# Statistical Demography Meets Ministry of Health: The Case of the Family Planning Estimation Tool

Leontine Alkema[1], Herbert Susmann[2], Evan Ray[1],
Kristin Bietsch[3], Priya Emmart[3], Rebecca Rosenberg[3], John Stover[3], Emily Sonneveldt[3],
Shauna Mooney[4], Niamh Cahill[4],
and co-authors[*]

*[1] University of Massachusetts Amherst, USA*
*[2] Department of Population Health, Grossman School of Medicine, NYU, USA*
*[3] Track20, Avenir Health, USA*
*[4] Maynooth University, Ireland*

Extended abstract for African Population Conference 2024
Not for circulation or citation
February 12, 2024

## Abstract
The Family Planning Estimation Tool (FPET) is used in low- and middle-income countries to produce estimates and short-term forecast of family planning indicators, such as modern contraceptive use and unmet need for contraceptives. Estimates are obtained via a Bayesian statistical model that is fitted to country-specific data from surveys and service statistics data. In this paper, we summarize the main features of the statistical model used in the Family Estimation Tool, explain how the tool is used in countries, and summarize recent updates related to subnational estimation, the use of service statistics data, and dealing with data outliers. We use our experience with FPET to discuss lessons learned and open challenges related to the broader field of statistical modeling for monitoring for demographic and global health indicators, to help further optimize the relevance of statistical modeling in practice.

[*] *The final paper will include additional co-authors involved in the Track20 project. Contact: lalkema@umass.edu.*

## Table of Contents

## Introduction

The elevation of family planning (FP) on the global stage in 2012 with the launch of FP2020 provided an opportunity to create new approaches to the annual monitoring of FP programs. The need for annual estimates to track progress of the initiative led to the development of standard indicators, approaches, and methodologies. For countries, having annual estimates of FP indicators was a priority to better gage progress towards their own strategies and objectives. After the FP2020 initiative concluded, the FP2030 initiative was started with a continued focus on measurement and country-specific commitment making.

Track20, implemented by Avenir Health and funded by the Gates Foundation, is a global family planning project aimed at improving global and country level use of data. The project's aims include the development and introduction of innovative, user friendly, methodologies, tools, and approaches that build capacity and broaden who can effectively engage with family planning data. Track20 has provided the annual estimates used for FP2020 related measurements and it continues to assess progress for FP2030 and the Ouagadougou Partnership. This is done through a bottom-up reporting process led by government monitoring and evaluation (M&E) officers, effectively linking country and global FP data.

The Family Planning Estimation Tool (FPET) is used by Track20 in low- and middle-income countries to produce estimates and short-term forecast of family planning indicators, such as modern contraceptive use and unmet need for contraceptives. Estimates are obtained via a Bayesian statistical model that is fitted to country-specific data from surveys and service statistics data. The model is a country-specific implementation of a global model for FP estimation, referred to as the global Family Planning Estimation Model (Alkema *et al.*, 2013; Cahill *et al.*, 2018; Kantorova *et al.*, 2020). The country-specific implementation was produced to introduce and facilitate in-country usage by the Track20 project.

In this paper, we summarize the main features of the statistical model used in the Family Estimation Tool, explain how the tool is used in countries, and how the model has evolved over the last decade based on usage in the Track20 project and user inputs. We use our experience with FPET to discuss lessons learned and open challenges related to the broader field of statistical modeling for monitoring for demographic and global health indicators.

## What FPET does

FPET produces estimates of FP indicators using available data for a population of interest, see Figure 1. A model-based approach to producing estimates of levels and trends is needed because data are not necessarily available for all years of interest, e.g., for years in the past, the period after the most recent data point until the current year, and future years. Moreover, data are subject to measurement error, which may be substantial, hence limiting the usefulness of considering just a recent survey or trend estimates based on survey data alone.

In FPET, the FP indicators considered are contraceptive use and unmet need for contraceptives. Methods are categorized into modern versus traditional methods. Modern methods of contraception include female and male sterilization, oral hormonal pills, the intra-uterine, device, male and female condoms, injectables, the implant (including Norplant; Wyeth-Ayerst, Collegeville, PA, USA), vaginal barrier methods, standard days method, lactational amenorrhea method, and emergency contraception. Traditional methods of contraception include abstinence, the withdrawal method, the rhythm method, douching, and folk methods. Unmet need for family planning is defined as the percentage of women who want to stop or delay childbearing but who are not currently using any method of contraception to prevent pregnancy. Also included are women who are currently pregnant or postpartum amenorrheic whose pregnancies were mistimed or unwanted.

FP indicators are estimated for women aged 15 to 49 years old, considering women's marital status. More precisely, estimates are produced separately for women who are married or in a union (referred to as married women in this paper), versus women who are not married. Sexual activity is considered in the calculation of the indicators, as summarized in Kantorova *et al*. (2020).

Data on FP are available from household surveys. Contraceptive prevalence data obtained from surveys refers to the percentage of women who report themselves or their partners as currently using at least one contraceptive method of any type (modern or traditional).

Data on FP can also be obtained from routine data collection. Service statistics data refer to data obtained from routine data collection. These systems may record various FP related outcomes, including family planning commodities distributed to clients or facilities, visits to FP facilities/providers, or family planning users. Track20 has developed a tool that converts service statistics data into a single metric representing volume of services, Estimated Modern Use (EMU), see Magnani *et al.* (2018). This "SS to EMU" tool converts different types of service statistics data into EMUs.

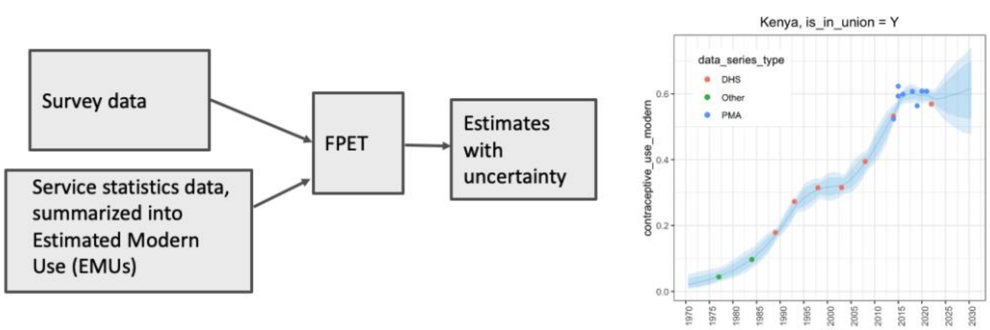



**Figure 1: The Family Planning Estimation Tool (FPET) produces estimates of FP indicators, based on survey data and observations on estimated modern use (EMUs), obtained from service statistics.** The graph shows survey data on modern contraceptive use for married women of reproductive age in Kenya, together with an FPET fit (point estimates are given by the solid blue line, shaded areas represent 80% and 95% credible intervals).

