

Bayesian time series regression methods for estimating national immunization coverage

Final Report

C. Edson Utazi, Sujit K. Sahu and Andrew J. Tatem

September 2020

Bayesian time series regression methods for estimating national immunization coverage

C. Edson Utazi (WorldPop, University of Southampton)
Sujit K. Sahu (Mathematical Sciences, University of Southampton)
Andrew J. Tatem (WorldPop, University of Southampton)

Address for Correspondence: Dr. C. Edson Utazi
Senior Research Fellow
WorldPop, School of Geography and
Environmental Science
Faculty of Environmental and Life Sciences
University of Southampton
SO17 1BJ
Southampton
Tel: +44(0)7956985992
Email: c.e.utazi@soton.ac.uk

September 2020

Acknowledgements

This project was funded by WHO and carried out at WorldPop, University of Southampton. All data were provided by WHO. The authors are grateful for the support provided to them by WorldPop colleagues and WHO working group during the implementation of this project.

Bayesian time series regression methods for estimating national immunization coverage

Summary

National estimates of immunization coverage are crucial for monitoring and evaluating coverage levels and trends, as well as immunization goals and targets, at the national and international levels. The World Health Organization (WHO) and the United Nations International Children's Emergency Fund (UNICEF) produce national estimates of immunization coverage annually for 195 countries and different vaccine-dose combinations from multiple data sources using a deterministic computational logic approach. However, this deterministic approach is incapable of characterizing uncertainties related to coverage measurement and estimation. In addition, the method provides no statistically-principled way of accounting for and exploiting the different sources of dependence that may exist in immunization coverage data collected for multiple vaccines, years and countries.

The aim of the work undertaken here is to develop a novel Bayesian statistical modelling approach for producing accurate estimates of national immunization coverage and associated uncertainties for multiple vaccines and all WHO countries. Our methodology comprises a robust, flexible framework implemented in two phases. First, *reported administrative coverage* estimates are processed using demographic data, recall-bias-adjusted survey data, official estimates and other contextual data. In the second phase, the logit-transformed processed *reported administrative coverage* data are modelled using Bayesian Gaussian time series regression models that explicitly account for temporal correlation, but also leverage correlations among multiple vaccines and countries to boost predictive performance. We developed five candidate models which are fitted independently for each WHO region. We tested and compared the models based on their ability to predict both in-sample and out-of-sample data using various model performance criteria. The methodology is fully implemented in R using the R-INLA package.

We present and discuss modelled coverage estimates and associated uncertainties for the period 2000 – 2017 and predictions for 2018 – 2019, using the best model parameterization. Results from different cross-validation exercises showed that the model performed well, both in terms of predicting in-sample (correlation ≥ 0.74) and out-of-sample (correlation ≥ 0.65) data. Across all five vaccines, substantial increases in coverage were generally observed between 2000 and 2010. However, these improvements in coverage appear to have levelled off, slowed down or, in some cases, regressed in the period from 2010 – 2017. Modelled estimates are shown to be precise, with $\approx 71\%$ of the in-sample predictions having 95% prediction interval widths $\leq 20\%$. In general, the approaches developed here are promising, producing plausible coverage estimates in most cases.

Building on this work, the processed data used for modelling can be further refined and improved through the inclusion of information on vaccine stocks or stock-outs (perhaps, for numerator adjustments) and independent reviews of the processed data.

Furthermore, the modelling framework can be extended to include covariate information that can help improve the modelled estimates, particularly in countries where there is evidence of over-smoothing or inadequate model performance. It is straightforward to scale up the methodology to include other vaccines estimated by WHO and UNICEF. If the approach developed here were to be adopted as a standard method for producing modelled estimates of national immunization coverage, the development of an R package would greatly enhance its use.

1. Introduction

National estimates of immunization coverage are crucial for monitoring and evaluating coverage levels and trends, as well as immunization goals and targets, at the national and international levels. Over the years, these statistics have guided and informed policies aimed at the control, elimination and eradication of vaccine-preventable diseases [1]. In particular, immunization coverage is an important indicator for measuring progress towards Sustainable Development Goal (SDG) 3, which seeks to achieve universal access to safe, effective, quality and affordable vaccines for all [2]. The World Health Organization (WHO) and the United Nations International Children's Emergency Fund (UNICEF) currently produce national estimates of immunization coverage annually for 195 countries and territories and different vaccine-dose combinations [1, 3].

Administrative estimates of national immunization coverage are reported as the percentage of the target population that has been vaccinated. The target population used in the calculation of coverage varies by vaccine and is usually dependent on the national immunization schedule [4]. Countries report administrative estimates of coverage annually to WHO and UNICEF through the Joint Reporting Form (JRF) [5]. This information is supplemented with survey data and other demographic and contextual data to produce WHO and UNICEF estimates of national immunization coverage (WUENIC) using a deterministic computational logic approach [3, 6]. This deterministic approach is, however, incapable of characterizing uncertainties related to both the input data sets and estimates of coverage that are produced. The method provides no statistically-principled way of accounting for and exploiting the different sources of variation/dependence that may exist in coverage data collected for multiple vaccines, years and countries. Furthermore, the approach includes no mechanism for the prediction of immunization coverage for future time points which may often be needed to guide program planning.

The overarching aim of this work is to develop a novel Bayesian statistical methodology that leverages temporal and spatial variation in immunization coverage and correlations among vaccines to produce accurate national-scale, annual point and interval estimates of immunization coverage. The methodology adapts specific elements of the WUENIC approach for processing *reported administrative coverage* data through supplementation and refinement using other available data sets. It then utilizes a Bayesian statistical modelling framework to smooth the processed data and make predictions for future time points, whilst accounting for the uncertainties associated with both the data and the predictions. We develop a model ensemble comprising five candidate models, which differ according to the complexities of the correlation structures used for modelling the different sources of variation in the data. The specific objectives of the work are:

1. Assemble, process and validate all available data for estimating immunization coverage at the national level.
2. Develop a robust and replicable statistical methodology for producing national estimates of immunization coverage from multiple input data.
3. Evaluate the performance of the proposed method for both in- and out-of-sample predictions using cross-validation techniques.

4. Develop R code for implementing the proposed methodology.
5. Delivery of modelled coverage estimates, software and documentation of the proposed methodology.

In this report, we address these objectives, from data assembly to model development and implementation, culminating in the production of modelled estimates of national immunization coverage for all WHO countries. We first describe the various data sets used in the work and detail how these were utilized to produce processed *reported administrative coverage* estimates which served as input data to the Bayesian models. Our analyses focus on five vaccines, namely DTP1, DTP3, MCV1, MCV2 and PCV3, and all WHO countries for which input data were available during 2000 – 2017 (see Figure A1 - Appendix). We then describe the modelling framework, detailing how the model is fitted using the integrated nested Laplace approximation (INLA) [7, 8] approach and metrics used for evaluating the performance of the models for predicting both in- and out-of-sample data using cross-validation techniques. We also present and discuss modelled estimates of national immunization coverage based on the best-fitting model, as well as predictions for 2018 and 2019.

2. Data description and processing steps

Publicly available WUENIC input data sets were assembled for this work. These include:

- i. Reported administrative coverage data (RADM) [9];
- ii. Reported country official coverage estimates (ROFF) [10];
- iii. Vaccination coverage survey data (SURV) [11];
- iv. United Nations Population Division's population estimates (UNPOP) [12];
- v. Year of vaccine introduction data (YOVI) [13].

For (i) – (iv), annual data covering the period 2000 – 2017 were obtained for all five vaccines. RADM are administrative estimates of vaccination coverage which countries report annually to WHO and UNICEF through the Joint Reporting Form [5]. ROFF are also administrative estimates which have been reviewed independently by countries and which represent each country's assessment of the most likely coverage estimates based on a combination of different data sets. Nationally representative household survey data used for WUENIC are obtained from three major sources: Expanded Programme on Immunization (EPI) cluster survey [14], the Demographic and Health Surveys (DHS) [15] and UNICEF Multiple Indicator Cluster Survey (MICS) [16]. The denominator data (UNPOP) relate mostly to surviving infants (i.e. under 1s) for DTP1, DTP3, MCV1 and PCV3; and various population groups for MCV2, depending on the national immunization schedule [4]. YOVI data were only available for MCV2 and PCV3 and these indicate the years during which these vaccines were introduced nationwide within countries. All countries among the 195 WHO countries (see Figure 1A – Appendix) spread across six WHO regions and for which data were available are included in this work. The following steps were implemented to process RADM using other available data sets mentioned above in preparation for model-fitting.

Step 1: Data cleaning and preparation

We first created standard vaccine-specific data files for the study period (i.e. 2000-2017) using the original input data files. Each of these files included all the relevant information needed for the analysis. For RADM, for example, there were five files, each of which included the “WHO region”, “ISO3 country code”, “Year”, “Number of doses administered” and “Target population” associated with the coverage estimates. ROFF, UNPOP and YOVI data were handled in a similar manner. However, for SURV, additional processing steps were carried out to undertake recall-bias adjustment and to select between multiple estimates where applicable.

Step 2: Recall-bias adjustment and additional processing of survey (SURV) data

Each coverage survey estimate was linked to a ‘birth cohort year’ which we used as the reference year for the estimate in this work. The birth cohort year was determined using the period of data collection and the age of the birth cohort that the survey estimate relates to [1]. Similar to the WUENIC approach, we applied a recall-bias adjustment to DTP3 and PCV3 survey estimates. Only estimates based on vaccination cards only or vaccination cards and recall were used for the adjustment. In the input data file, these estimates were labelled as: “Card”, “Card or History”, “C or H <12 months”, “card <12 months”, “C or H <15 months”, etc in the column for evidence of vaccination. For country-vaccine-year combinations with multiple estimates labelled as “crude” or “valid”, the “valid” estimates were retained in the analysis as these are considered more accurate. The formula used for the adjustment is:

$$VD3_{(Card+history)} = VD3_{(card\ only)} \times \frac{VD1_{(card+history)}}{VD1_{(card\ only)}}$$

where $VD3$ denotes the third dose of DTP or PCV vaccine and so on. We note that for each vaccine, the adjustment was applied only when all the data needed to compute it were available. After the adjustment, the original survey estimates of DTP3 and PCV3 were replaced with corresponding bias-adjusted estimates for further processing.

For a given vaccine, country and year, if one survey estimate was available, it was accepted if the sample size was greater than 300 or if the estimate was labelled ‘valid’. Otherwise, the estimate was not accepted. Where multiple estimates were available for the same vaccine, country and year, “Card or History” estimates were prioritized over “Card” only estimates, and these were accepted if the corresponding sample size was greater than 300 or if the evidence of vaccination was based on valid doses. This ensured that for DTP3 and PCV3, only the bias-adjusted estimates were used when available. When multiple “Card” or “Card or History” estimates were available from different surveys (or the same survey) for the same vaccine, country and year, the estimate based on valid evidence of vaccination was accepted. If there was no ‘valid’ estimate, the estimate with the largest sample size was accepted. If the sample sizes were missing, then one of the estimates was chosen – usually the first estimate available.

The resulting SURV estimates were used in the rest of the analyses.

Step 3: Denominator adjustment of reported administrative coverage (RADM) data using UNPD population (UNPOP) data and handling of estimates greater than 100%

Next, we applied a denominator adjustment to RADM estimates from Step 1 using UNPOP estimates. Essentially, the denominators used in calculating the original RADM estimates were replaced by UNPOP estimates to produce new RADM estimates. We proceeded to replace all denominator-adjusted RADM estimates greater than 100% with corresponding ROFF estimates. This step was implemented simultaneously for DTP1 and DTP3 to ensure the consistency (i.e. that $DTP1 \geq DTP3$) of the resulting estimates. Subsequently, RADM estimates greater than 100% with no corresponding ROFF estimates were treated as missing data.

Step 4: Benchmarking denominator-adjusted RADM estimates and substitution using survey (SURV) estimates

The RADM estimates from Step 3 were further compared with corresponding SURV estimates for validation. Similar to the WUENIC approach, for cases (country, vaccine and year) where the differences between SURV and RADM estimates were greater than 10%, the RADM estimates were replaced with the corresponding SURV estimates. As before, this process was implemented for DTP1 and DTP3 simultaneously.

We note that for country-vaccine-year combinations for which no corresponding survey data were available, the processed RADM estimates were the denominator-adjusted estimates where these were less than or equal to 100%.

Step 5: Data filtering using year of vaccine introduction (YOVI) data

This processing step was implemented for MCV2 and PCV3 for each country reporting data for both vaccines during the study period. For each country, all years before and including the year of introduction of these vaccines were excluded from the processed RADM data sets for both vaccines. This final processing step ensures that any missing values that remain in the data are 'valid', i.e., all things being equal, there ought to be observations for those cases.

Table 1 provides a summary of the processed RADM data for the five vaccines. The number of countries reporting coverage estimates during the study period varied among the vaccines, ranging from 182 countries for DTP1 and MCV1 to 120 countries for PCV3. DTP1 has the highest mean coverage of 87.56% and the least standard deviation of 14.08%. In contrast, PCV3 has the lowest mean coverage of 80.11% with the maximum standard deviation of 23.13%.

In Figures A2 – A7, we plot the differences between input RADM estimates and processed RADM estimates for all countries within each of the six WHO regions for the years 2000 – 2017. For most countries, there are small differences between these estimates, particularly in the EURO (except Serbia) and AMRO regions. The differences between the estimates appear more marked for some countries in AFRO (e.g. Nigeria and Chad) and WPRO (e.g. Tonga and Australia) regions. The processed RADM data are also included in some of the plots discussed in Section 5.

Table 1: Summary statistics for processed RADM data for the period 2000 – 2017

| Vaccine | No. of countries with input data (non-missing data points) | % of missing data | Summary statistics (%) | | | | | | |
|--------------|--|-------------------|------------------------|-------|-------|-------|-----------|-------|-------|
| | | | Min. | Q1 | Med. | Mean | Std. dev. | Q3 | Max. |
| DTP1 | 173 (2649) | 14.9 | 0.00* | 84.30 | 92.08 | 87.56 | 14.08 | 96.99 | 100 |
| DTP3 | 182 (2570) | 21.5 | 0.00* | 76.34 | 87.70 | 82.16 | 17.00 | 93.96 | 99.96 |
| MCV1 | 182 (2882) | 12.0 | 0.00* | 76.24 | 88.18 | 82.51 | 17.23 | 94.69 | 100 |
| MCV2 | 150 (1608) | 22.3 | 0.01 | 75.42 | 88.94 | 82.20 | 19.05 | 95.50 | 100 |
| PCV3 | 120 (577) | 15.5 | 0.07 | 78.36 | 88.30 | 80.11 | 23.13 | 94.77 | 99.99 |
| All vaccines | 182 (10286) | 18.5 | 0.00* | 78.40 | 89.27 | 83.55 | 17.29 | 95.18 | 100 |

*These data points with <0.01% coverage are for Namibia in 2000 (DTP1, DTP3 and MCV1) – see Figure A2 (Appendix).

3. Model specification

Let $c = 1, \dots, C$, $v = 1, \dots, V$, and $t = 1, \dots, T$ (*i.e.* 2000, ..., 2017) denote the countries, vaccines and years of interest, respectively. Also, let y_{cvt} denote the corresponding logit-transformed processed RADM estimates. This logit-transformation was necessary to constrain the modelled estimates to the unit interval when these are back-transformed, given that the data are modelled using a Gaussian distribution. For each WHO region, the multivariate Bayesian spatiotemporal Gaussian regression model with conditional mean, μ_{cvt} , and variance, σ_{cvt}^2 , is given by

$$\begin{aligned} y_{cvt} &\sim N(\mu_{cvt}, \sigma_{cvt}^2), \\ \mu_{cvt} &= \beta_0 + \phi_c + \nu_v + \alpha_t + \delta_{vt} + \gamma_{ct}, \end{aligned} \quad (1)$$

where β_0 is an intercept term, ϕ_c is a country-level random effect modelling variation between countries, ν_v is a vaccine-level random effect capturing variation between vaccines and α_t is a temporal term modelling dependence in time. We assume that ν_v is identically and independently distributed (iid) with variance σ_v^2 , *i.e.* $\nu_v \sim N(0, \sigma_v^2)$. We also assume that ϕ_c is either spatially unstructured or spatially structured in order to test for the presence of spatial autocorrelation in the data. The spatially unstructured term is given by $\phi_c \sim N(0, \sigma_c^2)$; while the spatially structured term is assigned a conditional autoregressive (CAR) model proposed by Leroux et al [17], which was found to outperform other choices in disease mapping studies [18]. That is, $\phi_1, \dots, \phi_C \sim N(\mathbf{0}, \sigma_\phi^2 \mathbf{Q}_\phi^{-1}(\mathbf{W}))$, where $\mathbf{Q}_\phi(\cdot)_{C \times C}$ is a precision matrix and σ_ϕ^2 is a variance parameter. More explicitly, $\mathbf{Q}_\phi(\mathbf{W}) = \rho_\phi (\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}) + (1 - \rho_\phi)\mathbf{I}_C$, where ρ_ϕ is a spatial autocorrelation parameter, $\mathbf{1}$ is a vector of C ones, \mathbf{I}_C is an identity matrix and \mathbf{W} is a binary matrix characterizing the neighbourhood structure of the countries. That is, $W_{ij} = 1$ if areas countries i and j share a common border and zero otherwise. In order to induce adequate temporal dependence in the model, we model α_t using a second order autoregressive (AR(2)) prior given by $\alpha_t | \alpha_{t-1}, \alpha_{t-2} \sim N(\rho_1 \alpha_{t-1} + \rho_2 \alpha_{t-2}, \sigma_\alpha^2)$, where ρ_1 and ρ_2 are autocorrelation parameters and σ_α^2 is the conditional

variance of the process (sample partial autocorrelation function plots for example countries and vaccines mostly indicated an $AR(1)$ prior).

