Do desired family size questions measure current demand for childbearing? An analysis of the Demographic and Health Surveys Data

Running head: Desired family size questions and current demand for childbearing

Word count: 5855

Abstract

We investigate whether the differential framing of questions on Desired Family Size (DFS) in the Demographic and Health Surveys (DHS) for women with surviving children (mothers) and childless women impacts reported DFS. The survey question for women with children asks them to consider the time before their first birth in stating their ideal number of children. Using the full matching method, we estimated the average treatment effect of this differential framing of the question for mothers vs. childless women on reported DFS. We also matched mothers with long durations since first birth to those with short durations and estimated the expected causal effect of duration since first birth on reported DFS of mothers. The results show that motherhood (or the differential framing of DFS questions) significantly impacts reported DFS. More importantly, for mothers, longer durations since first birth increases DFS even at low parities. These results provide strong evidence that the responses of childless women and mothers to the DFS questions are not equivalent; the responses of mothers refer to a different period in the past and not to the time of the survey. Therefore, the DFS of mothers cannot accurately measure current demand for childbearing.

Keywords

Desired family size, demand for childbearing, sub-Saharan Africa, Average Treatment Effect, Matching.

Introduction

Desired Family Size (DFS), also referred to as Ideal Family Size (IFS), is measured in major national surveys like the Demographic and Health Surveys (DHS) through *two different questions;* one for women with surviving children (mothers) and the other for women with no surviving child (childless women). For mothers, the question asks: "*If you could go back to the time you did not have any children* and could choose exactly the number of children to have in your whole life, how many would that be?" For childless women, the question is: "If you could choose exactly the number of children to have in your whole life, how many would that be?" Responses to these two questions are combined and used to estimate current demand for childbearing in surveys and in almost all studies that have measured DFS using these survey questions (Bankole & Westoff, 1995; Bongaarts 2001; Bongaarts & Casterline, 2018). Combining responses to the two questions to estimate DFS assumes that both questions are the same and that they both measure DFS at the time of the survey. This paper questions this assumption in the measure of DFS in the demographic literature. It contributes to the wealth of studies that have raised concerns about the DFS measure and calls for renewed attention to understanding and measuring this important concept. Given the preponderance of evidence on the instability of the DFS measure over time (Heiland et al. 2008; Müller et al. 2022; Sennott & Yeatman 2012), we argue that, for mothers, their preferences before the birth of their first child do not represent their current preferences at the time of the survey. This has not been examined previously in the demographic literature.

Our argument is simple: if mothers answer the DFS question as framed, then their responses will be partly influenced by the prevailing fertility norms of the reference period around the birth of their first child. Therefore, while the reported DFS of childless women can only be influenced by current norms at the time of the survey, those of mothers may be influenced by norms in the past. Similarly, among mothers, the framing of the question could result in different responses depending on how long before the survey they had their first birth. Mothers who had their first birth decades before the survey may respond differently from those who had their first birth only few years before the survey. This is particularly important for sub-Saharan Africa where DFS declined substantially between the 1980s and 2000s (Bongaarts 2011; Bongaarts & Casterline 2013), and where the importance of sociocultural factors (such as religious beliefs, lineage and kinship systems, access to resources especially land, marriage systems, and women's position) in shaping individual fertility preferences has long been established (Caldwell & Caldwell 1987; Lockwood 1995; Hayford 2009; Price & Hawkins 2007; Rodrigues et al. 2022). More importantly, these norms have also been undergoing fundamental shifts over the past 50 years, with significant implications for individual autonomy in reproductive preferences and behavior (Sennott & Yeatman 2012). The impact of societal norms prevailing at the reference point of reported DFS can be particularly important (Bacci 2001; Kebede et al. 2022).

The above argument represents a fundamental departure from most demographic literature on fertility preferences in at least two ways. First, it treats the preferences of mothers as relating to a different period than the time of survey. Second, and drawing from the first, it treats desired family size as a period measure. Most previous studies treat DFS as a cohort measure (Bongaarts 1990; Hagewen & Morgan 2005; Heiland et al. 2008; Yeatman et al. 2013). This approach of treating DFS as a cohort measure assumes that fertility desires, once formed, are firmly held and that changes in individuals' reported preferences represent true changes in their underlying preferences (Gray et al. 2013; Iacovou and Tavares 2011; Rackin & Bachrach 2016; Ryder 1973, 1980). Lee (1980) summarizes this thinking as follows: "Each couple formulate at marriage a desired completed family size (D), and pursue this relatively constant target throughout their reproductive life. Averaging across couples, a value of D may be obtained for the cohort which should be constant over time, as variations in individual values of D tend to cancel" (p. 205).

