ACCOUNTING FOR THE MISMATCH BETWEEN PREDICTED AND OBSERVED FERTILITY IN SUB-SAHARAN AFRICA

Introduction

Sub-Saharan Africa’s (SSA) fertility rates are expected to drive world population growth throughout the 21st century (United Nations, Department of Economic and Social Affairs, Population Division 2013). Among the biological and behavioral fertility drivers in SSA, contraceptive use is the most commonly recognized fertility-inhibiting factor that also possesses a policy lever through the family planning (FP) program. Demographic and reproductive analysts, as well as policymakers, have come to expect a robust, predictable, inverse relationship between contraceptive use and fertility (Tsui 2001), most commonly measured by the contraceptive prevalence rate (CPR) and the total fertility rate (TFR), respectively. However, questions arise about this relationship when new survey results do not meet expectations, producing a mismatch between expected and observed TFR.

One possible explanation for the mismatch between predicted and observed fertility levels is that the proximate determinants (PD) framework, which is often used to calibrate TFR expectations, may not be adapted to current reproductive behaviors, their corresponding impacts on other determinants, or today’s measurement realities and limitations (Stover 1998). As a result, the fertility-inhibiting effects of each of the PDs, particularly contraception, may be under- or over-estimated, thereby conflicting with on-the-ground realities. Therefore, the central question of our research is to what extent can the accuracy of the PD framework be improved in SSA—both in terms of predicting fertility levels and changes over time—by implementing revisions to the model, with emphasis on the contraception index? This research holds important implications for how SSA—and thus world—population may be expected to grow, and may also help set expectations for the demographic impacts of FP programs in the region.

Background: Mismatches between Contraceptive Use and Fertility in SSA, and the Role of Proximate Determinants

Our research focuses on SSA for several reasons. First, future world population growth is expected to be concentrated in SSA (United Nations, Department of Economic and Social Affairs, Population Division 2013), lending heightened importance to fertility projections for this region. Furthermore, various authors have posited ways in which the fertility transition in SSA may not follow the same pattern as in other regions. Notably, the SSA fertility transition began at higher fertility levels and has shown a slower pace of decline over time. Furthermore, desired family sizes remain relatively high compared to those of other regions at similar points in their
fertility transition (Bongaarts and Casterline 2012). Others have noted that birth intervals have increased over time in SSA, regardless of age and parity, suggesting fertility decline across all age groups instead of concentrating at the older ages as in other regions (Caldwell et al. 1992; Timaeus and Moultrie 2008). Finally, SSA has a relatively high level of unmet need and a rapidly changing FP landscape characterized by recent efforts focused on increasing contraceptive uptake, such as FP2020.

The genesis of this research was the repeated observation, in various forms, that TFR in SSA has not always changed as expected. Fertility expectations in SSA are often anecdotally based on the CPR-TFR trajectories of neighboring or comparable countries/regions, where a strong inverse relationship between contraceptive use and fertility is observed. In other cases, expectations are rooted in research estimating simple linear relationships between fertility and various explanatory variables, particularly CPR, based on international data. Published literature has long predicted fertility levels based on a single explanatory variable, particularly CPR. This body of work is distinguished by international regressions testing the idea that there is a simple, generalizable statistical relationship between CPR and TFR, thereby mirroring anecdotal expectations in direction and magnitude of fertility change. Over time, authors have developed variations of CPR-TFR regressions for SSA, with coefficients in the range of.04-.07. That is, TFR is expected to decrease by about .04-.07 for each percentage point increase of CPR (Liu et al. 2008; Mauldin et al. 1988; Westoff 1990). However, the large body of work exploring this relationship through regressions has commonly identified directional or level mismatches in SSA over time, indicating that fertility in the region is not behaving as expected based on trends in contraceptive use (Jain et al. forthcoming; Thomas and Mercer 1995; Westoff et al. 2013).

TFR expectations are also often based on levels and trends in the fertility determinants (including contraception) of the PD framework. Since the late 1970s, Bongaarts’ PD framework (Bongaarts 1978) has proven to be the most robust and commonly used method for examining fertility changes at country and regional levels. Unlike linear regressions, which focus on a single explanatory variable, Bongaarts’ original PD model profiles four key biological and behavioral factors that directly depress total fecundity (TF), the biological maximum number of children an average woman might have in her lifetime:

1. $C_m$, the index of marriage, or union
2. $C_l$, the index of lactational infecundability
3. $C_a$, the index of induced abortion
4. $C_c$, the index of contraception
In previous SSA TFR analyses using this framework, the combined impact of all indices fails to explain all variation in observed TFR (Weinberger and Coast 2014; Blanc and Grey 2002; Thomas and Mercer 1995). Several authors have proposed explanations to account for these mismatches.

