



Article

Spatio-Temporal Bayesian Models for Malaria Risk Using Survey and Health Facility Routine Data in Rwanda

Muhammed Semakula ^{1,2,3,4,*} , François Niragire ⁵  and Christel Faes ¹ 

¹ I-BioStat, Hasselt University, 3500 Hasselt, Belgium

² Centre of Excellence in Data Science, Bio-Statistics, College of Business and Economics, University of Rwanda, Kigali 4285, Rwanda

³ Rwanda Biomedical Center, Kigali 7162, Rwanda

⁴ KIT Royal Tropical Institute of Amsterdam, 1092 AD Amsterdam, The Netherlands

⁵ Department of Applied Statistics, University of Rwanda, Kigali 4285, Rwanda

* Correspondence: muhammed.semakula@uhasselt.be

Abstract: Introduction: Malaria is a life-threatening disease occurring mainly in developing countries. Almost half of the world's population was at risk of malaria in 2020. Children under five years of age are among the population groups at considerably higher risk of contracting malaria and developing severe disease. Most countries use Demographic and Health Survey (DHS) data for health programs and evaluation. However, malaria elimination strategies require a real-time, locally-tailored response based on malaria risk estimates at the lowest administrative levels. In this paper, we propose a two-step modeling framework using survey and routine data to improve estimates of malaria risk incidence in small areas and enable quantifying malaria trends. Methods: To improve estimates, we suggest an alternative approach to modeling malaria relative risk by combining information from survey and routine data through Bayesian spatio-temporal models. We model malaria risk using two steps: (1) fitting a binomial model to the survey data, and (2) extracting fitted values and using them in the Poisson model as nonlinear effects in the routine data. We modeled malaria relative risk among under-five-year old children in Rwanda. Results: The estimation of malaria prevalence among children who are under five years old using Rwanda demographic and health survey data for the years 2019–2020 alone showed a higher prevalence in the southwest, central, and northeast of Rwanda than the rest of the country. Combining with routine health facility data, we detected clusters that were undetected based on the survey data alone. The proposed approach enabled spatial and temporal trend effect estimation of relative risk in local/small areas in Rwanda. Conclusions: The findings of this analysis suggest that using DHS combined with routine health services data for active malaria surveillance may provide more precise estimates of the malaria burden, which can be used toward malaria elimination targets. We compared findings from geostatistical modeling of malaria prevalence among under-five-year old children using DHS 2019–2020 and findings from malaria relative risk spatio-temporal modeling using both DHS survey 2019–2020 and health facility routine data. The strength of routinely collected data at small scales and high-quality data from the survey contributed to a better understanding of the malaria relative risk at the subnational level in Rwanda.

Keywords: malaria; survey data; routine data; spatio-temporal models



Citation: Semakula, M.; Niragire, F.; Faes, C. Spatio-Temporal Bayesian Models for Malaria Risk Using Survey and Health Facility Routine Data in Rwanda. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4283. <https://doi.org/10.3390/ijerph20054283>

Academic Editors: Lilia González-Cerón and Patricia Penilla-Navarro

Received: 18 January 2023

Revised: 24 February 2023

Accepted: 26 February 2023

Published: 28 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Malaria is a life-threatening disease occurring mainly in developing countries. Almost half of the world's population was at risk of malaria in 2020. Some population groups are at considerably higher risk of contracting malaria and developing severe disease: infants, children under five years of age, pregnant women, and patients with HIV/AIDS, as well as people with low immunity moving to areas with intense malaria transmission such

as migrant workers, mobile populations, and travelers. The World Health Organization reported an increase of 6.2% in cases from 2019 (227 million cases) to 2020 (241 million cases). In the same period, the number of deaths due to malaria increased by 12.4% as compared to the previous year. African countries contribute about 95% of all malaria cases and 96% of deaths. Children under five are more vulnerable compared to the elderly, with 80% of all malaria deaths occurring in Africa [1].

Four countries in Africa count over half of malaria deaths globally: Nigeria (31.9%), the Democratic Republic of the Congo (DRC) (13.2%), the United Republic of Tanzania (4.1%), and Mozambique (3.8%) [1]. Those four countries count for 80% of all malaria deaths in Africa. East African countries such as Burundi, the DRC, Kenya, Rwanda, South Sudan, the United Republic of Tanzania, and Uganda are among the second most vulnerable countries [1]. The deaths and burden of malaria cases among children under five years is a leading public health concern. Malaria prevalence among children under five years in Burundi was 21.8% in 2012, and a recent Demographic and Health Survey (DHS) in 2017 showed an increase to 37.9%. For Rwanda, the last two DHS reported a decrease in malaria prevalence among children under five from 7.8% in 2015 to 2.7% in 2020 [2]. In 2015 malaria was the second cause of morbidity in Rwanda, representing 7.4% of outpatient consultations, and the sixth cause of mortality, representing 4.3% of total mortality. However, the Malaria Indicator Survey (MIS) conducted in 2017 showed a high prevalence of 11.8% in Rwanda. In Tanzania, the prevalence was 14.4% for DHS 2016 and slightly lower at 7.3% for MIS 2017. Uganda estimated a prevalence of 30.4% based on DHS of 2016, and MIS reported 18.2% in 2019 [2]. The prevalence at the subnational level is locally higher within countries based on DHS, as Figure 1 shows. In the Burundi's northeastern regions bordering Rwanda and northeastern Uganda bordering South Sudan, malaria prevalence among under-five-year old children is above 60%.

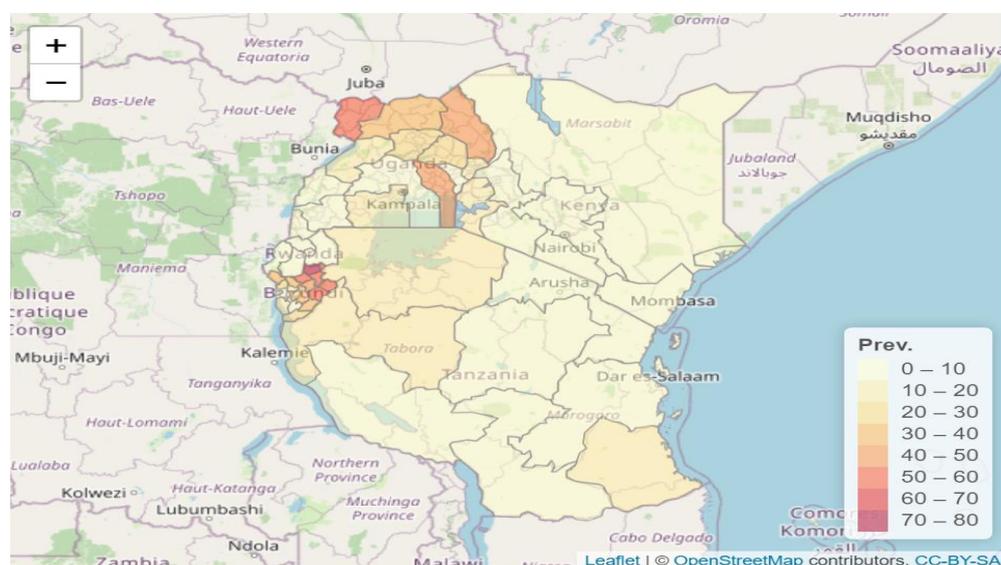


Figure 1. Percentage of children of age 6–59 months that tested positive for malaria in East Africa (Burundi, Kenya, Rwanda, Tanzania, and Uganda at subnational. Source: DHS data).

