes. *PLoS Negl. Trop. Dis.* **6**, e1884. (doi:10.1371/journal.pntd.0001884)
38. Minnesota Population Center. 2019 *IPUMS International: version 7.2*. See https://www.ipums.org/projects/ipums-international/d020.v7.2.
39. Dias JCP. 2015 Evolution of Chagas disease screening programs and control programs: historical perspective. *Glob. Heart* **10**, 193–202. (https://pubmed.ncbi.nlm.nih.gov/26407516/)
40. Lang M *et al*. 2019 mlr3: A modern object-oriented machine learning framework in R. *J. Open Source Softw*. **4**, 1903. (doi:10.21105/joss.01903)
41. Rousseeuw PJ, Croux C. 1993 Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **88**, 1273–1283. (doi:10.1080/01621459.1993.10476408)
42. Cucunubá ZM, Okuwoga O, Basáñez M-G, Nouvellet P. 2016 Increased mortality attributed to Chagas disease: a systematic review and meta-analysis. *Parasit. Vectors* **9**, 42. (doi:10.1186/s13071-016-1315-x)
43. Stevens B, Verdian L, Tomlinson J, Zegenhagen S, Pezzullo L. 2016 The economic burden of heart diseases in Colombia. World Congress of Cardiology Scientific Sessions 2016, poster PM021. *Glob. Heart* **11**(Suppl.), e73–e74. (doi:10.1016/j.gheart.2016.03.257)
44. Bocchi EA, Arias A, Verdejo H, Diez M, Gómez E, Castro P, Interamerican Society of Cardiology. 2013 The reality of heart failure in Latin America. *J. Am. Coll. Cardiol*. **62**, 949–958. (doi:10.1016/j.jacc.2013.06.013)
45. Pérez-Molina JA, Molina I. 2018 Chagas disease. *Lancet* **391**, 82–94. (doi:10.1016/S0140-6736(17)31612-4)
46. Kim JYH, Ledien J, Rodriguez-Monguí E, Dobson A, Basáñez M-G, Cucunubá ZM. 2019 Global trends of seroprevalence and universal screening policy for Chagas disease in donors: a systematic review and meta-analysis. *medRxiv*, 2019.12.25.19015776. (doi:10.1101/2019.12.25.19015776)
47. Carlier Y, Torrico F, Sosa-Estani S, Russomando G, Luquetti A, Freilij H, Vinas PA. 2011 Congenital chagas Disease: recommendations for diagnosis, treatment and control of newborns, siblings and pregnant women. *PLoS Negl. Trop. Dis.* **5**, e1250. (doi:10.1371/journal.pntd.0001250)
48. Rincón-Acevedo CY, Parada-García AS, Olivera MJ, Torres-Torres F, Zuleta-Dueñas LP, Hernández C, Ramírez JD. 2021 Clinical and epidemiological characterization of acute Chagas disease in Casanare, eastern Colombia, 2012–2020. *Front. Med.* **8**, 1178. (doi:10.3389/fmed.2021.681635)
49. Bruneto EG *et al.* 2020 Case-fatality from orally-transmitted acute Chagas disease: a systematic review and meta-analysis. *Clin. Infect. Dis.* **72**, 1084–1092. (doi:10.1093/cid/ciaa1148)
50. Filigheddua MT, Górgolasb M, Ramos JM. 2017 Enfermedad de Chagas de transmisión oral [Orally-transmitted Chagas disease]. *Med. Clin. (Barc.)* **148**, 125–131. (doi:10.1016/j.meddi.2016.10.038) [In Spanish.]