First, the model may either overlook or not accurately capture fertility-regulating behaviors specific to SSA. Recent SSA analyses have posited other behaviors as important determinants, such as the effect of gender-based violence (Odimegwu et al. 2014), adjustment of coital frequency (Blanc and Grey 2002), the practice of polygyny and its positive impact on union formation and community fertility (Cahu et al. 2014), and conflicted fertility desires leading to high short-term demand for contraception in contrast with long-term demand for larger families (Rossier et al. 2014). More generally, other investigations of SSA’s CPR and TFR patterns have explored the possibility that contraceptive use may not be as influential or as closely related to fertility changes as it has been for other regions’ fertility transitions (Guengant and May 2009).

Second, the limiting assumptions of the PD framework may account for some of the mismatch between predicted and observed TFR. For instance, the framework treats women who are married/in-union as the only population at risk of pregnancy. As a result, the model may underestimate the degree of fertility inhibition of $C_m$ in countries where exposure to pregnancy occurs outside of marriage/union and vice versa. Thanks to recent data collection efforts, new definitions of sexual activity have been proposed to more accurately capture women at risk of pregnancy (Stover 1998). In other cases, the model may overestimate the degree of fertility inhibition of a certain index. For instance, women who are both contracepting and postpartum insusceptible are captured in both the $C_i$ and $C_c$ indices, thereby double counting the protective effects (Boohene and Dow 1987; Stover 1998). One analysis found that this overlap may account for up to half the difference between predicted and observed fertility values in the past (Adamchak and Mbizvo 1990).

Finally, obstacles related to the direct measurement of each index and fertility itself—including data quality issues—may account for the mismatches (Stover 1998; Blanc and Grey 2002). For example, birth histories, contraceptive use, abortion, amenorrhea, and abstinence may be impacted by self-reporting errors and respondent bias, which can lead to over- or under-estimation of predicted TFR.

Despite wide recognition of discrepancies between predicted and observed fertility levels, no standard guidance exists for judging mismatches and the level of acceptable residuals. In some cases, arbitrary cutoffs, like half or one child above or below observed fertility, have been used (Jain et al. forthcoming). There have also been few attempts at minimizing residuals through
changes to Bongaarts’ original PD framework and index measurement. In recent analyses of predicted TFR and the relative influence of each determinant in SSA, any residuals between calculated and observed fertility rates were absorbed into the abortion index (Tsui et al. 2010), into the total fecundity (TF) term (Guengant and May 2009), or left as an error term (Madhavan and Guengant 2013). In one case (Madhavan and Guengant 2013), the residual error term ranged in magnitude from negligible to larger than any of the PD indices. A 2014 UN report utilizing the PD framework notes that “estimates of TF may exceed its theoretical limits because of a large error term in the proximate determinants framework, attributed in a large part to the effects of unmeasured factors exogenous to the framework” (Kisambira 2014). These large error terms—whether presented as such or absorbed into total fecundity or the abortion index—suggest that existing theories regarding fertility determinants, fertility data, PD data or the PD framework itself, may have limitations, especially within the SSA context.

Thus, there is an unfulfilled need for better explaining, understanding, and communicating how to predict fertility in SSA. Working within Bongaarts’ PD framework—which is entrenched in the FP literature and field of demography—and two alternative variations of the framework, we aim to 1) implement select adjustments to the framework; 2) establish a criterion for classifying acceptable residuals; 3) evaluate the performance of adjustments; 4) provide guidance to technical audiences, program planners and in-country analysts on expectations related to fertility reduction due to contraceptive use; and 5) suggest avenues for future research. Adjustments to the PD framework necessarily involve examination of all of the indices, although our focus remains heavily on the contraception aspect, because that is the starting point of our inquiry and currently an area of policy focus, as well as a concern for national and international funding of FP programs.