East African community (EAC) member states (Burundi, Kenya, Rwanda, Tanzania, Uganda, South Sudan, and the DRC) have invested in malaria prevention and treatment strategies for decades and have implemented various prevention strategies. The strategies focused on distributing insecticide-treated mosquito nets (ITNs) and employing indoor residual spraying (IRS) within the households, and some countries facilitate citizens to access free or minimized primary healthcare services through community health insurance schemes. However, the malaria risk is still high, and countries must accelerate the implementation of life-saving strategies. The World Health Organization elaborated a global technical strategy for malaria elimination by 2030 that was built on three pillars: (1) Ensure

universal access to malaria prevention, diagnosis, and treatment; (2) accelerate efforts towards elimination and attainment of malaria-free status; and (3) transform malaria surveillance into a core intervention.

Malaria surveillance is a core component of accelerating progress toward malaria elimination by 2030. All East African member states are malaria-endemic countries, and the level of malaria prevention and control measures depends on country-specific strategies. Figure 1 shows that areas of Burundi bordering Rwanda are high-risk malaria zones as well as areas bordering Uganda, South Sudan, and Kenya. Since malaria can easily cross borders, it is vital to establish joint effective and active surveillance that enables early detection of outbreaks at subnational levels and continuous assessment of interventions to guide changes for improved interventions.

All EAC member states use DHS findings to plan for malaria elimination including for localized regions within the country. However, the DHS does not provide information about time since it is a cross-sectional survey [3]. Indeed, DHS surveys are methodologically designed to provide estimates statistically powered at a specific domain, not necessarily at the lowest subnational levels (DHS, most of the time, is limited to the second subnational level). The locally-tailored response requires having real-time malarial estimates at the lowest administrative levels. In addition, DHS or Malaria Indicator Survey (MIS) surveys do not allow temporal analysis because DHS is often conducted every five years and MIS every two years. However, malaria patterns vary weekly or monthly, which implies immediate action to control malaria outbreaks. DHS/MIS are of high-quality standards and are representative through random selection and two-stage sampling approaches.

This paper combines information from Rwanda Demographic and Health Survey (RDHS) data from 2019–2020 and health facility routine data from 2019–2020. Individual malaria cases of each sector are not routinely recorded in Rwanda so the accurate number of malaria cases is unknown, even though a five-digit number is given in the Table 1, and instead, estimates done by each sector are kept in a health information system as the routine data, and solely those estimates have been used in other malaria analyses. Part of the DHS data was collected during the COVID-19 pandemic, and little is known about the disruption and how the exercise was affected. Similar to the health facility routine data, part of malaria cases were reported during the COVID-19 period. The COVID-19 pandemic resulted in patients increasingly seeking care in the community and posed challenges to maintaining the delivery of malaria services in Rwanda [4].

Table 1. Datasets used for children under five years old.

| Type | Cases | Sample | Domain | Freq | Covariates |
|-----------------|--------|--------|---------------|-----------------|-----------------------|
| Routine | 39,936 | | Sector (416) | Daily/Month | Gender, location |
| DHS | 99 | 3665 | District (30) | Cross-sectional | Gender, HHD, location |
| Shapefile | | | Sector (416) | | |
| Population size | | | Sector (416) | | |

Rwanda developed a theory of change progressing from a lack of access to data and the absence of a culture of learning to a state of routine data utilization to identify and drive change in public health decision making [5]. Both DHS and health facility routine data include community-reported information. Figure 2 shows Rwanda's primary data sources structure. The survey data are often considered the gold standard compared to routine data from health facilities. The routine data are used within health facilities for disease monitoring. Still, most countries use only data from surveys for public health decisions [6], despite the availability of a huge amount of routine data at a local level.

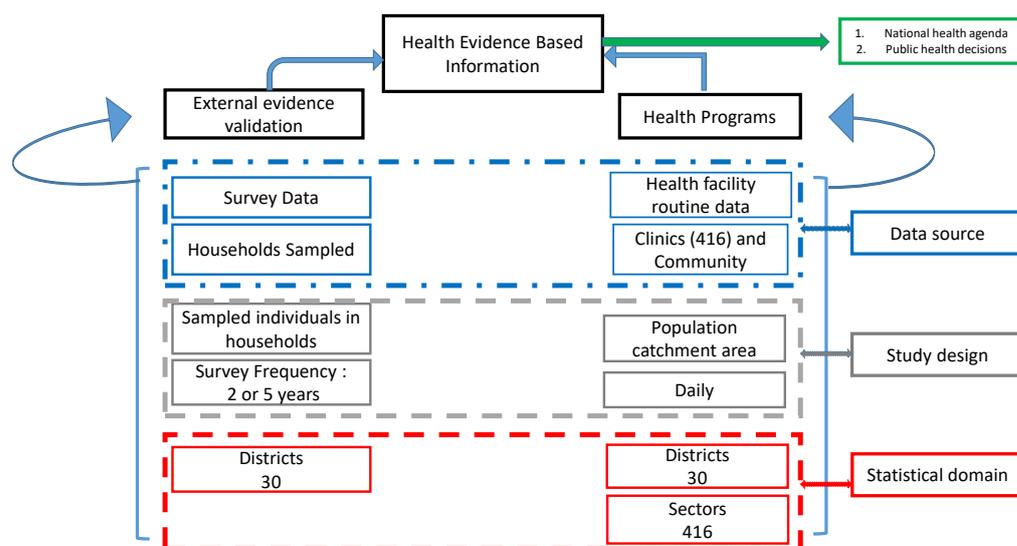


Figure 2. Data sources to inform evidence-based decision in Rwanda.

In this paper, we propose a method to account for both national surveys and routinely collected health facility data to estimate malaria relative risk incidence in small areas. Surveys are specifically designed to collect data on a specific topic and can be conducted with high levels of precision and accuracy. Updating routine data estimates with survey data can lead to more accurate and reliable estimates, especially for smaller geographic regions or subpopulations that may not be well represented in routine data. Updating routine data estimates with survey data can help to fill in gaps in the data and provide a more complete picture of the situation. Surveys can be designed to collect data on specific subpopulations or geographic regions that are not well represented in routine data. Updating routine data estimates with survey data can help to increase the coverage and representativeness of the data. Surveys can be designed to collect detailed information on a specific topic, allowing for more nuanced and in-depth understanding of the phenomenon being studied. Updating routine data estimates with survey data can help provide a more complete picture of the situation and support more informed decision-making.