In contrast, we argue that when mothers are asked: "*If you could go back to the time you did not have any children …"* they do indeed think back to a period in the past in framing their responses to the DFS question. The shortest period they would think back to would be at the time of the birth of their first child. The birth of a first child is a defining moment in the lives of women, and if women do recall their fertility preferences at any time in the past, they would most likely recall what it was around the birth of their first child. By extension, if women's reported fertility preferences vary depending on the reference period they consider in their responses, then their preferences cannot be assumed to be fixed; rather, they are dynamic, adapting and responding to various "opportunities and experiences" over their life (Bernheim et al. 2021; Kodzi et al. 2010).

This paper seeks to explore and to quantify whether mothers, who retrospectively answered the DHS question on ideal family size, respond differently to childless women who stated their ideal family size prospectively. Additionally, the paper also quantifies the impact of longer duration since first birth on DFS among mothers. The time of first birth refers to a specific period characterized by a set of prevailing norms and individual attributes. Since the question on DFS for mothers specifically references *before* first birth in its framing, the period of first birth can be taken as the shortest period women with surviving children could consider in answering the DFS question. In this paper, *mothers* and *childless women* are used to represent the two formulations of the DFS question in the DHS. *Women* refers to both mothers and childless women.

There are three broad sets of factors that influence women's reported DFS. First is their individual characteristics and experiences. Women's age, cohort, education, religion, place of residence, household wealth status, etc. are all known to be related to reported DFS (Isiugo-Abanihe 1994a; McCarthy & Oni 1987). Second is ex-post rationalization, especially among high parity mothers (Bongaarts 1990; Knodel & Prachuabmoh 1973; Yeatman et al. 2013). Third is the prevailing societal norms on childbearing around the reference period of the reported DFS (Caldwell & Caldwell 1978; Dharmalingam 1996; Isiugo-Abanihe 1994b). The effect of this third mechanisms will be zero if all women reported their DFS with respect to the same reference period. For mothers, therefore, duration since the birth of their first child should not matter if the framing of the question does not constrain them to consider a different period in their responses.

Methods and Data

Identification and estimation strategy

In statistical terminology, the desired family size effect of having a surviving child that we attempt to measure is the Average Treatment Effect on the Treated (ATT). Using the Neyman-Rubin potential outcome framework (Rubin 2005), the expected causal effect of *motherhood* or *having a surviving child* can be defined as:

$$ATT = \mathrm{E}[Y_i(1) - Y_i(0) | T_i = 1] ----------------- [1]$$

Where, $Y_i(1)$ and $Y_i(0)$ are the potential desired family size for a treated mother $(i \in N)$ that can be observed in the presence of the treatment, $(T = 1)$ and under the absence of the treatment $(T = 0)$, respectively. The ATT compares the observed outcomes for a mother with the counterfactual outcome that would be observed had she not had a surviving child. That is, what her DFS is based on answering the question for *mothers* versus what it would have been if she answered the question for *childless women*. Thus, to estimate the motherhood effect on women with surviving children, one must establish what would have happened to the mothers in the absence of the surviving child. However, for a given woman (i), one could observe only one of the two potential outcomes. This invokes the following identification assumptions.

The key assumption to the causal identification of the ATT is that the treatment is assigned independent of the potential outcomes. Under random treatment assignment, one would expect the same average desired number of children for *mothers* and *childless women* if both groups were to receive the treatment.

$$\mathrm{E}[Y_i(1)|T_i = 1] = \mathrm{E}[Y_i(1)|T_i = 0] = E[Y_i(1)] - - - - - - - - - - - - - - - - - - [2]$$

Similarly, under random treatment allocation, both groups would have the same average desired family size in the absence of the treatment.

$$\mathrm{E}[Y_i(0)|T_i = 1] = \mathrm{E}[Y_i(0)|T_i = 0] = E[Y_i(0)] - - - - - - - - - - - - - - - - - - [3]$$

The randomization implies that in studies where one cannot observe the counterfactual outcomes for the treated units, the average outcome for the control units can act as a surrogate to the counterfactual for the treatment units. This assumption is called the ignorability of treatment, independence of treatment, or unconfoundedness assumption (Huber 2020).