Methodology

For this analysis, we implemented identical adjustments to three variations of the PD framework:

1. Bongaarts’ Original PD model (Bongaarts 1978), considering only married/in-union women;
2. A Sexually Active variation (henceforth SA) of the PD model, identical to the original model with the exception of customization for sexually active women1 rather than those who are in-union/married only; and

1 Those women who were married/in-union at the time of interview or sexually active in the month prior to survey.
3. A John Stover variation (JS), which features Stover’s revisions (Stover 1998) to the original indices to better account for actual fertility-inhibiting effects.

We made three adjustments to each PD framework variation:

1. Eliminating CPR-postpartum insusceptibility overlap;
2. Accounting for the timing mismatch in the measurement of CPR and TFR; and
3. Customizing TF to each country.

We computed predicted TFRs—both TFR levels and changes in TFR between surveys—for the three PD variations described above, both before and after our three adjustments. We also separately evaluated the impact of each individual adjustment. We then compared each TFR prediction with the TFR reported by the DHS—the observed TFR—and analyzed residuals against the reported standard error of the DHS observed TFR estimate.

We assessed a total of 65 SSA DHS surveys on the accuracy of predicted TFR level and 40 SSA DHS surveys on the accuracy of predicted changes in TFR between surveys. Our exclusion criteria were: a) countries with populations under one million; b) islands, except Madagascar; c) surveys with poor data quality or missing variables (Nigeria 1999, Eritrea 2002 and Rwanda 2007-2008); d) surveys conducted prior to 1990 due to limited data on never-married women; e) countries with only one post-1990 survey, which would hinder analysis of intersurvey change; f) surveys with no standard errors provided. Poor data quality was assessed based on expert opinion. When we encountered inconsistencies in DHS data between reports, StatCompiler, and data files, we deferred to the data files or the DHS reports.

Defining the Population at Risk of Pregnancy

Most analyses of fertility must define the population exposed to the risk of pregnancy. This population is always confined to women of reproductive age (WRA)—generally between the ages of 15 and 49—and often includes women who are either married or in-union (cohabitating). Because the births that contribute to the TFR may come from any WRA, regardless of marital or cohabitation status, the SA and JS variations define more precisely which women are exposed to the risk of pregnancy due to recent sexual activity. The JS variation considers WRA who were sexually active within the last month, pregnant, or postpartum abstaining to be exposed to the risk of pregnancy. In the case of the SA variation, the WRA we considered to be exposed to the risk of pregnancy were those who were married/in-union or sexually active (regardless of union/marital status) in the preceding month. Unlike the PD revisions proposed by Stover (1998), we did not consider current pregnancy or postpartum status alone as sufficient evidence of current sexual activity, because of the possibility of inconsistent sexual activity over time.
Furthermore, many postpartum women are amenorrheic, abstaining, or otherwise insusceptible precisely because of their postpartum status.

Proximate Determinants Adjustments

*CPR-Postpartum Insusceptible Overlap*

In Bongaarts’ original PD framework, contraceptive use and postpartum insusceptibility both have protective effects against pregnancy. Some women use contraception during the period of postpartum insusceptibility; in this case, their contraceptive use provides very little, if any, additional protection from the risk of pregnancy. Nonetheless, the contraception index does not exclude insusceptible contraceptive users from being counted. For this reason, we removed the postpartum insusceptible population from the CPR estimate.

*CPR 27 months prior to survey*

In the DHS, TFR is calculated based on births during the three years preceding a woman’s interview date; these conceptions took place between 44 and 9 months prior to interview date. In contrast, contraceptive use is measured by “current use” at the time of the interview. Therefore, a timing mismatch exists between the measurement of fertility and measurement of contraceptive use. Unless CPR has remained constant over the time period, a current-status estimate of CPR may not represent contraceptive use during the period when the births that contribute to the TFR were conceived. Therefore, we adjusted the measurement of CPR so that it more accurately aligns to the midpoint of conception of the TFR births: 27 months—or 2.25 years, prior to the survey. In order to synchronize the timing of measurement for CPR and TFR, we interpolated contraceptive prevalence between surveys and used the point 27 months (the TFR midpoint between 9 and 44 months) prior to date of interview.

We also calculated CPRs at 9, 27, and 44 months prior to the survey date, as well as a measure of the total person-months of contraceptive use over the 9-44 month timespan, both based on calendar data. However, we present here the only interpolated 27-month CPR due to limited availability of calendar data and concerns about the reliability of respondent recall on the calendar-data CPR estimates.