Further, the vaccine- and country-time interaction terms - δ_{vt} and γ_{ct} - are assumed to follow Gaussian distributions with precision matrices given by $\sigma_\delta^{-2}\mathbf{Q}_\delta$ and $\sigma_\gamma^{-2}\mathbf{Q}_\gamma$, respectively. The parameters σ_δ^2 and σ_γ^2 are variance parameters, while the structure matrices \mathbf{Q}_δ and \mathbf{Q}_γ specify the nature of interdependence between the elements of $\{\delta_{vt}; v = 1, \dots, V; t = 1, \dots, T\}$ and $\{\gamma_{ct}; c = 1, \dots, C; t = 1, \dots, T\}$ respectively. These interaction terms are used to model additional variation that cannot be explained by the main effects - ϕ_c, ν_v and α_t . The structure matrices can assume different forms as given in Clayton [19]. For the space-time interactions, γ_{ct} , we assume Type I and Type II interaction models of Knorr-Held [20]. For δ_{vt} , we assume a Type II interaction. We did not consider a Type I interaction for δ_{vt} as it neither yielded consistent estimates of DTP1 and DTP3 nor improved predictive performance significantly over the other terms included in the model (see Table 2). Also, our choice of interaction terms are based on a prior assumption that it is unlikely that any structured spatial dependence exists in the data. Based on different combinations of the main effects and interaction terms, the different models considered are outlined in Table 2.

Table 2: Model parameterizations considered for each WHO region

| Model | μ_{cvt} |
|-------|---|
| Ia | $\beta_0 + \phi_c + \nu_v + \alpha_t$ $\phi_c \sim IID; \nu_v \sim IID; \alpha_t \sim AR(2)$ |
| Ib | $\beta_0 + \phi_c + \nu_v + \alpha_t$ $\phi_c \sim IID; \nu_v \sim CAR; \alpha_t \sim AR(2)$ |
| II | $\beta_0 + \phi_c + \nu_v + \alpha_t + \gamma_{ct}$ $\phi_c \sim IID; \nu_v \sim IID; \alpha_t \sim AR(2); \gamma_{ct} \sim IID$ |
| III | $\beta_0 + \phi_c + \nu_v + \alpha_t + \gamma_{ct}$ $\phi_c \sim IID; \nu_v \sim IID; \alpha_t \sim AR(2); \gamma_{ct} \sim IID \times AR(2)$ |
| IV | $\beta_0 + \phi_c + \nu_v + \alpha_t + \gamma_{ct} + \delta_{vt}$ $\phi_c \sim IID; \nu_v \sim IID; \alpha_t \sim AR(2); \gamma_{ct} \sim IID; \delta_{vt} \sim IID \times AR(2)$ |
| V | $\beta_0 + \phi_c + \nu_v + \alpha_t + \gamma_{ct} + \delta_{vt}$ $\phi_c \sim IID; \nu_v \sim IID; \alpha_t \sim AR(2); \gamma_{ct} \sim IID \times AR(2); \delta_{vt} \sim IID \times AR(2)$ |

Model Ia is the main effect model in which $\delta_{vt} = \gamma_{ct} = 0$. Model Ib differs from Model Ia by specifying a CAR prior for the country-level random effect ϕ_c . This model enables us to assess the presence of structured spatial variation in the data when compared with Model Ia (see Section 5.1), as highlighted previously. All other models include the main effects from Model Ia and different interaction terms. Specifically, the interaction term in Model II (Type I) adjusts for unobserved covariates for each country-time combination that do not have any structure in space and time; while that of Model III (Type II) accounts for structured temporal variation that differs from country to country but without any structure in space. The additional interaction term in Model IV - δ_{vt} (Type II) - captures structured temporal variation for each vaccine that is independent

of other vaccines. Finally, Model V includes both Type II country-time and vaccine-time interactions.

When γ_{ct} and δ_{vt} are modelled as Type II interactions as given in Table 1, their structure matrices can be parameterized as $\mathbf{Q}_\gamma(\rho_1^\gamma, \rho_2^\gamma)$ and $\mathbf{Q}_\delta(\rho_1^\delta, \rho_2^\delta)$, where $\rho_1^\gamma, \rho_2^\gamma, \rho_1^\delta$ and ρ_2^δ are autoregression parameters which arise from the autoregression components of the interactions. These structure matrices can be expressed, more explicitly, as the Kronecker product of the identity matrix of the corresponding IID term and a neighbourhood structure matrix specified through an $AR(2)$ process [19, 21]. With a Type I interaction, no additional parameters are involved since $\mathbf{Q}_\gamma = \mathbf{I}$, where \mathbf{I} is the identity matrix (see [20]).

We note that it is straightforward to incorporate covariate information in Models I-V. However, this may simplify the residual structure of the models (or the structures of the random effects) and some of the interaction terms and/or main effects may become redundant. In this setting, the model choice criteria discussed in Section 4.2 will be useful to determine important residual terms to retain in the models.

4. Bayesian inference and computation

Models I - V were fitted in a Bayesian framework using the INLA approach. INLA is a fast deterministic algorithm for performing approximate Bayesian inference which avoids the convergence problems often encountered when using MCMC methods. Models that can be fitted in INLA, such as those proposed here, can be reformulated as latent Gaussian models with the general structure given by:

$$\begin{aligned} \mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\theta}_1 &\sim p(\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\theta}_1), \\ \boldsymbol{\eta}|\boldsymbol{\theta}_2 &\sim p(\boldsymbol{\eta}|\boldsymbol{\theta}_2) = N(\mathbf{0}, \boldsymbol{\Sigma}), \\ \boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^T &\sim p(\boldsymbol{\theta}), \end{aligned} \quad (2)$$

where \mathbf{y} denotes the observed data, $\boldsymbol{\eta}$ is the latent field comprised of the joint distribution of all the parameters in the linear predictor $\boldsymbol{\mu}$ in the models (the intercept term and the random effects), $\boldsymbol{\theta}_1$ are the hyperparameters of the likelihood (the variance of the Gaussian observations, \mathbf{y}) and $\boldsymbol{\theta}_2$ are the hyperparameters of the latent field (e.g. the variances and autocorrelation parameters of the random effects). As shown in equation (2), $\boldsymbol{\eta}$ is assumed to follow a multivariate Gaussian density with additional conditional independence, also known as Markov property, which yields a sparse precision matrix [22]. The joint posterior distribution of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ can be expressed as:

$$p(\boldsymbol{\eta}, \boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta}) \times p(\boldsymbol{\eta}|\boldsymbol{\theta}_2) \times p(\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\theta}_1). \quad (3)$$

The goal is to accurately approximate the marginal posterior distributions of the components of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$. For a given η_i , for example, this is given by

$$p(\eta_i|\mathbf{y}) = \int_{\boldsymbol{\theta}} \int_{\boldsymbol{\eta}_{-i}} p(\boldsymbol{\eta}, \boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\eta}_{-i} d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}} p(\eta_i|\boldsymbol{\theta}, \mathbf{y}) \times p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}, \quad (4)$$

which is approximated by INLA as

$$\sum_k \tilde{p}(\eta_i | \theta_k, \mathbf{y}) \tilde{p}(\theta_k | \mathbf{y}) \Delta_k, \quad (5)$$

where \tilde{p} denotes an approximation, and $\{\Delta_k\}$ is a set of weights corresponding to the integration points $\{\theta_k\}$. In the second term in (5), a Laplace approximation is used, whereas the first term depends on the implementation of the INLA program, with a simplified Laplace approximation being the default strategy [8], which is used in our work. The marginal posterior distributions of the components of θ can be obtained in a similar manner from $\tilde{p}(\theta | \mathbf{y})$. As discussed in Rue et al [8] and shown in different applications, INLA has a small approximation error compared to Monte Carlo error and is negligible in practice [23, 24].

The Bayesian model specification is completed by placing appropriate prior distributions on the parameters. We placed a weakly-informative inverse gamma $IG(10, 100)$ prior on $\sigma_{cv_t}^{-2}$, σ_v^{-2} , σ_c^{-2} , σ_δ^{-2} and σ_γ^{-2} . For β_0 , we use the default non-information prior $\beta_0 \sim N(0, 10^6)$. For the parameters of α_t and the autoregression parameters of the Type II interactions, we use default priors as these parameters are parameterized differently in R-INLA [7]. For example, R-INLA estimates the marginal precision (or variance) of an $AR(2)$ process and reparameterizes the autoregression parameters using the partial autocorrelation function; see R-INLA documentation [25]. However, the autoregression parameters can be calculated post-model estimation using appropriate transformations. All models were fitted in R using the R-INLA package [7].

4.1 Prediction and forecasting

Typically, Bayesian in-sample and out-of-sample prediction (the latter is also known as forecasting) are both based on the posterior predictive distribution. For example, for one-step-ahead prediction, the posterior predictive distribution of $\tilde{y}_{cv(t+1)}$ can be obtained by integrating over η and the hyperparameters, θ , as follows:

$$p(\tilde{y}_{cv(t+1)} | \mathbf{y}) = \int_{\theta} \int_{\eta} p(\tilde{y}_{cv(t+1)} | \eta, \theta) \times p(\eta, \theta | \mathbf{y}) d\eta d\theta. \quad (6)$$

In R-INLA, prediction is done during model fitting. All desired predictions are included in the data as missing values (i.e. NA's) which forces the prediction when the model is run.

4.2 Model comparison, evaluation and validation

To assess the goodness-of-fit and predictive ability of the models, we conducted two kinds of validation, namely

- i. Leave-one-out and leave-two-out cross-validation, and
- ii. One- and two-step-ahead predictions.

With validation exercise (i), we assessed the ability of the models to predict in-sample data, whereas with (ii), we evaluated how well the models can predict future observations. We used the first six years (i.e. 2000 – 2005) as base years in (ii), which

means that the models were evaluated for each country-vaccine combination using predictions for the years 2006 – 2017 wherever applicable with this approach. We used the following metrics to quantify predictive performance:

$$\text{Average relative bias: } ARB = \frac{1}{n} \sum_{k=1}^n (\hat{y}_k - y_k) / y_k;$$

$$\text{Root mean square error: } RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (\hat{y}_k - y_k)^2};$$

$$\text{Mean absolute error: } MAE = \frac{1}{n} \sum_{k=1}^n |\hat{y}_k - y_k|;$$

$$\text{95\% coverage: } \text{Coverage} = 100 \times \sum_{k=1}^n I(\hat{y}_i^l \leq y_i \leq \hat{y}_i^u);$$

and the Pearson's correlation between observed and predicted values. In the formulae above, n is the total number of observations (across all vaccines, countries and relevant years) used for validation, \hat{y}_k and y_k are the predicted (i.e. the posterior means) and observed (i.e. processed RADM estimates) values, \hat{y}_i^l and \hat{y}_i^u are the lower and upper 95% credible limits of the predictions and $I(\cdot)$ is an indicator function. All validation metrics were calculated on the percentage scale using the back-transformed data. The 95% coverage rates assess the accuracy of the uncertainty intervals associated with the predictions while all the other metrics evaluate the accuracy of the point estimates. The closer the achieved 95% coverages are to the nominal value of 95%, the better the predictions. Similarly, the closer the $RMSE$, MAE and ARB (in absolute value) are to zero, the better the predictions. Correlations close to 1 indicate strong predictive power.

Additionally, we computed the deviance information criterion (DIC) [26] and the Watanabe-Akaike information criterion (WAIC) [27] of the models, which we used primarily to assess the presence of spatial autocorrelation in the data by comparing the fits of Models Ia and Ib. The model with the smaller DIC and WAIC gives a better fit.

For residual analysis, we examined the autocorrelations of the residuals for each country-vaccine combination that had at least 5 observations during the study period (other thresholds or complete observations are possible). The maximum autocorrelation at lags 1-3 was extracted to check for significant autocorrelation in the residuals in each case. The autocorrelation is significant if it falls outside the 95% confidence limits (i.e. $-\frac{1}{n} \pm \frac{2}{\sqrt{n}}$, where n is the number of observations for the given country-vaccine combination [28]) of an autocorrelation function (acf) plot and this indicates inadequate model performance for the given country and vaccine. Finally, we examined the plots of observed and predicted coverage estimates for each vaccine to determine vaccines and, perhaps, countries with large residuals, i.e. cases where predicted values differed greatly from observed values.

5. Results

5.1 Testing for spatial autocorrelation

We fitted Models Ia and Ib for each WHO region to examine the presence of spatial autocorrelation in the data. Table 3 provides a summary of the model comparison statistics. These statistics are very similar for both models across all regions, except for the EURO region where Model Ia is clearly the better model. These results suggests a lack of meaningful spatial autocorrelation in the data, and hence justify our earlier decision to model the variations between countries as random in Models II-V.

Table 3: DIC and WAIC statistics for Models Ia and Ib

| Region | Model Ia - Non-spatial | | Model Ib - Spatial | |
|--------|------------------------|------|--------------------|------|
| | DIC | WAIC | DIC | WAIC |
| EMRO | 7346 | 7346 | 7346 | 7346 |
| AFRO | 7503 | 7503 | 7503 | 7503 |
| EURO | 7098 | 7103 | 7731 | 7734 |
| AMRO | 6139 | 6141 | 6139 | 6141 |
| WPRO | 4977 | 4974 | 4977 | 4974 |
| SEARO | 1806 | 1805 | 1805 | 1805 |

5.2 Model comparison and validation

Tables 4 and 5 report the performance evaluation metrics for Models Ia, II-V based on the different validation exercises. There are similar patterns in model performance with each validation approach. In particular, Model IV consistently achieved the best 95% coverage but Model V is the preferred model for both in-sample (Table 4) and out-of-sample (Table 5) prediction, based on the *ARB*, *RMSE*, *MSE* and correlation statistics. The *ARB* statistics show that all the model parameterizations are, on average, more likely to overestimate immunization coverage. However, the *RMSE* values show that the models have smaller prediction errors compared to a standard deviation of 17.29% for the full data (see Table 1), with Model V's errors shown to be < 11.3% for in-sample data and <12.4% in all cases. This shows that the fitted models reduced the amount of variability in the data, which is desirable. The reported *MAE* statistics, which are all <7.6% for Model V, further demonstrate the accuracy of the predictions. It is also interesting to note that Model V yields correlations of ≥ 0.74 for in-sample data and ≥ 0.65 for out-of-sample data, both of which suggest a good predictive power. Based on these results, which are in agreement with the DIC and WAIC statistics reported in Tables 3 and A1, the rest of the analyses reported here were carried out using Model V.