However, the fundamental problem in our treatment effect estimation is that the treatment (motherhood) was not assigned randomly or independent of the observed desired family size outcomes. Conversely, the two comparison groups, childless women and mothers, attain different demographic and socio-economic attributes that could predict the treatment status (having a surviving child) and their desired family size. For example, mothers are generally married, older, less educated, and live in rural areas than their childless counterparts. Therefore, to judge whether having a surviving child(ren) (that is, answering the question for mothers), would causally lead to a different response to the question on DFS, the appropriate comparison must be between mothers with similar baseline characteristics as childless women.

We use the **full matching** method to select an appropriate comparison of childless women for mothers with surviving children. The matching exercise aimed to control for observable differences in baseline characteristics that affect the stated ideal number of children and childlessness status. By creating balance across covariates, matching mimics randomization in the treatment assignment, which ensures that the remaining differences in the DFS are due to differences in treatment status (having a surviving child or answering a different DFS question). We assume that post-matching, the average desired family size of *mothers* and *childless women*, in the absence or presence of the treatment, is the same. That means that the average DFS for matched childless women represent the treated mother's counterfactual desired family size. This stronger identification assumption, also called the conditional ignorability or the conditional independence assumption can be stated formally as:

$$\mathrm{E}[Y_i(1)|T_i = 1, \mathrm{X}] = \mathrm{E}[Y_i(1)|T_i = 0, \mathrm{X}] = \mathrm{E}[\mathrm{Y_i}(1)|\mathrm{X}] -------------- [4]$$

$$\mathrm{E}[Y_i(0)|T_i = 1, \mathrm{X}] = \mathrm{E}[Y_i(0)|T_i = 0, \mathrm{X}] = \mathrm{E}[\mathrm{Y_i}(0)|\mathrm{X}] -------------- [5]$$

Where X is the set of pretreatment characteristics on which women are matched (see the data section for the list of covariates included in the matching exercise), $Y_i(1)$ is the outcome when the woman has a surviving child, $Y_i(0)$ is the outcome in the absence of a surviving child, T is treatment (T=1 if a woman has a surviving child). The conditional independence assumption states that the potential desired family size outcomes are jointly independent of the treatment assignment conditional on the set of matching covariates: $Y(1), Y(0) \perp T|X$

We use the full matching method, which assigns each treated and control woman to a subclass and receives at least one match. In addition to the advantage that no treated or controlled woman would be discarded in the matching exercise, the full matching method provides an excellent covariate balance (see Appendix Table 2).

Post-matching, we fit a linear regression model for the desired family size as a function of the treatment status and covariates included in the matching exercise. In the next step, we estimate the marginal effect of the treatment by comparing the weighted (matching weights) average of the estimated desired family size for a treated unit under the treatment and control status. Finally, since non-uniform weights were used in the estimation, we estimate robust standard errors using the R package '*sandwich*'.

The causal interpretation of the estimated effect hinges on the conditional independence assumption that units have comparable observed and unobserved pretreatment characteristics, which determine the desired family size and assignment to the treatment. Although the matching exercise accounts for heterogeneity in key observable covariates, one cannot rule out the influence of unmeasured confounders. Therefore, we use the sensitivity test suggested by Cinelli and Hazlett (2020) to examine the robustness of our estimates to a possible effect of unobserved heterogeneity. We also use similar matching and post-matching estimation methods to explore the impact of duration since first birth on differences in the stated desired family size among mothers. We hypothesize that mothers who had to refer to an earlier period when responding to the question on ideal family size (thus longer duration since first birth) would state a different ideal family size than mothers with similar characteristics who recently gave birth to their first. For this exercise, we define the control as mothers who gave birth to their first in the last ten years, while the treated are women with more than ten years duration since their first birth.

Data

The study is based on data from the Demographic and Health Surveys (DHS) program. The DHS surveys use a two-stage cluster sampling design and standard questionnaires to collect comparable nationally representative data on demographic, socioeconomic, and health characteristics of eligible households and individual household members, including women of reproductive age (15-49) (further details are available in Rutstein and Rojas 2006). This study uses data from all available DHS surveys, as of December 2022, conducted in sub-Saharan Africa since 2003. Surveys conducted before 2003 did not collect information on household wealth index, which is one of the key matching covariates in our analysis**. Appendix Table 1** provides information on the DHS surveys included in the analysis and the number of pre-matching observations across all surveys in each country.

In our matching analysis, a treated woman is one with surviving child(ren) at the interview date, while women with no surviving child at the interview date were defined as control units. Before matching, our analysis pools relevant information on 281,215 childless women and 693,059 mothers from 81 surveys conducted in 37 sub-Saharan African countries.