This allows for both timely and local surveillance of malaria. A Bayesian spatio-temporal model is used for the analysis of the data.

2. Materials and Methods

2.1. Data Source

We extracted the routine health facility data from the Rwanda Health Information Management System (HMIS). These data include the number of malaria cases by health facility per month. The malaria cases are disaggregated by sex and age. More than 416 health facilities in Rwanda record malaria-positive cases in their respective catchment areas in the sector. This study focused on cases of malaria among children who are under five years old in each sector of Rwanda (416 sectors) for 2019 and 2020. The Rwanda Biomedical Centre granted permission to use the de-identified data for this research.

Rwanda Demographic Health Survey (RDHS) is a second data source. DHS is a nationally representative two-stage cluster sample designed to provide national, province, and district-level population and health indicators. Villages were the primary sampling unit (PSU), with 500 PSUs sampled, stratified by district. A total of 12,949 households were sampled. Our analysis focused on children under five who tested positive during the survey. The participants provided informed consent before participating in the survey. We were granted permission by the MEASURE DHS project (Authorization letter number 113068) to use these de-identified data for this analysis [7]. Table 1 summarizes the characteristics of the data used in this paper.

HHD stands for household details; the DHS collects household information, mainly demographic, education of parents, mosquito net use, etc.

2.2. Statistical Analysis

We suggest modeling malaria relative risk using surveys and routine data together to improve estimates. We modeled malaria risk in a two-step procedure: (1) fitting a binomial model using survey data, then extracting \hat{P}_{x_i} ; (2) assigning mean \hat{P}_{x_i} at a location (administrative level) in the Poisson model as an effect.

2.2.1. Binomial Model

For modeling malaria prevalence using DHS, we specified the model to predict malaria prevalence in Rwanda using the stochastic partial differential equation (SPDE) approach and integrated nested Laplace approximation with Bayesian inference.

The number of positive children under five years old Y_i out of N_i sampled people follows a Binomial distribution, with $P(x_i)$ being the prevalence at location i ($i = 1, 2, 3, \dots, n$)

$$Y_i | P(x_i) \sim \text{Binomial}(N_i, P(x_i)),$$

$$\text{logit}(P(x_i)) = \beta_0 + \beta_1 \text{Altitude}_i + S(x_i). \tag{1}$$

where β_0 denotes the intercept, β_1 is the coefficient of altitude, and $S(\cdot)$ is a spatial random effect that follows a zero-mean Gaussian process with Matérn covariance function $\text{Cov}(S(x_i), S(x_j)) = \frac{\sigma^2}{2^{v-1}\Gamma(v)} (\kappa \|x_i - x_j\|)^v K_v(\kappa \|x_i - x_j\|)$. In this formula, $K_v(\cdot)$ is the modified Bessel function of second kind and order $v > 0$ is the smoothness parameter, σ^2 denotes the variance, and $\kappa > 0$ is related to the practical range $\rho = \sqrt{\frac{8v}{\kappa}}$, which is the distance at which the spatial correlation is close to 0.1.

We refer to this model as model (1) (M1).

2.2.2. Modeling Routine Data Only

For modeling malaria relative risk using health facilities routine data only, we followed the approach proposed in the Bayesian spatio-temporal model for malaria risk in Rwanda [8].

Let Y_{jgr} be a number of under-five-year old malaria cases, the period in months $j = 1, \dots, 12$, gender $g = 1, 2$, and sectors $r = 1, \dots, 416$. The malaria cases data were counted and assumed to follow a Poisson distribution with the rate $n_{jgr}\lambda_{jgr}$, where n_{jgr} is the associated population count. Thus, the linear predictor is given by the following equation:

$$\eta_{jgr} = \log(\lambda_{jgr}) = \theta_g + \varphi_{jg} + u_r + v_r \tag{2}$$

where θ_g is the gender-specific effect, and φ_{jg} is a gender-period effect.

The last two terms are related to the spatial domain. Structured and unstructured spatial components were considered in this model; thus, $u = 1, \dots, 416$ stands for spatial structured and $v = 1, \dots, 416$ for unstructured components. This model is referred to as model (2) (M2).

The unstructured spatial random effects, $\mathbf{v} = (v_1, \dots, v_{416})^T$, are assumed to be independent and identical (iid) normally distributed with an unknown variance σ_v . Thus, the formulation can be written as iid $v_r \sim N(0, \sigma_v^2)$ [9].

The neighboring regions might have a similar malaria risk. Therefore we model $\mathbf{u} = (u_1, \dots, u_{416})^T$ using an intrinsic Gaussian Markov random field (GMRF). The Besag model includes a spatial random effect. The spatial random effects are assigned an intrinsic conditional autoregressive (ICAR) prior, which can be specified in a general form as $u_i | u_j, i \neq j, \tau \sim N(\frac{1}{n_i} \sum_{i \sim j} u_j, \frac{1}{n_{i,\tau}})$, where n_i is the number of neighbors of node i , and $i \sim j$ indicates that the two nodes are neighbors. We defined the neighboring matrix as spatial

proximity of the locations where malaria cases have been observed or are being predicted. The neighboring matrix was used to define a spatial weighting matrix, which is used to weight the contributions of neighboring locations to the prediction of malaria risk at a given location. For more details see [10,11].

For unstructured spatial effect \mathbf{v} , we assigned a gamma distribution with a shape equal to 0.5 and rate equal to 0.00149, such as $\sigma_V^2 \sim G(0.5, 0.00149)$. For the other parameters, we used marginal standard deviations to assign a prior distribution to scaled precision parameters.

2.2.3. Modeling Routine Data with Survey Data

To improve the estimates of malaria in children under five years old in Rwanda, we added information from RDHS to the analysis of routine data. We did this by adding the estimated prevalence based on DHS as a non-linear function in model M2. To do so, we needed an estimate of malaria prevalence \hat{P} from model M1 at the sector level. This was obtained by averaging the posterior mean from model M1 at locations $x \in r$, i.e., $\hat{P} = \frac{1}{n(xcr)} \sum_{xcr} \hat{P}(x)$, and the model is given as:

$$\eta_{jgr} = \log(\lambda_{jgr}) = \theta_g + \varphi_{jg} + f(\hat{P}_r) + u_r + v_r \quad (3)$$

where $f(\hat{P}_r)$ is modeled as an RW2 model after grouping \hat{P}_{x_i} in 20 bins.

This model is referred to as model (3) (M3).