Table 4: Model validation statistics based on leave-one- and leave-two-out cross-validation. Bold values indicate the best model in each case.

| Model | ARB | RMSE | MAE | Coverage | Correlation |
|---------------------------------------|--------------|---------------|--------------|--------------|--------------|
| Leave-one-out cross-validation | | | | | |
| Ia | 0.919 | 12.208 | 7.695 | 36.50 | 0.688 |
| II | 0.838 | 13.518 | 8.626 | 98.59 | 0.606 |
| III | 0.674 | 10.946 | 6.765 | 71.46 | 0.758 |
| IV | 0.907 | 12.405 | 7.851 | 93.63 | 0.675 |
| V | 0.668 | 10.927 | 6.741 | 72.10 | 0.759 |
| Leave-two-out cross-validation | | | | | |
| Ia | 0.918 | 12.557 | 7.997 | 38.61 | 0.664 |
| II | 0.846 | 13.748 | 8.800 | 98.59 | 0.591 |
| III | 0.686 | 11.405 | 7.160 | 73.40 | 0.735 |
| IV | 0.911 | 12.653 | 8.039 | 93.55 | 0.658 |
| V | 0.680 | 11.264 | 7.022 | 74.40 | 0.742 |

Table 5: Model validation statistics based on one- and two-step-ahead predictions. Bold values indicate the best model in each case.

| Model | ARB | RMSE | MAE | Coverage | Correlation |
|----------------------------------|--------------|---------------|--------------|--------------|--------------|
| One-step-ahead prediction | | | | | |
| Ia | 0.509 | 13.411 | 8.299 | 45.92 | 0.589 |
| II | 0.490 | 13.526 | 8.471 | 94.77 | 0.582 |
| III | 0.481 | 12.394 | 7.519 | 79.06 | 0.657 |
| IV | 0.492 | 13.474 | 8.409 | 94.93 | 0.585 |
| V | 0.479 | 12.346 | 7.475 | 80.00 | 0.660 |
| Two-step-ahead prediction | | | | | |
| Ia | 0.526 | 13.487 | 8.428 | 46.08 | 0.586 |
| II | 0.504 | 13.656 | 8.627 | 94.29 | 0.577 |
| III | 0.500 | 12.507 | 7.690 | 79.25 | 0.651 |
| IV | 0.509 | 13.491 | 8.460 | 94.85 | 0.584 |
| V | 0.501 | 12.362 | 7.557 | 80.34 | 0.658 |

5.3 Estimates of parameters of the fitted models

Estimates of the parameters of Model V are presented in Tables A2 – A7 (Appendix) for all six WHO regions. When considering the main effects - ϕ_c, ν_v and α_t - these estimates indicate that the vaccine random effect, ν_v , accounts for much ($\geq 55\%$) of the total variation ($\hat{\sigma}_\alpha^2 + \hat{\sigma}_\nu^2 + \hat{\sigma}_c^2$) explained by these terms across all the regions. This demonstrates substantial variation in coverage levels between the vaccines. There is also considerable variation in coverage among countries within each region. However, the temporal main effect, α_t , explains very little variation in the data (except for the AFRO region), which is likely due to the significant effect of temporally-correlated interaction terms in the model. Similarly, when considering the estimated variances of the interaction terms, it can be seen that the vaccine-time interaction term, δ_{vt} , explains more of the variation in the data compared to the country-time interaction, γ_{ct} . Estimates of the autoregression parameters for α_t - $\hat{\rho}_1$ and $\hat{\rho}_2$ - are not significant across all the regions, further indicating the negligible contribution of this term in the fitted models (although it's inclusion supports the structure of the interaction terms

[20]). For the interaction terms, only the first autoregression parameter estimates, $\hat{\rho}_1^\delta$ and $\hat{\rho}_1^Y$, are consistently significant for all the regions. This indicates the presence of strong vaccine- and country-specific temporal trends in the data. The estimates of the second autoregression parameters, $\hat{\rho}_2^\delta$ and $\hat{\rho}_2^Y$, are significant for the AFRO region only.

5.4 Modelled coverage estimates

In Figures 1 and A8-A9, we present the modelled coverage estimates for some example countries.

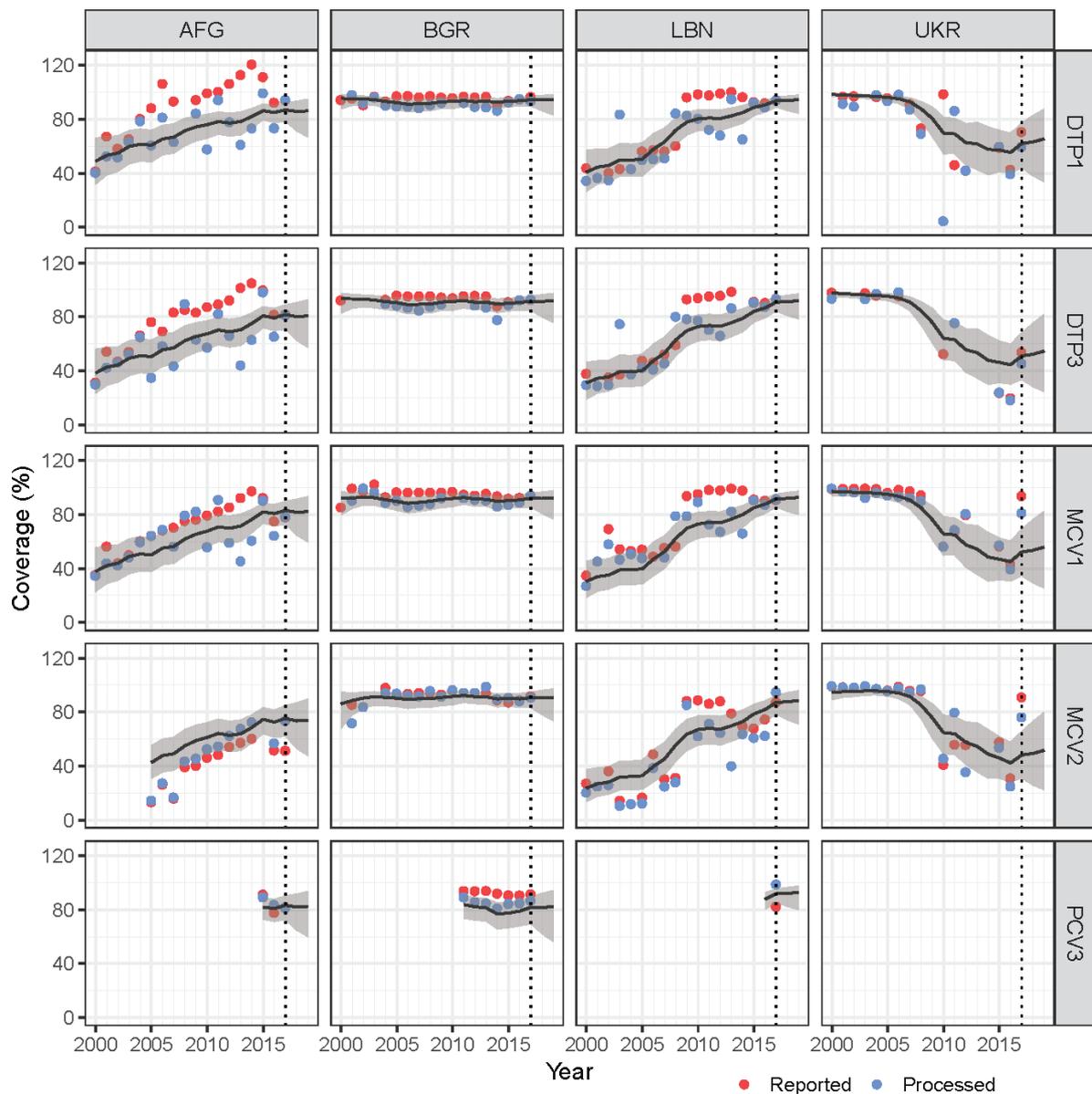


Figure 1: Modelled coverage estimates and corresponding 95% prediction bands for Afghanistan (AFG), Bulgaria (BGR), Lebanon (LBN) and Ukraine (UKR). Predictions for 2018 and 2019 are shown on the right side of the dotted vertical lines. The red and blue dots are the input and processed RADM estimates, respectively.

These figures show that the modelled estimates produced reasonable fits to the processed RADM data for these countries. The modelled estimates are smoother than the processed RADM data and there is evidence that the models did not overfit the data. The prediction intervals associated with the estimates are generally narrower when there is less variability in the data, and wider for out-of-sample predictions and when data are sparse and/or more variable, as expected. We note that although the predictions for 2018 and 2019 seem visually appropriate in most cases, in practice, influences from external factors not considered in the model could lead to coverage estimates completely different from these predictions.

In Figure 2, we illustrate multi-vaccine trends for some countries for the study period and prediction years. Overall, all five vaccines exhibit similar trends within countries, although MCV2 and PCV3 tend to follow slightly or completely different patterns in some countries. Additionally, Figures A10 – A14 map the modelled coverage estimates for all five vaccines and WHO countries. Generally, there are substantial increases in coverage between 2000 and 2010. However, these improvements in coverage appear to have levelled off, slowed down or, in a few cases, regressed in the period from 2010 – 2017. Here again, the uncertainties associated with the predictions, shown as the widths of the 95% prediction intervals are small both for the in-sample data and out-of-sample predictions. More precisely, for in-sample estimates, the widths of these intervals are $\leq 20\%$ for $\approx 71\%$ of the data. Whereas for out-of-sample predictions, the widths are $\leq 40\%$ for $\approx 77\%$ of the data. We, however, note that these smaller uncertainties associated with the modelled estimates is model-specific – see Tables 4-5. For example, Model IV will produce wider and more accurate prediction intervals compared to Model V used in these analyses.

5.5 Residual analysis results and further evaluation of processed RADM and modelled estimates

The fitted models accounted for much of the temporal trends in the country-vaccine-specific data, as can be seen in Figure A15 (Appendix). However, there are a few countries where it appears that the model did not fully capture vaccine-specific trends – these are countries whose maximum lags 1-3 residual autocorrelation fell outside the blue lines in Figure A15. These countries are listed in Table A10 for each vaccine (where the maximum lags 1-3 residual autocorrelation (in absolute value) was > 0.5). These range from 5 countries for PCV3 to 28 countries for DTP3. However, visual examination of the fits of the models for each country-vaccine combination listed in the table revealed that the models produced reasonable estimates of the data for most of these countries - see, for example, Figures A16 – A18 (Appendix). Hence, the estimated residual autocorrelations may have been influenced by potential outlying observations in those cases. However, there are countries where the models fit the data poorly for some vaccines (e.g. Germany (PCV3) and Jamaica (MCV1) in Figure A18). This indicates that the models possibly over-smoothed the data in these cases, perhaps due to the influence of country- or region-level trends.

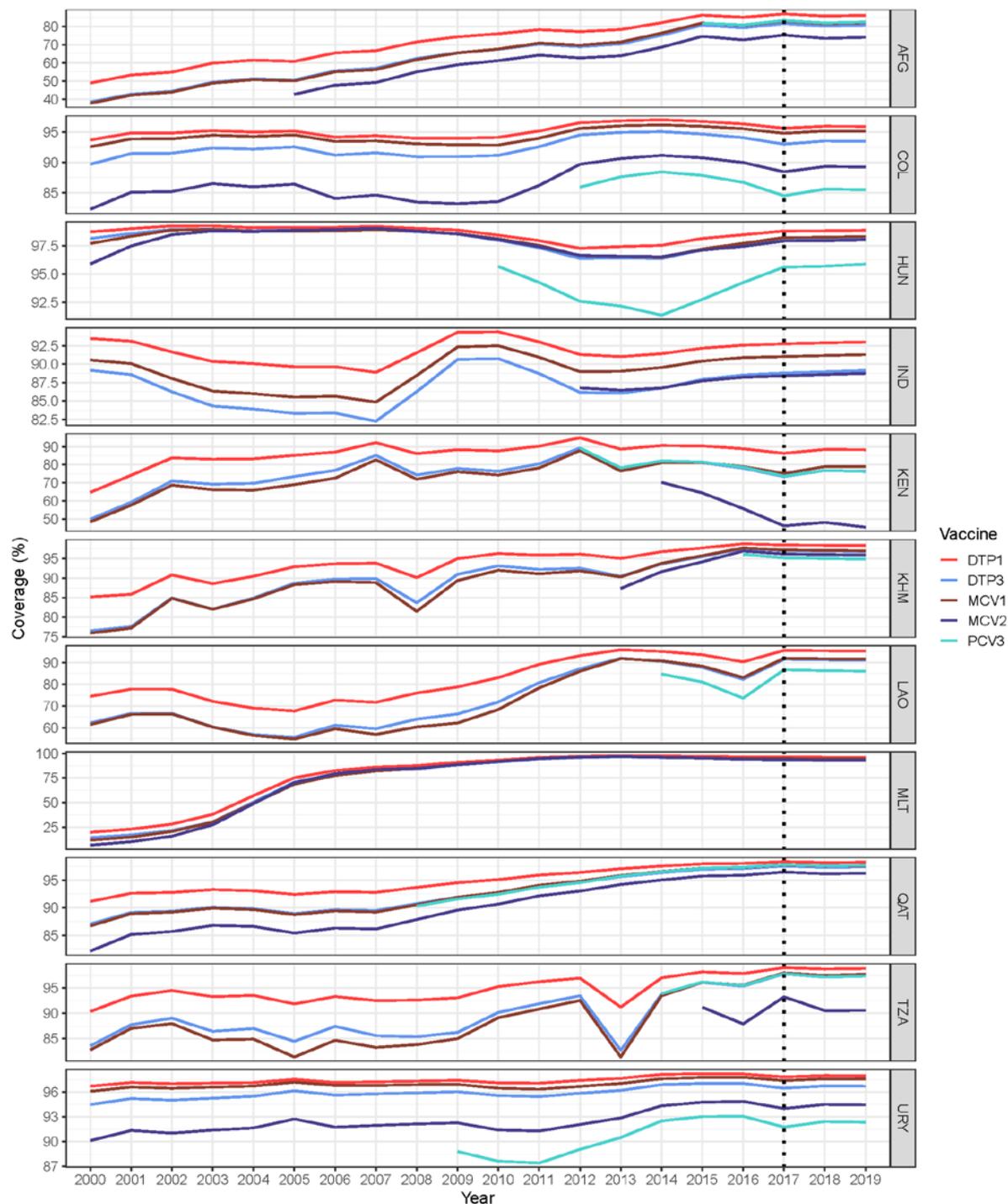


Figure 2: Trends in modelled coverage estimates for (top-bottom) Afghanistan, Colombia, Hungary, India, Kenya, Cambodia, Laos, Malta, Qatar, Tanzania and Uruguay.

We further examined the correspondence between processed RADM estimates and the modelled estimates across all countries for each vaccine. The plots in Figure A19 reveal a good correspondence in all cases. These plots also show that DTP1 (corr. = 0.86), DTP3 (corr. = 0.92) and MCV1 (corr. = 0.88) are better estimated, compared to MCV2 (corr. = 0.74) and PCV3 (corr. = 0.77). The greater bias in the estimates of

MCV2 and PCV3 is likely due to smoothing and comparatively smaller amounts of data used to estimate both vaccines (see Table 1).

6. Discussion

In this report, we have presented a novel Bayesian statistical methodology for producing accurate estimates of national immunization coverage. A major advantage of the methodology is its fast implementation in the R computing environment using the INLA package. For example, it took less than 4 minutes to run the most complex model, i.e. Model V, for all WHO regions on an Intel Core i7 2.60GHz 16GB RAM laptop. Project outputs that accompany this report include R scripts and modelled estimates of coverage for all five vaccines and WHO countries.

Our models were fitted independently for each WHO region. This multi-vaccine, multi-country approach was deemed plausible as it allows borrowing strength across vaccines and countries, in addition to leveraging temporal dependence, to estimate and predict immunization coverage. This approach also has other advantages, which include more robust parameter estimation given that data are only available for a maximum of 18 time points (much less for MCV2 and PCV3) for each country-vaccine-dose combination and considering the presence of missing values in the data. We had previously tested country-level models during model development - these models are modified versions of Model IV (see Table 2) which do not include the country-time interaction term, γ_{ct} . We observed (results not shown) that these country-level models had similar achieved 95% coverage rates as Model IV (regional). However, country-level models are computationally less efficient and were found to perform either worse or equally well, in comparison with Model V (best regional model), based on different model evaluation criteria (results not shown).

Furthermore, at the regional level, we tested an additional model parameterization that included the main effects in Model Ia and the interaction term, δ_{vt} , which was modelled as a Type II interaction, but this generally yielded poorer results than Models II-V across all the regions based on the DIC and WAIC statistics. We also investigated the possibility of modelling temporal autocorrelation using random walk models as in [20, 29]. The autoregressive models used here outperformed these models in all cases.

There are methodological caveats that need to be taken into account when evaluating the modelled estimates. The steps taken to process RADM data to produce the data that are modelled are a crucial part of our methodology, not least because the accuracy of the modelled estimates are largely dependent on the accuracy of the input data. Some of the decisions made during this phase of the analyses, e.g. substitution of denominator-adjusted RADM estimates with recall-bias-adjusted (where applicable) survey estimates where there was a difference of >10% between both estimates, were borrowed from WUENIC methodology. Other decisions were based on either statistical considerations or expert judgement. We recommend an independent assessment of these processing steps and the outputs to ensure that these are consistent with local or expert knowledge, particularly for those countries or cases (e.g. Namibia in Table 1) where the processed estimates appeared to have been problematic. In addition, the

current analyses do not include information on vaccine stocks or stock-out data. Such data can be useful for numerator adjustments and/or for refining the coverage estimates when incorporated into the fitted models. We plan to undertake this in future work.

Other directions for future work include exploring whether covariate information can help improve the modelled estimates. Candidate covariates for consideration are factors related to health system performance such as prevalence of vaccine-preventable diseases, health spending, conflict-related metrics, health access and quality index [30], etc. It is straightforward to produce regional and global estimates and associated uncertainties within our methodology through population-weighted aggregation of the country-level estimates. Additionally, the methodology can be readily extended to include other vaccines produced as part of WUENIC.

In conclusion, our methodology provides a robust, easy to implement and readily scalable framework for producing accurate estimates of national immunization coverage.