A secondary analysis was also conducted to explore the impact of duration since the first birth among mothers. The duration variable measures the number of years that have elapsed since the respondent gave birth to her first child. It was calculated by subtracting the respondent's age at first birth from their current age at the time of the survey. This new measure represents the minimum reference point women with at least one surviving child would have considered when responding to the question on DFS. In this secondary analysis, a control woman is the one who recently (within the last ten years) gave birth to her first child, while those with more than ten years duration since first birth are treated group. Our analytic samples exclude women who provided non-numeric responses to the questions about ideal family size.

As discussed in the methods section, the causal identification of the effect of motherhood or duration requires controlling for confounding variables: factors that affect the likelihood of having a surviving child(ren) and the desired number of children. To block confounders' influence while identifying the average treatment effect, we matched units on the following demographic and socioeconomic factors known to influence DFS and childbearing: birth cohort, area of residence, educational attainment, household wealth quintile index, marital status, country, and survey year. In the secondary analysis of the duration effect, in addition to the above listed variables, we also controlled for parity as a categorical variable (women with 1-2, 3-4, and 5+ children). Comparing the control and treated units across this rich set of variables would meet the conditional independence assumption required for causal identification.

Results

Effect of different framing of DFS question on reported desired family size

Table 1 shows the pre-matching summary statistics of key variables included in the analysis. As would be expected, childless women are younger, less likely to be in a union, and more educated than women with surviving children. They are also more likely to live in urban areas and in richer households. Among all women, childless women reported a mean DFS that is more than one child lower than women with at least one surviving child. Among mothers, those who had their first birth more than 10 years ago reported a mean DFS that is more than one child higher than those who gave birth to their first child less than 10 years ago. They are also older, have more surviving children, and slightly less educated compared to those who had their first birth more recently.

Table 1 is about here

The above comparisons reveal that the distributions of baseline characteristics that could led to differences in the mean DFS between mothers and childless women are different, which calls for matching methods to eliminate the influence of observable baseline confounders. **Table 2** presents a post-matching average treatment impact estimates from a ‘full matching” exercise with regression bias adjustments. The result shows that, for all women, mothers have a higher mean DFS of about 0.62 children compared to childless women. It may be argued that mothers and childless women, in addition to answering different DFS questions, also differ based on their motherhood status and the factors that affect their motherhood status may also influence their reported DFS. If mothers and childless women differ on their DFS based on their motherhood status, we would expect such differences to exist at every age group. We therefore examined the relationship separately among younger women (15-19) and older women (20-49). The assumption is that if motherhood status matters, it should matter equally for both younger and older women. That is, younger mothers (15-19) should report higher DFS than childless women 15-19, like the differences among older mothers (20-49) and older childless women. However, if it is the framing of the question that matters, then, we should expect differences in DFS to matter less among younger women than older women. Although younger mothers answered the retrospective DFS question, since they would have had their first birth more recently, the reference period for their DFS would be very short and should be influenced by similar societal norms as those of childless women who reported their DFS prospectively. The result is very interesting: following the full matching, mothers aged 15-19 desire fewer children than childless women aged 15-19. Among older women aged 20-49, however, being a *mother* is associated with a 0.68 higher DFS than being a *childless woman*.

Table 2 is about here

Figure 1 further examined the above relationship between motherhood and desired family size across each of the 37 countries included in the sample. The result revealed that, except for three small countries with only one survey each since 2003 (Eswatini, Sao Tome and Principe, and Comoros), the estimated effect of the different framing of DFS questions on DFS is positive and significant in all countries. In countries such as Niger and Uganda, the mean difference in the desired family size between *mothers* and *childless women* is particularly high (more than a child per woman). The estimated impact is more pronounced among older women aged 20-49. For younger women aged 15-19, the effect is non-significant in about half of the countries and negative in the rest (see **Appendix Figure 1**).

Figure 1 is about here

Rival explanations: duration effect or post-facto rationalization?

There are two possible explanations for the estimated positive effect of motherhood on reported DFS. The first is parity or post facto revision of DFS and the second is duration since the birth of the first child. The first explanation supports the argument that mothers, especially those with high number of children, rationalize their DFS to accommodate all the children they already have. The duration explanation argues that when *mothers* are asked to think back to a time when they did not have any children, those with longer durations think back to an earlier period when prevailing societal norms supported larger family sizes. To test the effects of these competing explanations, we limited the analysis to only women with *at most* two children. The argument is that the effect of post facto revision of DFS will be minimal among women with two or fewer children. However, if duration matters, it should matter at all parity levels, including at low parities.

