## Literature Review

### Existing Methods for Estimating the MMR

#### The United National Maternal Mortality Estimation Inter-Agency Group (MMEIG) Estimates

The MMEIG is a collaboration between United Nations Member States, the WHO, the World Bank Group, the United National Population Fund (UNFPA), the United National Agency for Children (UNICEF), and the United National Department of Economic and Social Affairs, Population Division (UNDESA/Population Division). This collaboration produces estimates of regional and country-specific maternal mortality rates (MMR) between 2000 and 2023 for women between the ages of 15 and 49 [1]. The group has published multiple reports at different intervals in this time frame, with each report adding MMR estimates for country-years elapsed since the previous report and, in some cases, updating the model used to estimate the MMR [1].

The MMEIG considers data from a variety of sources to produce these MMR estimates. More specifically, it uses data from Member States’ civil registration and vital statistics (CRVS) systems, as well as information from specialised maternal mortality studies, surveys, censuses, and national surveillance data, as described in the background information. It only uses nationally representative data and thus does not consider data from health facility studies [3, 12]. Observations from sources other than CRVS systems and specialised studies were increased by 10% to account for underreporting due to deaths occurring early in pregnancy that may have been missed. Additionally, the MMEIG uses data from other United National agencies as inputs and covariates in its MMR model [1]. For example, the MMEIG sourced estimates of all-cause deaths among women aged 15 to 49, live births, total fertility and age-specific fertility rates, gross domestic product (GDP), skilled birth attendance rates, and deaths due to HIV from other agencies.

The MMEIG’s current MMR estimates are produced by a combination of two models. First, the Bayesian maternal mortality misclassification (BMis) model calculates adjustment factors for the provided CRVS data to account for under-reporting of maternal deaths and reporting errors [1,3]. These reporting errors are due to either incorrect medical classification of cause of death or incorrect application in applying the correct ICD code [1, 3]. Errors in CRVS data were defined in terms of sensitivity (probability of correctly classifying a maternal death) and specificity (probability of correctly classifying non-maternal deaths). Global and country-specific adjustment factors were calculated using specialised studies, as specialised studies are considered ‘gold-standard data’, and can thus provide a benchmark for the accuracy of CRVS data [3]. Previous studies have found that the level of misclassification in CRVS systems compared to specialised studies is typically around 50% [2]. Given that 53% of the total data input into the MMEIG’s maternal mortality estimate model comprised of CRVS data, these adjustment factors are essential to being able to accurately predict MMR [3]. Global estimates were calculated by fitting the BMis model to specialised study data from all countries and were used to determine the adjustment factors for countries with no specialised studies [3]. For countries with at least specialised study, the BMis model is fit to country-specific data, but with hyperparameters from the global model [3].

After using the BMis model to adjust the CRVS for errors, the MMEIG estimates the MMR per country per year using the Bayesian maternal mortality estimation (BMat) model [1]. In the latest iteration of the BMat model, the MMEIG estimates MMR for WHO Member States with populations of at least 100,000 people [3]. The BMat calculates MMR as the sum of non-HIV-related maternal deaths and HIV-related maternal deaths, where death was due to pregnancy-induced aggravation on the existing HIV/AIDS condition [1]. To reduce errors caused by under-reporting of all-cause deaths, the model first calculated the proportion of maternal deaths among all-cause deaths, then converted this measurement into the MMR [2].

The BMat estimates non-HIV-related MMR using a Bayesian hierarchical regression model [3]. Briefly, a hierarchical model determines a general trend and individual-specific deviations from the trend, which are referred to as random effects [15]. In the case of the BMat model, the non-HIV related maternal deaths were estimated with region and country specific trends, indicating the belief that countries in the same geographic have similar trends in their MMR [2]. The general trend was the global parameter values [2]. The model generates the general and individual trends using prior knowledge, with each of the ‘trends’ defined by a parameter that is drawn from a prior probability distribution [15]. These prior distributions are used to generate model parameters before fitting the model with data, with the size of the distribution indicating certainty in the parameter’s value [15]. The parameters are updated via Bayes’ update rule upon observation of data, with less data causing the parameters to more closely resemble the parameters drawn from the prior distribution [15].

The BMat model’s covariates are GDP per capita, general fertility rate, and presence of a skilled birth attendant. These covariates were representatives of three broad predictive groups: social and economic development (GDP), process (skilled care), and risk exposure (fertility) variables [1]. The expected non-HIV MMR was then multiplied by a data-driven multiplier, which adjusted the expected non-HIV MMR based on how closely the data indicated that the non-HIV MMR tracked with the covariates. For example, if the data indicates that the non-HIV MMR decreased more slowly than predicted by the covariates alone, the data-driven multiplier would reduce the MMR estimate [1]. The multiplier was estimated using an autoregressive ARIMA process [3]. The less data available for a specific country-year, the more strongly the covariates affect the non-HIV MMR estimate and the less strongly the effect of the data-driven multiplier [1]. This inclusion of a data-driven multiplier was a relatively new addition to the model, with versions of the model used before 2014 not including the multiplier [2]. The use of the data-driven multiplier was motivated by criticism that the earlier models which solely based their MMR predictions on country-specific covariates did not always echo country-level trends in maternal mortality, which were important reflections of a country’s progress toward meeting the Millenium Development Goals [2].

While the model was fit to all data provided by the country, it placed higher weight on values with lower error variances, which were derived from calculating the random error in the data collection processes. As a result of incorporating error, the final BMat estimates had smaller uncertainty intervals for countries with higher-quality data [1].