## Overview of Track20's approach

The Track20 project works to build the capacity of countries to generate and use data to inform their programming, improve impact, and accelerate progress toward their FP goals. One of the pillars of the Track20 approach is the cultivation of a network of M&E Officers dedicated to increasing the quality and use of FP data. These M&E Officers are working within or in partnership with Ministries of Health in 34 countries. This approach positions governments to be the driving force of FP data instead of relying on implementing partners to produce needed analyses and findings. Track20 provides ongoing, on demand, technical support to these countries. This includes an annual training that brings all M&E Officers together to learn additional analytical skills, engage in south-to-south learning, and to identify opportunities for new analyses and tools based on current government challenges. This direct exchange with multiple countries enables Track20 to develop tools that directly speak to current government priorities. It also provides an opportunity for any new tools to be vetted and tried across many countries, providing valuable feedback that improves Track20's work.

Track20 M&E officers use FPET to assess progress towards country strategies and commitments. The officers are trained in FPET and use it to produce FP estimates for their country. These FPET estimates are reviewed, vetted, and approved through annual consensus meetings. These meetings are convened by governments and attended by partners and donors, providing a time to reflect on progress and identify any barriers or opportunities moving forward.

## FPET model specification

### Evolution of FPET

The estimation of FP indicators in FPET has evolved based on user inputs and further model developments. The first version was based off a global modeling approach to estimate and project FP indicators for married women aged 15-49 in all countries in the world, using survey data (Alkema *et al.*, 2013). For in-country usage in the Track20 project, we produced FPET as local version of the global model, to fit to data from one country only (New *et al.* 2017). This version expanded upon the global model by allowing for using service statistics data.

Subsequently, FPET evolved to better capture local contexts. Model updates were introduced to improve predictive performance and to better account for survey data quality issues (Cahill *et al.*, 2018), and to improve the use of service statistics data (Cahill *et al.*, 2020). The population considered was extended from married women to all women, using a local implementation of the global model for estimation among unmarried women (Kantorova *et al.,* 2020; Guranich *et al.*, 2020).

In 2024, several updates were introduced to further improve FPET's utility. We have relaxed the parametric assumption made regarding contraceptive use transitions (Susmann and Alkema, 2023a). We have updated the way in which survey data are used to inform estimates, to improve estimation in the presence of data outliers. We have also updated the approach to include service statistics to better account for uncertainty associated with EMUs and to capture country-specific differences in data quality (Mooney et al, 2024+). Finally, we extended FPET to allow for the production of subnational estimates.

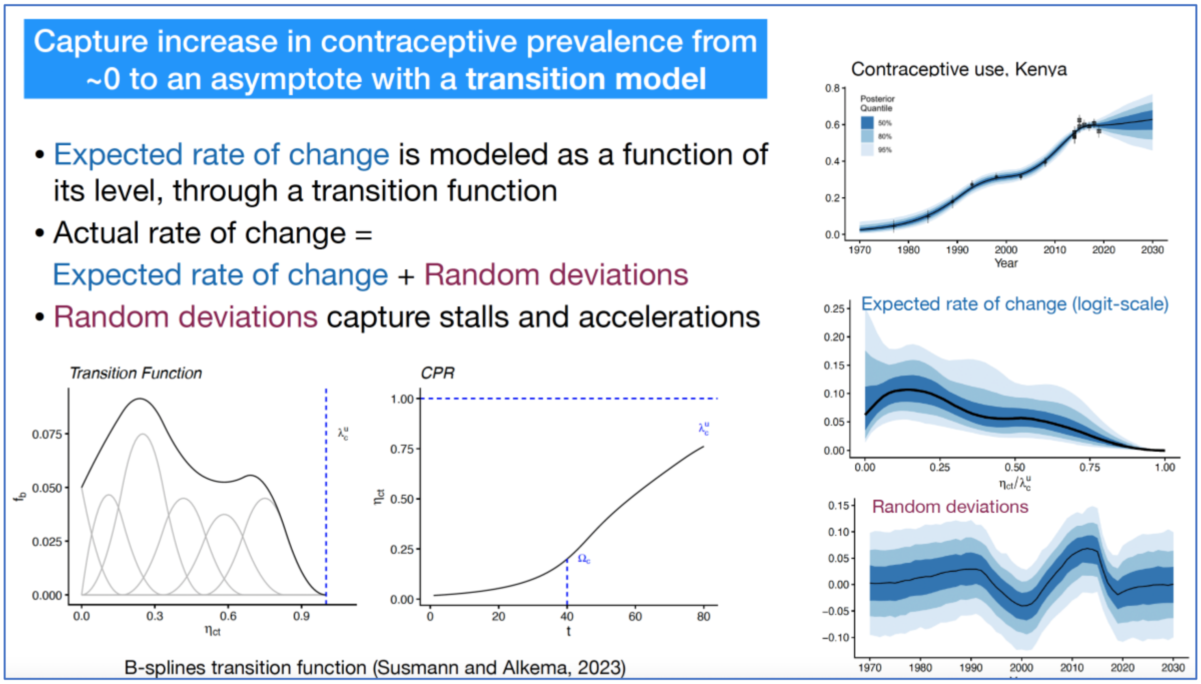
In the remainder of this section, we present the model specification for the most recent version of FPET, used for 2024 reporting activities. To produce estimates and forecasts for FP indicators, we made assumptions about (1) how FP indicators may change over time and vary across different populations, and (2) how available data relate to the indicators. In model descriptions, we distinguish between these two sets of assumptions and refer to the first set as the process model and the second set as the data model assumptions.

## Process model: How do FP indicators vary over time and differ across populations?

### Capturing contraceptive transitions

In the process model, we consider how FP indicators change over time. The main assumptions underlying FPET are based on capturing a contraceptive transition (Alkema et al, 2013). Specifically, over longer time periods, we assume that demand for contraceptives follows an S-curve (see Figure 2 for illustrations in selected countries), where increases start slowly at low demand, accelerate once growth has been initiated, and then rates of change slow down as demand reaches high levels, which we refer to as an asymptote. Similarly, we assume that demand satisfied with modern methods also follows an S-curve. Based on combing the two transitions, we obtain modern contraceptive use (mCPR) and unmet need for modern methods. Resulting estimates show a typical pattern in unmet need whereby unmet need increases at the start of a transition, when supply is lagging demand, followed by a decrease once supply catches up with demand. We obtain estimates on traditional use from an additional model placed on the ratio of traditional use to unmet need for modern methods.