Table 3 presents the results of the analysis. Column 1 compares the ATT of *mothers* with only 1-2 surviving children to *childless women*. The mean treatment effect of having a child (or answering the retrospective DFS question) is small but statistically significant. This small but positive effect of motherhood in the above comparison is consistent with the duration hypothesis. Women with only 1-2 surviving children are relatively young, with short durations since first birth; therefore, the overall duration effect (relative to childless women) is expected to be small. To further test the duration effect, we restricted the analysis to *mothers* with 1-2 children who had their first child more than ten years ago (versus childless women) in Column 2 of Table 3. The result shows that these low parity mothers with longer durations since their first birth reported a mean DFS of 0.38 children higher than childless women.

Table 3 is about here

Further test for duration effect

To explore, more explicitly, the role of duration on mothers' response to the question on DFS, we limited the analysis to only mothers who answered the retrospective DFS question. We matched mothers with longer duration (more than 10 years) since first birth with those who started childbearing more recently, cross-classified by parity. **Table 4** shows the average treatment effect on mothers with longer duration (versus those with shorter duration) by parity. As expected, across all parity levels, the results show that mothers with longer duration have significantly higher estimated DFS compared to those with shorter durations.

Table 4 is about here

Sensitivity tests

We further explored the robustness of the above results to changes in matching methodology, a possible deviation from the conditional ignorability assumption, and potential rival explanations. **Table 5** displays the treatment effect on the treated women using two alternative matching methods, the nearest neighbor with replacement and the exact matching methods, compared to the full matching method. The nearest neighbor with replacement method selects the closest control unit based on propensity score distance to be paired with each treated unit. It could find a match to each treated unit but discarded a significant number of control units from the analysis. Exact matching is the most powerful method where only units with the same covariate values can be paired, and cases that lack a match with the same covariate values are dropped. The estimates from these alternative methods are compared with the full matching results in Table 5. While the exact matching method produced results quite comparable to the above-discussed estimates, the nearest neighbor method produced more substantial effects of motherhood and duration on DFS.

Table 5 is about here

Robustness test

The causal interpretation of the estimated treatment effects discussed above hinges on a strong identification assumption that the treatment assignment (motherhood) is as good as random assignment, conditional on selected matching covariates. However, despite the rich set of covariates included in our matching exercise, the estimates would be biased if we were to believe that motherhood is linked to some unobserved factors (such as family background), which would influence not only the likelihood of having a surviving child, but also women's family size preferences.

To test the robustness of our results to a possible deviation from the conditional ignorability assumption, we follow the sensitivity test proposed by Cinelli and Hazlett (2020) [implemented using the *sensemakr* package in R]. The method assesses how strong an unobserved confounder would have to be to change our main conclusion regarding the motherhood and duration effects discussed above. The test is based on an indicator called robustness value $RV_{q=1,\alpha=0.05}$, which measures the residual variance of both the treatment and the outcome explained by an unobserved confounder that brings the estimate to a range where it is no longer 'statistically different from 0'. A large robustness value indicates the robustness of the research conclusion to an unmeasured confounder. They also propose a method to assess the question of how strong a confounding factor would need to be, relative to the strength of observed covariates, to alter our conclusions.

Table 6 presents the sensitivity test results for the estimated treatment effect of motherhood and duration reported in Table 2 and Table 4, respectively. It reveals that the robustness values ($RV_{q=1,\alpha=0.05}$) that would reduce the point estimates of motherhood and duration effects to zero at a 5 percent significance level are 11.15% and 15.8%, respectively. This can be interpreted as follows: the strong positive motherhood and duration effects on desired family size would be maintained unless an unmeasured confounder explains at least 11 percent and 16 percent, respectively, of the residual variance of both the treatment and outcome variables.