References

- [1] Burton A, Monash R, Lautenbach B, Gacic-Dobo M, Neil M, Karimov R, et al. WHO and UNICEF estimates of national infant immunization coverage: methods and processes. *Bulletin of the World Health Organisation*. 2009;87(7):535 - 41.
- [2] United Nations. Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a data revolution for the SDGs; 2015: Available from: <https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=2013&menu=35>. [Accessed on 03/08/2020].
- [3] Burton A, Kowalski R, Gacic-Dobo M, Karimov R, Brown D. A Formal Representation of the WHO and UNICEF Estimates of National Immunization Coverage: A Computational Logic Approach. *PLOS ONE*. 2012;7(10):e47806.
- [4] WHO. Immunization Schedule; 2020: Available from: https://www.who.int/immunization/monitoring_surveillance/data/en/. [Accessed on 03/09/2020].
- [5] WHO. WHO/UNICEF Joint Reporting Process; 2020: Available from: https://www.who.int/immunization/monitoring_surveillance/routine/reporting/reporting/en/. [Accessed on 03/09/2020].
- [6] WHO. WHO and UNICEF estimates of national immunization coverage; 2020: Available from: https://www.who.int/immunization/monitoring_surveillance/routine/coverage/en/index4.html. [Accessed on 03/09/2020].
- [7] Lindgren F, Rue H. Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*. 2015;63(19):25.
- [8] Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009;71(2):319-92.
- [9] WHO. Reported administrative coverage data; 2019: Available from: https://www.who.int/immunization/monitoring_surveillance/data/en/. [Accessed on 01/07/2019].
- [10] WHO. Official country reported coverage estimates; 2019: Available from: https://www.who.int/immunization/monitoring_surveillance/data/en/. [Accessed on 01/07/2019].
- [11] WHO. Coverage Survey data; 2019: Available from: https://www.who.int/immunization/monitoring_surveillance/data/en/. [Accessed on 01/07/2019].
- [12] United Nations DoEaSA, Population Division. *World Population Prospects 2019*; 2019: Available from: <https://population.un.org/wpp/>. [Accessed on 01/11/2019].
- [13] WHO. Year of vaccine introduction data; 2019: Available from: https://www.who.int/immunization/monitoring_surveillance/data/en/. [Accessed on 01/07/2019].
- [14] World Health O. *Immunization coverage cluster survey : reference manual*. Geneva: World Health Organization; 2005.
- [15] ICF International. The DHS Program website. Funded by USAID: Available from: <http://www.dhsprogram.com>. [Accessed on 21/04/2020].
- [16] UNICEF. Multiple Indicator Cluster Surveys: Available from: <http://www.micsunicef.org>. [Accessed on 22/05/2020].
- [17] Leroux B, Lei X, Breslow N. Estimation of disease rates in small areas: A new mixed model for spatial dependence. In: Halloran ME, Berry D, editors. *Statistical*

- Models in Epidemiology, the Environment and Clinical Trials. New York, NY: Springer New York; 2000. p. 179-91.
- [18] Lee D. A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology*. 2011;2(2):79-89.
- [19] Clayton DG. Generalized linear mixed models. In: Gilks WR, Richardson S, Spiegelhalter DJ, editors. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall; 1996. p. 275-301.
- [20] Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*. 2000;19(17-18):2555-67.
- [21] Blangiardo M, Cameletti M. *Spatial and Spatio-temporal Bayesian Models with R-INLA*. United Kingdom: John Wiley & Sons; 2015.
- [22] Rue H, L H. *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman & Hall; 2005.
- [23] Schrödle B, Held L, Riebler A, Danuser J. Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2011;60(2):261-79.
- [24] Riebler A, Held L, Rue H. Estimation and extrapolation of time trends in registry data--Borrowing strength from related populations. *Ann Appl Stat*. 2012;6(1):304-33.
- [25] R-INLA documentation: Available from: <https://inla.r-inla-download.org/r-inla.org/doc/latent/ar.pdf>. [Accessed on 01/04/2020].
- [26] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002;64(4):583-639.
- [27] Watanabe S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J Mach Learn Res*. 2010;11(3571–94).
- [28] Cowpertwait PSP, Metcalfe AV. *Introductory Time Series with R*. New York, USA: Springer; 2009.
- [29] Mercer LD, Wakefield J, Pantazis A, Lutambi AM, Masanja H, Clark S. Space-time smoothing of complex survey data: Small area estimation for child mortality. *Ann Appl Stat*. 2015;9(4):1889-905.
- [30] Fullman N, Yearwood J, Abay SM, Abbafati C, Abd-Allah F, Abdela J, et al. Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: a systematic analysis from the Global Burden of Disease Study 2016. *The Lancet*. 2018;391(10136):2236-71.

Bayesian time series regression methods for estimating national immunization coverage

Supplementary Appendix

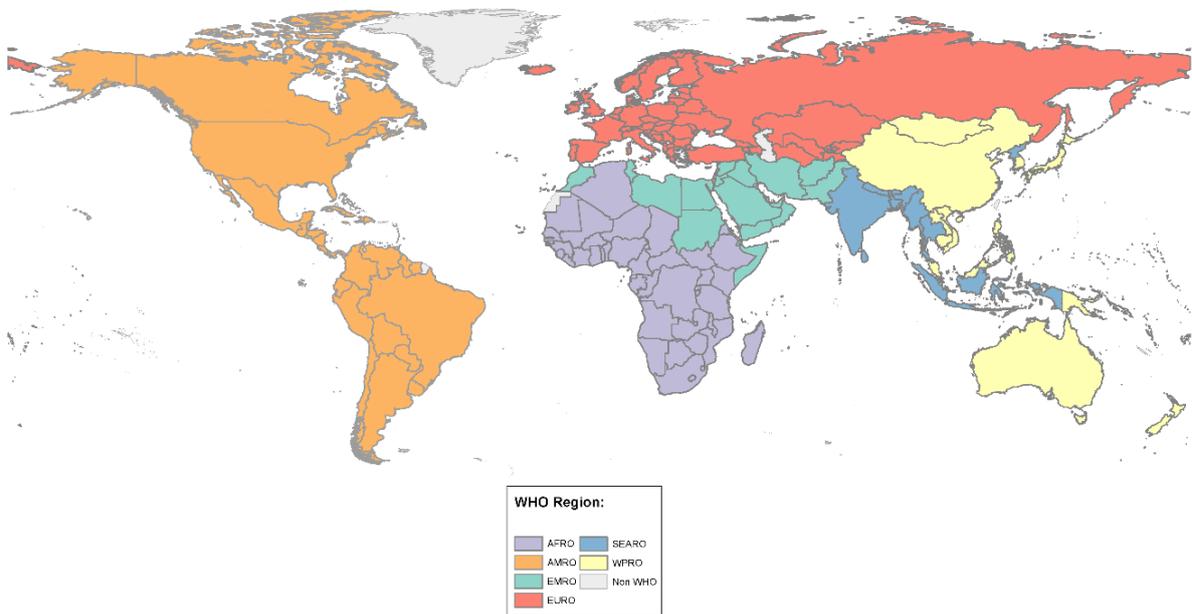


Figure A1: A map of WHO countries and regions.

Plots of differences between input reported administrative coverage estimates (RADM) and processed RADM estimates for each WHO region

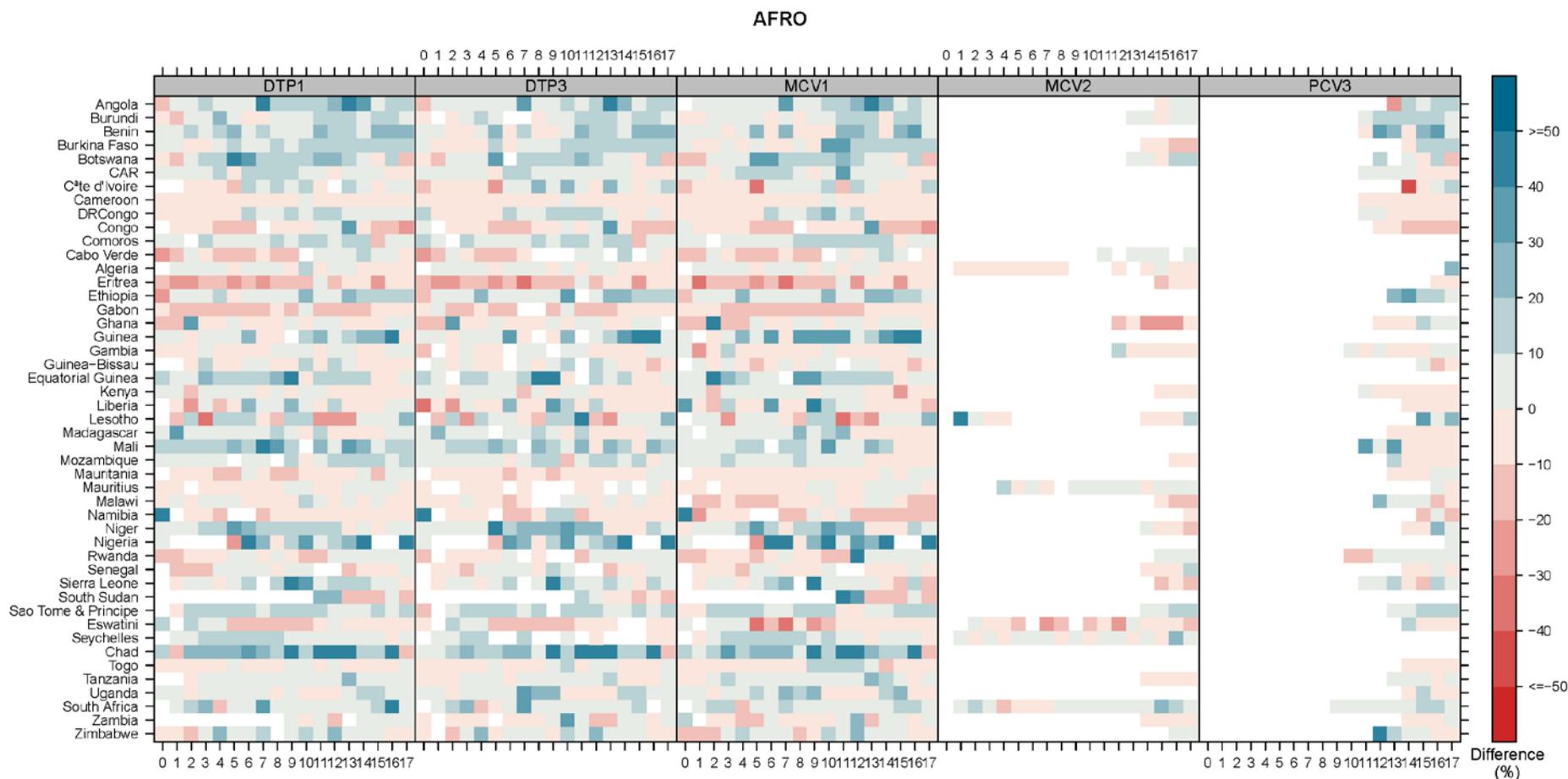


Figure A2: Differences between input reported administrative coverage estimates (RADM) and processed RADM estimates (i.e. input RADM – processed RADM) for WHO AFRO region for the period 2000 – 2017. The white spaces indicate missing data.

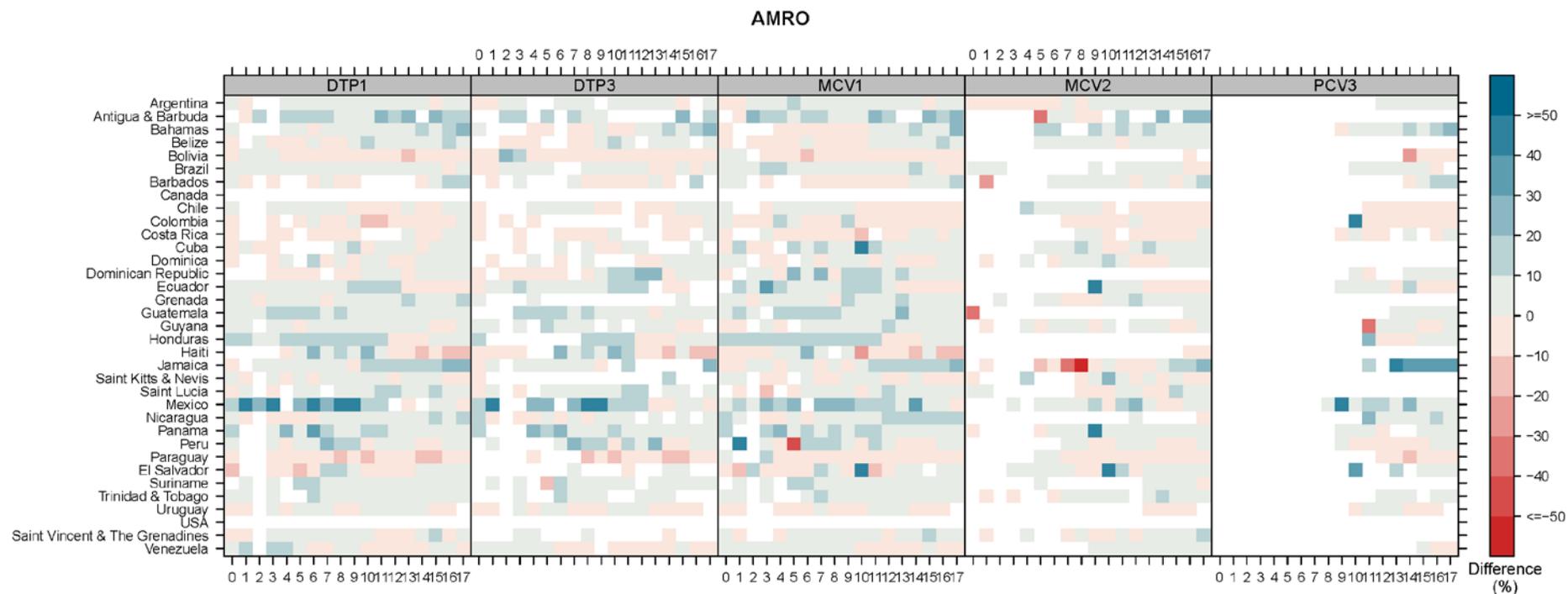


Figure A3: Differences between input reported administrative coverage estimates (RADM) and processed RADM estimates (i.e. input RADM – processed RADM) for WHO AMRO region for the period 2000 – 2017. The white spaces indicate missing data.

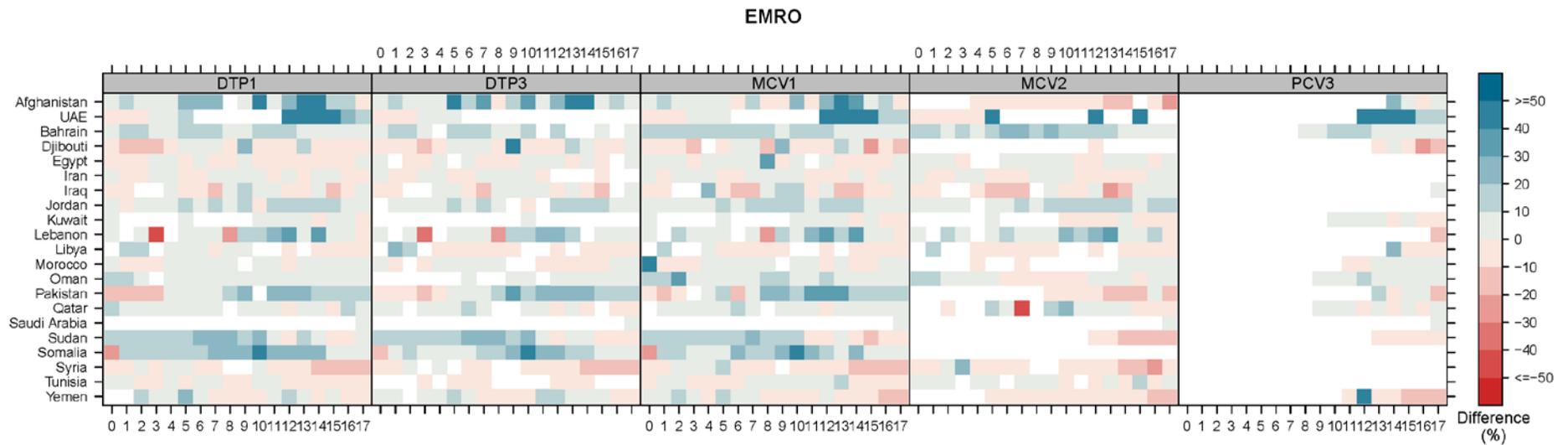


Figure A4: Differences between input reported administrative coverage estimates (RADM) and processed RADM estimates (i.e. input RADM – processed RADM) for WHO EMRO region for the period 2000 – 2017. The white spaces indicate missing data.