The increase in demand and demand satisfied with modern methods from low levels to an asymptote are modeled with transition models. More details on transition models are provided in Box 1. A main feature of a transition model is that the rate of change can be modeled as a function of its level through a transition function. Population-year deviations away from smooth transition functions are incorporated to capture short-term accelerations or stalls, as suggested by the data. In initial work, such deviation terms were added to the transition function directly, so to prevalence levels (on logit-transformed scales), as described in Alkema *et al*. (2013). To improve predictive performance, we updated this assumption to have the deviation terms be added to the rate of change instead (Cahill *et al*., 2018). The parametrization of the transition function, how the expected rate of change varies as a function of prevalence levels, has evolved as well. Initially, this relation was based on assuming a logistic function (Alkema *et al.*, 2013; Cahill *et al., 2018*). In recent work, we have relaxed the parametric assumption made regarding the rates of change as a function of the level, by using a splines model to represent this relationship (Susmann and Alkema, 2023a).




**Box 1: Illustration of the transition model specification and fit for Kenya.** Transition models are used to capture increases in an outcome of interest from close to zero to an asymptote. In a transition model, the expected rate of change is modeled as a function of its level. The illustration shows a transition model applied to contraceptive use in Kenya (figures taken from Susmann and Alkema, 2023a). The two figures on the left illustrate an example transition function (modeled with B-splines) and the associated CPR over time. The figures in the right column illustrate a transition model fit for Kenya. Data and mCPR estimates over time are shown in the top plot. The rates of change that are estimated are composed of an expected rate of change (middle plot) and random deviation terms (lower plot). The expected rate of change decreases with the mCPR level while the random deviation terms capture the slow-down around 2000.

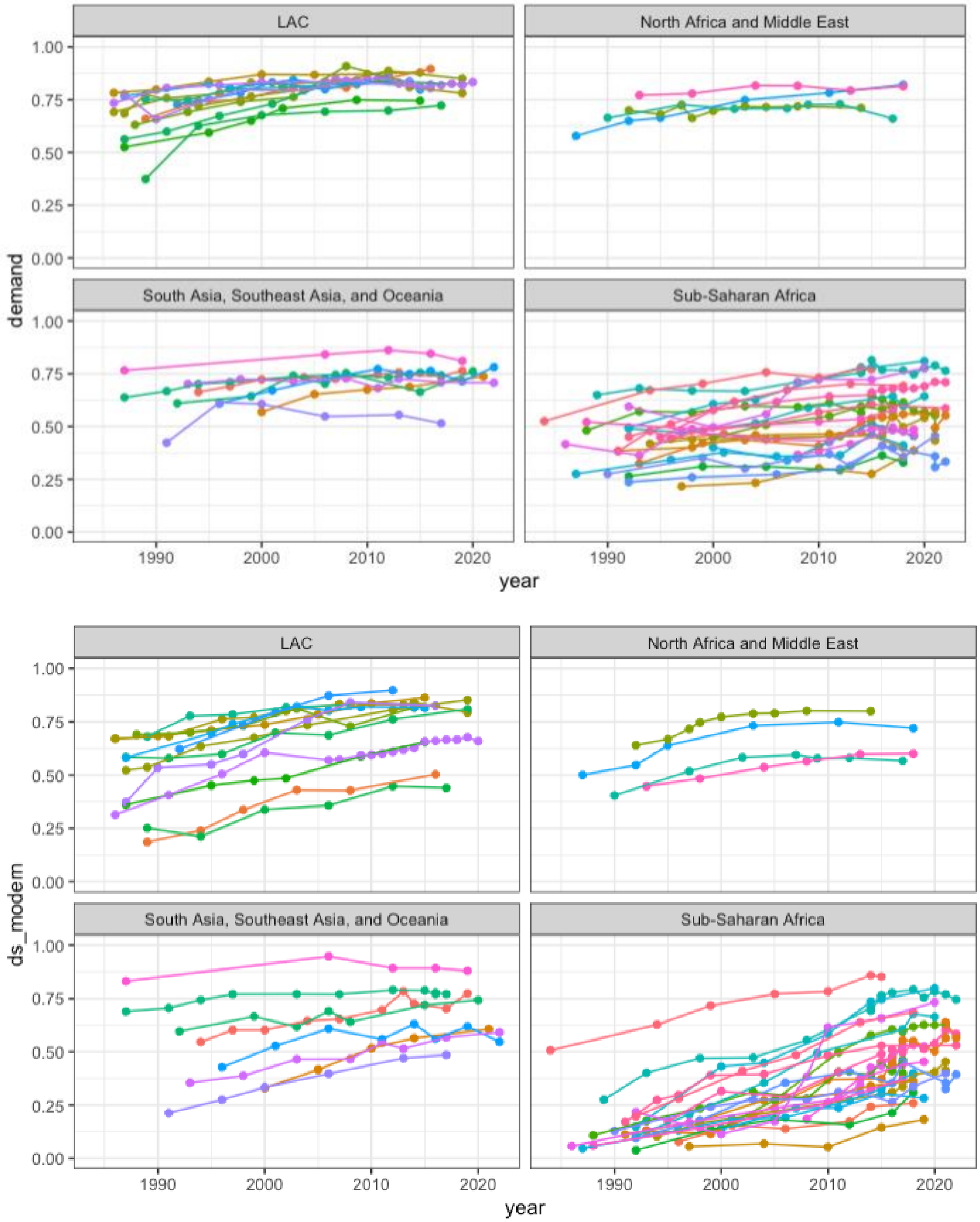



**Figure 2: Observed data on demand and demand satisfied for modern methods for selected countries over time.** Observations refer to prevalence among married women of reproductive age. Countries are included in the selected main regions if they have more than 4 observations.

## Sharing information across populations

Our goal is to estimate and forecast FP indicators from 1970 until 2030 for all countries globally, for women of reproductive age, separately by marital/in-union status. However, a country's data may not be sufficient to estimate FP for the time period of interest. For example, for countries that are at the start of their transition, country data alone do not allow for estimating its asymptote. Some countries may have very limited data on FP among unmarried women.

To overcome data limitations, we fit a global model in which information is shared across populations using hierarchical (or multilevel) models. We fit the model separately for married versus unmarried women, given that transitions differ between these groups. For each model, we group countries based on geo-political characteristics into clusters and subclusters of countries, such that information is shared among countries within subclusters. In addition, for the modeling of FP among unmarried women, we further extend the groupings to account for differences in cultural norms related to FP and sexual activity among unmarried women, following the approach by Kantorova *et al.* (2020).

Illustrative results are shown in Figure for Togo. We see that the model can produce estimates and forecasts for all years. uncertainty associated with FP indicators increases with the projection horizon.