Table 6 is about here

Although the robustness value ($RV_{q=1,\alpha=0.05}$) is a useful summary indicator of the robustness of the estimated effects in the presence of unobserved confounding, it is not easy to make judgments based on absolute values. Therefore, Cinelli and Hazlett (2020) proposed a relative measure of the strength of the hypothesized unobserved confounder based on contour plots. The contour plot visualizes how much variations in the treatment and outcome variables are explained by the unobserved confounder (Z), and the corresponding point estimates. In Figure 2 and Figure 3, these hypothetical residual shares of variations ($R^2{}_D \sim Z|X, and\ R^2{}_Y \sim Z|X$) are presented in the horizontal and vertical axis, respectively. The dashed red curves represent all possible combinations of the explained variations for which the ATT is equal to zero at a 5 percent level of significance (i.e., all possible combinations that produce *insignificant* motherhood and duration effects). One can also assess the robustness of the estimated effects (the ATT) to unobserved covariates, which is K times as strong as a given observed covariate in explaining the treatment and the outcome. In Figure 2 and Figure 3, we use women's education as a benchmark covariate and examine whether the estimated coefficients reduced to zero for unmeasured covariates up to four times as strong as the reference variable, which is whether a woman attained at least secondary education or not. The plot reveals that, at the 5% significance level, we rejected the zero-effect hypothesis given a confounder that is once, twice, three times, and even four times as strong as the effect of women's education. In conclusion, our finding is that having a surviving child and longer duration since first birth would be altered only in the presence of exceptionally strong covariates not included in the matching exercise and whose effects are more than four times the effect of female education.

Figures 2 and 3 about here

Discussion

This study examined whether the current framing of DFS questions for childless women and mothers in the DHS surveys measure the same construct – i.e., desired family size at the time of the survey. Although several studies have noted problems with the DFS measure, none has attempted to estimate how the different framing of the question for mothers and childless women could affect their reported DFS nor how duration since first birth could affect the reported DFS of mothers. Using matched samples of mothers and childless women from all publicly available DHS data from 37 sub-Saharan African countries, we show that *mothers*, who reported their DFS retrospectively, reported larger DFS that *childless women* who reported their DFS prospectively. For mothers, the further back in time they go in reporting their DFS, the higher their DFS. Matching women on all the covariates that have been used previously to explain differences in DFS does not remove this significant difference between the responses of *mothers* and *childless women* to the DFS question. These results, although not surprising because the questions measuring DFS are different for childless women and mothers (Casterline et al. 2009), they offer new insights into why the DFS measure has been seen as highly unstable. While the responses of childless women can be assumed to reflect their preferences at the time of the survey (Bongaarts 1990), the lack of temporality in the responses of mothers questions the use of such responses as a measure of current demand for childbearing. Combining responses to these two questions to get a composite measure of DFS is atypical of demographic measurements which recognize the importance of time and periodicity in measured outcomes.

For mothers, duration since first birth impacts their reported DFS. While it is understood that recall bias will generally increase with the duration respondents are asked to recall an incident or a preference (Tinker et al. 2013; Bradburn et al. 1987; Müller et al. 2022), first birth is a defining moment in women's lives, and it represents an event most women will more easily recall (Oakley 2016). In "going back to the time they did not have any children," it is possible that mothers are not merely recalling their individual preferences then but may be considering the prevailing childbearing norms in their community at the time of their first birth.

In previous attempts to reconcile perceived inconsistencies in women's reported DFS, researchers identified different concepts to explain how women's reported DFS could be biased; their reported DFS, it is argued, could be influenced by ex-post rationalization of their current number of children as wanted births and, hence, the fluidity of DFS over the life course (Bongaarts 1990; Rackin & Morgan 2018). Our results suggest that duration since first birth may be the mechanism that explains differences between mothers' reported DFS and their parity.

During the 1970's and 1980's, the World Fertility Survey utilized a single-question DFS measure. The question began as "How many children in all do you want to have?" and later changed to "If you could choose exactly…" as some respondents had difficulties comprehending the notion of desired number of children with the former question wording (Lightbourne &

Macdonald 1982). In designing the WFS questionnaire, the DFS question was intended to measure the respondent's personal DFS if economic constraints were not a factor, and the additional number of wanted children question was expected to align more closely with the number of children women desire, taking into account the real-world costs. The DHS 2-question measure was developed with the aim to reduce rationalization, particularly at higher parities. A study comparing the DHS and WFS questions concluded that the DHS 2-question measure is preferable given that it is less affected by post facto rationalization, as evidenced by weaker correlation between the two-question DFS measure and number of children (Goldman 1989). However, the study did not examine how altering the question for mothers could impact the face validity of the DFS measure. Decades later, there are no further studies that have examined the validity of the two-question DFS measure.

Limiting the analysis to only women with surviving children, we show that longer durations are associated with higher DFS across all parity levels. This result suggests that, rather than post facto rationalization, mothers may be fully considering and answering the DFS question as it was posed to them; thinking back to earlier periods characterized by societal norms that supported larger family sizes. Combining their responses and those of childless women to estimate DFS at the time of the survey is unsound conceptually and technically.