HIV-related MMR was estimated separately, as evidence indicates that HIV/AIDS is a prominent cause of maternal mortality in countries with ongoing HIV/AID epidemics, with studies showing that women infected with HIV have approximately 8 times higher risk of pregnancy-related death [1, 19]. The HIV-related MMR estimate was calculated using data about the proportion of maternal and pregnancy-related deaths caused by HIV [2]. This calculation relies on a constant that defines the relative risk of dying from HIV/AIDS for a pregnant versus non-pregnant women, which is estimated in conjunction with experts and thus is semi-subjective [1,2]. The subjectivity in this constant is demonstrated through the change in its value between the 2014 and most recent BMat models [18, 2].

The MMR estimates produced by the MMEIG also included maternal deaths that occurred between 2019 and 2023 due to the COVID pandemic [1].

##### Limitations of the UN MMEIG Estimates

While the UN MMEIG’s estimates are used by the international community to inform policy, they have limitations. The model’s use of sparse and low-quality maternal mortality data has meant that the MMR estimates generated by the model are associated with substantial uncertainty [2,3]. Unfortunately, the underreporting and misclassification errors that reduce data quality can be difficult to correct, as there is limited data about the extent of errors in different countries and systems, reducing the efficacy of the BMis model [3]. This is especially true in countries without CRVS systems, which have lower quality and quantity data [13]. Wide uncertainty bounds around the MMR estimates reduces the estimates’ utility and ability to inform policy, penalising countries with less developed data collection systems, which generally have higher MMRs and would thus benefit greatly from better, evidence informed programs [3, 13]. Additionally, low-quality data can cause the BMat model to produce misleading results, again limiting the estimates’ ability to effectively inform policy.

Low-quality training data can limit the usefulness of UN MMEIG MMR estimates in other ways. The Bayesian hierarchical technique used by the BMat model is sensitive to the choice of prior distribution, where lack of information about a parameter can force researchers to use wide prior distributions [15]. When there is little empirical data available to inform the posterior distribution, causing the parameter’s value to largely rely on the prior distribution, this can result in incorrect results with large uncertainty, impacting the validity of the model’s results [15]. This is especially the case for low-income countries with high MMR values, as they would have little information to inform the priors.

Additionally, when there is no data, or only very low-quality data, for a country, the BMat model estimates its non-HIV related maternal deaths solely using the regional trend and covariates [2]. However, there are a few limitations of this approach. Within regional heterogeneity can result in the country’s estimates being pulled towards a regional-level estimate that does not reflect the local reality, resulting in inaccurate country-level predictions. Additionally, the model assumes that the relationship between MMR and the covariates is globally applicable, but this assumption is not always true. For example, one of the three covariates used in the BMat model is the presence of a skilled birth attendant (SBA). However, research has found that SBA only significantly reduces MMR when SBA coverage across the country is at least 40% [16]. Thus, countries with very low rates of SBA coverage would have a different relationship with MMR than countries with 90% coverage, which could contribute to inaccurate model predictions.

The authors of BMat model describe the need for further exploration of alternative predictor variables back in their 2014 paper [18]. There are a wide variety of factors that influence maternal health outcomes. For example, non-communicable disease (NCDs) are a leading cause of maternal mortality, with cardiovascular disease being one of the major causes of maternal mortality between 2018 and 2020 in Australia [27, 28]. Other NCDs like diabetes, asthma and mental health conditions also commonly affect pregnant women, with anemia increasing probability of postpartum hemorrhage, the primary cause global of maternal mortality [27]. NCD related risk is due to ongoing chronic conditions exascerbated by pregnancy and conditions developed due to pregnancy [27]. Additionally, studies have highlighted the relationship between medical care and maternal mortality burden, with emphasis on increasing quality of care and reducing unnecessary medical procedures as well as increasing coverage [29]. Researchers have also found that addressing socio-demographic trends can substantially reduce maternal mortality, with ensuring that all women finish high school producing a reduction in maternal mortality of a similar size to a strategy that increases health coverage [31]. The literature also gives evidence for how excess maternal mortality can be linked to the incidence of infectious diseases like malaria, availability of contraception, financial constraints, violence in the woman’s region, the woman’s geographic remoteness, racism in the health system, and gender inequities that influence the gender-based violence and the woman’s ability to make decisions about childbearing and medical care [c]. Studies have also found that women are more vulnerable to climate-related hazards, such as air pollution and extreme temperatures [30].

The wide variety of variables relating to socio-economic, health outcomes, and health system quality indicates that the BMat model’s consideration of only three covariates limits its accuracy. Additionally, it may be difficult to use the model to determine the effects of different candidate policies [9]. For example, targeting aggregate covariates like GDP per capita may not be the best approach, as the mechanism through which GDP per capita impacts maternal mortality involves a variety of other factors.

Researchers found that the BMat model may overestimate recent decreases in maternal mortality in low-income countries when tested on out-of-sample data from more recent years [2]. This overestimation may be due to the limitations in the modelling process discussed above. As a result, the authors indicated the need for further investigation into the modelling techniques used, with the suggestion that this investigation may result in their choosing a new modelling technique in the future [2].

#### The Global Burdan of Disease Estimates

Coordinated by the Institute of Health Metrics and Evaluation (IHME), the Global Burden of Disease Study (GBD) is an international scientific initiative that benchmarks major diseases, risk factors, and clinical intermediate outcomes [5]. It was partially motivated by the need for population health estimates to be produced by an agency independent of the WHO to encourage scientific debate via use alternative modelling techniques [22]. Additionally, unlike the UN MMEIG estimates, which are produced under consultation with Member States, the GBD estimates are produced independently, which researchers have pointed out could remove the prospect of biased results due to pressure from the Member States [21].