*Total Fecundity*

For the final adjustment, we calculated country-specific TFs rather than assume a global constant level of fecundity, such as the 15.3 proposed in the original PD framework, or the 21 proposed by Stover (Stover 1998). We used the PD framework equation to solve for TF, using both
observed CPR and TFR from the survey data. We averaged the calculated TFs from each survey within a country, and used that country-specific average for each survey from that country. This was done separately for each variation of the PD framework. We searched for literature about variations in fecundity levels between countries, and did not find any guidance. However, an ANOVA comparing the estimated total fecundities from all surveys by country produced a highly significant p-value, which supports the hypothesis that there is important country variation in estimated TF. This adjustment customizes the PD to account for unobserved country-level characteristics; we were not able to assess whether these inter-country variations truly reflect variation in underlying biological fecundity or variation in factors exogenous to the PD framework. We used the multi-survey national average, rather than calculated total fecundity from each survey for two reasons: (1) to smooth out survey-specific or period-specific features of any one total fecundity estimate; and (2) using the total fecundity estimate from each survey would be contrary to our research question of predicting TFR for the survey in question.

Defining Accuracy: TFR Confidence Intervals

As discussed in the background section, existing analyses of TFR estimates are based on an arbitrary band of accuracy around the observed TFR, such as +/- 0.5. We calculated 95% confidence intervals of the DHS TFRs in order to more rigorously assess TFR predictions and establish criterion for classifying acceptable residuals; if the predicted TFR fell within the confidence interval, we deemed that an accurate prediction. These confidence intervals were calculated from DHS published standard errors and we selected an alpha value of 0.05. The confidence interval was calculated as TFR +/- 1.96*Standard Error.

Since there are no published standard error values for intersurvey TFR change, we consulted DHS staff and estimated the standard error of intersurvey TFR change as \( \sqrt{(SE1^2 + SE2^2)} \), where SE1 is the standard error of the earlier survey and SE2 the standard error of the more recent survey. This method was selected because successive surveys are assumed to be independent. The confidence interval for intersurvey change was thus calculated as \((TFR1-TFR2) +/- 1.96*(\sqrt{(SE1^2 + SE2^2)})\).

Age-Specific Analysis

In addition to the aforementioned adjustments, we also selected one country for an age disaggregated PD calculation. We selected Zimbabwe because it is a country whose TFR is consistently poorly predicted. We used age-specific inputs, where available, for all indices. The only input for which we did not have any age specific data was the Total Abortion Rate. Median
duration of postpartum insusceptibility was only available for two age groups: 15-29 and 30-49. All other inputs were specific to each five-year age group.

$C_m$ was calculated as a fertility-weighted average of the proportion of women in union or sexually active (whichever the relevant subset was) in each age group. We took two different approaches to aggregating the age-specific data into TFR estimates. In the first approach, we calculated age-specific PD indices, which were used to calculate age-specific fertility rates; the sum of these age-specific fertility rates was the TFR estimate. In the second approach, we calculated overall (for ages 15-49) PD indices using the averages of the age-specific inputs; the PD indices were then used to estimate the TFR. We examined the results both with and without a set of simple age-specific fecundity weights, which were based on Hutterite natural fertility data.

**Intermediary Descriptive Results**

Proximate Determinants Adjustments

In this section we present intermediary descriptive results comparing the impact of our PD framework adjustments on the individual indices; in the subsequent results section, these adjustments feed into the predicted TFRs.

*CPR Excluding Postpartum Insusceptible Users*

As expected, we observed a decline in CPRs after removing postpartum insusceptible (amenorrheic or abstaining) contraceptive users from the CPR estimate (see Fig. 1). We detected some country-specific patterns, with larger CPR declines in southeastern African nations—specifically Malawi, Zambia, Zimbabwe and Namibia—compared to other SSA regions, pointing to higher incidence of dual protection from pregnancy because of overlapping contraceptive use and postpartum insusceptibility. The magnitude of decline was near-uniform across the three variations of the population of women at risk of pregnancy.

**Fig.1** Total and Adjusted CPR Following Postpartum Insusceptible User Exclusion, Country Averages Across Surveys
**Interpolated CPR**

We observed declines in contraceptive use when interpolating CPR to 27 months prior to survey date, with few exceptions. This pattern was an expected outcome, given the overall trend of increases in CPR over time. The exceptions—instances where interpolated CPRs were higher than the point (survey) measurement of CPR—represent cases where contraceptive use was on the decline. The percentage point changes were again near-uniform across the three variations, with an average CPR decline of 1.6 percentage points.