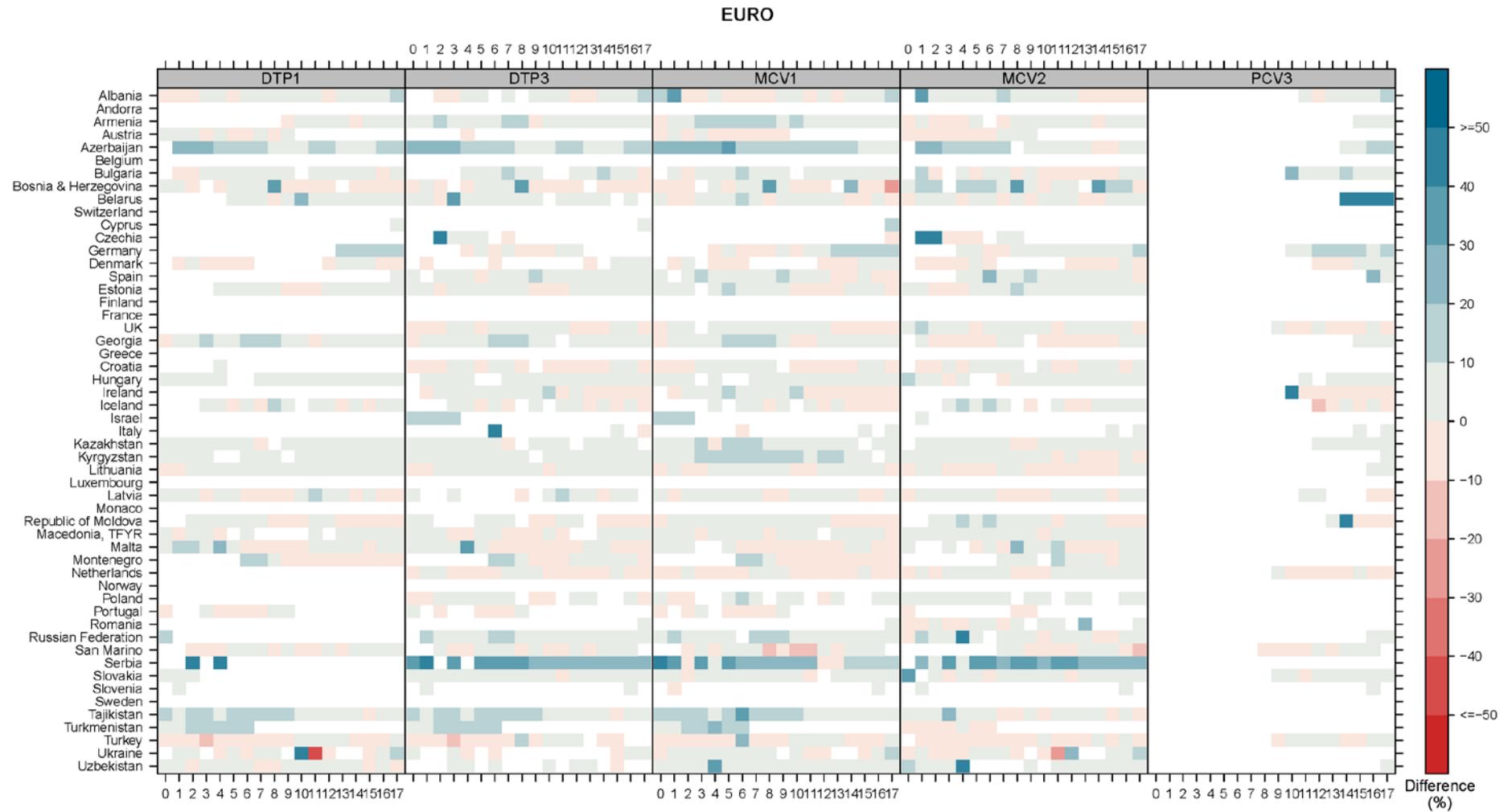


Figure A5: Differences between input reported administrative coverage estimates (RADM) and processed RADM estimates (i.e. input RADM – processed RADM) for WHO EURO region for the period 2000 – 2017. The white spaces indicate missing data.

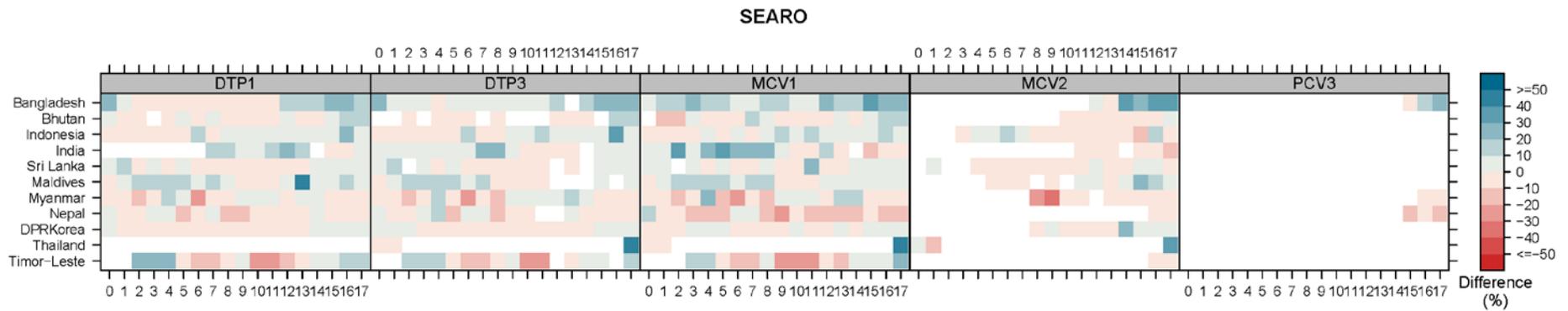


Figure A6: Differences between input reported administrative coverage estimates (RADM) and processed RADM estimates (i.e. input RADM – processed RADM) for WHO SEARO region for the period 2000 – 2017. The white spaces indicate missing data.

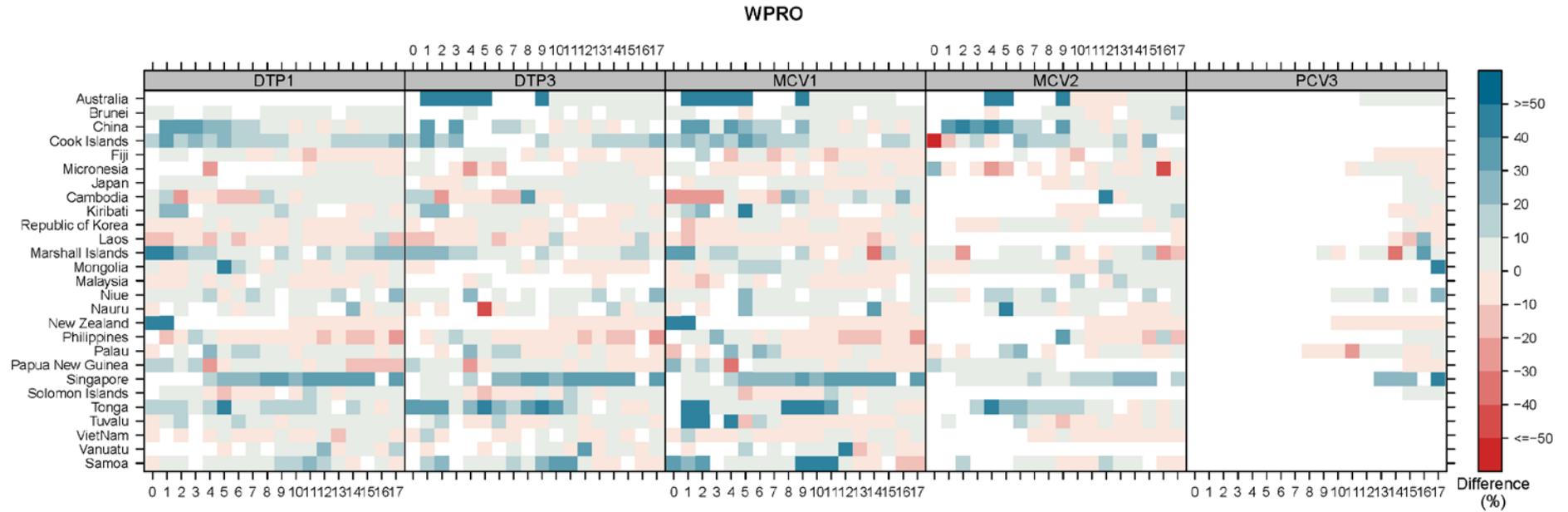


Figure A7: Differences between input reported administrative coverage estimates (RADM) and processed RADM estimates (i.e. input RADM – processed RADM) for WHO WPRO region for the period 2000 – 2017. The white spaces indicate missing data.

**Additional model evaluation statistics and estimates of parameters of Model V
for each WHO region**

Table A1: DIC and WAIC statistics for Models II – V. The bold figures indicate the best model in each case.

| Region | Model II | | Model III | | Model IV | | Model V | |
|--------|----------|------|-------------|-------------|----------|-------------|-------------|-------------|
| | DIC | WAIC | DIC | WAIC | DIC | WAIC | DIC | WAIC |
| EMRO | 3432 | 3382 | 3297 | 3284 | 3433 | 3379 | 3295 | 3284 |
| AFRO | 6427 | 6336 | 6421 | 6365 | 6300 | 6207 | 6299 | 6239 |
| EURO | 6801 | 6785 | 6505 | 6507 | 6772 | 6752 | 6476 | 6477 |
| AMRO | 5994 | 5962 | 5805 | 5802 | 5994 | 5960 | 5805 | 5802 |
| SEARO | 1819 | 1783 | 1746 | 1733 | 1818 | 1780 | 1747 | 1733 |
| WPRO | 4443 | 4394 | 4339 | 4313 | 4438 | 4385 | 4332 | 4305 |

Table A2: Posterior estimates of the parameters of Model V for the AFRO region

| Parameter | Mean | Std. dev. | 2.5% | 50% | 97.5% |
|-------------------------|--------|-----------|--------|--------|--------|
| β_0 | 1.219 | 2.064 | -2.850 | 1.218 | 5.287 |
| $\hat{\sigma}_{cvt}^2$ | 0.560 | 0.019 | 0.525 | 0.559 | 0.598 |
| \hat{s}_α^{2*} | 3.216 | 0.912 | 0.070 | 3.393 | 4.277 |
| $\hat{\sigma}_v^2$ | 9.919 | 3.071 | 5.473 | 9.365 | 17.414 |
| $\hat{\sigma}_c^2$ | 4.626 | 0.909 | 3.148 | 4.515 | 6.704 |
| $\hat{\sigma}_\delta^2$ | 9.171 | 2.749 | 5.122 | 8.699 | 15.822 |
| $\hat{\sigma}_\gamma^2$ | 4.175 | 0.853 | 2.846 | 4.050 | 6.173 |
| $\hat{\rho}_1$ | 0.920 | 0.483 | -0.414 | 0.992 | 1.624 |
| $\hat{\rho}_2$ | -0.185 | 0.322 | -0.757 | -0.189 | 0.422 |
| $\hat{\rho}_1^\delta$ | 1.709 | 0.182 | 1.253 | 1.747 | 1.939 |
| $\hat{\rho}_2^\delta$ | -0.710 | 0.182 | -0.940 | -0.748 | -0.254 |
| $\hat{\rho}_1^\gamma$ | 0.592 | 0.053 | 0.497 | 0.592 | 0.692 |
| $\hat{\rho}_2^\gamma$ | 0.365 | 0.055 | 0.260 | 0.365 | 0.464 |

*This is the marginal variance of α_t

Table A3: Posterior estimates of the parameters of Model V for the AMRO region

| Parameter | Mean | Std. dev. | 2.5% | 50% | 97.5% |
|-------------------------|--------|-----------|--------|--------|--------|
| β_0 | 2.028 | 2.038 | -1.985 | 2.028 | 6.037 |
| $\hat{\sigma}_{cvt}^2$ | 1.142 | 0.040 | 1.067 | 1.141 | 1.223 |
| \hat{s}_{α}^{2*} | 0.020 | 0.017 | 0.004 | 0.015 | 0.065 |
| $\hat{\sigma}_v^2$ | 9.833 | 2.968 | 5.371 | 9.360 | 16.931 |
| $\hat{\sigma}_c^2$ | 5.852 | 1.308 | 3.761 | 5.683 | 8.872 |
| $\hat{\sigma}_\delta^2$ | 9.607 | 2.867 | 5.402 | 9.109 | 16.556 |
| $\hat{\sigma}_\gamma^2$ | 5.618 | 1.278 | 3.548 | 5.464 | 8.543 |
| $\hat{\rho}_1$ | 0.332 | 0.506 | -0.619 | 0.353 | 1.263 |
| $\hat{\rho}_2$ | -0.142 | 0.292 | -0.705 | -0.128 | 0.392 |
| $\hat{\rho}_1^\delta$ | 0.989 | 0.162 | 0.680 | 0.988 | 1.305 |
| $\hat{\rho}_2^\delta$ | 0.011 | 0.162 | -0.305 | 0.012 | 0.320 |
| $\hat{\rho}_1^\gamma$ | 1.107 | 0.169 | 0.796 | 1.103 | 1.446 |
| $\hat{\rho}_2^\gamma$ | -0.116 | 0.170 | -0.453 | -0.111 | 0.193 |

*This is the marginal variance of α_t

Table A4: Posterior estimates of the parameters of Model V for the EMRO region

| Parameter | Mean | Std. dev. | 2.5% | 50% | 97.5% |
|-------------------------|--------|-----------|--------|--------|--------|
| β_0 | 2.106 | 2.099 | -2.023 | 2.106 | 6.231 |
| $\hat{\sigma}_{cvt}^2$ | 0.868 | 0.037 | 0.796 | 0.867 | 0.942 |
| \hat{s}_{α}^{2*} | 0.011 | 0.012 | 0.001 | 0.008 | 0.042 |
| $\hat{\sigma}_v^2$ | 9.639 | 2.890 | 5.240 | 9.200 | 16.499 |
| $\hat{\sigma}_c^2$ | 7.150 | 1.799 | 4.405 | 6.869 | 11.424 |
| $\hat{\sigma}_\delta^2$ | 9.746 | 2.867 | 5.375 | 9.311 | 16.545 |
| $\hat{\sigma}_\gamma^2$ | 7.129 | 1.739 | 4.306 | 6.925 | 11.098 |
| $\hat{\rho}_1$ | -0.007 | 0.319 | -0.652 | 0.001 | 0.620 |
| $\hat{\rho}_2$ | 0.103 | 0.262 | -0.354 | 0.089 | 0.606 |
| $\hat{\rho}_1^\delta$ | 0.980 | 0.273 | 0.492 | 0.973 | 1.527 |
| $\hat{\rho}_2^\delta$ | 0.020 | 0.273 | -0.528 | 0.027 | 0.508 |
| $\hat{\rho}_1^\gamma$ | 1.461 | 0.255 | 0.833 | 1.519 | 1.824 |
| $\hat{\rho}_2^\gamma$ | -0.468 | 0.256 | -0.831 | -0.527 | 0.164 |

*This is the marginal variance of α_t

Table A5: Posterior estimates of the parameters of Model V for the EURO region

| Parameter | Mean | Std. dev. | 2.5% | 50% | 97.5% |
|-------------------------|--------|-----------|--------|--------|--------|
| $\hat{\beta}_0$ | 2.157 | 2.034 | -1.846 | 2.157 | 6.156 |
| $\hat{\sigma}_{cvt}^2$ | 1.010 | 0.033 | 0.946 | 1.010 | 1.076 |
| \hat{s}_α^{2*} | 0.003 | 0.005 | 0.000 | 0.002 | 0.016 |
| $\hat{\sigma}_v^2$ | 10.160 | 3.090 | 5.355 | 9.734 | 17.391 |
| $\hat{\sigma}_c^2$ | 5.046 | 1.021 | 3.297 | 4.960 | 7.290 |
| $\hat{\sigma}_\delta^2$ | 9.947 | 3.158 | 5.035 | 9.519 | 17.323 |
| $\hat{\sigma}_\gamma^2$ | 4.696 | 0.969 | 3.146 | 4.568 | 6.935 |
| $\hat{\rho}_1$ | -0.017 | 0.352 | -0.731 | -0.009 | 0.655 |
| $\hat{\rho}_2$ | 0.043 | 0.215 | -0.360 | 0.037 | 0.469 |
| $\hat{\rho}_1^\delta$ | 1.118 | 0.238 | 0.691 | 1.095 | 1.596 |
| $\hat{\rho}_2^\delta$ | -0.119 | 0.238 | -0.597 | -0.095 | 0.308 |
| $\hat{\rho}_1^\gamma$ | 1.308 | 0.130 | 1.038 | 1.312 | 1.553 |
| $\hat{\rho}_2^\gamma$ | -0.321 | 0.131 | -0.566 | -0.325 | -0.046 |

*This is the marginal variance of α_t

Table A6: Posterior estimates of the parameters of Model V for the SEARO region

| Parameter | Mean | Std. dev. | 2.5% | 50% | 97.5% |
|-------------------------|--------|-----------|--------|--------|--------|
| $\hat{\beta}_0$ | 2.399 | 2.316 | -2.159 | 2.399 | 6.953 |
| $\hat{\sigma}_{cvt}^2$ | 1.274 | 0.083 | 1.122 | 1.270 | 1.447 |
| \hat{s}_α^{2*} | 0.101 | 0.221 | 0.000 | 0.042 | 0.560 |
| $\hat{\sigma}_v^2$ | 9.965 | 2.995 | 5.320 | 9.546 | 16.987 |
| $\hat{\sigma}_c^2$ | 8.395 | 2.370 | 4.800 | 8.025 | 14.039 |
| $\hat{\sigma}_\delta^2$ | 9.694 | 2.921 | 5.341 | 9.214 | 16.714 |
| $\hat{\sigma}_\gamma^2$ | 7.881 | 2.193 | 4.534 | 7.545 | 13.087 |
| $\hat{\rho}_1$ | 0.527 | 0.444 | -0.504 | 0.612 | 1.219 |
| $\hat{\rho}_2$ | -0.012 | 0.235 | -0.481 | -0.005 | 0.403 |
| $\hat{\rho}_1^\delta$ | 0.984 | 0.258 | 0.516 | 0.978 | 1.474 |
| $\hat{\rho}_2^\delta$ | 0.015 | 0.258 | -0.474 | 0.021 | 0.484 |
| $\hat{\rho}_1^\gamma$ | 0.998 | 0.147 | 0.717 | 0.998 | 1.278 |
| $\hat{\rho}_2^\gamma$ | -0.007 | 0.148 | -0.292 | -0.006 | 0.275 |