The GBD uses a standardised approach to be able to compare its estimates across time, geographic populations, and health conditions [5]. GBD studies have been ongoing since 1993, with the estimates being reviewed by independent experts as well as the WHO Headquarters and Regional Offices [5]. Additionally, many of the studies have been published in prestigious journals like *Lancet* after being subject to peer-review cycles [5]. The GBD Studies’ strong reputation is apparent in their use in national planning by a variety of governments, such as the United Kingdom, Norway, and China [5].

The 2021 GBD Study produced estimates of maternal mortality for 204 countries between 1990 and 2021 using data from a diverse range of sources, including CRVS, verbal autopsy, sibling histories, surveillance, survey and census data, as well as police records and open-source databases [6]. Unlike the UN MMEIG estimates, the GBD estimates can be informed by sub-national data [6]. The data is then cleaned, standardised, and any deaths reported with an unclear or incorrect cause-of-death are redistributed to a more likely cause of death [8]. This redistribution is performed by first assigning the garbage codes to a group of possible underlying causes of death. The garbage-coded deaths are then probabilistically redistributed among these different causes [8]. Data quality is quantified by a star rating system, where higher star estimates are associated with more complete, available data with a smaller presence of garbage codes [8].

The GBD Study estimated maternal mortality using cause of death ensemble modelling (CODEm) [4]. This involves generating multiple statistical models to capture the large number of factors that can co-vary with maternal mortality [4]. CODEm uses linear mixed effects regression (LMER) and spatiotemporal Gaussian process regression (ST-GPR) models, unlike the MMEIG estimates, which instead rely on Bayesian hierarchical regression [4]. Each of the models are used to test candidate covariates to maternal mortality, with each combination of covariates tested for plausibility based on the literature as well as statistical significance [4]. The candidate covariates are split into three groups depending on the level of evidence in the literature that supports a causal relationship between the covariate and the mortality measure [7]. Level 1 covariates have a strong causal relationship based on the biological disease process, while level 2 covariates have strong supporting evidence but no direct biological relationship [7]. Level 3 covariates have weak evidence supporting a causal relationship with the mortality measure [7].

The GBD 2021 Study used more covariates than the MMEIG, with the former using 19 covariates while the latter only used 3 (GDP, skilled birth attendance, and general fertility rate) [4, 1]. More specifically, the GBD study estimated maternal mortality using covariates including, but not limited to, age-specific fertility, maternal education, neonatal mortality ratio, skilled birth attendance, age-specific HIV mortality in females 10 to 54 years old, and age-standardised wasting [4]. Models based on this best set of covariates were then generated and tested on out-of-sample data [4].

The models’ weights in the ensemble were determined by their out-of-sample predictive performance [4]. The final mortality prediction is the mean of 1000 ensemble predictions, with each prediction being generated by one individual, component model. The likelihood of each model being chosen is determined by its weight [4]. The 1000 draws also allow the construction of a 95% uncertainty interval [4].

##### Limitations of the GBD Estimates

Unfortunately, the GBD estimates are associated with many of the same limitations as the estimates produced by the UN MMEIG. The primary limitation is inaccuracy cause by their use of sparse and low-quality data, which the GBD authors have noted as being especially sparse for south Asia and African sub-regions [6].

Additionally, there are limitations associated with each of the models chosen for use in the CODEm ensemble. Briefly, the LMER models have similar weaknesses to the BMat model, where they rely on fixed effects from chosen covariates with random effects introduced by super-region, region, country, and age [7]. As above, if a country lacks empirical data, estimates of its MMR are pushed toward the region’s mean MMR. This ignores region-level heterogeneity and could particularly affect countries in south Asia and African sub-regions, which the GBD authors have highlighted as having particularly sparse and low-quality data [6]. This smoothing effect could also be pronounced for age groups with lower pregnancy MMRs or about which MMR is less frequently measured, such as in teenagers and older women. Moreover, this model formulation assumes that relationships between the covariates and MMR are constant, which may be untrue, as discussed above. Again, these limitations would reduce the accuracy of the model’s estimates.

The ST-GPR used in the CODEm ensemble incorporate information about how maternal mortality varies across age, space, and time by analysing errors between the LMER models’ predictions and the ground truth [7]. In the regression analysis, each datapoint’s individual error is replaced by the error the weighted average of the errors in its neighbourhood [7]. This neighbourhood is defined by distance over time, geographical space, and age [7]. The datapoint weighting smooths the errors to reduce noise, with the extent of smoothing increasing for countries with less data [7]. While this reflects uncertainty in the data, it can also create inaccuracy if there is heterogeneity between the country being modelled and the data being weighted more highly in the smoothing process. When there is no data for a particular country, the model even uses information from the super-region, which may be unrepresentative of the specific country for which maternal mortality is being estimated [7]. Additionally, the smoothing pushes the estimate toward lack of data is due to an abrupt change or crisis. In this case, the model could lose important information that would have a material effect on its predictions of maternal mortality.

Another limitation of the approach used by the GBD is the lack of interpretability associated with its CODEm model, where using a combination of complex models encoding a variety of relationships can make it difficult to interpret the underlying associations between covariates and maternal mortality. This could hinder the estimates’ ability to effectively inform health policy.

While the GBD study predicts maternal mortality using more covariates than the UN MMEIG (19 covariates versus 3), only two of the covariates used are not related to quality of care, fertility or mortality rates [4]. As discussed in the previous section, there are a wide variety of socio-economic factors that have been shown to impact maternal mortality outcomes. Unfortunately, the desire to avoid multicollinearity reduces the number of possible covariates that can be used in this model [4]. Thus, the model cannot consider all the factors that may affect maternal mortality rates, both limiting the accuracy of its mortality estimates and reducing its ability to inform policymakers about the most important areas to target to reduce mortality. Additionally, the process for testing candidate covariates for multicollinearity with maternal mortality I itself computationally intensive, adding complexity to the model.