The dissatisfaction with the measurement of DFS in the DHS has led to the use of different tools, and in some contexts, to the development of new measures altogether. For instance, researchers based in Iran developed and validated a new tool composed of 27 items (Naghibi et al. 2019). How such tools can be further refined and abbreviated to develop new measures is an important area for further research. On the other hand, the Eurobarometer survey utilizes different formulations of questions to measure current desired family size. One of their questions asks, "And for you personally, what would be the ideal number of children you would like to have or would have liked to have had?" (Testa 2006). To some extent, this question appears to be a double-barreled question enquiring about both past and present desires in the same question. The concern is whether non-literate women can understand such a double-barreled question and how would researchers interpret responses to the question.

Even though the current measures of fertility desires used in DHS are seen as standard measures, other researchers (Machiyama et al. 2017) criticize the current DHS measures for its failure to capture the complexity and emotional components of the variable and hence recommended further research to add more items to measure emotional component of this attitudinal variable (Gibby & Luke 2019). These dissatisfactions with the measure among scholars, together with the limited confidence in the validity and reliability of the existing measure of DFS, calls for immediate action to develop and validate a better tool that is suitable for national surveys. The starting point will be to validate these results by clarifying with mothers what period they think of when answering the current DFS question and how different their reported DFS would be if they were to respond with reference to the current period. We have used the birth of the first child to represent the shortest interval mothers may consider in framing their DFS. Duration since first cohabitation could represents another milestone in women's lives that they could have considered when answering the DFS question. If women consider periods further away from the birth of their first child, then the differences observed here could be much larger. If women also differ in the period they think about, then such differences could also impact their reported DFS. Indeed, inconsistencies in reported DFS may simply result from differences in the period before their first birth that mothers consider when answering the question on DFS. The same mother could consider different periods at different interviews and therefore provide different responses.

We need to develop and test alternative formulations of questions that would measure DFS at the time of the survey. It will be important to define what is the construct we are seeking to measure and how best to operationalize it. Does DFS measure an individual attribute or a normative, community-level attribute? Should the woman respond bearing in mind her lived experiences and current realities or in her ideal situation or circumstance? Each of these clarifications will make it easier to formulate the question in a way that most women can understand and interpret the question the same way so that responses can be compared within and across samples of women.

Conclusion

Our analysis indicates significant differences in reported DFS between childless women and mothers. Among mothers, our results also show a strong impact of duration since first birth on reported DFS. This strong impact of duration implies that women might be referring to some period before their first birth when responding to the DFS question. The results suggest that the current DHS formulation of the DFS question for mothers, who constitute the vast majority of all women of reproductive age in sub-Saharan Africa, is not a good proxy for women's prevailing desired family size at the time of the survey. Therefore, new measures are needed to understand this important indicator of current demand for childbearing in a population.

Abbreviations

Average Treatment Effect on the Treated (ATT)

DFS: Desired family size

DHS: Demographic and Health Surveys

IFS: Ideal Family Size

SSA: Sub-Saharan Africa

Ethical considerations

Not applicable

Consent for Publication

Not applicable

Data availability

This work was produced through the analysis of DHS data, which are already publicly available.

All data described in the manuscript have been presented in the form of figures and tables.

Author Contributions

All authors made substantial contributions during the design, conduct, analysis, and report writing.

Disclosure statement

The authors declare that they have no conflicts of interest for this work.

References

Bacci, M. L. (2001). Comment: Desired family size and the future course of fertility. *Population and Development Review*, *27*, 282-289.

Bernheim, B. D., Braghieri, L., Martínez-Marquina, A., & Zuckerman, D. (2021). A theory of chosen preferences. *American Economic Review*, *111*(2), 720-54.

Bankole, A., & Westoff, C. F. (1995). *Childbearing attitudes and intentions* (Vol. 17). Macro International, Incorporated.

Bongaarts, J. (1990). The measurement of wanted fertility. *Population and Development Review***,** 487-506.

Bongaarts, J. (2001). Fertility and reproductive preferences in post-transitional societies. *Population and Development Review*, *27*, 260-281. http://www.jstor.org/stable/3115260

Bongaarts, J. (2011). Can family planning programs reduce high desired family size in sub-Saharan Africa? *International Perspectives on Sexual and Reproductive Health*, *37*(4), pp.209-216.