2.3. Bayesian Inference

The Bayesian hierarchical model in which prior distributions were assigned to all model parameters was applied. Independent diffuse priors were used for gender-specific intercepts, and independent smoothing priors for the gender effects. Then, a smoothing prior based on the second-order walk was assigned to each parameter. Recall that a second-order random walk penalized deviations from a linear trend. It was used to increase the precision [12]. The Deviance Information Criterion (DIC) proposed by Spiegelhalter [13] as a tool for model selection in a Bayesian context was used in this analysis. It is written as:

$$DIC = D(\bar{\theta}) + 2pD \quad (4)$$

where $D()$ is the deviance and $\bar{\theta}$ is the posterior mean vector of unknown parameters of the model. The deviance measures how well the model fits the data, with larger values corresponding to a worse model. The parameter pD is the adequate number of parameters and measures the complexity of the model. The parameter is defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters. The rule of thumb for using DIC in model selection is roughly the same as for Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC); namely, a difference in DIC of more than ten rules out the model with higher DIC, while with a difference of less than five, there is no clear winner [14]. In order to verify whether the results obtained are not dependent on underlying prior specifications, sensitivity analysis was performed. Four different configuration priors were constructed to see whether estimates are not sensitive to priors. The conditional predictive ordinate for each observation was computed to check for the possible outlying observations in y_j [15].

3. Results

3.1. Data Exploratory Analysis

The malaria incidence rate for children under five years old increased among both males and females in 2020 as compared to the incidence rate in 2019. The incidence rate increased from 50 to about 250. Figure 3 shows a consistent pattern of the positivity rate among males and females. The positivity rate was relatively constant in 2019. In 2020, the pattern of malaria transmission began to rise from September, reached its highest point in November. The implementation of indoor residual spraying (IRS) in the three districts of

Rwanda with the highest risk of malaria, Kirehe, Ngoma, and Nyagatare, in September 2019 and January 2020, has been reported to have a significant impact on reducing the incidence of new malaria cases in a short amount of time [16,17].

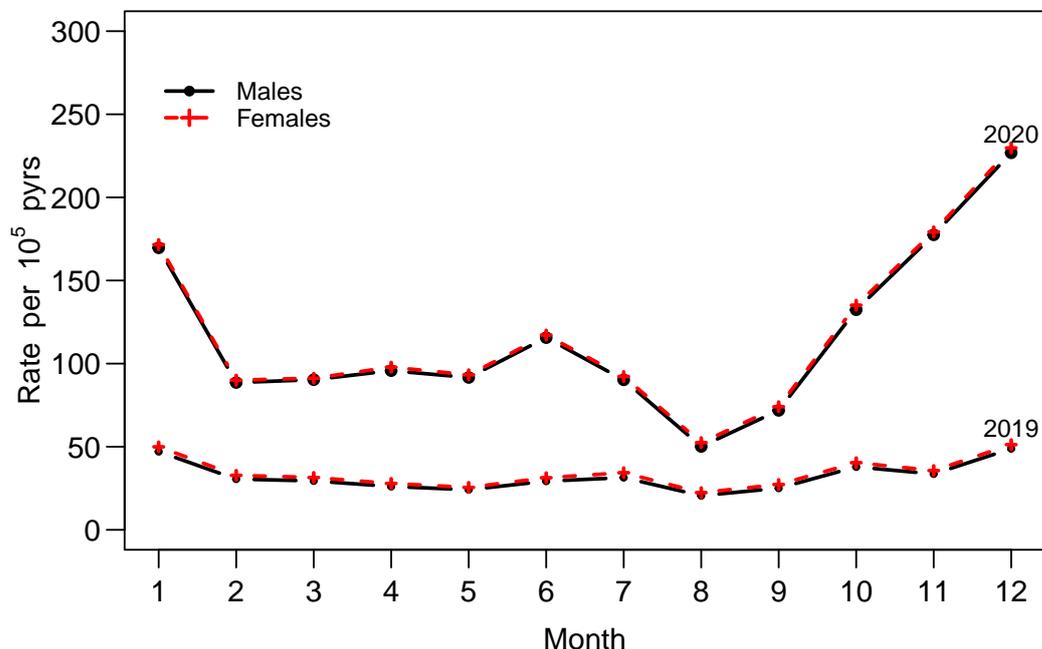


Figure 3. Incidence rate of malaria among children under five years old per 100,000 population. Health facility routine data for year 2019 and 2020.

For DHS 2019–2020, A total of 3665 children under five years old were selected in households surveyed and tested for malaria. Out of 3665 children tested, only 99 tested positive for malaria in Rwanda.

3.2. Main Findings

The main findings are presented in two parts. The first part provides malaria estimates from the binomial model, and the second provides estimates from joint Spatio-temporal models. We discussed the advantages of using survey data to improve health facility routine data estimates for decision-making in the era of the malaria elimination goal.

For the survey data, model M1 was used. Note that this is a spatial model, as DHS is a cross-sectional design. For the hospital routine data, models M2 and M3 were compared using DIC. M2 had a DIC of 831,071, while M3 had a DIC of 822,394. Therefore, model M3 is a more preferred model. The model used in this study is spatio-temporal, as monthly routine data were collected and analyzed over both space and time. In order to address the limitations of routine data, the study also incorporated less frequently obtained Demographic and Health Survey (DHS) data. The spatio-temporal M3 model, which incorporated the DHS data, was found to be superior to the M2 model that did not use the DHS data.

The M3 model provides a more comprehensive understanding of malaria trends, as it can capture both short-term and long-term changes over space and time. Therefore, the spatio-temporal M3 model is a valuable tool for monitoring and predicting malaria incidence in a given area.

Table 2 shows the parameter estimates of these final models.

Table 2. Malaria risk parameter estimates for M1 and M3.

| Models | M1 | | | M3 | | |
|---------------------|----------------|--------|--------|----------------|--------|--------|
| Parameters | Est (sd) | LL | UL | Est (sd) | LL | UL |
| Fixed Effect | | | | | | |
| μ_0 | -2.584 (1.069) | -4.796 | -0.588 | | | |
| μ_{Male} | | | | -5.892 (0.047) | -5.984 | -5.800 |
| μ_{Female} | | | | -5.865 (0.047) | -5.957 | -5.773 |
| $\beta_{1Altitude}$ | -0.002 (0.001) | -0.003 | 0.001 | | | |
| Random Effect | | | | | | |
| u_r | -4.62 (0.268) | -5.14 | -4.62 | 0.584 (0.124) | 0.380 | 0.866 |
| v_r | 2.69 (0.186) | 2.33 | 3.06 | 1.242 (0.192) | 0.905 | 1.658 |
| φ_j | | | | 15.59 (4.72) | 8.025 | 26.420 |
| $varf(\hat{P}_r)$ | | | | 3.072 (0.975) | 1.538 | 5.336 |

The posterior summaries, including the posterior mean (estimate) and its standard deviation (sd), lower limit (LL), and upper limit (UL), are presented.

Figure 4 shows spatial variation of malaria prevalence among under-five-year old children in Rwanda.