As a final point, researchers have noted that the authors of the GBD study do not have access to all available national and sub-national data due to data privacy restrictions [20]. This can result in further discrepancies between estimates produced by the GBD study and those produced by a country’s government [20]. For example, researchers found that the estimates from the German federal health reporting system of the number of diabetes-related deaths were outside the uncertainty intervals of the GBD’s estimates [20]. This reduces the usefulness of the GBD statistics to inform national health policy [20].

#### The Global Maternal Health Microsimulation Model (GMatH) Estimates

The authors of the GMatH model motivated their approach by describing how the models produced by the MMEIG and GBD Study may inadequately describe country-specific trends [9]. More specifically, they described how the MMEIG and GBD estimates were based on statistical relationships between aggregate country-level factors and MMR, preventing them from modelling variation within a specific country [9]. In contrast, the GMatH model simulates individual women’s reproductive lifecycles to determine estimates of maternal mortality, with differences in how those lifecycles are simulated used to reflect heterogeneity across the country [9]. This allowing the model to better represent country-specific variation [9]. More specifically, the model uses monthly cycles to simulate each stage of pregnancy to determine the incidence of complications related to pregnancy and childbirth [9]. As part of this simulation, the model must estimate the probabilities of pregnancy, termination, and complications as a result of individual-level and institutional risk factors [9].

Parameters governing transition probabilities to different states within the model were estimated from probability distributions based on empirical data where possible, and on expert opinion where data was unavailable [9]. Relationships between parameters were similarly derived through a mixture of empirical data and expert opinion [9]. The uncertainty in the models’ inputs was captured setting the parameters’ prior probability distributions based on a 5-level hierarchical model. The model was then fit to empirical data [9]. This empirical data was sourced from UN agencies, randomized clinical trials, observational studies, census data, survey data, expert opinion, and meta-analyses [9].

The model used 5 sets of parameters, categorised into biological parameters, family planning parameters, health system parameters, obstetrical complications, and clinical interventions [9]. Examples of biological parameters include age-specific probability of pregnancy, anaemia status, and risk of miscarriage, while examples of family planning parameters include contraceptive preferences. Health system parameters include the type of care available at birth, the probability of recognition of complications, and underreporting of maternal deaths. Examples of parameters that capture obstetrical complications are the risk of postpartum haemorrhage and risk of indirect obstetric causes, such as deaths due to HIV aggravated by pregnancy. Finally, examples of the parameters representing clinical interventions are the efficacy and availability of interventions for certain complications, such as hypertension management, as well as use of elective interventions, such as caesarean sections [9].

The authors argue that their use of a diverse range of parameters that simulate causal relationships between risks and the stage of a woman’s reproductive lifecycle is more informative than the correlation-based approach used by the MMEIG and GBD [9]. Consequently, the GMatH model can produce more robust predictions based on causal-inference, allowing policymakers to simulate programs that target maternal health [9]. Additionally, the breadth of parameters used by the GMatH model allows a wider variety

of policies to be modelled and health system barriers to be explored [9]. For instance, the GMatH model has been used to estimate the difference between maternal mortality outcomes for women from rural versus urban backgrounds, as well as for women with different levels of education [10]. Through this analysis, the study authors found that, in 2022, the global MMRs for women with a low, middle, and high education level were 536, 143, and 85, respectively [10]. Thus, the authors identified the importance of addressing women’s education as an avenue for improving maternal mortality outcomes [10]. In contrast, it is difficult to produce an effective policy to reduce maternal mortality based on the observation that the MMEIG’s maternal mortality outcomes are primarily predicted by GDP, which is a difficult outcome for politicians to change [9].

Through this process, the GMatH has been used to estimate annual values for maternal mortality indicators like cause-specific and total maternal deaths, MMR, live births, and the lifetime risk of maternal death [9]. These estimates were produced for 200 countries and territories between 1990 and 2050. Additionally, the calibrated model was used to make projections for each year up to 2050 [9].

To test the model’s predictive accuracy, the authors calibrated the model’s maternal death estimates using CRVS data collected between 1990 and 2015, then compared the model’s estimates for 2016 to 2020 to the CRVS estimates for the same time period. The mean absolute error for the total number of maternal deaths in test set was 47.5.

##### Limitations of the Global Maternal Health Microsimulation Model Estimates

Similar to the limitations mentioned above, one of the major limitations of the GmatH model was sparse and low-quality data, again impacting the accuracy of its predictions [9]. Additionally, lack of data reduced the authors’ ability to choose informative priors for the model’s parameters. For example, there were multiple instances where the authors had to apply the same prior to high-income countries that it calculated for middle-income countries [23]. These parameters were generally informed by Demographic and Health Surveys, which included data solely from lower income countries, preventing informative priors from being generated for high-income countries [23]. Examples of these parameters include the number of antenatal care visits, number of living children, skilled birth attendance given home birth, and unmet need for family planning [23]. Other parameters that lacked supporting empirical evidence were instead informed by expert opinion, which may not reflect the local reality [23]. These uninformative priors, or priors based on data about another income group, could reduce the accuracy of the model’s estimates, thus decreasing its ability to inform policy. Additionally, these parameters were often estimated using hierarchical models [23]. As described above in relation to the other models, parameter values for countries with sparse and low-quality data will rely more heavily on the prior distributions and other countries within its geographic area, potentially over-smoothing heterogeneity between the country and its neighbours [23].