*This is the marginal variance of α_t

Table A7: Posterior estimates of the parameters of Model V for the WPRO region

| Parameter | Mean | Std. dev. | 2.5% | 50% | 97.5% |
|-------------------------|--------|-----------|--------|--------|--------|
| $\hat{\beta}_0$ | 1.829 | 2.115 | -2.339 | 1.830 | 5.988 |
| $\hat{\sigma}_{cvt}^2$ | 0.924 | 0.038 | 0.852 | 0.923 | 1.002 |
| \hat{s}_α^{2*} | 0.316 | 0.400 | 0.001 | 0.115 | 1.120 |
| $\hat{\sigma}_v^2$ | 9.682 | 2.963 | 5.372 | 9.155 | 16.897 |
| $\hat{\sigma}_c^2$ | 6.399 | 1.465 | 4.052 | 6.211 | 9.778 |
| $\hat{\sigma}_\delta^2$ | 9.534 | 2.822 | 5.241 | 9.102 | 16.237 |
| $\hat{\sigma}_\gamma^2$ | 6.145 | 1.428 | 3.700 | 6.032 | 9.269 |
| $\hat{\rho}_1$ | 0.833 | 0.469 | -0.500 | 0.913 | 1.504 |
| $\hat{\rho}_2$ | -0.108 | 0.283 | -0.626 | -0.103 | 0.417 |
| $\hat{\rho}_1^\delta$ | 1.057 | 0.277 | 0.529 | 1.046 | 1.561 |
| $\hat{\rho}_2^\delta$ | -0.057 | 0.277 | -0.564 | -0.047 | 0.471 |
| $\hat{\rho}_1^\gamma$ | 0.966 | 0.095 | 0.769 | 0.968 | 1.148 |
| $\hat{\rho}_2^\gamma$ | 0.007 | 0.097 | -0.182 | 0.002 | 0.209 |

*This is the marginal variance of α_t

Additional plots of modelled estimates of national immunization coverage for select countries

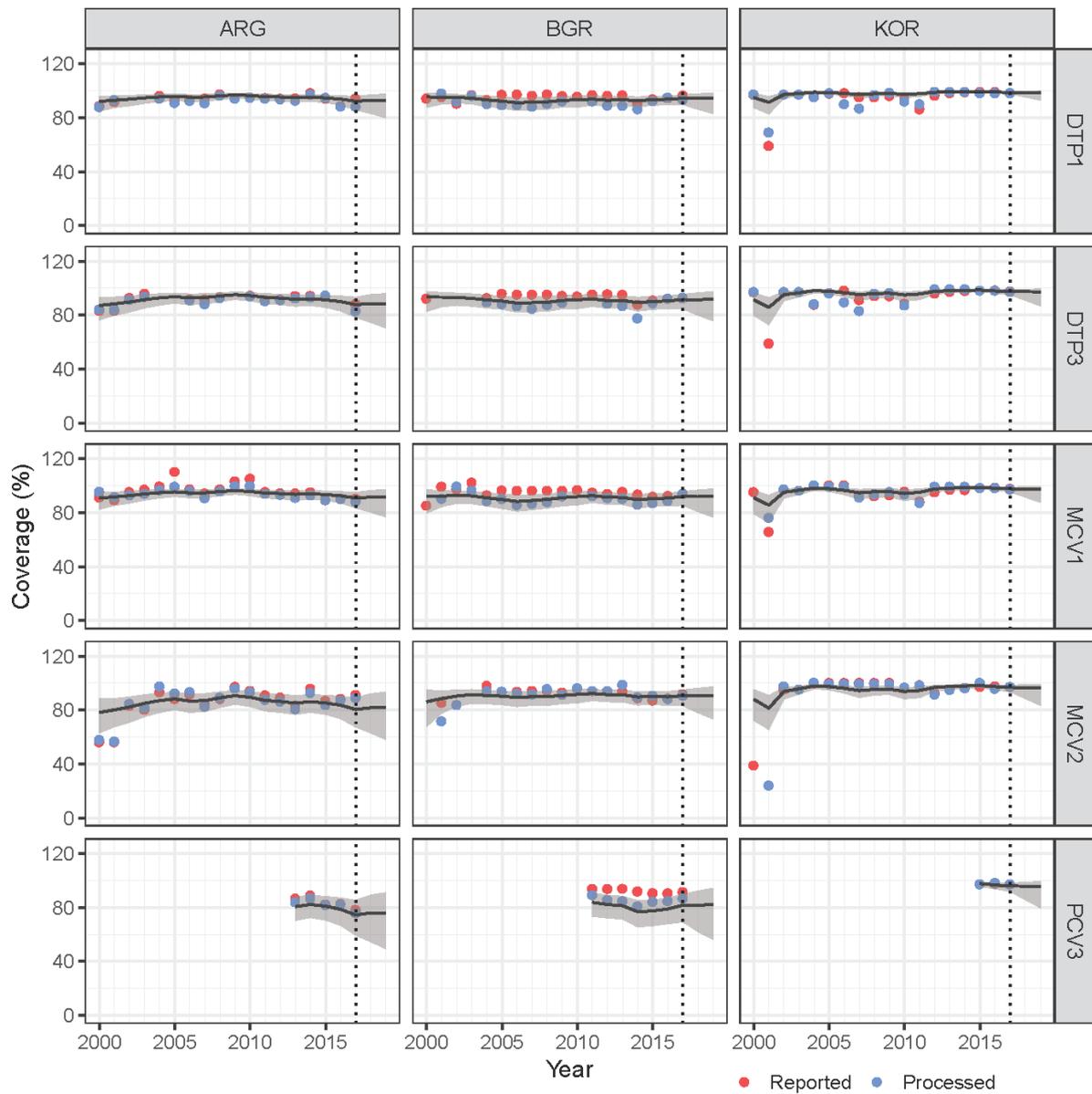


Figure A8: Modelled coverage estimates and corresponding 95% prediction bands for Argentina (ARG), Bulgaria (BGR) and Korea (KOR). Predictions for 2018 and 2019 are shown on the right side of the dotted vertical lines. The red and blue dots are the input and processed RADM estimates, respectively.

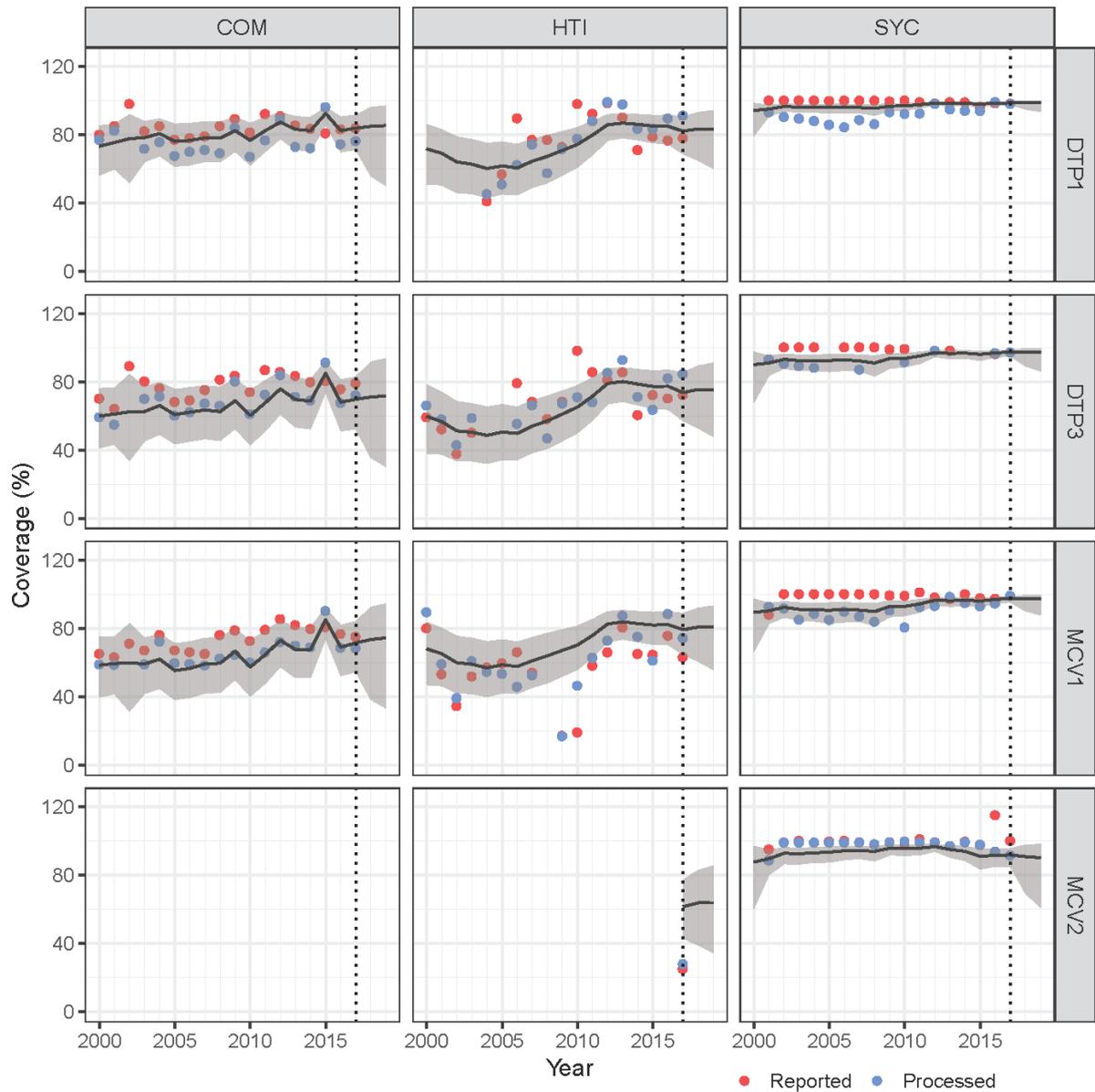


Figure A9: Modelled coverage estimates and corresponding 95% prediction bands for Comoros (COM), Haiti (HTI) and Seychelles (SYC). Predictions for 2018 and 2019 are shown on the right side of the dotted vertical lines. The red and blue dots are the input and processed RADM estimates, respectively.

Maps of modelled estimates of national immunization coverage for all countries

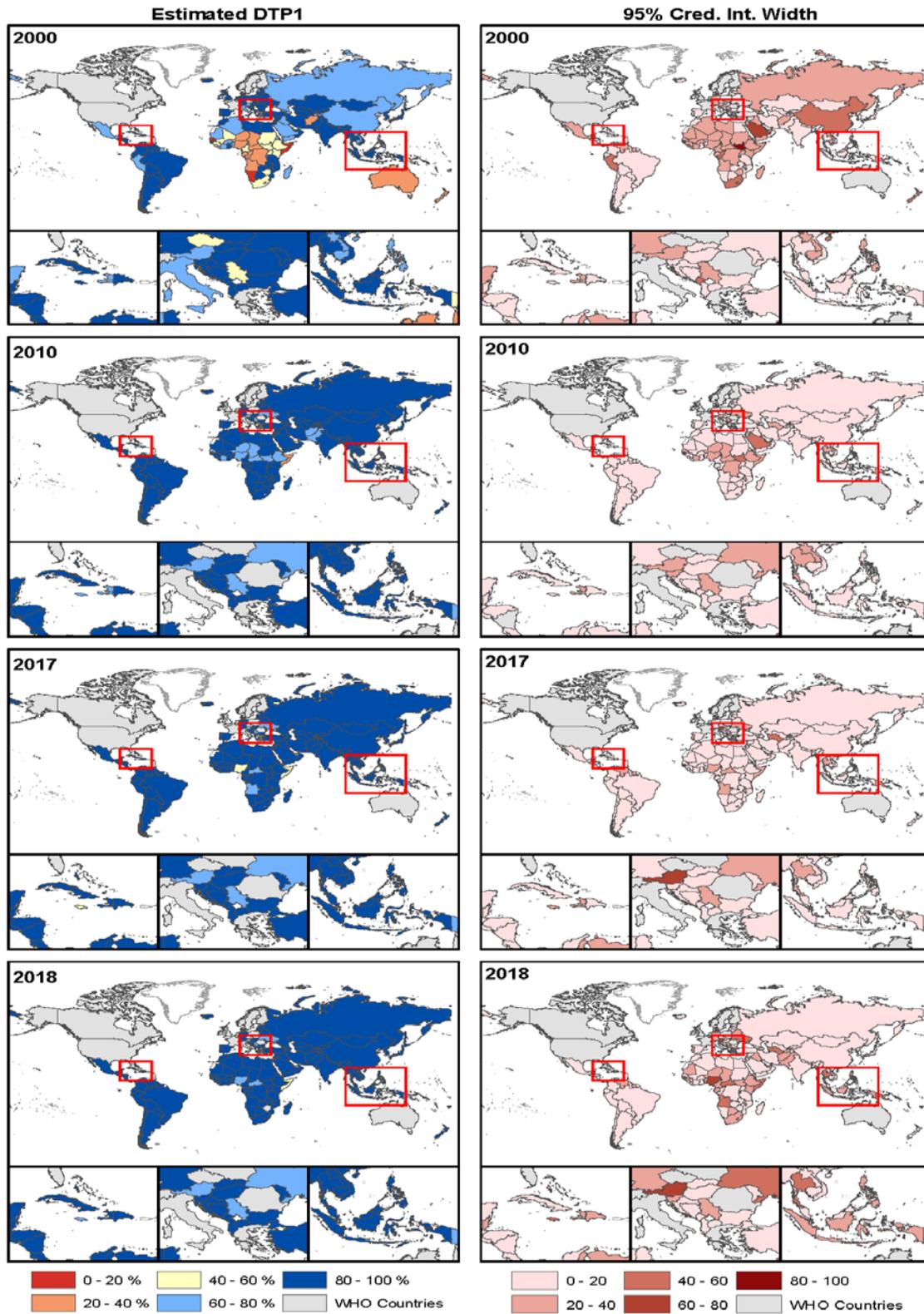


Figure A10: Modelled estimates of DTP1 coverage (i.e. the posterior means) and associated uncertainties shown as the widths of the 95% prediction intervals.

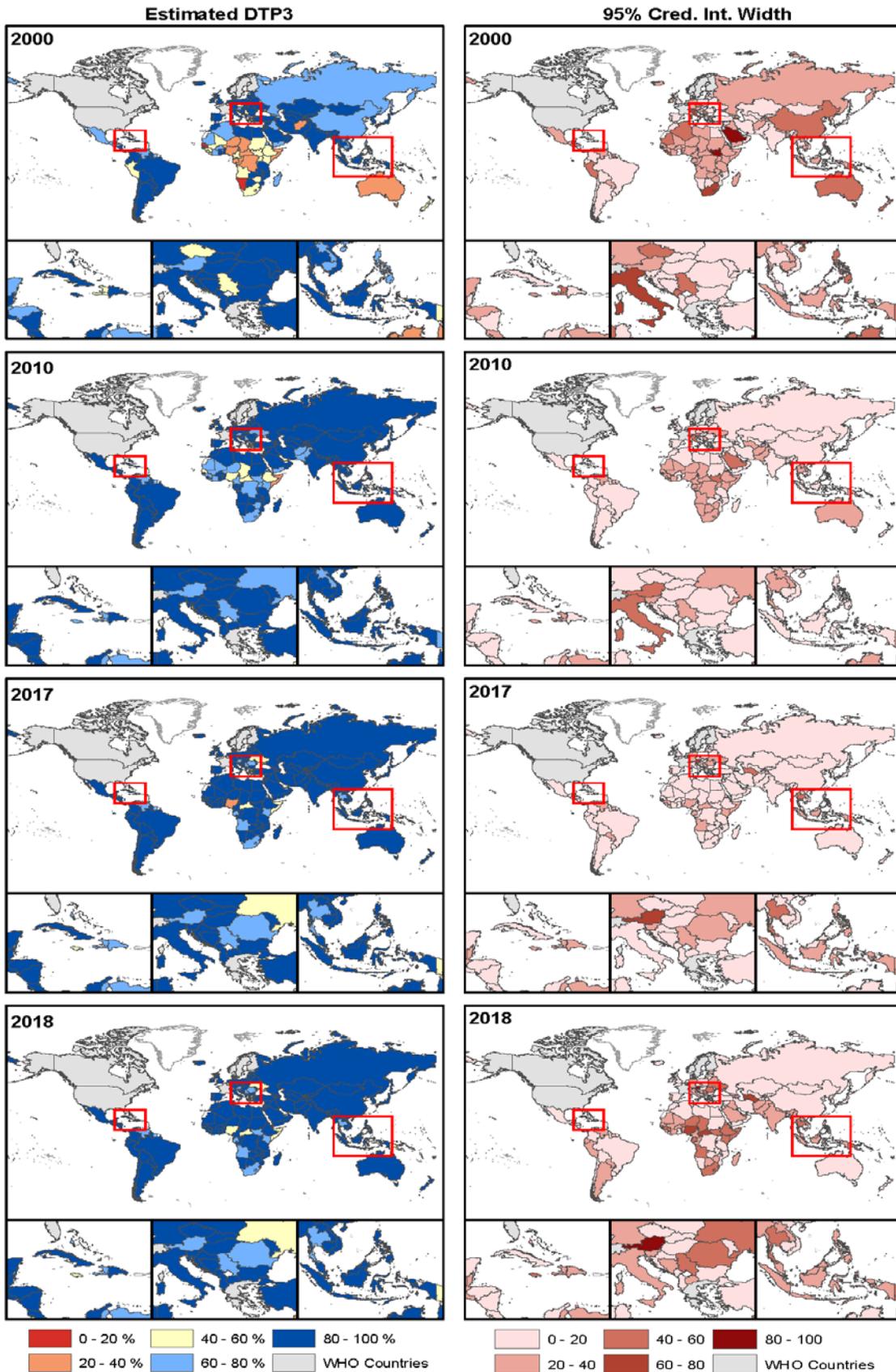


Figure A11: Modelled estimates of DTP3 coverage (i.e. the posterior means) and associated uncertainties shown as the widths of the 95% prediction intervals.