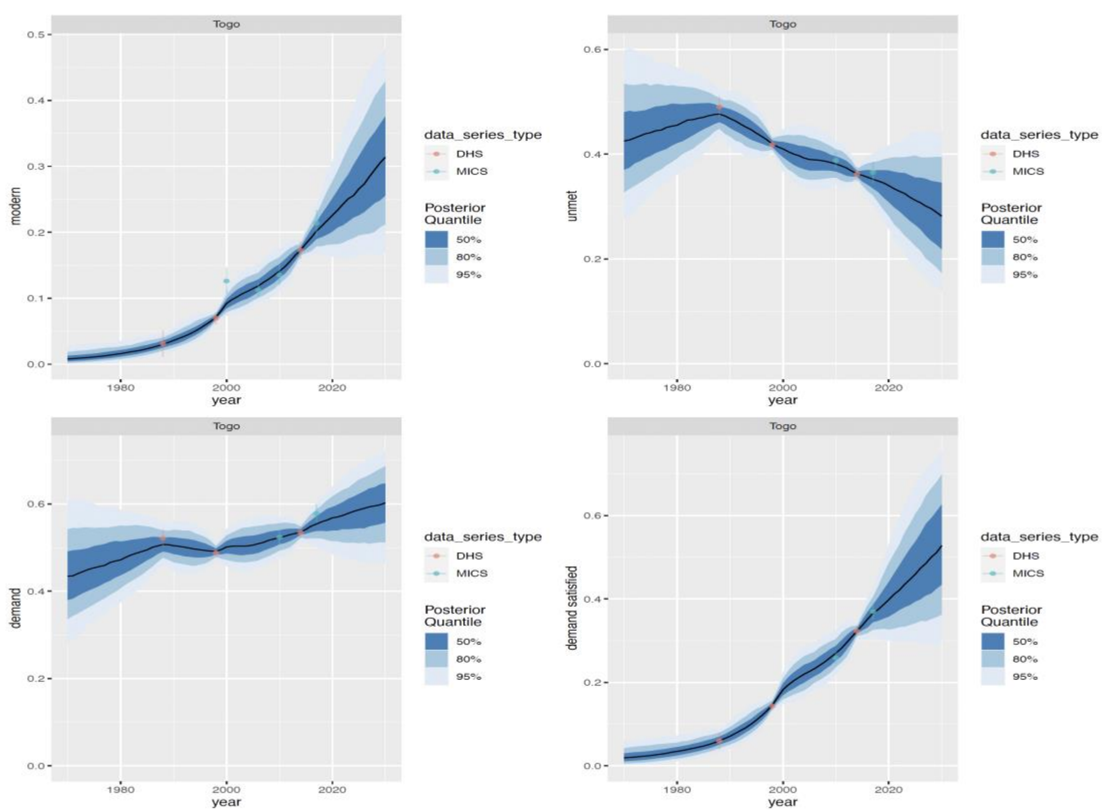



**Figure 3: Estimates and forecasts from 1970 until 2030 for Togo.**

## Data model

### Usage of survey data

FPET is fitted to survey data on mCPR, traditional use, and unmet need. Survey data are subject to various sources of error: sampling error due to survey design as well as additional errors due to data collection approach, reporting errors, characteristics of the surveyed population, etc.

When fitting FPET to survey data, we account for uncertainty associated with data errors, to produce estimates that are aligned with observations that are very certain, while smoothing over outlying data points that are very uncertain. Specifically, we account for (i) sampling error variance, (ii) source-type specific non-sampling error, and (iii) observation-specific outlier errors. Sampling error variance is calculated from survey micro data. Source-type specific non-sampling error is assumed to be zero for observations from DHS and estimated for each of the other main source types (MICS, PMA, national surveys, and other surveys). Observation-specific outlier errors are added to deal with outlying erroneous data points. These error terms are estimated for each observation. They are assigned shrinkage priors such that most errors are shrunk to zero, except for observations that are outlying. This set up results in smoothing over outlying observations.

An example fit is shown for Burundi in Figure 4. We see that with the data outlier error set up, the estimates smooth over the second-last data point.

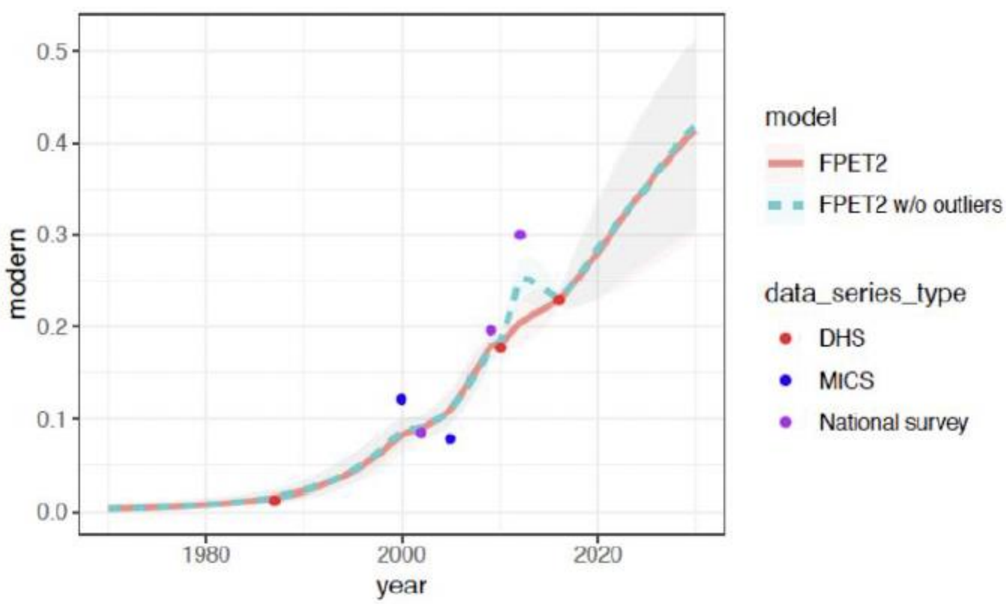



**Figure 4: FPET fits in Burundi.** The graph illustrates fits with and without the data outlier set up.

## Usage of service statistics data

Service statistics data (e.g., commodities supplied to clients) are translated in Estimated Modern Use (EMUs), as explained in an earlier section. We use EMUs in FPET to inform projections past the most recent survey.

Given that EMUs may be biased in level, we account for such biases when including EMUs in FPET. In initial work, level biases were estimated in a prior run and then used in a subsequent run (New *et al*., 2017). In a subsequent implementation, we used the observed EMU-based rate of change (Cahill *et al., 2020*). In this set up, we accounted for varying levels of uncertainty associated with each type of service statistic (e.g., commodities supplied versus client visits) based on a global assessment. Figure 5 illustrates country fits with and without EMUs from this set up.