There are also a variety of limitations unique to the microsimulation model in addition to potential misspecification of causal relationships between maternal mortality and feature variables [24]. While the model has the ability to consider a wide variety of explanatory variables, each variable is associated with uncertainty, especially in the case of countries with little empirical data for parameter-tuning [24]. By progressively adding variables, the model may become overly influenced by uncertainty and noise, with its estimates becoming dominated by the parameters’ uncertainty [24]. As a result, the variation in the estimates may increase beyond the point at which the estimates themselves are informative, as they cover too wide a range of outcomes [24]. This can be observed in the Figure below, which demonstrates the wider uncertainty intervals of the GmatH predictions than for UN MMEIG or GBD estimates. Small inaccuracies in the parameter values may also accumulate given the large number of parameters in the GmatH model, further decreasing accuracy [26]. Moreover, the more parameters included in the model, the greater its complexity, and thus the greater its likelihood of overfitting to the data sample [25]. This is particularly problematic when the data sample does not contain much information about low-income, data sparse countries, reducing the model’s ability to generalise to these countries.

Another potential limitation of the GmatH model is how it orders events in the simulated reproductive lifecycles. According to the original paper, the model “progresses in monthly intervals”, indicating that the simulated women’s states are updated at discrete timesteps [9]. Many of the parameters in the model are inter-related, and as a result the order in which they are updated can make a material difference in the model’s final estimates [24]. For example, if a woman gets pregnant and gets married with the month, the probability of marriage may increase in the case of pre-marital pregnancy, but if marriage occurs first, the likelihood of falling pregnant increases [25].

The validity of the GmatH model’s estimates is also affected by the starting state of its simulated population, which is a sampled population and thus could result in the model’s final estimates being unrepresentative [9, 24].

#### Comparison of the MMEIG, GBD, and GMatH Maternal Mortality Estimates

The GMatH global MMR estimates were similar to the MMEIG’s global estimates but notably higher than the GBD’s global estimates [9]. These global differences were shown in the figure below. Additionally, there were substantial country-level differences in the estimates produced by the three models [9]. The deviations in the models’ estimates could result in the development of differing policies depending on which model is used as evidence for the policy [9].

A graph of the number of pregnant women

AI-generated content may be incorrect.

<https://www.nature.com/articles/s41591-023-02310-x/figures/3>

The inter-model variation is likely a result of their different methodologies and input datasets. Given these differences, the authors of the GMatH model describes their hope that their intrinsically different modelling approach could provide insight into the reason why the models results have diverged [9]. Use of an alternative type of model is also motivated by the observation that all three models are computationally intensive, as they each require significant calibration and data processing as well as input from a variety of sources [25, 26]. This computation effort is compounded by the need to not only develop the sophisticated statistical models, but also to transform low quality data into a usable form and embed uncertainty in the data from different sources. As described above, the MMEIG produced an extra model for the sole purpose of adjusting the CRVS data to account for mistakes while the GBD estimates were produced after categorising data quality using a star-based system and implementing complex algorithms to re-label unspecific or incorrect causes of death within the input data [2, 4, 8]. The GMatH model introduced separate parameters solely to model site-specific underreporting of maternal deaths [9].

### Data Modelling versus Decision-Tree Based Machine Learning

The methods described in the previous section fit into the ‘data modelling culture’, where they estimate maternal mortality by trying to understand and model the processes that generate the observed data [32]. This is approach is reflected in their assumptions about the structure of the data and the relationships that exist within the data, which informs their choice of priors and covariates. In contrast, decision-tree based machine learning techniques are part of the ‘algorithmic modelling culture’, where the data generation process is considered a ‘black box’, and instead of trying to learn how the data is generated, the model simply focuses on predicting the outcome of interest [32]. Having an algorithmic modelling approach would avoid sources of error from generating informative priors and assuming small regional heterogeneity in maternal mortality rates when predicting for data-sparse countries.

Additionally, data modelling approaches can be affected by multicollinearity, which occurs when feature variables are linearly dependent [35]. This increases the difficulty of attributing change in maternal mortality to a specific feature variable and is problematic in models based in some way on linear regression, such as some of the models discussed above [35]. Additionally, by having multiple variables giving very similar information, the model is more likely to learn noise in the dataset and thus overfit and generalise less easily [35]. Researchers often select a considering a subset of relevant features to reduce multicollinearity, a process explained explicitly in the GBD model’s documentation [32]. However, this approach can result in the model losing valuable information that would improve its predictive accuracy. This is particularly relevant for the prediction of maternal outcomes, which are associated with a wide variety of socio-economic and health-related variables, as described previously. Studies have highlighted the importance of analysing these relationships to aid in crafting country-specific strategies for reducing maternal mortality [10, 33, 34].

This limitation of data modelling approaches is a strength of algorithmic modelling. Decision-tree based methods are particularly suited to working with high-dimensional data, as splits in the individual decision trees are determined by the feature partition that produces the largest decrease in error [36, 37]. Thus, if three feature variables are highly correlated and if a split is already based on one of the variables, the others are much less likely to be chosen for future splits because they would not add additional information. Therefore, decision-tree based methods can include all possible feature variables that could influence maternal mortality, unlike data modelling approaches. Additionally, if a variable does not help reduce predictive error because it is not correlated with maternal mortality, it will not be used in any splits and thus will be ignored [36, 37]. Thus, error in redundant and/or uninformative variables will have a much smaller contribution to uncertainty in the final prediction than in micro-simulation models, which rely on all variables, as discussed below. As a result of these strengths, decision-tree based methods can work well with high-dimensional datasets.