**Country-Specific TF**

We identified considerable inter-country variation in TF by customizing fecundity rather than assuming a constant value for all countries. The resulting TF ranged from 12-13 for countries like Senegal (Bongaarts Original) and Gabon (SA Variation and JS Variation) to 25 in Namibia and Ethiopia (SA Variation and JS Variation). The largest TF was Namibia for the Bongaarts original framework, attributable to the country’s low proportion of married/cohabiting women—the only subset of women considered to be at risk of pregnancy in Bongaarts’ original...
framework. As a result of the smaller population considered at risk for pregnancy, estimated TFs for Bongaarts’ Original model tend to be higher (average 19.9) as compared to SA and Stover variations (averages 17.3 and 18, respectively). We did not observe major regional patterns in TF.

**Fig. 2** Country-Specific Total Fecundity, Country Averages Across Surveys

---

**Results**

**TFR Level**

In this section we present all results pertaining to predicting TFR level. To gauge overall accuracy, we analyzed the predicted TFR levels, after all three adjustments combined, to see if they fell in the 95% confidence interval of the observed TFR. Next, we compared the predicted TFR levels when applying each of the individual adjustments independently, to see if they yielded a TFR prediction that fell within the 95% confidence interval.

We also calculated and inspected residuals—that is, the distance between the observed and predicted TFRs—for any patterns by survey year, fertility level or country. Finally, we compared the three variations of the PD framework to identify any systematic differences in accuracy between the three variations.
Overall Accuracy

Our adjustments substantially improved the predictive accuracy of the PD framework. Accuracy was evaluated in two ways: 1) the percent of predicted TFRs within the 95% confidence interval of the observed TFR and 2) average residuals between the observed and predicted TFRs.

Before our adjustments, between 12% and 17% of predicted TFRs across all three PD variations were within the 95% confidence interval of the observed TFRs (Table 1). After our adjustments, between 45% and 63% of the predicted TFRs were within the 95% confidence interval of the observed TFRs. Because the TFR confidence intervals are a more narrow band of acceptability (average width of +/- 0.25) than those previously used in other analyses, Table 1 also includes the percent of surveys within a +/- 0.5 and +/- 1 band of the observed TFR for comparability with other studies.

Implementing our adjustments decreased the average residual. The absolute values of the residuals were approximately three times smaller after our adjustments, indicating more accurate TFR predictions.

Table 1 Accuracy of Predicted TFR Level Before and After Adjustments

<table>
<thead>
<tr>
<th></th>
<th>Bongaarts Original</th>
<th>SA Revision</th>
<th>JS Revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Adjustments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>12%</td>
<td>17%</td>
<td>14%</td>
</tr>
<tr>
<td>Average Residual (absolute)</td>
<td>1.27</td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td>After Adjustments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>55%</td>
<td>63%</td>
<td>45%</td>
</tr>
<tr>
<td>Average Residual (absolute)</td>
<td>0.40</td>
<td>0.32</td>
<td>0.37</td>
</tr>
<tr>
<td>Accuracy (+/- .5)</td>
<td>75%</td>
<td>82%</td>
<td>71%</td>
</tr>
<tr>
<td>Accuracy (+/-1)</td>
<td>94%</td>
<td>98%</td>
<td>91%</td>
</tr>
</tbody>
</table>

Disaggregated Impacts of the PD Adjustments

We evaluated the isolated impact of each individual adjustment on the predicted TFR to determine whether some adjustments improved accuracy more than others. We found that all three adjustments improved the accuracy of predicted TFR, though to varying degrees. As shown in Table 2, the country-specific TF adjustment produced the greatest improvement on the the accuracy of the predicted TFR, with up to 58% of predicted TFRs falling within respective confidence intervals. Before the adjustment, a standard TF of 15.3 (Bongaarts 1978) or 21 (Stover 1998) was used and country-specific patterns were captured in residuals. For example, all residuals in Ethiopia and Zimbabwe were negative before the adjustments, while all the
residuals in Mali were positive. Allowing the total fecundity to vary by country addressed such country-specific patterns by shifting all predicted TFRs upwards or downwards, thereby decreasing the magnitude of residuals. As a result, the residuals for countries like Ethiopia, Zimbabwe, and Mali were far less skewed following country TF customization. As noted in the Methodology section, we do not know if this is due to true variation in underlying biological fecundity or if it captures fertility determinants exogenous to PD.