In recent work, we have extended the approach to using EMUs to account for the uncertainty associated with inputs and country-specific variation in data quality. The final paper will include further details and illustrative findings of this approach.

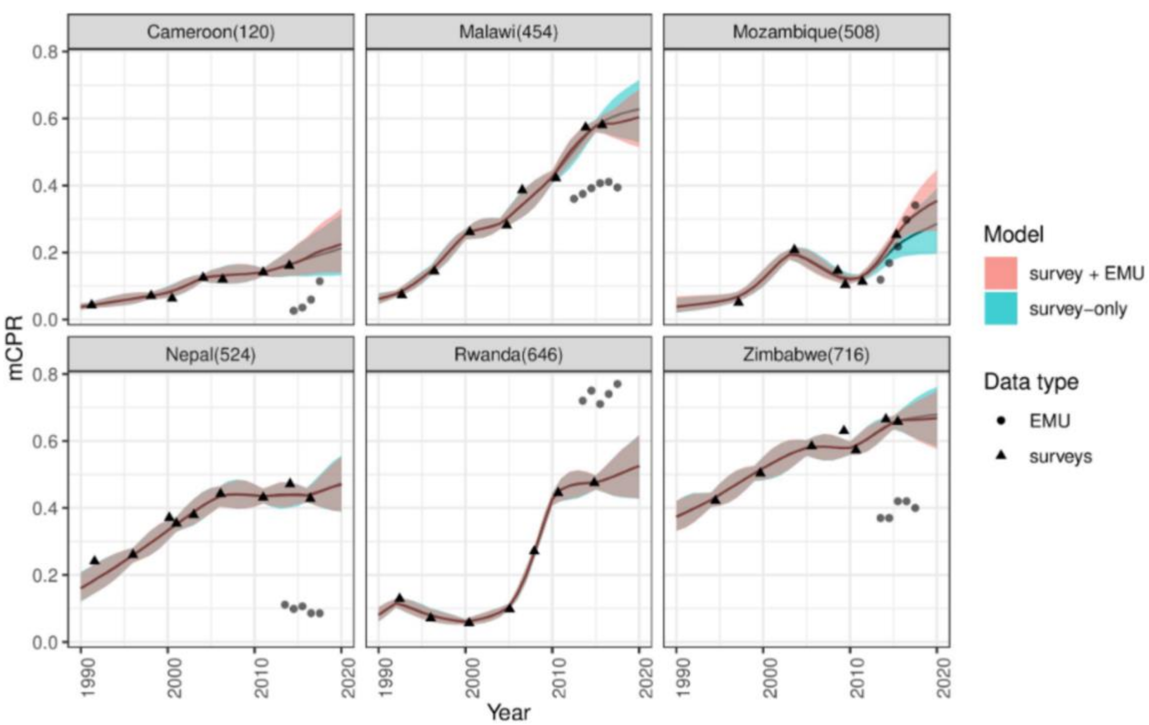



**Figure 5: FPET estimates and projections of modern contraceptive use (mCPR) based on survey data and Estimated Modern Use observations (EMUs).** Solid lines indicate point estimates of mCPR and the shaded regions show the 95% uncertainty intervals. The results in green are from the EMU+survey model and the results in red are from the survey-only model. Shapes of data points distinguish survey observations from EMU. Figure taken from Cahill *et al., 2020.*

## Local fitting

To allow for fitting to data from one population, e.g., a country, we derived a 1-population model set-up from the global model. This 1-population model is referred to as a local model. In summary, point estimates for model parameters that are not country-specific are obtained from the global model fit and used in the local model. Further details are given in Figure 6. The final paper will include further explanation on how hierarchical structures are used in the local model.

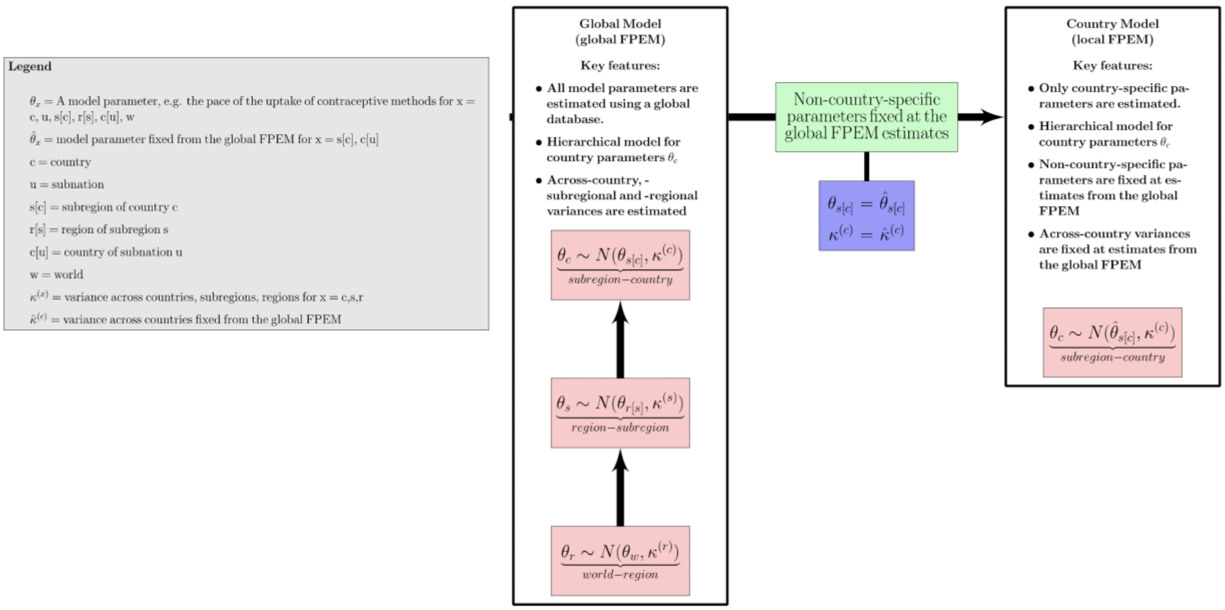



**Figure 6: Flowchart to illustrate the relation between the global implementation of the family planning model (global FPEM) to the local FPET.** Figure taken from New et al. 2017 distributed with a CC BY license.

## Subnational estimation

Earlier versions of FPET have been used for subnational estimation (see, for example, New *et al.*, 2017). In the prior approach, FPET was fit to one region at the time. In the updated version of FPET, we have extended the model to allow for fitting to multiple regions simultaneously and to allow for information in one region to inform other regions. In addition, with the new implementation of FPET, we can use data specific to regions as well as any additional data available at greater levels of aggregation only. This version allows for production of aggregated, bottom-up, estimates with uncertainty assessments. An illustration is given in Figure 7.