Another primary motivator for use of decision-tree based methods is their treatment of missing values. They use surrogate splits or a default direction to predict the outcome despite prescence of missing data, as described in the background information. This prevents them from needing to use imputation methods that may introduce error into their estimates or relying on a framework like Bayesian hierarchical models that borrow strength from potentially heterogeneous region-wide parameters.

As a result of these benefits of the algorithmic modelling approach and the limitations of the data modelling technique, researchers in maternal outcomes have stated the need to explore the application of these algorithmic modelling approaches to improve maternal outcome predictions [34]. Thus, I propose the use of a decision-tree based machine learning approach to estimate maternal mortality. The use of an alternative model will also provide an alternate set of estimates to help form a consensus from the range of estimates presented in the literature.

### Algorithmic Machine Learning Techniques in Public Health

To further motivate my choice of decision-tree based machine learning methods to predict maternal mortality outcomes, I will present exemplar studies of how these approaches and other algorithmic modelling approaches have been used in public health research. In this section, I will refer to all algorithmic models as machine learning models to use their common name despite the data modelling approaches discussed above technically also being machine learning models.

#### General Overview

Machine learning (ML) models are used for a wide variety of purposes in public health as researchers take advantage of the large quantities of health data being generated by wearable devices, clinical records, and social media [40]. For example, they can be used for image-based medical diagnostic tasks, improving operational efficiency, predicting patient-specific risks, and drug discovery [40]. Machine learning approaches can be categorised as ‘white box’, ‘grey box’, and ‘black box’ models depending on their level of interpretability, where ‘white box’ models are the most interpretable and ‘black box’ are the least [39]. ‘Black box’ models include modelling approaches like deep learning and neural networks and are frequently used in image-based diagnosis [39]. However, I will not focus on such models because their complexity and lack of interpretability make them unsuited for predicting maternal mortality outcomes in a way that allows identification of the most important feature variables to be able to inform policy [39]. In contrast, ‘white-box’ models like decision tree-based algorithms can be used to predict medical risk factors and complications in an interpretable manner, meaning the model user can understand the factors used by the model to produces its results [39].

Mahajan et al.’s (2023) review discussed the use of ensemble-based models in public health. It found that while boosting algorithms were the most popular in the literature (used in 37 of 45 studies), they were only the most accurate algorithm evaluated in a study 40.5 percent of the time [48]. In contrast, stacking and voting were less frequently used (23 and 41 out of 45 studies), but they had the highest accuracies 82.6% and 71.4% of the time, respectively [48]. Stacking models’ high performance was attributed to its ability to use each individual base estimator’s strength while minimising its limitaitons [48]. As a result, it has been used to predict incidence of diabetes, heart disease, liver disease, and skin cancer [48]. Bagging was the most accurate model in only 26.8% of studies even though it was used in 41 of 45 studies [48].

#### Estimations of Cause Specific Maternal Mortality and Risk of Mortality

Many of the papers that use decision-tree based methods to predict maternal health outcomes focus on predicting patient risk and cause-specific maternal mortality. As a result, much of the machine learning research in this domain uses classification models, which can categorise a woman’s mortality risk. While I use a regression model in my thesis, I include examples of classification models to provide context for how machine learning is being used to reduce maternal mortality outcomes.

Akazawa et al. (2021) studied machine learning models could accurately classify whether a woman would experience postpartum haemorrhage, a leading cause of maternal mortality, to inform the woman’s treatment plan [41]. They compared the performance of a logistic regression model, decision tree model, random forest model, boosted tree model and a deep learning model for this task when trained on 11 clinical variables [41]. While they found the boosted tree model had the highest accuracy, they tested and trained the model on data from the same institution, potentially leading to overestimation of the model’s generalisability [41]. Similarly, researchers have also used machine learning techniques to predict a woman’s risk of pre-eclampsia, another leading risk of maternal mortality, to improve identification and treatment of high-risk pregnancies [42]. The study trained Elastic Net and gradient boosting models on a range of medical and socio-demographic covariates, with similar performance between the two models [42]. However, the study was limited by a large amount of missing data, which it chose to impute, and the fact that it was trained and tested on data from only one hospital [42]. Given this was a referral hospital with a consequently higher proportion of high-risk patients, their model may not extrapolate to the wider population [42]. As a final example, Sylvain et al. (2025) study predicted adverse pregnancy outcomes in Rwanda using logistic regression, decision trees, random forests, gradient boosting models, support vector machines, and neural networks [46]. The random forest and gradient boosting models had the highest accuracy (90.6% and 88.49%, respectively) [46]. The study also determined the most predictive variables, highlighting the importance of gestational age [46]. However, the study predicted a pregnancy having an adverse outcome as a binary variable, where this variable included a wide variety of both maternal and neonatal health outcomes [46]. This could result in the model truing to predict too many diverse outcomes, reducing its nuance and utility to inform treatment plans [46]. Another limitation of this study was that the model was solely trained on data from district hospitals, which are estimated to only be responsible for 35% of births in Rwanda, with samples further excluded if they have too much missing data [46]. This may limit the study’s generalisability, especially in more rural regions and in hospitals with lower resources and thus more missing data [46].

Khadidos et al. (2024) used a stacking-ensemble model to classify maternal health risk in Bangladesh, where they trained gradient tree booster, random forest, decision tree, and k-Nearest Neighbours models as the base estimator, with each base estimator trialled as the meta-learner [47]. Using the gradient boosting tree as the meta-learner had the highest precision (0.86), with all decision-tree based stacking ensemble outperforming sole use of bagging or boosting [47].