**Table 2** Accuracy of Individual TFR Level Adjustments

<table>
<thead>
<tr>
<th></th>
<th>Bongaarts Original</th>
<th>SA Variation</th>
<th>JS Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>After Adjustments</strong></td>
<td>TF Adjustment</td>
<td>51%</td>
<td>58%</td>
</tr>
<tr>
<td>Interpolated CPR</td>
<td>9%</td>
<td>26%</td>
<td>12%</td>
</tr>
<tr>
<td>CPR-PPI Overlap</td>
<td>14%</td>
<td>23%</td>
<td>18%</td>
</tr>
</tbody>
</table>

**Residual Patterns**

We did not identify any patterns in the residuals that might suggest systematic bias or provide direction for further investigation. We hypothesized that TFRs from more recent surveys might be better predicted because of improved data quality, increased sample size, or improved data collection techniques. However, we did not find any time-varying pattern to the residuals. Conversely, we also hypothesized that our accuracy rate might have decreased over time due to smaller standard errors in more recent surveys with larger sample sizes and thus smaller confidence intervals in more recent surveys; in fact, the standard errors and confidence intervals have not systemically decreased with time. We also analyzed the residuals by TFR level, to explore the possibility that TFR could be better predicted at low or high TFR levels; we found no such pattern.

The only residual pattern before our adjustments was country-specific, with some countries consistently having positive or negative residuals, across multiple surveys. After the adjustments, these country-specific patterns were greatly attenuated by the TF adjustment. We looked for but did not find other residuals patterns, such as by TFR level or by survey year.

**Comparison of Three PD Variations**

We compared the three variations of the PD framework to determine whether the use of one variation yielded more consistently accurate results than the other variations. The SA variation performed the best when considering the accuracy rate and average residuals (Table 1). The only difference between the Bongaarts original and the SA variation was the group of women
considered to be at risk for pregnancy. In contrast, the JS variation included other modifications to the framework. Because the results for the JS variation often did not deviate notably from the results from the other two variants, we believe the question of what group of women should be considered at risk for pregnancy is perhaps more salient than some other methodological tweaks to the PD indices.

TFR Intersurvey Change

In this section we present all results predicting TFR change between survey years. We first analyzed predicted TFR change, after all three adjustments combined, to see if they fell within the 95% confidence interval of the observed TFR change. Next, we compared the predicted TFR change when applying each of the individual adjustments independently, to see if they yielded a TFR change prediction that fell within the 95% confidence interval. Finally, we compared the predicted to observed TFR change to identify instances when the direction of change was predicted incorrectly: for example, when the predicted TFR change was positive and the observed change was negative, or vice versa.

Overall Accuracy

Our adjustments did not notably improve the accuracy of TFR change predictions. As shown in Table, of 40 intersurvey intervals, approximately half of the predicted TFR changes fell within corresponding confidence intervals before our adjustments. This relatively high accuracy rate should not be compared to that of TFR level (see Table) because of two factors: 1) the relatively large confidence intervals (average width of 0.72, compared to the average width of 0.5 for TFR level); and 2) the smaller absolute value of intersurvey change (an average of 0.37), compared to the absolute value of TFR level. As seen in Table, our adjustments only produced an increase in the predictive accuracy of TFR change for the JS PD variation.

Table 3 Accuracy of Predicted TFR Intersurvey Change Before and After Adjustments

<table>
<thead>
<tr>
<th></th>
<th>Bongaarts Original</th>
<th>SA Revision</th>
<th>JS Revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Adjustments</td>
<td>Accuracy</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>After Adjustments</td>
<td>Accuracy</td>
<td>50%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Disaggregated Impacts of the Individual Adjustments

Unlike fertility level predictions, the gains in accuracy from our adjustments when applied separately were small or negative (see Table 4). Of the three adjustments, the interpolated CPR
adjustment produced the largest gains in accuracy, especially for the PD variations dealing with sexually active women; this may be because measuring CPR at a more precise time may be especially important to predicting TFR change over time. The TF adjustment decreased the accuracy of the TFR change predictions. Since TF was held constant in each country between survey years both before and after the TF adjustment, albeit at different levels, we would expect this adjustment to have minimal impact, and believe the decreased accuracy rate is random, rather than systematic.