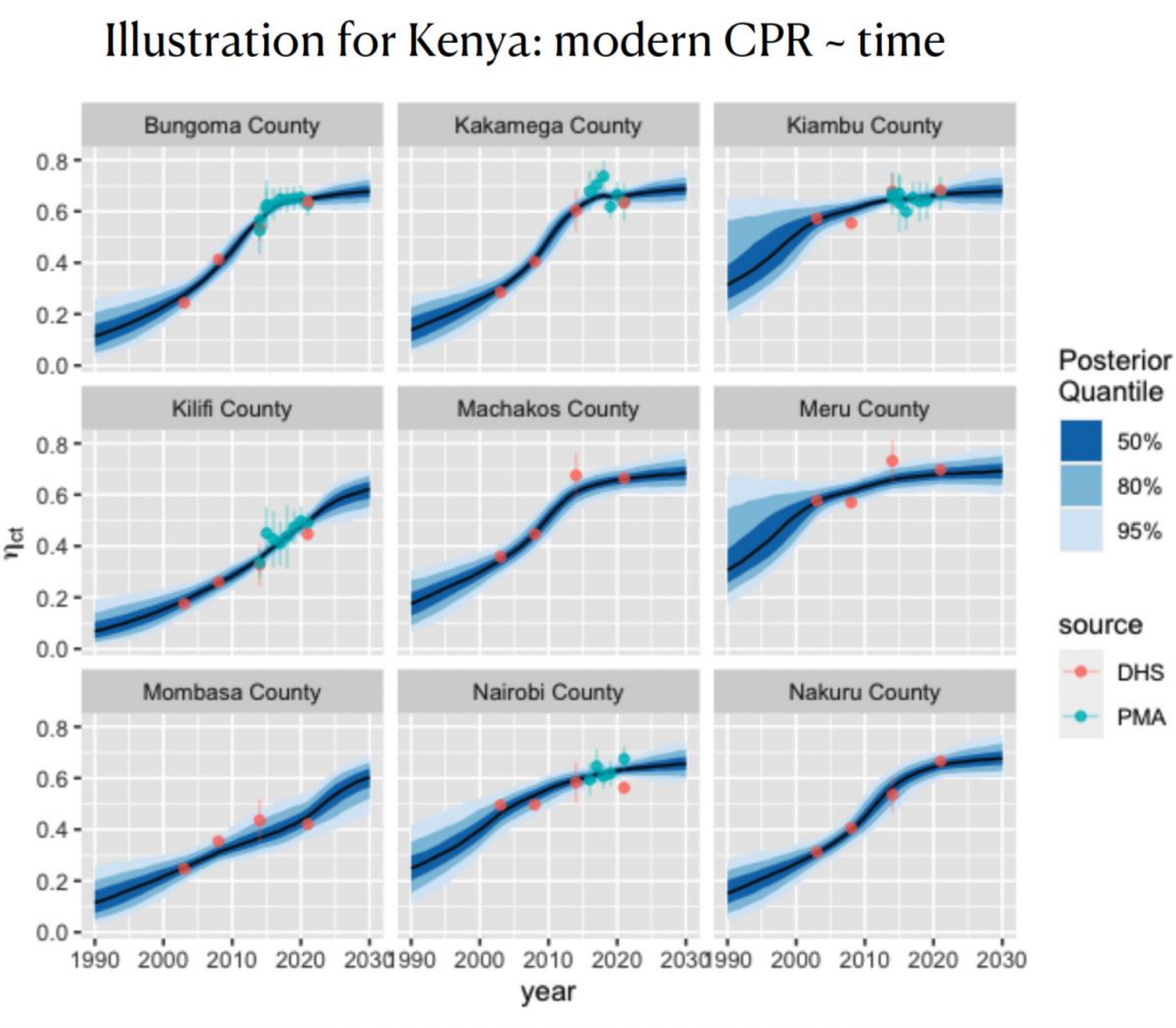



**Figure 7: Illustration of FPET subnational estimates for counties in Kenya**.

### Statistical computing

FPEM and local FPET are Bayesian models. Sampling from the posterior distribution of model parameters is performed using Markov Chain Monte-Carlo (MCMC) algorithms. All models are available in open-source R packages (Wheldon et al., 2019; Guranich et al., 2020; Susmann and Alkema, 2023b). The original version of the global model and FPET are implemented in JAGS. The updated transition model is implemented in Stan.

For the local tool, a web-based application was developed with a user-friendly interface, available at https://fpet.track20.org/fpet/.

## Discussion

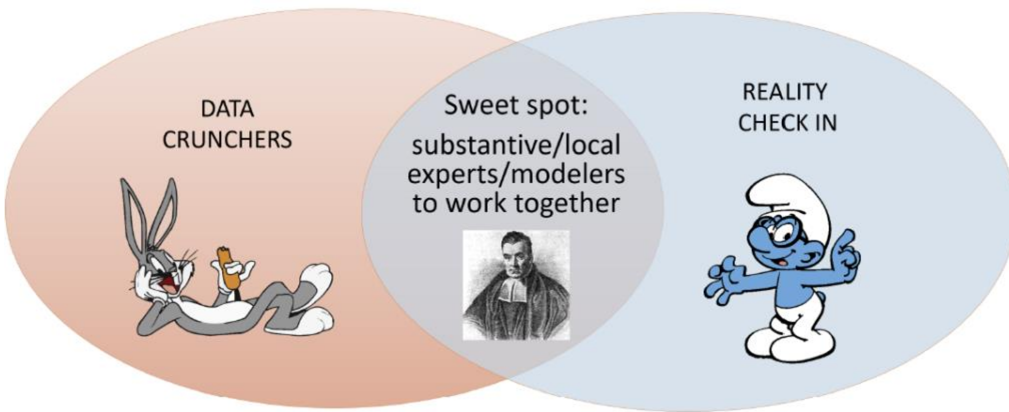
In this paper, we described the Family Estimation Tool and the interplay between statistical modeling and usage in practice. We now turn to discussing take aways and lessons learned from FPET that we believe are relevant to statistical modeling for demographic and global health indicators more generally. We end with some open challenges that also generalize to the broader field.

### Take aways and lessons learned

#### *1 Collaborations between users and modelers are a key to success*

Based on our experience with statistical modeling for family planning, as well as other reproductive, maternal, and child health indicators, we have concluded that the most important (but often overlooked) modeling challenge is to produce results that are reasonable for *all* populations and settings. An 80-20 rule in the context of statistical modeling for global health indicators would be follows: getting reasonable estimates and forecasts for 80% of population-periods is easy, the remaining 20% is hard. While an easy-to-build-low-on-assumptions-type model, for example, commonly used generic spatio-temporal smoothing methods, may predict well on when considering average results across populations and settings, it typically produces unrealistic results for a subset of settings. Moreover, the 20% of population-periods where models may fail to produce appropriate estimates are often those settings where the estimation matters most, as data are typically sparse or uncertain, or the outlyingness of outcomes is something that needs to be reflected in the model-based estimates.