Machine learning classifiers are also used to predict qualities of health systems. For instance, Taye et al. (2025) used a random forest classifier to predict whether a birth was attended by a skilled birth attendant using a mixture of socio-economic and health system quality variables, as well as to indicate the most predictive variables [44]. While the model had 92% accuracy, it was trained and tested on survey data, which is known to be of lower quality, potentially limiting the model’s true accuracy [44]. This data was imputed, which may introduce bias into the model’s estimates [44]. Another example is Fredriksson et al.’s (2022) paper that compared the performances of a random forest model and artificial neural network to more classical statistical models to classify a woman as likely to deliver her baby in a health facility [45]. The random forest model had the highest classification accuracy (74%), with the paper also reporting the most predictive variables [45]. The lower accuracy may be related to the authors’ use of imputation [45].

#### Estimation of Maternal Mortality

There is a severe paucity of studies that use algorithmic modelling techniques to estimate MMR and the number of maternal deaths. The only relevant study found in my literature review was published in 2025 and only estimated maternal mortality rates for Bangladesh [38]. The authors compared the performance of a Bidirectional Recurrent Neural Network (BRNN) and an Elastic Neural Network (ENET) to predict the MMR ratio, with the associated root mean square errors being 3.30 and 3.44 per 100,000 live births, respectively [38]. This study was severely limited by its dataset size, as it reported having only 21 observations of MMR, which is generally insufficient to train a robust model [38]. No critique can be made on its feature variables, as these did not appear to be reported. Despite the limitations of this study, it serves as proof of concept for use of algorithmic modelling techniques to estimate global maternal mortality. Additionally, the lack of robust studies doing this type of analysis highlights a gap in the literature that my thesis aims to fill.

### Conclusion

In conclusion, the UN MMEIG and Global Burden of Disease models are the primary modelling techniques for estimating global maternal mortality ratios. The Global Maternal Health Microsimulation model is a newer method for MMR prediction. These three methods are limited by their low-quality, sparse input data. As a result, estimates for countries with less data may generated using parameters generated by uninformative priors. The parameter values and overall MMR estimates may also be forced towards average regional estimates by the Bayesian hierarchical model, ignoring potential regional heterogeneity. Given that countries with less data tend to have higher MMR, this can reduce the estimates’ ability to inform national policy to successfully reduce MMR. The models are also limited in their consideration of only a small subset of socio-economic and health-related variables that impact maternal mortality. Using decision-tree based models would circumvent the need to model the data generating process, eliminating assumptions about data distribution and the need for priors. Additionally, these models can handle a wide range of feature variables, allowing them to make more comprehensive estimates, with evaluation of feature importance covering a wider range of domains. After a literature search, I only found one algorithmic modelling approach to estimating MMR, with the study performed within a single country. Other applications of machine learning to the reduction of maternal mortality were concentrated in the classification of a woman’s overall and cause-specific maternal mortality risk level. Thus, there is a gap in the literature about how a decision-tree based machine learning method can be used to estimate maternal mortality ratios at a global level.

As a final note, building a model with an entirely new methodology will produce another set of MMR estimates. My model’s estimates can be compared to the literature and potentially contribute to resolving some of the lack of consensus around current MMR estimates (see Figure X) and promote scientific discussion.