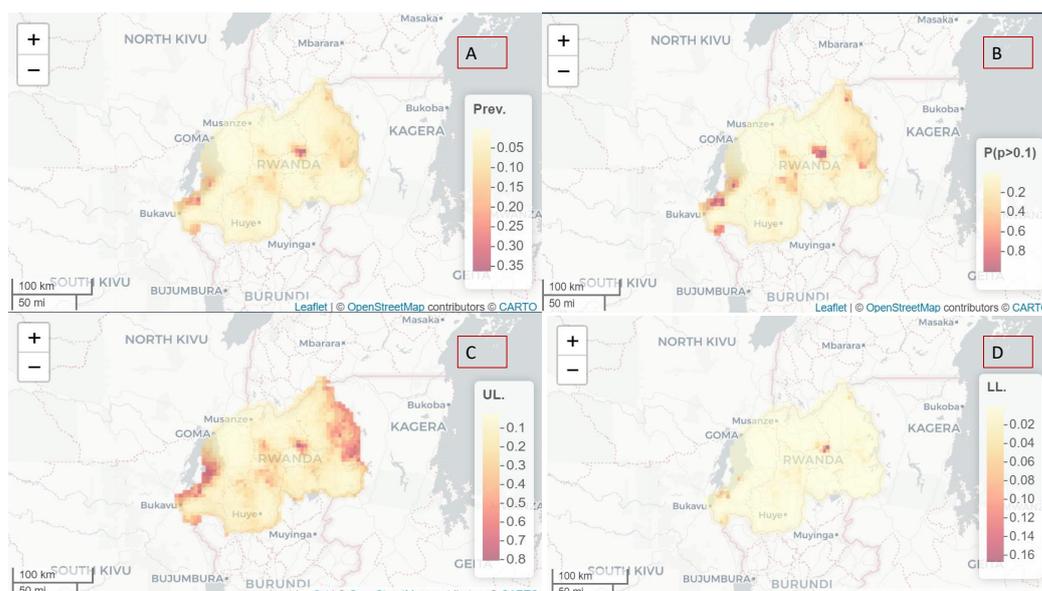


Figure 4. Spatial variation of malaria prevalence among children under five years old in Rwanda, RR posterior means of spatial effects (A) with 95% UL upper limit (C). Lower limit (D). Panel (B) shows exceeding probability of prevalence above 10% (B). Source: DHS 2019–2020.

The estimation of malaria prevalence among those under five years old using Rwanda demographic and health survey data 2019–2020 showed a higher prevalence in the south-west, central, and northeast of Rwanda compared to the rest of the country. The upper limit (UL) prevalence map shows a similar pattern to the prevalence estimated though the intensity is different. The UL shows the worst scenario. The hot spot areas were identified using the exceeding probability of prevalence above 10%.

Figure 5 shows spatial variation of malaria relative risk incidence among under-five-year old children in Rwanda using DHS and routine health clinical data combined. Figure 5 provides additional clusters as compared to the results in Figure 4; the last one did not show the southern areas of Rwanda among the high-risk malaria zones. Using M3, we identified new malaria clusters in the southern area of Rwanda for the years 2019 and 2020. Those clusters had a more significant RR than 35 (RR > 35), with an increase from 2019 to 2020

in those specific clusters. The relative risk (RR) quantifies whether the area i has a higher or lower occurrence of cases than the expected from reference rate. The estimated RR quantifies the heterogeneity of the risk and highlights unusual patterns of risks. The upper and lower limits are provided through the 95% credible interval.

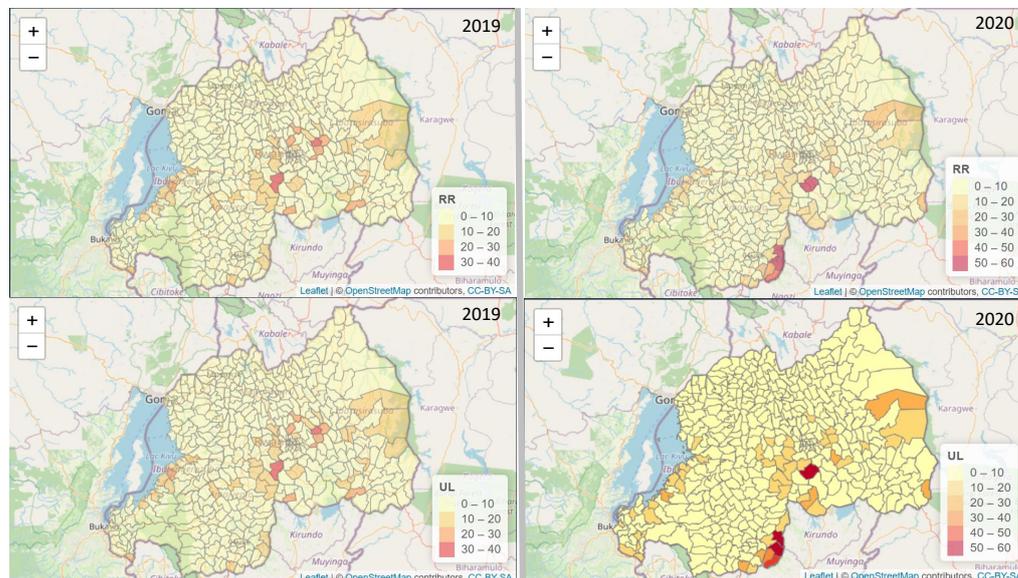


Figure 5. Spatial variation of malaria relative risk incidence among children under five years old in Rwanda (RR), posterior means of spatial effects with 95% upper limit (UL). Source: health facility routine data from 2019 and 2020.

Figure 5 shows Rwanda's southern, southeastern, southwestern, and central regions as high-risk malaria areas. The estimates are very localized at the sector level with greater precision than the DHS data. The M3 provides the advantage of detecting new clusters not identified using only DHS survey data. We identified more than nine sectors (Mamba, Gishugi, Muganza, Mukingo, Mugombwa, Nyanza, Nyanza, Kansi, Nora, Nyanza) with $RR > 40$, and two sectors (Gishubi, Muganza) with an $RR \geq 50$ in southern areas of Rwanda that were not detected using DHS survey 2019–2020 data.

This modeling approach enables the use of all available information and minimizes the loss of information. For this case, we can also evaluate malaria's relative risk over months. Malaria risk is very localized and requires more localized interventions. We compared malaria risk in January, June, and December for 2019 and 2020. We observed that overall malaria incidence increased in 2020, but sectors at risk were fewer (30 sectors) than those at risk in 2019 (52 sectors).

We assessed the posterior temporal trend effect for malaria relative risk ($\exp(\phi_t + \gamma_t)$) for condition autoregressive (CAR) for 2019 and 2020. Figure 6a,b shows CAR models with a 95% credible interval.

The general trend of malaria risk reveals a rise in malaria risk from September to December 2019, and then a decrease from April to July 2020. However, there was an increase in cases in August, reaching a peak in November 2020.