**Table 4** Accuracy of Individual TFR Intersurvey Change Adjustments

<table>
<thead>
<tr>
<th>After Adjustments</th>
<th>Bongaarts Original</th>
<th>SA Variation</th>
<th>JS Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF Adjustment</td>
<td>42.5%</td>
<td>42.5%</td>
<td>50%</td>
</tr>
<tr>
<td>Interpolated CPR</td>
<td>50%</td>
<td>57.5%</td>
<td>57.5%</td>
</tr>
<tr>
<td>CPR-PPI Overlap</td>
<td>52.5%</td>
<td>50%</td>
<td>45%</td>
</tr>
</tbody>
</table>

A directional mismatch occurred when the observed TFR declined between surveys, but the predicted TFR increased, or vice versa. Twenty-eight percent of our TFR predictions were directional mismatches, while 72% predicted TFR change in the correct direction. Fig. 3 plots the observed vs. predicted TFR changes. Points in upper right and lower left quadrants represent intersurvey changes that were predicted in the correct direction.

**Fig. 3** Predicted vs. Observed Intersurvey TFR Change
Age-Specific Analysis

We examined the TFR predictions for both the TFR level and the TFR intersurvey change for four Zimbabwe surveys (2010-2011, 2005-2006, 1999, and 1994). Overall, the residuals of the age-specific analyses were on a scale similar to the residuals of the 15-49 residuals. We also did not see any residual patterns between the two different versions of the age-specific analyses, nor between the versions with and without the fecundity weighting.

Discussion

The unadjusted PD framework does not predict TFR level well, according to our criterion of the 95% confidence interval, regardless of how the population of women at risk for pregnancy is defined. Two of the individual adjustments to the PD framework (removing the CPR-postpartum insusceptible users overlap and estimating the CPR at the midpoint of the conception of the TFR...
births) yielded generally positive, but inconsistent results (see Table). Although these two adjustments are commonly suggested in both the literature and by conventional wisdom among demographic researchers, their application did not always render notable gains in predictive accuracy, resulting instead in either marginal or no improvement.

The individual adjustment with the largest improvement on the predicted TFR level was customizing TF by country. Allowing TF to vary by country accounted for cross-national variation, as evidenced by the resulting wide range of estimated TF values. It is unlikely that the full range of these estimated TF values is attributable to true variation in underlying biological fertility alone. More likely, the wide range of TF values points to the importance of country-level factors exogenous to the PD framework. When the adjustments are implemented simultaneously, the accuracy rate notably improved for all three variations of the PD framework (see Table). This is largely attributed to the use of the country-customized TF values and emphasizes the importance of better understanding and evaluating inter-country variation.

Approximately half of all unadjusted TFR intersurvey change predictions fell within the relevant 95% confidence interval. After applying the adjustments, there was no substantial improvement in predictive accuracy. When applied and analyzed individually, the adjustments to the PD framework also did not produce major improvements in accuracy of predicted TFR change. While applying a country-specific TF as an adjustment improved TFR level predictions, it did not similarly improve TFR change predictions. This is because prior to the fecundity adjustment, TF was held at a standard 15.3 (Bongaarts) or 21 (Stover) for all countries and across all survey years. After the TF adjustment, a customized national average TF was applied to each survey year for a country; thus from survey year to survey year within a country, the TF remained constant. Since TF was held constant between surveys both before and after the TF adjustment, though at different levels, this adjustment expectedly had minimal or inconsistent impact on improving the accuracy of predicted TFR change between surveys.

While our adjustments yielded substantial improvement over the unadjusted PD framework when predicting TFR level and marginal improvement when predicting TFR change, the overall accuracy rates still may not be suitable for the degree of certainty that policy and program planners would prefer. The key to improving accuracy may lie in better understanding country-specific patterns, as evidenced by the range of estimated TF values. Thus, our work points to the need for research directly addressing the question of whether there is true population-level variation in biological fecundity between countries in sub-Saharan Africa. Additionally, research should further explore cross-national variation in behaviors and sociological patterns that may
also impact fertility, such as spousal separation and coital frequency, and better methods for measuring these determinants.