1. Citation: Trends in maternal mortality estimates 2000 to 2023: estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division. Geneva: World Health Organization; 2025. Licence: CC BY-NC-SA 3.0 IGO
2. The Annals of Applied Statistics 2017, Vol. 11, No. 3, 1245–1274 DOI: 10.1214/16-AOAS1014 © Institute of Mathematical Statistics, 2017
   1. 2017 un mmeig estimates
3. <https://www.tandfonline.com/doi/full/10.1080/2330443X.2023.2286313#abstract>
   1. Bmis model information
4. <https://www.thelancet.com/cms/10.1016/S0140-6736(24)00367-2/attachment/b7b1d025-a598-408a-b62e-025ebc1f08d1/mmc1.pdf>
   1. Supplement to: GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024; published online April 3. https://doi.org/10.1016/ S0140-6736(24)00367-2.
5. <https://www.nature.com/articles/s41591-022-01990-1>
   1. Gbd at 30 years
6. <https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)00367-2/fulltext>
   1. Gbd 2021
7. <https://pophealthmetrics.biomedcentral.com/counter/pdf/10.1186/1478-7954-10-1.pdf>
   1. Foreman KJ, Lozano R, Lopez AD, Murray CJL. **Modeling causes of death: an integrated approach using CODEm**. Population Health Metrics. 2012; 10:1.
8. <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01501-1>
   1. Garbage codes in the gbd studies
9. <https://www.nature.com/articles/s41591-023-02310-x>
   1. Ward, Z.J., Atun, R., King, G. *et al.* Simulation-based estimates and projections of global, regional and country-level maternal mortality by cause, 1990–2050. *Nat Med* **29**, 1253–1261 (2023). <https://doi.org/10.1038/s41591-023-02310-x>
10. <https://www.sciencedirect.com/science/article/pii/S2589537024002323>
11. <https://www.sciencedirect.com/science/article/pii/S2666827025000714>
12. <https://academic.oup.com/bmb/article/121/1/121/2926155>
13. <https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(24)00560-6/fulltext>
    1. Lancet global determinants of MMR
14. <https://www.mdpi.com/1999-5903/15/8/255>
    1. Arima
15. <https://link.springer.com/article/10.3758/s13428-023-02204-3#Sec36>
    1. Bayesian hierarchical model tutorial
16. <https://pubmed.ncbi.nlm.nih.gov/17274999/>
    1. Maternal mortality, stillbirths, and measures of obstetric care
17. Onambele, L.; Guillen-Aguinaga, S.; Guillen-Aguinaga, L.; Ortega-Leon, W.; Montejo, R.; Alas-Brun, R.; Aguinaga-Ontoso, E.; Aguinaga-Ontoso, I.; Guillen-Grima, F. Trends, Projections, and Regional Disparities of Maternal Mortality in Africa (1990–2030): An ARIMA Forecasting Approach. Epidemiologia 2023, 4, 322–351. https://doi.org/ 10.3390/epidemiologia4030032
18. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3886639/>
    1. 2014 un mmeig model
19. Zaba B, Calvert C, Marston M, Isingo R, Nakiyingi-Miiro J, Lutalo T et al. Effect of HIV infection on pregnancyrelated mortality in sub-Saharan Africa: secondary analyses of pooled community-based data from the network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA). Lancet.
    1. 2013;381(9879):1763-71 (<https://doi.org/10.1016/s0140-6736(13)60803-x>).
20. <https://link.springer.com/article/10.1007/s00103-018-2793-0>
    1. Germany gbd
21. <https://link.springer.com/article/10.1186/s13690-020-00458-3?fromPaywallRec=true>
    1. Who pressure
22. <https://www.tandfonline.com/doi/epdf/10.1016/S0968-8080%2811%2937550-7?needAccess=true>
    1. Ihme vs un
23. <https://static-content.springer.com/esm/art%3A10.1038%2Fs41591-023-02310-x/MediaObjects/41591_2023_2310_MOESM1_ESM.pdf>
    1. Gmath methods
24. <https://www.jstor.org/stable/pdf/2998681.pdf>
    1. Jstor microsimulation paper
25. <https://researchsystem.canberra.edu.au/ws/portalfiles/portal/8849112/rm2013000054_journal_article.pdf>
    1. UC microsimulation
26. <https://journals-sagepub-com.virtual.anu.edu.au/doi/epdf/10.1177/0894439310370085>
    1. More microsimulation
27. <https://onlinelibrary.wiley.com/doi/10.5694/mja2.52452>
    1. Non-communicable disease and pregnancy
    2. Ramson et al., 2024
28. Akselrod S, Banerjee A, Collins TE, Acharya S, Artykova N, Askew I, Berdzuli N, Diorditsa S, Eggers R, Farrington J, Jakab Z, Ferreira-Borges C, Mikkelsen B, Azzopardi-Muscat N, Olsavszky V, Park K, Sobel H, Tran H, Vujnovic M, Weber M, Were W, Yaqub N, Berlina D, Dunlop CL and Allen LN (2023) Integrating maternal, newborn, child health and non-communicable disease care in the sustainable development goal era. Front. Public Health 11:1183712. doi: 10.3389/fpubh.2023.1183712
29. [**https://pubmed.ncbi.nlm.nih.gov/27642018/**](https://pubmed.ncbi.nlm.nih.gov/27642018/)
30. [**https://pmc.ncbi.nlm.nih.gov/articles/PMC11117177/pdf/jogh-14-04128.pdf**](https://pmc.ncbi.nlm.nih.gov/articles/PMC11117177/pdf/jogh-14-04128.pdf)
31. [**https://www.sciencedirect.com/science/article/pii/S2589537024002323**](https://www.sciencedirect.com/science/article/pii/S2589537024002323)
32. [**https://projecteuclid.org/journals/statistical-science/volume-16/issue-3/Statistical-Modeling--The-Two-Cultures-with-comments-and-a/10.1214/ss/1009213726.full**](https://projecteuclid.org/journals/statistical-science/volume-16/issue-3/Statistical-Modeling--The-Two-Cultures-with-comments-and-a/10.1214/ss/1009213726.full)
    1. **Two cultures: statistics vs ML**
33. <https://www.thelancet.com/action/showPdf?pii=S2214-109X%2823%2900468-0>
    1. Global determinants of maternal health -> lancet
34. [**https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-025-23528-z**](https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-025-23528-z)
    1. Maternal hypertensive disorder projections
35. <https://www.mdpi.com/2227-7390/10/8/1283>
    1. Multicollinearity
36. <https://projecteuclid.org/journals/annals-of-statistics/volume-43/issue-4/Consistency-of-random-forests/10.1214/15-AOS1321.full>
    1. RF and high-dimensionality
37. <https://onlinelibrary.wiley.com/doi/epdf/10.1111/insr.12016>
    1. Decision tree info
38. Molla MM, Hossain MS, Ali MA, Islam MR, Sultana MP, Roy DC (2025) Exploring the achievements and forecasting of SDG 3 using machine learning algorithms: Bangladesh perspective. PLoS ONE 20(2): e0314466. <https://doi.org/10.1371/journal.pone.0314466>
39. Digital Object Identifier 10.1109/ACCESS.2024.3399827
    1. Box models
40. <https://eurjmedres.biomedcentral.com/articles/10.1186/s40001-025-02680-7>
    1. Brief ml pub health overview
41. <https://www.nature.com/articles/s41598-021-02198-y>
    1. Postpartum haemorrhage prediction
42. <https://www.sciencedirect.com/science/article/pii/S2589933320300306?via%3Dihub>
    1. Pre-eclampsia prediction
43. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9379432/pdf/PAMJ-42-89.pdf>
    1. Medical supplies demand
44. <https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-025-22007-9#Sec2>
45. <https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2022.855236/full>
46. <https://link.springer.com/article/10.1186/s12911-025-02921-z?fromPaywallRec=false>
47. <https://www.nature.com/articles/s41598-024-71934-x#Abs1>
48. <https://pmc.ncbi.nlm.nih.gov/articles/PMC10298658/>