The 95% credible trend effects intervals were narrow for 2019 and 2020. We observed no difference in risk between males and females. This approach to modeling malaria relative risk indicates a bottleneck analysis of establishing a Malaria trend surveillance system that provides evidence that public health experts can use to elaborate an area-specific strategy. It is crucial to use M3, which combines survey and health facility routine data to enhance or improve detailed and localized estimates at sector levels.

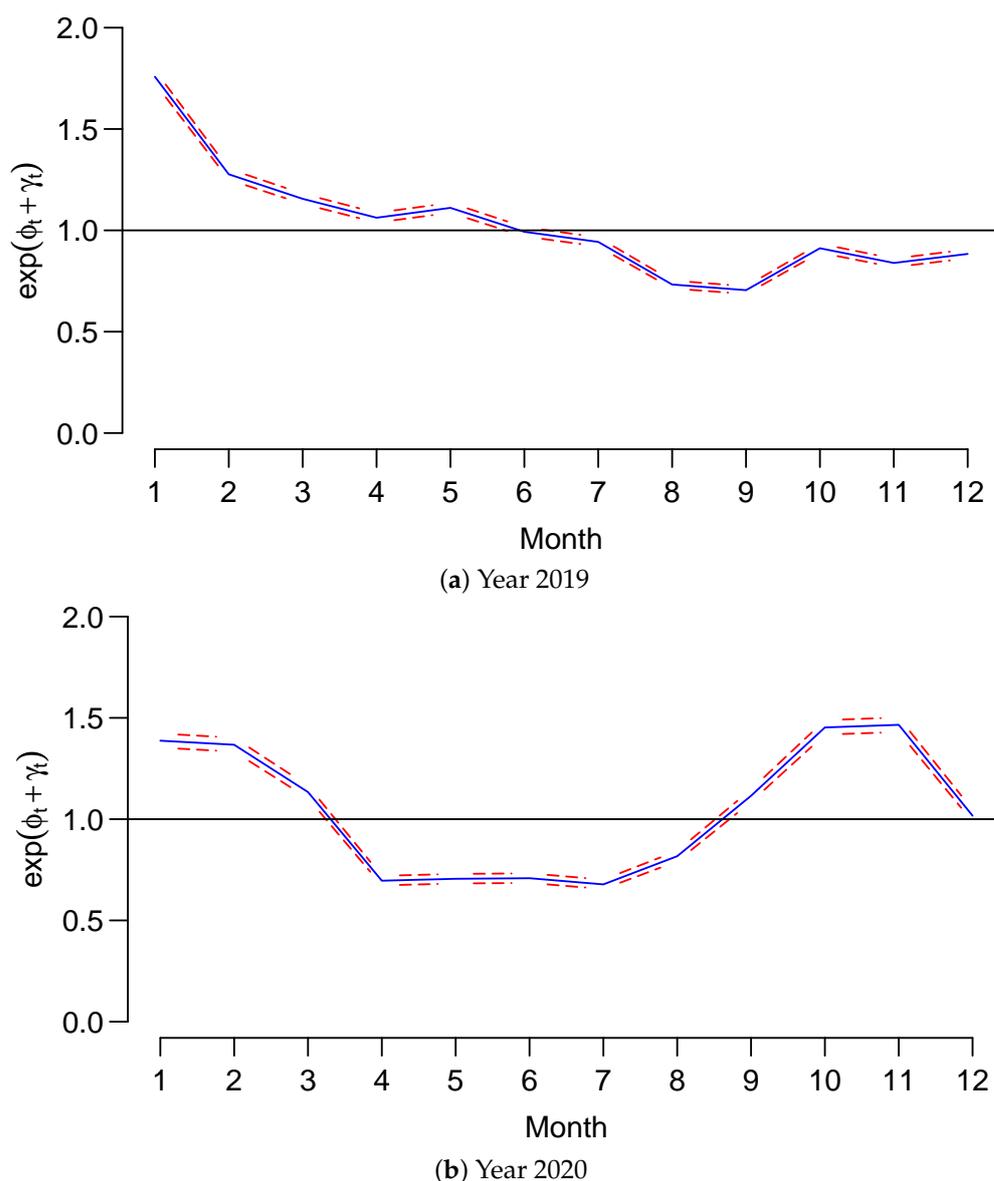


Figure 6. Posterior temporal trend effect for malaria relative risk: $\exp(\phi_t + \gamma_t)$ with 95% credible interval, year 2019.

4. Discussion

Our results present a novel approach to modeling health facility data and survey data to improve estimates of malaria relative risk in small areas. Most developing countries use DHS data as a primary data source to estimate key health indicators such as malaria prevalence. In the last decade, there have been emerging findings recommending the use of routinely reported clinical data for program planning and evaluating impact. However, most of those studies reported data quality concerns as a limitation, mainly data completeness, timeliness, accuracy, consistency, and poor utilization of Health Management Information System (HMIS) tools [18]. The introduction of web-based information systems for health facility data and the implementation of universal health policy improved the completeness and accuracy of data at the local level and population-based statistics based on those data. The most recent studies have reported improved data quality and accuracy of HMIS in some countries [8,19,20]. Sub-Saharan Africa (SSA) countries have adopted the HMIS that includes data quality control checks to ensure high quality. Rwanda, as one of the East African countries, has been using HMIS since 2011. Generally, data from health facilities are of high quality due to consistent data quality checks and feed-

back between national and decentralized levels [8]. Ideally, health programs are built on evidence-based practice through successfully integrating data into health policy. Decision making throughout the health system is an ongoing challenge.

Low- and middle-income countries use DHS data as an alternative source for planning and evaluating impact. DHSs are cross-sectional studies conducted every two or five years. The DHS data are of high quality and enable comparison of studies using DHS data to ascertain commonalities or unusual patterns and to compare countries or regions. DHS surveys provide population-based coverage data of key health service indicators. These indicators can be disaggregated, and differentials can be assessed according to geographic, biodemographic, and socioeconomic characteristics. However, DHS survey data have some limitations. Desegregation in small geographic areas is desirable in malaria elimination stages since the unit of implementation of health programs or intervention is dynamic downwards. DHS surveys are generally not designed to yield estimates of health indicators at the sector level since this is too costly. The frequency of availability of new data is between two and five years. Lastly, DHS data does not allow trend analysis [3,21]. DHS surveys remain an important source of health data in developing countries at the national and international levels.

Different solutions were proposed to improve estimates from health facility routine data, from more straightforward approaches such as data triangulation, developing a robust algebra estimator by adjusting for testing practices, care-seeking behavior, and reporting completeness to complex models that account for uncertainties [8,22–24]. Both health facility routine and survey data are important sources of health data; therefore, using both data sources is crucial.