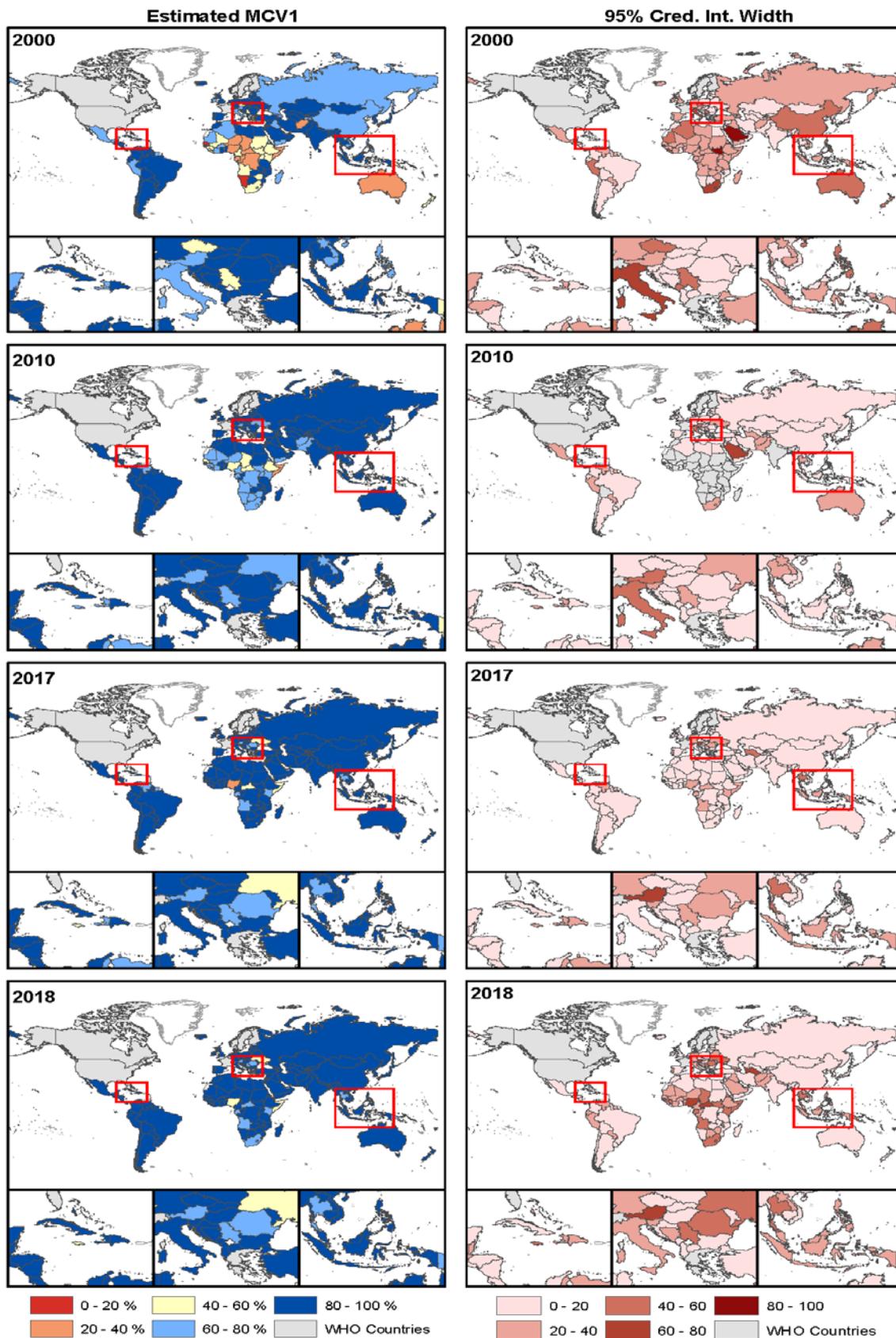


Figure A12: Modelled estimates of MCV1 coverage (i.e. the posterior means) and associated uncertainties shown as the widths of the 95% prediction intervals.

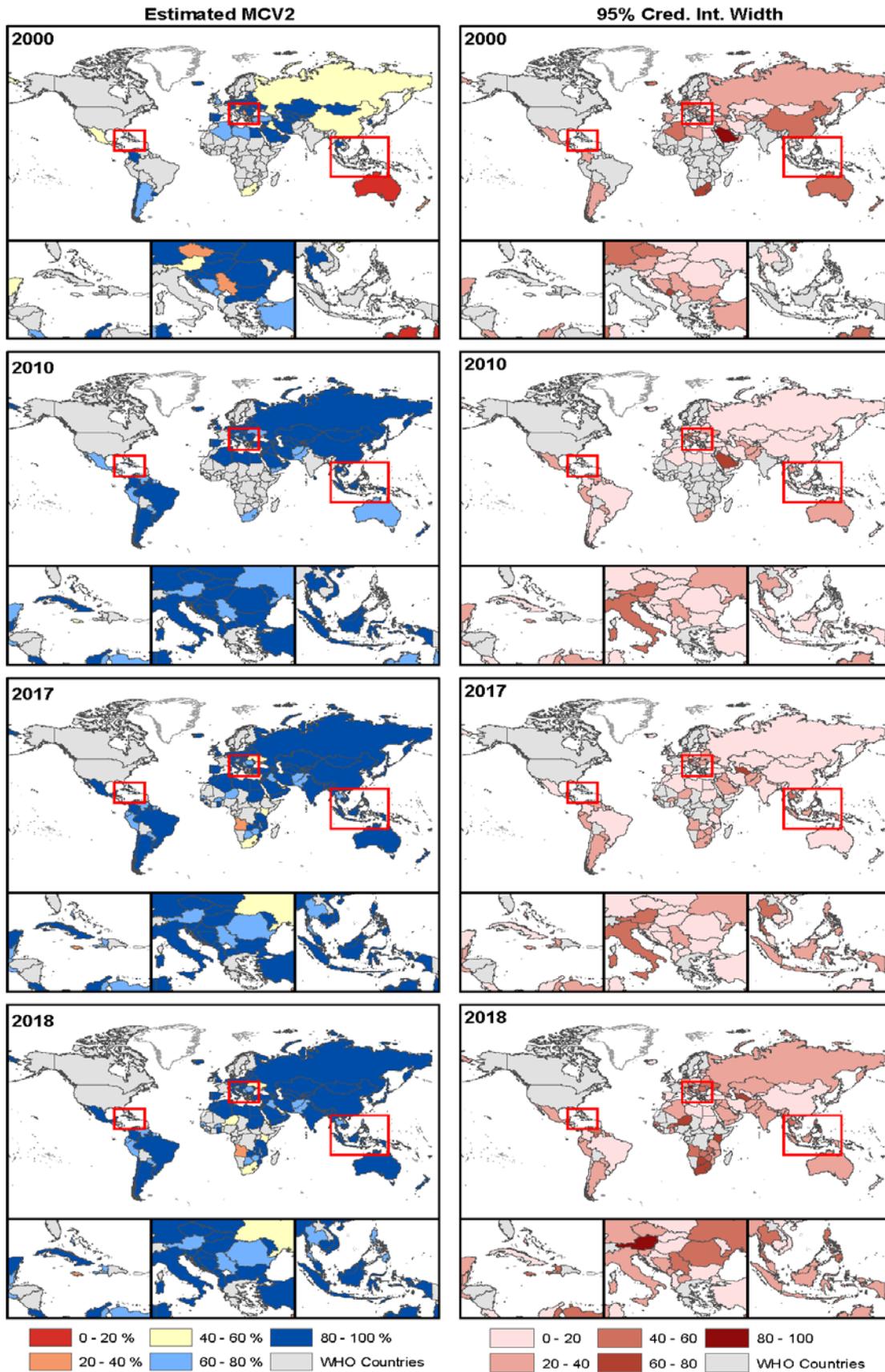


Figure A13: Modelled estimates of MCV1 coverage (i.e. the posterior means) and associated uncertainties shown as the widths of the 95% prediction intervals.

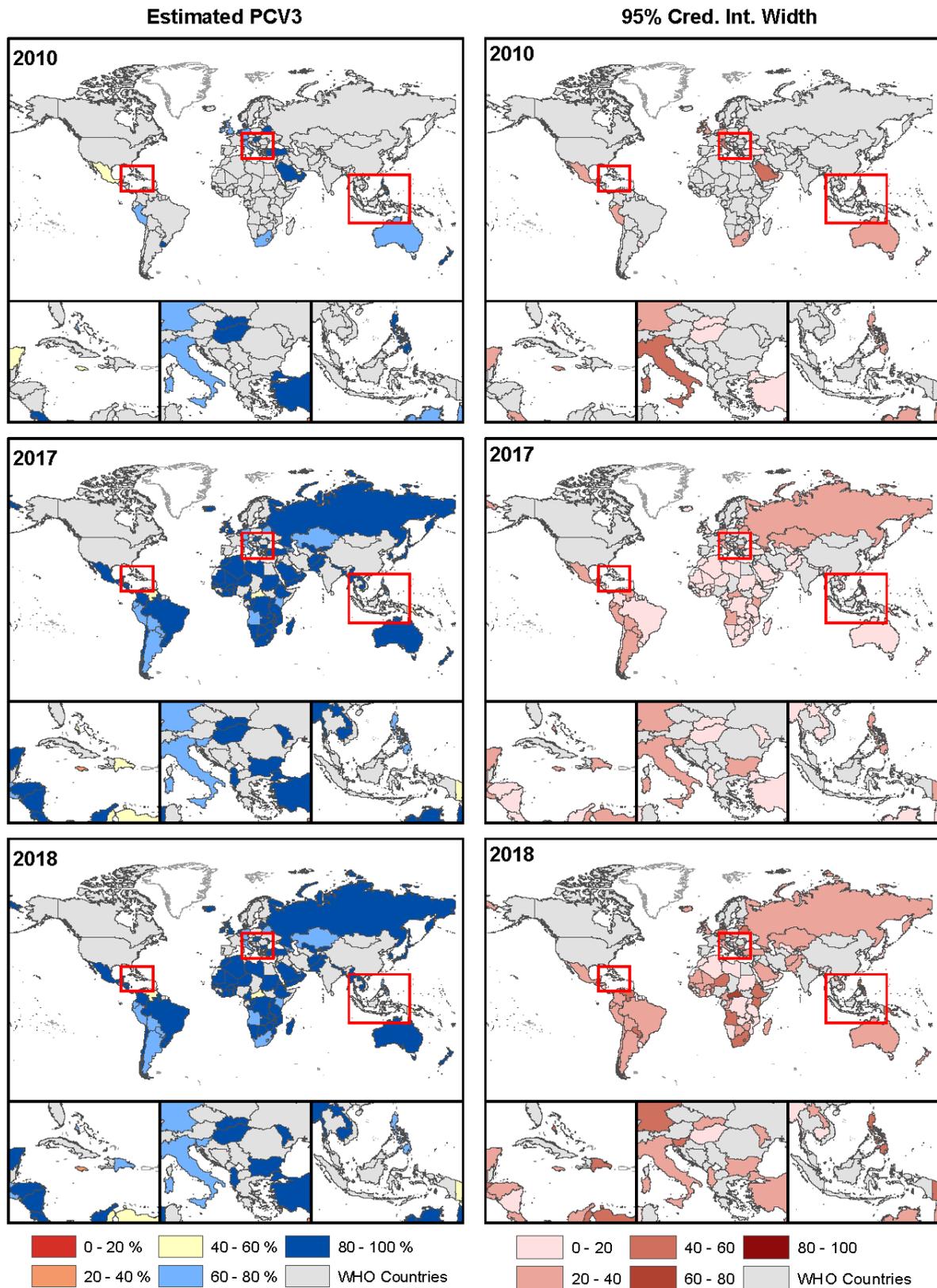


Figure A14: Modelled estimates of PCV3 coverage (i.e. the posterior means) and associated uncertainties shown as the widths of the 95% prediction intervals.

Additional figures and tables for residual analysis

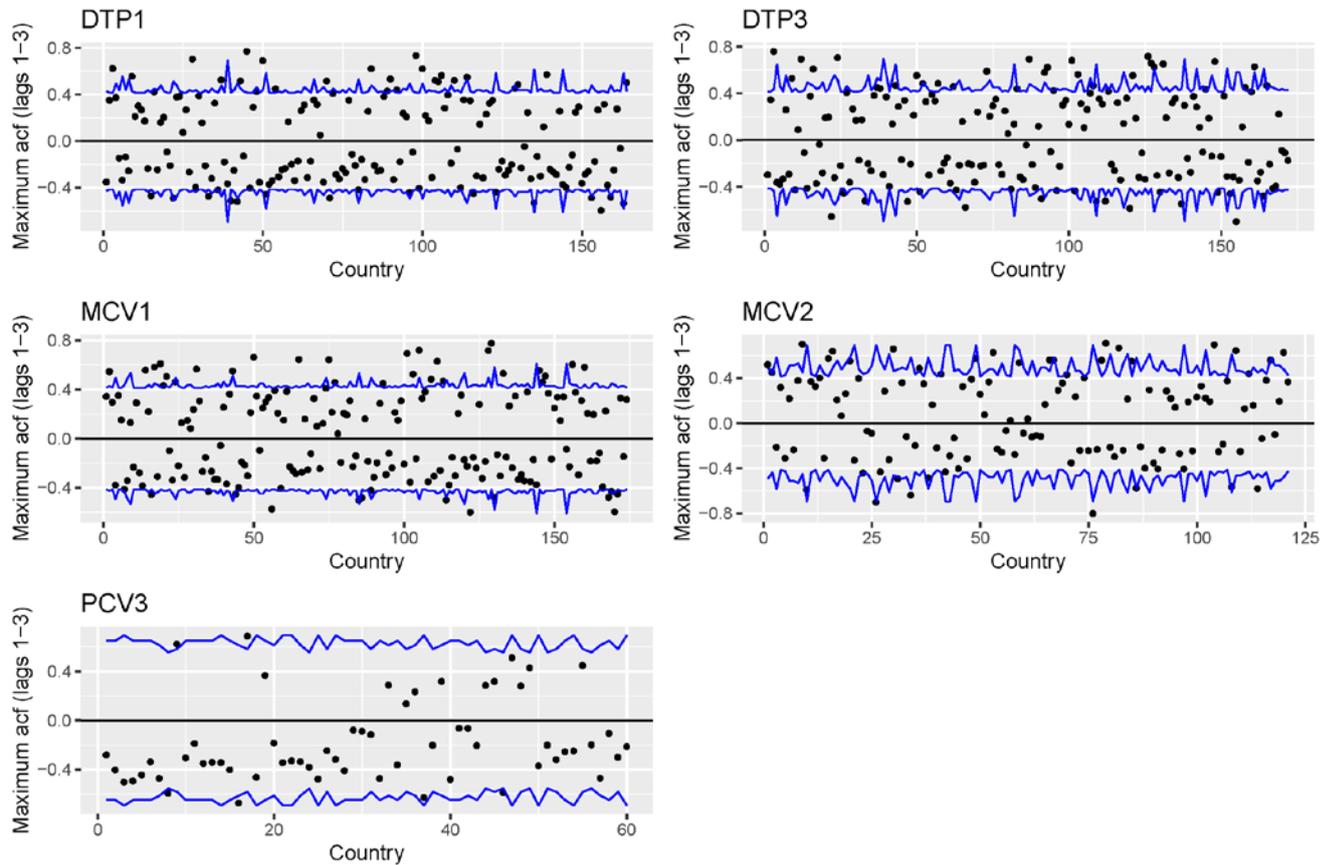


Figure A15: Residual autocorrelation plots for countries with at least 5 observations during 2000 – 2017. The blue lines indicate the 95% confidence limits for the autocorrelations.