To obtain realistic model-based estimates for all settings of interest, including the 20% of population-periods that are unusual in some sense or where data are limited, models need to be built by a collaboration of statisticians and subject experts. Described in a less formal way (see Box 2): to obtain realistic model-based estimates for all settings of interest, models need to be built by a collaboration between “data crunching” Bugs Bunnies and “this is how the world works” Brainy Smurfs. While the Bugs Bunnies can “easily” introduce powerful statistical tools, the Brainy Smurfs are able to provide reality check-ins regarding appropriateness of assumptions made and resulting findings. Indeed, the evolution of FPET, in terms of its set-up and assumptions made, was driven by a collaboration of data crunchers and substantive experts through the Track20 project. Active usage and feedback on modeled results resulted in the various updates that we described in earlier sections.




**Box 2: To obtain realistic model-based estimates for all settings of interest, models need to be built by a collaboration between "data crunching" Bugs Bunny and "this is how the world works" Brainy Smurf. The Bayesian framework can be considered for producing model-based findings.**

### *2 Local tools are a key to success*

We believe that local tools are incredibly important for successfully producing locally-relevant estimates. At a minimum, these tools allow for production of estimates based on data from a setting alone. It is typically easier to get buy-in for a set of estimates that is produced in-house, based on a fit to data that has been vetted, as opposed to estimates that have been produced in a top-down approach.

But the use and usefulness of local models go well beyond the production of just one set of estimates. Having a model that can be run annually builds a mechanism for data review, of surveys but most importantly service statistics. Track20's experience suggests that countries have improved service statistics over the project because they are reviewing them annually, usually in a consensus meeting with many stakeholders, and it prompts follow-up action. Local tools also enable sensitivity analyses of various forms: models can be fitted to specific data sets or based on changing model assumptions. These kinds of fits can be helpful to the analyst to better understand the data and model. Track20 has successfully used this approach in consensus meetings, as well as in additional technical meetings with various government and non-government stakeholders. For example, in India and Kenya, the tool has been used to analyze data outliers and inform decisions related to data exclusion.

In summary, by enabling in-house production of estimates and additional assessments, local tools empower governments and move away from a commonly used approach of externally produced global health estimates.

### *3 Communication and training needs to be geared at different audiences*
Given the range of audiences involved in modeling exercises, clearly communication needs to be geared at these different audiences and, depending on the audience, vary in terms of technical detail. To make this possible, communication of model assumptions and explanations of "how the model works" is best produced by different actors as well. In our case, the statisticians produced technical writings while Track20 focused on producing materials and supportive documentation that was focused on less statistical audiences. For communication, the local tool can be helpful too, through sensitivity analyses.

### *4 Go Bayesian!*
We have found that the use of Bayesian models enabled the production and evolution of FPET. The Bayesian framework provides flexibility of translating assumptions into probabilistic models. The wide availability of Bayesian modeling software facilitates the easy implementation of such models and active development of computational approaches have resulted in reduced computation times and expanded the types of models that can be fitted. Last but not least, the Bayesian paradigm also allows for uncertainty to be accounted for in, and propagated across, various model components. In our set-up, we derived a local tool from a global model, as a pragmatic, solution-focused approach. The Bayesian framework facilitates this type of set up.

## What's next: open challenges
### *1 How to allow for context-specific adjustments to model-based estimates?*
The use of model-based estimates can be a double-edged sword: On the one hand, model-based estimates can provide insights but on the other hand, model-based estimates may be based on assumptions that do not hold true or estimates may be subject to substantial levels of uncertainty, limiting their usefulness for action. Moreover, as mentioned in relation to the 80-20 rule, the population-periods where models may fail to produce appropriate estimates are often those settings where the estimation matters most, as data are typically sparse or uncertain, or the outlying-ness of outcomes is something that needs to be reflected in the model-based estimates. It is worth noting that the increased focus on capturing within-country differentials compounds these issues: the interest in smaller subgroups brings about greater data needs and – when data availability does not match these needs – exacerbates the importance as well as the potential limitations of models.

When dealing with a setting where model-based estimates conflict with substantive knowledge, i.e. when model-based estimates do not reflect the truth, what can a user or modeler do? Examples from FPET fits are settings where either estimates smooth over a data point (consider it an outlier) while that data point is deemed to capture the situation very well, or vice versa, where data of dubious quality or from populations that may differ from the population of interest have a great influence on the estimates. In earlier work, we came up with some approaches to deal with some of these settings. For example, users can indicate whether data points refer to populations that may not be representative of the population of interest, and the model fit takes that into account (Alkema *et al.*, 2013). In ongoing work with FPET models that include a shock term as well as a more flexible data model, we are exploring options to allow for user input to distinguish between true shocks versus data quality issues. Specifically, we have found that, when allowing flexibility in the truth as well as how data relate to the truth, the model alone may not be able to distinguish between a true shock or a data outlier. In those cases, if substantive information on shocks versus data quality issues is available, then allowing for user inputs is beneficial (or necessary) to help distinguish between data errors versus true outlying levels or trends. These FPET approaches are examples of set-ups where users can provide input to improve upon estimates.

While giving the user options to provide additional context related to data or the outcome of interest is a positive way forward to producing high-quality estimates, the open challenge relates to what information to ask for, and how that information can be used. While some context-specific questions may have reasonably clear answers that can be used in modeling, others may not. For example, when considering data quality and errors, it is likely to be easier to answer the question whether a survey is representative of the entire country or limited to specific regions as opposed to a question related to data quality more general. In addition to difficulties obtaining relevant information, another challenge is related to the need for standardization of estimation procedures and keeping this process as objective as possible. The option of user inputs should not enable manual tweaking of estimates to optimize gains that are not statistical accuracy but instead, could relate to political agendas. Given these challenges, the process of obtaining context-specific information (and documentation) and its usage needs to be carefully considered.

### *2 How to improve the process of model development and communication?*

Model development and communication of strengths and weaknesses of model-based estimates is an area of great importance. We believe that the community of statistical modelers have a key role to play in improving on the status quo. The current literature on statistical modeling of global health indicators lacks standardization of the specification and communication of model assumptions. Because of this gap, we find that models are difficult to compare and modeling assumptions may be difficult to understand. In addition, guidance to systematize model building is very limited. This may result in model choice being based on prior experiences of the research team involved as opposed to suitability considerations.