In this paper, we proposed a novel approach of extracting fitted values from a binomial model using survey data and using them as random effect covariates in the poisson model to estimate the relative risk of malaria at sector administrative levels and evaluate trends over time. The M3 model incorporating DHS estimates in modeling malaria relative risk incidence using routine case data yielded a substantial variation within and between sectors' relative risk. The overall trend effect of malaria relative risk varied by month. We observed two sensitive malaria risk periods in Rwanda, namely April–June and September–December. The model allows to evaluate malaria relative risk per month, in contrast to cross-sectional surveys. The model takes advantage of cross-sectional survey data to improve area-specific and time-specific estimates from routine data.

Though the time effect component is vital to understand malaria epidemiology, area-specific risks can inform the national malaria control program to conduct effective and efficient strategies for reducing malaria incidence at the subnational level. Our findings show malaria relative risks per sector (administration level 3) in Rwanda.

The areas detected through combining DHS survey and routine data (Mamba, Gishugi, Muganza, Mukingo, Mugombwa, Nyanza, Nyanza, Kansi, Nora, Nyanza, Gishubi, Muganza, and Mukingo in the southern areas of Rwanda) neighbor high-risk malaria areas in Burundi, as Figure 1 shows. Not detecting a high-risk area might lead to classifying southern areas of Rwanda among areas at low risk of malaria during the planning process and a failure to develop area-specific interventions. The findings show an overall increase in malaria in 2020, but the distribution of cases was more localized in a few sectors as compared to the previous year, 2019. DHS survey data are often used in Rwanda for planning and evaluating the impact of interventions conducted to prevent malaria [25]. The malaria incidence from a model using surveys and routine health facility data allows a comparison of incidence between sectors and an evaluation of trends over time. The two-step modeling approach allowed us to quantify the trends in changes in relative risk incidence in every small area over time.

The modeling approaches presented in this paper enable estimates at small administrative areas that only DHS data cannot provide because DHS surveys are designed and empowered to provide estimates at the domain level. Most of the time, the domain levels in the survey are second-higher administrative entities, such as districts in Rwanda. A survey

such as DHS is costly; sampling in small areas might not be practical due to time and cost related. DHS data collection does not include all months of the year; it usually targets high and low-malaria-risk months in a year. Though the approach of targeted sample collection on space and time is important, the left-out areas and months during sample collection for seasonal diseases might be a limitation in capturing the true picture of the disease in small areas. The Rwanda DHS 2019–2020 data collection process occurred in November and December 2019 and January, February, March, June, and July 2020. The posterior trend effect showed that malaria risk rose in April, May, September, and October. The peak of malaria risk was in June and December. The two-step modeling approach can provide a complete picture of malaria in space and over time using both survey and routine data. The strength of both sources of data overcomes the limitations of both sources. The health facility routine data have a limitation in quality and completeness. Routine data might not include those who did not seek health services in health facilities and in the community. Malaria is primarily managed by Community Health Workers (CHWs) at the village level who are trained to provide essential health services in their communities in Rwanda [26]. DHS is a household survey, and there is an equal chance of sampling both people who seek health services and those who do not. Therefore, taking advantage of routine and DHS survey data through statistical modeling contributes to a better understanding of the malaria burden. The proposed modeling approaches enable monthly trend effect analysis and small-area disease mapping.

The World Health Organization (WHO) issued a new final phase of malaria elimination and related appropriate strategies in June 2022. Three categories of possible interventions were identified: (1) “mass” strategies applied to the entire population of a delimited geographical area such as a district, sector, cell, or village; (2) “targeted” strategies applied to people at increased risk of infection compared to the general population; and (3) “reactive” strategies implemented in response to individual cases [27]. All three strategies would benefit from active surveillance with real-time data. The WHO and groups of experts recommend the establishment of a malaria real-time data surveillance system to eliminate malaria [8,27]. It is crucial to empower routine data with DHS estimates to establish an active malaria surveillance data system with real-time data.

The modeling approach presented in this paper contributes to the utilization of all data sources to gain improved estimates at small areas and quantify the trend effects.

5. Conclusions

In summary, we recommend using both the DHS survey and routine data in a two-step modeling framework for spatio-temporal models as one of the tools for active malaria surveillance to identify high-risk areas, monitoring malaria trends, prediction of new cases and evaluation of malaria incidence at subnational levels. This approach enables the use of cross-sectional survey data and health facility routine data in modeling relative malaria risks, as well as generating malaria risk maps that provide information on the spatio-temporal trend effects of malaria at subnational levels. The proposed approach is not limited to malaria cases and can also be applied to other infectious disease monitoring and evaluation at small areas.

We discussed the advantage of using multiple data sources in a two-step modeling framework. We modeled malaria relative risk among under-five-year old children in Rwanda. We compared findings from geostatistical modeling of malaria prevalence among under-five-year old children using demographic and health survey 2019–2020 (M1) data and modeling malaria relative risk using both DHS survey 2019–2020 and health facility routine data. The model provided the estimates at sector levels and enabled analysis of malaria relative risk trend effects. The findings show that the model that uses DHS survey and routine data (M3) identified more high risk areas in the southern province of Rwanda that were not detected using only survey data.

Both survey and routine data are used to establish efficient permanent monitoring of the variation of malaria risks in areas and evaluate interventions towards the final phase

of malaria elimination. The strength of routinely collected data at small scales and high-quality data from the survey should contribute to a better understanding of the malaria burden and support evidence-based decision making.

Author Contributions: Conceptualization, M.S.; methodology, M.S. and C.F.; software, M.S.; validation, C.F., F.N. and M.S.; formal analysis, M.S.; investigation, M.S.; resources, M.S.; data curation, M.S.; writing—original draft preparation, M.S.; writing—review and editing, C.F.; visualization, M.S.; supervision, C.F.; project administration, M.S.; funding acquisition, C.F. All authors have read and agreed to the published version of the manuscript.

Funding: The last author acknowledges funding from the European Union’s Horizon 2020 research and innovation program—project EpiPose (No. 101003688).

Institutional Review Board Statement: The use of the routine program data was approved by the Rwanda National Ethics Committee hosted under The Rwanda Ministry of Health, which also granted approval for data access and use to the principal investigator (M.S.). The data presented in this study were accessed directly in the national reporting system. A consent waiver was obtained from Rwanda Ethics Committee. For DHS data, participants provided informed consent before participating in the survey; we were granted permission by the MEASUREDHS Project to use these de-identified data for this analysis.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the DHS survey.

Data Availability Statement: All data used in this paper can be accessed at <https://github.com/semakulam> (accessed on 5 May 2022).