Table A10: Countries with maximum lags 1-3 residual autocorrelation (in absolute value) greater than 0.5 based on at least 5 observations during the period 2000 - 2017. Bolded countries also have significant positive residual autocorrelation based on complete observations.

| Vaccine | Country | Total |
|---------|--|-------|
| DTP1 | Albania , Azerbaijan, China, Costa Rica, Djibouti , Denmark, Dominican Republic, Ecuador , Fiji, Jamaica, Libya, Morocco, Malta , Montenegro, Mauritius, Malawi , Malaysia, Nicaragua, New Zealand, Seychelles, Saint Vincent and the Grenadines, Zimbabwe | 22 |
| DTP3 | Albania, Azerbaijan , Burkina Faso , Bahamas, Barbados, Bhutan, Cook Islands, Egypt, Estonia , Haiti, Istanbul, Libya, Lithuania , Latvia, Morocco, Malta , Montenegro, Nigeria, Pakistan , Panama, Poland, Korea , Portugal, Russia, Sierra Leone, Syria, Trinidad and Tobago, Ukraine | 28 |
| MCV1 | Angola, Bangladesh , Bahamas, Belarus , Bolivia, Cameroon , Denmark, Estonia, Georgia , Honduras , Jamaica , Macedonia, Malta, Montenegro, Malaysia, Nigeria, Pakistan, Poland, Korea , Russia, Sao Tome and Principe , Slovakia, Tonga, Tuvalu, Samoa | 25 |
| MCV2 | Afghanistan, Azerbaijan, Belarus , Belize, Cape Verde, Germany, Ecuador, Jamaica, Sri Lanka, Myanmar, Montenegro , Mauritius, Netherlands , Oman , Switzerland, Tonga, Ukraine, Yemen | 18 |
| PCV3 | Bahrain, Bahamas, Costa Rica, Germany, Oman | 5 |

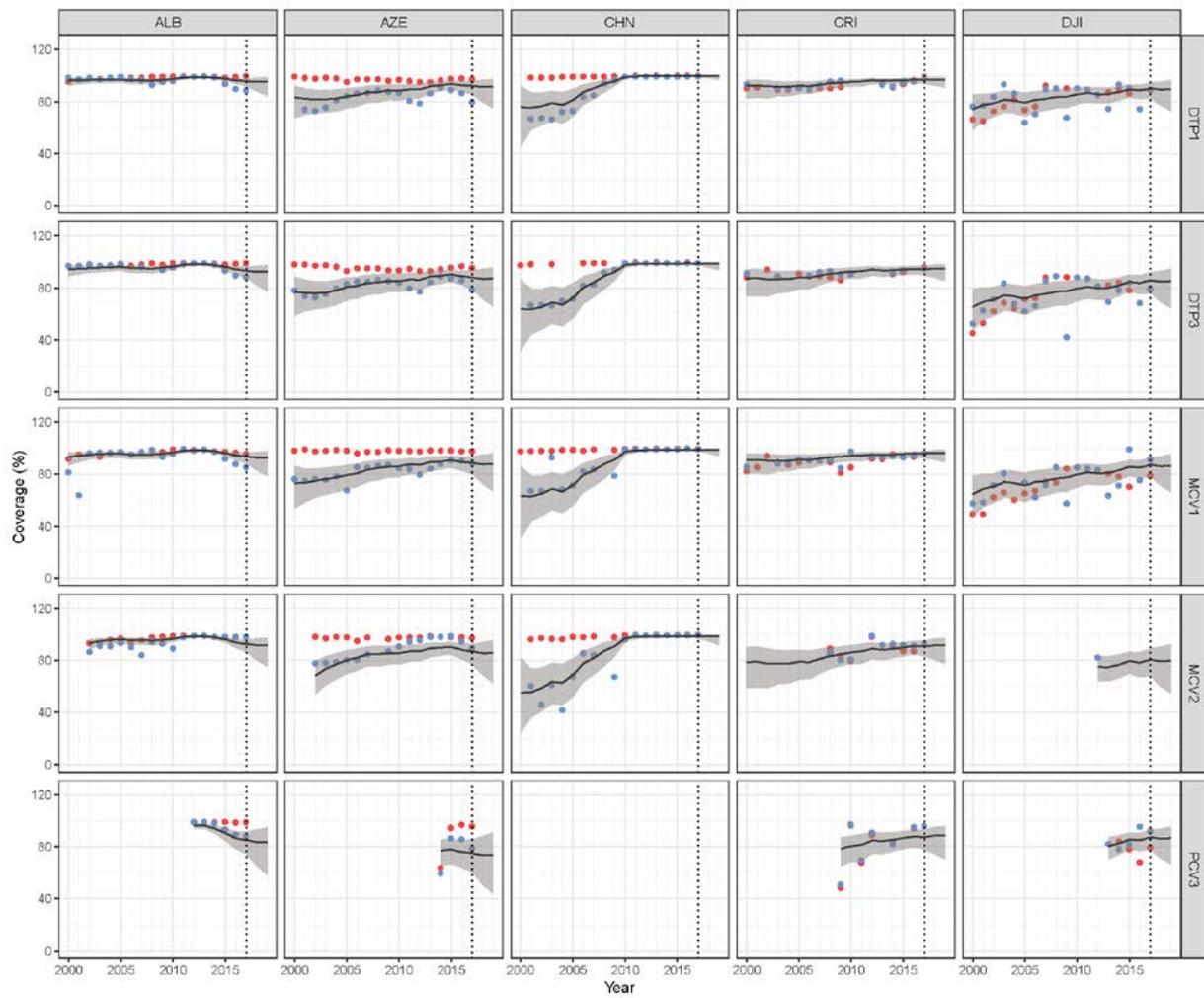


Figure A16: Plots of some countries with maximum absolute lags 1-3 residual autocorrelation greater than 0.5 (left – right: Albania, Azerbaijan, China, Costa Rica and Djibouti). These countries had at least 5 observations during 2000 – 2017.

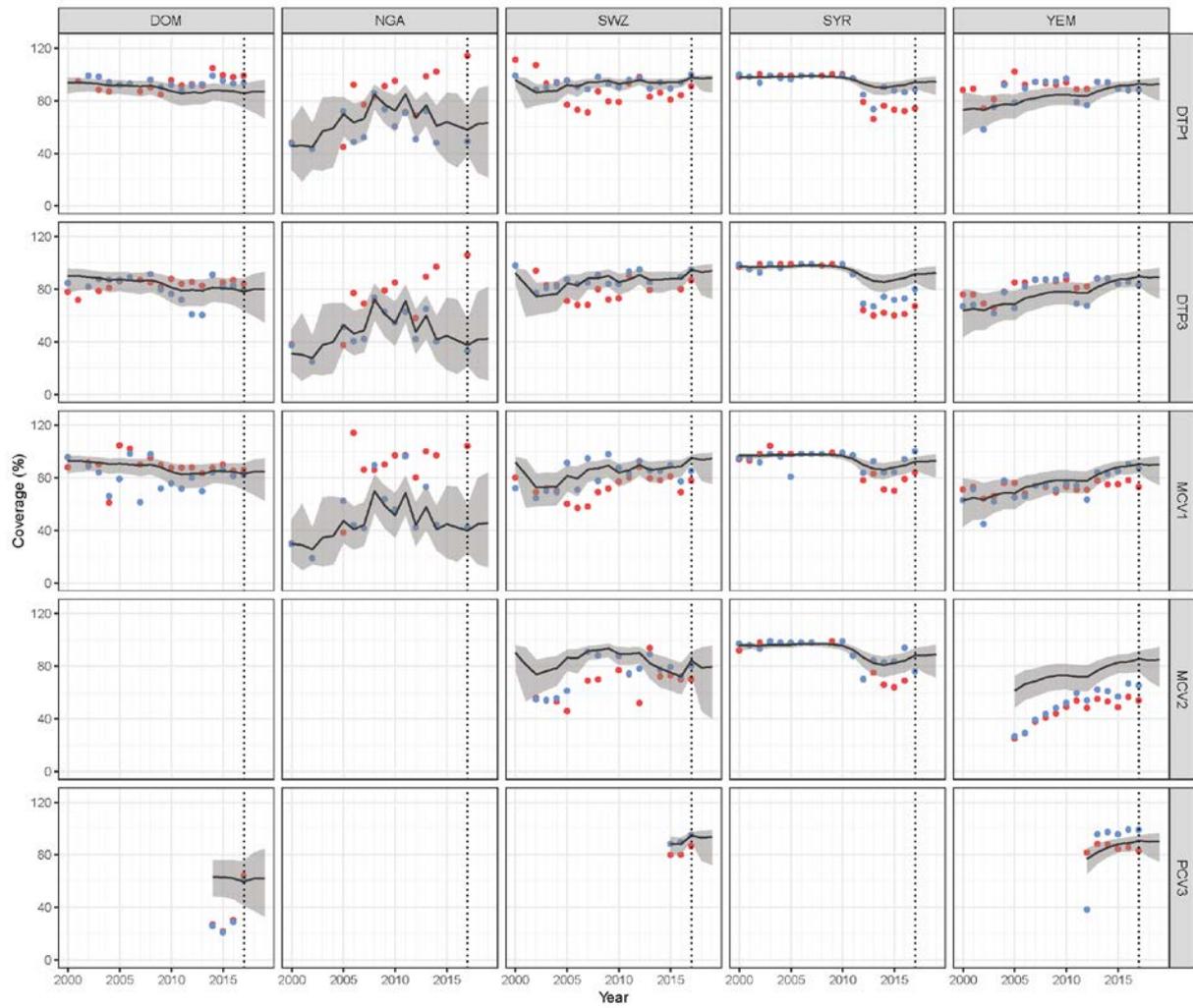


Figure A17: Plots of some countries with maximum absolute lags 1-3 residual autocorrelation greater than 0.5 (left – right: Dominican Republic, Nigeria, Switzerland, Syria and Yemen). These countries had at least 5 observations during 2000 – 2017.

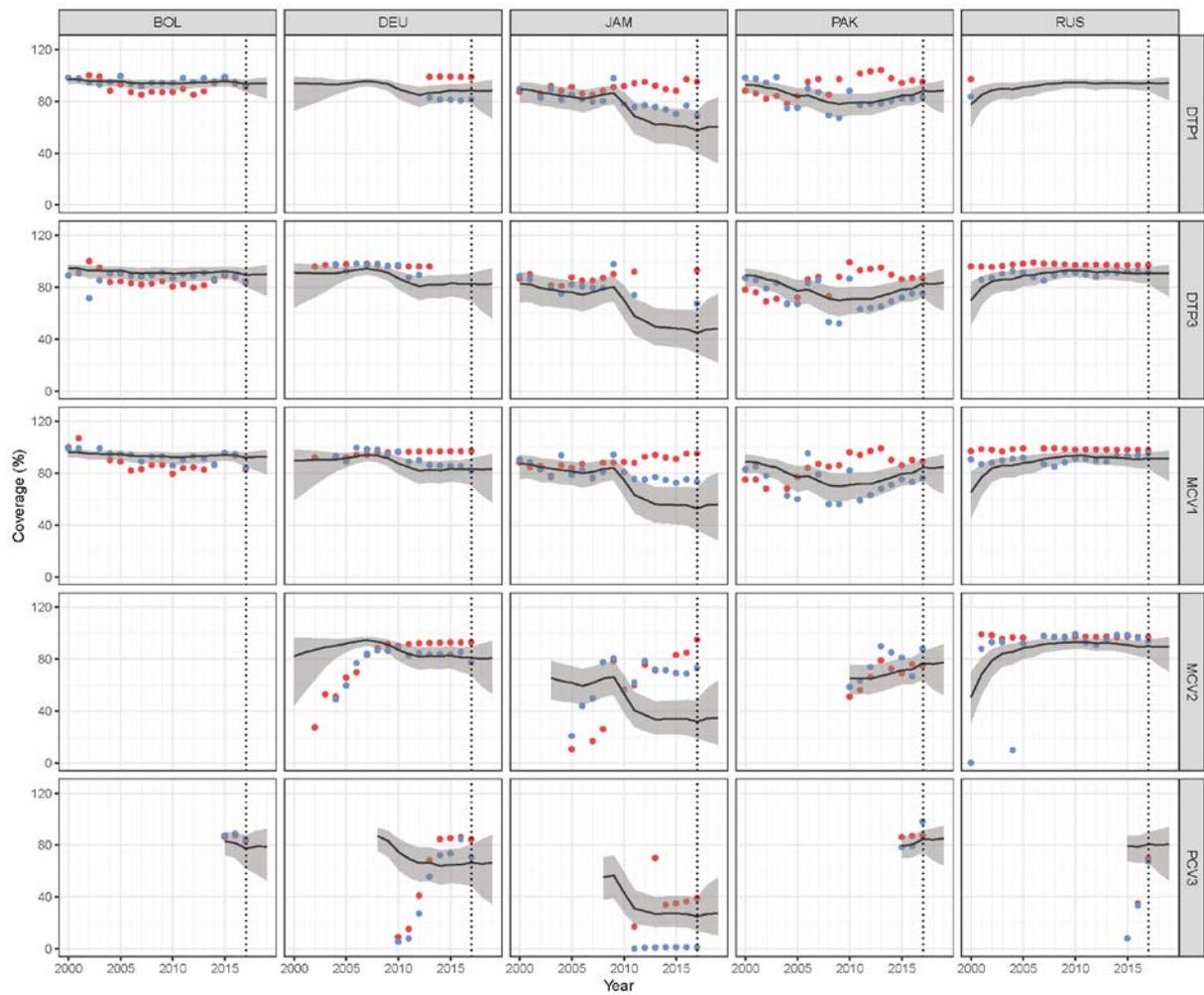


Figure A18: Plots of some countries with maximum absolute lags 1-3 residual autocorrelation greater than 0.5 (left – right: Bolivia, Germany, Jamaica, Pakistan and Russia). These countries had at least 5 observations during 2000 – 2017.

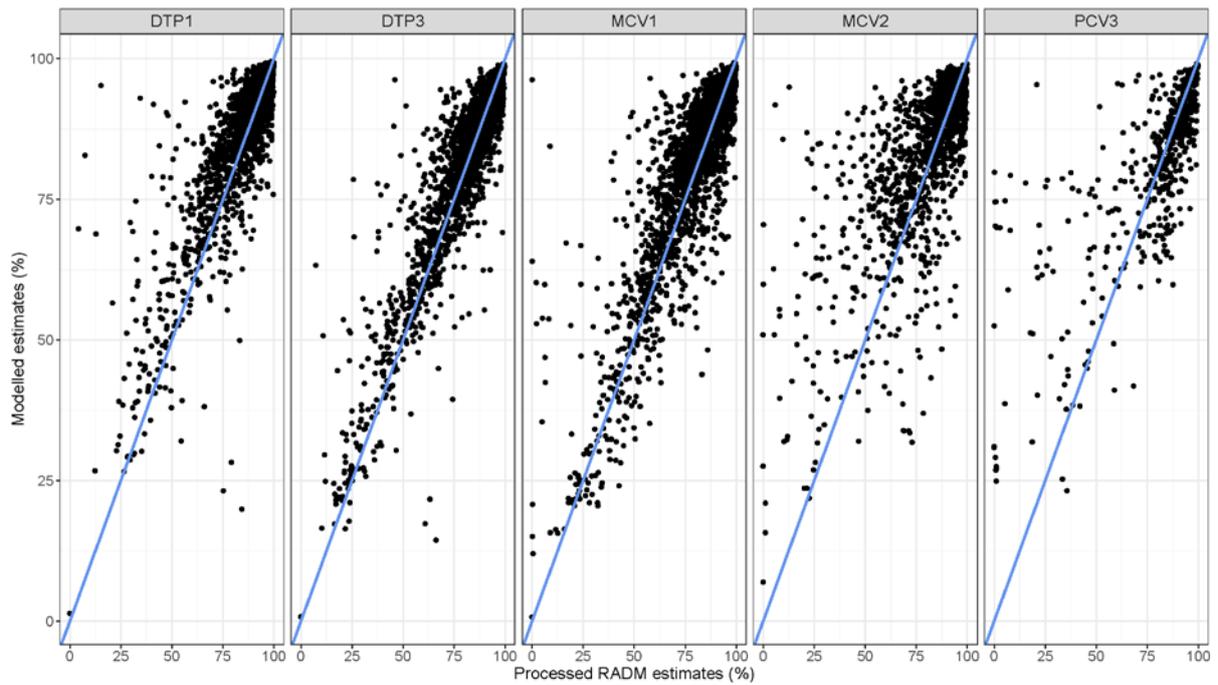


Figure A19: Plots of processed reported administrative estimates (RADM) versus modelled estimates for all five vaccines. The data plotted are for 2000-2017 and these include all countries in each case.