To advance the development and use of model-based estimates, we see a need to develop material to help standardization of the specification and communication of model assumptions. To work towards that goal, we proposed the use of a new model class, referred to as Temporal Models for Multiple Populations (TMMPs, Susmann *et al.*, 2022). This model class facilitates documentation of model assumptions in a standardized way and facilitates comparison across models. We show how existing models for a variety of indicators can be written as TMMPs and how the TMMP-based description can be used to compare and contrast model assumptions. This approach, or other efforts, can be considered to help communicate model assumptions.

There are opportunities to move toward systematizing model building. Given the typical challenges associated with the 80-20 rule, i.e., getting the 20% right, it does not seem realistic

to aim for a standard recipe for model building that can be applied across a range of indicators and settings and research questions. However, we do see an opportunity to consider the process of model building in a more systematic manner, through ideas that have been recently put forward under the Bayesian workflow umbrella. In our recent work on FPET, we have introduced elements of this type of workflow that are relevant to our modeling goals.
Last but not least, modelers cannot (and should not aim to) pretend that model-based estimates have the final word in settings where additional data are needed to answer open questions. A more targeted approach to data collection, one that is informed by statistical models, could result in great gains in insights.

### *3 How to increase diversity in the modeling community?*

To maximize the value of statistical models for demographic monitoring, action is needed to expand and diversify the community of actors involved, ranging from end users of model outputs to advanced users (applied statisticians and data scientists who understand the detailed workings of models), and modelers who develop and update models.

To date, the great majority of major modeling efforts are based on a set up where academics or international organizations based in the global north develop models and often also produce estimates. The Track20 project with FPET can serve as an example of how to diversify this community, through local capacity building that enabled local usage of tools and local production and ownership of estimates, and through improved interaction between local users and other actors involved. Additional opportunities are to diversify communities of modelers and advanced users so that there are more statistical models being produced in the global south. The expansion of the community of advanced users is especially useful when thinking about rolling out tools that have more advanced usage options.

The opportunities to diversify communities of modelers and advanced users are contingent upon outreach to, and capacity building of under-represented groups, in particularly based on geography. The good news is that with the availability of high-quality open-source software tools such as R and Stan, and the increased availability and quality of online training material in statistics and data science in various forms, access and training are no longer limited to audiences that are geographically close.

With the combination of initiatives like Track20, that diversify the community involved in model-based estimates on specific areas, and the opportunities provided through open-source software and expanded communities of individuals trained in statistics and data science, a more diverse modeling community could become reality in the short-term.

## Conclusions

Using our experience with the Family Planning Estimation tool, we discussed take aways and lessons learned related to the field of statistical modeling for monitoring for demographic and global health indicators. We discussed that (1) Interaction between modelers and users/local

experts is essential to producing high quality model-based estimates; (2) local tools are a key to success; (3) Communication and training needs to be geared at different audiences; and (4) going Bayesian allows for modeling flexibility.
We ended with a more general call to action to the community of statisticians and statistical demographers in the form of steps to consider, to optimize the relevance of statistical modeling in practice. Our call to action focuses on open challenges: (1) How to allow for context-specific model adjustments, for improving estimates in specific settings, (2) How to improve the process of model development and communication, and (3) How to diversify the modeling community. While challenges remain, so do continued advancements in the field of statistics and data science, improvements in access and training opportunities, and the need for high quality measurements. Given all that, we are very optimistic about the future of field of statistical modeling for monitoring for demographic and global health indicators.

## References

L. Alkema, V. Kantorova, C. Menozzi, A. Biddlecom (2013). National, regional and global rates and trends in contraceptive prevalence and unmet need for family planning between 1990 and 2015: a systematic and comprehensive analysis. *The Lancet* 381(9878): 1642–1652.

N. Cahill, E. Sonneveldt, P. Emmart, J. Williamson, R. Mbu, A. Barrière Fodjo Yetgang, I. Dambula, G. Azambuja, A. Mahumane Govo, B. Joshi, S. Felix, C. Makashaka, V. Ndaruhutse, J. Serucaca, B. Madzima, B. Muzavazi, L. Alkema (2021). Using family planning service statistics to inform model-based estimates of modern contraceptive prevalence. *Plos One* 16(10), e0258304.

N. Cahill, E. Sonneveldt, J. Stover, M. Weinberger, J. Williamson, C. Wei, W. Brown, L. Alkema (2018). Modern contraceptive use, unmet need, and demand satisfied among women of reproductive age who are married or in a union in the focus countries of the Family Planning 2020 initiative: a systematic analysis using the Family Planning Estimation Tool. *The Lancet* 391(10123): 870 – 882.

G. Guranich, N. Cahill, L. Alkema (2021). Fpemlocal: Estimating family planning indicators in R for a single population of interest. *Gates Open Research* 5 (24).

G. Guranich, M. Wheldon, N. Cahill, L. Alkema (2020). *fpemlocal*: The local implementation of the Family Planning Estimation Model. Available at https://github.com/AlkemaLab/fpemlocal

V. Kantorová, M.C. Wheldon, P. Ueffing, A.N.Z. Dasgupta (2020). Estimating Progress towards Meeting Women's Contraceptive Needs in 185 Countries: A Bayesian Hierarchical Modelling Study." *PLOS Medicine* 17(2): e1003026. https://doi.org/10.1371/journal.pmed.1003026.

R.J. Magnani, J., Ross, J. Williamson, M. Weinberger (2018). Can Family Planning Service Statistics Be Used to Track Population-Level Outcomes? Global Health: Science and Practice; 6(1):93–102.

J.R. New, N. Cahill, J. Stover, Y.P. Gupta, L. Alkema (2017). Subnational Rates and Trends in Contraceptive Prevalence and Unmet Need for Family Planning from 1990 to 2020: An Analysis for All 29 States in India. *The Lancet Global Health* 5(3): e350-e358.

H. Susmann, M. Alexander, L. Alkema (2022). Temporal models for demographic and global health outcomes in multiple populations: Introducing a new framework to review and standardize documentation of model assumptions and facilitate model comparison. *International Statistical Review* 90(3): 437–67. https://doi.org/10.1111/insr.12491.

H. Susmann and L. Alkema (2023a). Flexible Modeling of Demographic Transition Processes with a Bayesian Hierarchical Penalized B-Splines Model. Available at https://doi.org/10.48550/arXiv.2301.09694.

H. Susmann and L. Alkema (2023b). *BayesTransitionModels*: R package for fitting Bayesian transition models for demographic and health indicators. Available at https://github.com/AlkemaLab/BayesTransitionModels

M. Wheldon, J.R. New, N. Cahill, G. Guranich, C. Wei, A. Tait, K. Bietsch, L. Alkema (2019). *FPEMglobal*: An R Package for producing global estimates and projections of contraceptive use, unmet need for family planning, met need for family planning, and other family planning indicators for all (married/partnered and unmarried) women. Available at github.com/FPcounts/FPEMglobal.