Acknowledgments: The authors are grateful to the Rwanda Biomedical Centre (RBC), national institute of statistics of Rwanda and DHS Macro for making available the data. We are grateful to the Rwanda Ministry of Health and RBC for authorizing the principal investigator to access data and sharing additional materials to understand strategies used to control malaria in Rwanda. Author M.S. acknowledges University of Rwanda, African centre of excellence in data science for supporting his PhD researches.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. World Malaria Report 2021. Global Report. 2021. Available online: <https://www.who.int/publications/i/item/9789240040496> (accessed on 1 May 2022).
2. ICF International. Spatial Data Repository. The Demographic and Health Surveys Program. Available online: <https://spatialdata.dhsprogram.com> (accessed on 16 September 2022).
3. Ties, Boerma, J.; Sommerfelt, A.E. Demographic and health surveys (DHS: Contributions and limitations. *World Health Stat. Q.* **1993**, *46*, 222–226.
4. Dieudonne, H.; Ntizimira, C.; Mbituyumuremyi, A.; Hakizimana, E.; Mahmoud, H.; Birindabagabo, P.; Musanabaganwa, C.; Gashumba, D. The impact of COVID-19 on malaria services in three high endemic districts in Rwanda: A mixed-method study. *Malar. J.* **2022**, *21*, 48.
5. Wagenaar, B.H.; Hirschhorn, L.R.; Henley, C.; Gremu, A.; Sindano, N.; Chilengi, R. Data-driven quality improvement in low-and middle-income country health systems: Lessons from seven years of implementation experience across Mozambique, Rwanda, and Zambia. *BMC Health Serv. Res.* **2017**, *17*, 830. [[CrossRef](#)] [[PubMed](#)]
6. WHO. *Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment*; World Health Organization: Geneva, Switzerland, 2017.
7. National Institute of Statistics of Rwanda; Ministry of Health [Rwanda]; ICF International. *Rwanda Demographic and Health Survey 2019–2020*; NISR: Kigali, Rwanda; MOH: Kigali, Rwanda; ICF International: Reston, VA, USA, 2020.
8. Semakula, M.; Niragire, F.; Faes, C. Bayesian spatio-temporal modeling of malaria risk in Rwanda. *PLoS ONE* **2020**, *15*, e0238504. [[CrossRef](#)] [[PubMed](#)]
9. Birgit, S.; Leonhard, H. Spatio-temporal disease mapping using INLA. *Environmetrics* **2011**, *22*, 725–734.
10. Håvard, R.; Sara, M.; Nicolas, C. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2009**, *71*, 319–392.
11. Julian, B.; Peter, G.; David, H.; Kerrie, M. Bayesian computation and stochastic systems. *Stat. Sci.* **1995**, *10*, 3–41.
12. Andrea, R.; Leonhard, H.; Håvard, R.; Matthias, B. Gender-specific differences and the impact of family integration on time trends in age-stratified Swiss suicide rates. *J. R. Stat. Soc. Ser. A (Statistics Soc.)* **2012**, *175*, 473–490.

13. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. (Stat. Methodol.)* **2002**, *64*, 583–639. [[CrossRef](#)]
14. Lesaffre, E.; Lawson, A.B. *Bayesian Biostatistics*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
15. Pettit, L.I. The conditional predictive ordinate for the normal distribution. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **1990**, *52*, 175–184. [[CrossRef](#)]
16. Hemingway, J.; Shretta, R.; Wells, T.N.; Bell, D.; Djimdé, A.A.; Achee, N.; Qi, G. Tools and strategies for malaria control and elimination: What do we need to achieve a grand convergence in malaria? *PLoS Biol.* **2016**, *14*, e1002380. [[CrossRef](#)] [[PubMed](#)]
17. Rwanda Biomedical Centre. *Indoor Residual Spraying (IRS) and Long-Lasting Insecticide-Treated Nets (LLINs) Distributed in High Malaria Burden Districts*; RBC Report; Rwanda Biomedical Centre: Kigali, Rwanda, 2020.
18. Rumisha, S.F.; Lyimo, E.P.; Mremi, I.R.; Tungu, P.K.; Mwingira, V.S.; Mbata, D.; Malekia, S.E.; Joachim, C.; Mboera, L.E.G. Data quality of the routine health management information system at the primary healthcare facility and district levels in Tanzania. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 340. [[CrossRef](#)] [[PubMed](#)]
19. Sabella, M.; Pepela, W.; Soti, D.; Hillary, K.; Benson, D.; Ties, B. Using health-facility data to assess subnational coverage of maternal and child health indicators, Kenya. *Bull. World Health Organ.* **2017**, *95*, 683–694.
20. Agiraembabazi, G.; Ogwal, J.; Tashobya, C.; Kananura, R.M.; Boerma, T.; Waiswa, P. Can routine health facility data be used to monitor subnational coverage of maternal, newborn and child health services in Uganda? *BMC Health Serv. Res.* **2021**, *21*, 512. [[CrossRef](#)] [[PubMed](#)]
21. Wilson, E.; Hazel, E.; Park, L.; Carter, E.; Moulton, L.H.; Heidkamp, R.; Perin, J. Obtaining district-level health estimates using geographically masked location from Demographic and Health Survey data. *Int. J. Health Geogr.* **2020**, *19*, 2. [[CrossRef](#)] [[PubMed](#)]
22. Adane, A.; Adege, T.M.; Ahmed, M.M.; Anteneh, H.A.; Ayalew, E.S.; Berhanu, D.; Berhanu, N.; Beyene, M.G.; Bhattacharya, A.; Bishaw, T.; et al. Routine health management information system data in Ethiopia: Consistency, trends, and challenges. *Glob. Health Action* **2021**, *14*, 868–961. [[CrossRef](#)] [[PubMed](#)]
23. Sawadogo-Lewis, T.; Keita, Y.; Wilson, E.; Sawadogo, S.; Térera, I.; Sangho, H.; Munos, M. Can We Use Routine Data for Strategic Decision Making? A Time Trend Comparison Between Survey and Routine Data in Mali. *Glob. Health Sci. Pract.* **2021**, *9*, 869–880. [[CrossRef](#)] [[PubMed](#)]
24. Thwing, J.; Camara, A. C.; rinho, B.; Zulliger, R.; Colborn, J.; Painter, J.; Plucinski, M.M. A Robust Estimator of Malaria Incidence from Routine Health Facility Data. *Am. J. Trop. Med. Hyg.* **2020**, *102*, 811–820. [[CrossRef](#)] [[PubMed](#)]
25. Nzabakiriraho, J.D.; Gayawan, E. Geostatistical modeling of malaria prevalence among under-five-year old children in Rwanda. *BMC Public Health* **2021**, *21*, 369. [[CrossRef](#)] [[PubMed](#)]
26. Masimbi, O.; Schurer, J.M.; Rafferty, E.; Ndahimana, J.D.A.; Amuguni, J.H. A cost analysis of the diagnosis and treatment of malaria at public health facilities and communities in three districts in Rwanda. *Malar. J.* **2022**, *21*, 150. [[CrossRef](#)] [[PubMed](#)]
27. World Health Organization. *Global Malaria Program*; WHO: Geneva, Switzerland, 2022.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.