Although there are a variety of avenues for further investigation, it is reasonable to conclude that until more research is done, it remains difficult to predict short-term fertility change to a high level of accuracy. It is therefore important to adjust expectations about the certainty of fertility predictions among demographic researchers, donors, policy makers and program planners in the field. Managing expectations about the impact of contraceptive uptake on fertility will be a challenging task—one that needs to highlight the health and economic benefits of voluntary FP uptake, as well as FP provision as a fundamental component of healthcare.

**Limitations**

Due to data and time limitations, we were not able to perfectly customize each index for each PD variation to the chosen population exposed to the risk of pregnancy. For instance, the index of insusceptibility was calculated based on data from all women because we did not obtain data on median duration of postpartum insusceptibility by subgroup of women. The abortion index was also calculated based on regional data and does not differ by subgroup of women.

We made similar assumptions for the Zimbabwe age-specific analysis, where abortion rates did not vary by age or by the population considered at risk of pregnancy, and the median duration of postpartum insusceptibility did not vary by five-year age groups.

Our analysis also does not address non-sampling error that can result from subject or interviewer misinterpretation of interview questions, or errors in data entry and processing. Furthermore, questionable self-reporting of recent sexual activity, due to either social desirability bias among unmarried or postpartum women, or recall bias, may also misidentify the population of women at risk of pregnancy.

**Conclusions**

Predicting national fertility levels and fertility change to a pragmatically acceptable degree of accuracy using the PD framework is challenging given issues around defining the population at risk of pregnancy, as well as country-specific factors or behaviors currently unaccounted for by the framework.

Defining the population at risk of pregnancy is not a straightforward process for many reasons. For example, while most definitions include women in union and/or sexually active in the four
weeks preceding the interview, some also include other proxies, such as current or recent pregnancy status. Another challenge to defining the population at risk is that sexual activity and its proxies are measured within a month of the interview, while the TFR accounts for all births in the past three years. It is likely that many women will have changed their sexual activity status in that time period. Furthermore, measurements of sexual activity involve the possibility of respondent bias or recall problems. Finally, because comprehensive definitions of the population at risk of pregnancy are not standardized by the DHS, estimates of related metrics (for example, CPR or the proportion of postpartum insusceptible women) must be calculated manually from data files; publication of these as standard summary metrics would greatly enhance their accessibility to a wider range of analysts.

With this in mind, we recommend that the DHS standardize a definition of the population at risk of pregnancy and provide key variables for defining this subset of women. We also recommend that the DHS publish values for key fertility determinants (such as the CPR, the method mix, the median duration of postpartum insusceptibility) for varying subsets of women (all women, married/in union women, sexually active women, etc.), as well as the proportion of women in each relevant subset. This would greatly facilitate fertility analyses among a wider audience, and also allow analysts to examine which subset of women is most appropriate to consider for their national context. A further recommendation is that inconsistencies between the DHS data and summary statistics found in the report, StatCompiler, and the data files be harmonized to avoid unnecessary confusion; we found a number of inconsistencies between the various sources reporting results from the same surveys.

Further research should explore the importance of country-specific factors. The possibility of cross-national variation in biological fecundity should be examined. Other sociological and behavioral factors, like coital frequency and spousal separation, which are exogenous to the PD framework should also be considered. Inter-country variation in such factors could explain the large variations we saw in our country-specific TF estimates, and therefore help refine the framework for improved predictive accuracy.

One contribution of our work is the proposal of using standard errors as the basis for a standardized and evidence-based criterion for TFR analyses. Previous analyses have used arbitrary criteria, such as +/- 0.5. While one could expand the confidence interval used—for example, to 90% —our analysis underscores the importance of using these underutilized sampling error data to frame our thinking about the uncertainty of all survey estimates, including the TFR. Further analyses could explore using the sampling errors provided for CPR and other metrics in FP and fertility analyses.
Our research has allowed us to comment on the effectiveness of various theories regarding how to make fertility predictions more accurate using the PD framework. We conclude that while the PD framework may be used to analyze and compare the contribution of different fertility determinants from an order-of-magnitude perspective, there are limitations to its use for making precise predictions. Demographic researchers working with policy makers and program planners should better communicate the limitations of basing expectations of future fertility levels or fertility change on either a simple linear regressions approach or the PD framework. Despite the challenges involved, the PD framework, often populated with DHS data, continues to dominate theory, literature, and practice. Other approaches to examining the fertility-contraception relationship should be evaluated in detail in the context of SSA.

References