Bongaarts, J., & Casterline, J. (2013). Fertility Transition: Is sub-Saharan Africa different? *Population and Development Review*, *38*(Suppl 1), 153–168. https://doi.org/10.1111/j.1728-4457.2013.00557.x

Bongaarts, J. & Casterline, J.B., (2018). From fertility preferences to reproductive outcomes in the developing world. *Population and Development Review*, pp.793-809

Bradburn, N. M., Rips, L. J. & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236***,** 157-161.

Caldwell, J. C., & Caldwell, P. (1978). The achieved small family: early fertility transition in an African city. *Studies in Family Planning*, *9*(1), 2–18. https://doi.org/10.2307/1966293

Caldwell, J. C., & Caldwell, P. (1987). The cultural context of high fertility in sub-Saharan Africa. *Population and Development Review*, 409-437.Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *82*(1), 39-67.

Dharmalingam A. (1996). The social context of family size preferences and fertility behaviour in a south Indian village. *Genus*, *52*(1-2), 83–103.

Gibby, A. L. & Luke, N. (2019). Exploring multiple dimensions of young women's fertility preferences in Malawi. *Maternal and Child Health Journal*, 23**,** 1508-1515.

Goldman, N., et al. (1989). Peru Experimental Study. An Evaluation of Fertility and Child Health Information. Princeton, New Jersey, USA, Office of Population Research and Institute for Resource Development/Macro Systems.

Gray, E., Evans, A., & Reimondos, A. (2013). Childbearing desires of childless men and women: When are goals adjusted? *Advances in Life Course Research*, *18*(2), 141-149.

Hagewen, K. J., & Morgan, S. P. (2005). Intended and ideal family size in the United States, 1970–2002. *Population and Development Review*, *31*(3), 507-527.

Hayford, S. R. (2009). The evolution of fertility expectations over the life course. *Demography*, *46*(4), 765-783.

Heiland, F., Prskawetz, A., & Sanderson, W. C. (2008). Are individuals' desired family sizes stable? Evidence from West German panel data: La taille de famille désirée est-elle stable? Analyse de données de panel en Allemagne de l'Ouest. *European Journal of Population/Revue Européenne de Démographie*, *24*, 129-156.

Huber, M. (2020). Mediation analysis. *Handbook of labor, human resources and population economics*, 1-38.

Iacovou, M., & Tavares, L. P. (2011). Yearning, learning, and conceding: Reasons men and women change their childbearing intentions. *Population and Development Review*, *37*(1), 89-123.

Isiugo-Abanihe, U. C. (1994a). Reproductive motivation and family-size preferences among Nigerian men. *Studies in Family Planning*, *25*(3), 149–161. https://doi.org/10.2307/2137941

Isiugo-Abanihe U. C. (1994b). The socio-cultural context of high fertility among Igbo women. *International Sociology: Journal of the International Sociological Association*, *9*(2), 237–258. https://doi.org/10.1177/026858094009002008

Kebede, E., Striessnig, E., & Goujon, A. (2022). The relative importance of women's education on fertility desires in sub-Saharan Africa: A multilevel analysis. *Population Studies*, *76*(1), 137-156. https://doi.org/10.1080/00324728.2021.1892170

Knodel, J., & Prachuabmoh, V. (1973). Desired family size in Thailand: Are the responses meaningful? *Demography*, *10*(4), 619–637. https://doi.org/10.2307/2060887

Kodzi, I. A., Casterline, J. B., & Aglobitse, P. (2010). The time dynamics of individual fertility preferences among rural Ghanaian women. *Studies in Family Planning*, *41*(1), 45-54.

Lee, R. D. (1980). Aiming at a moving target: Period fertility and changing reproductive goals. *Population Studies*, *34*(2), 205-226. (Page 205).

Lightbourne, R. E., & Mac Donald, A. L. (1982). Family size preferences. International Statistical Institute.

Lockwood, M. (1995). Structure and behavior in the social demography of Africa. *Population and Development Review*, 1-32.

Machiyama, K., Casterline, J. B., Mumah, J. N., Huda, F. A., Obare, F., Odwe, G., ... & Cleland, J. (2017). Reasons for unmet need for family planning, with attention to the measurement of fertility preferences: protocol for a multi-site cohort study. *Reproductive Health, 14*(1), 1-11.

McCarthy, J., & Oni, G. A. (1987). Desired family size and its determinants among urban Nigerian women: A two-stage analysis. *Demography, 24*(2), 279–290. https://doi.org/10.2307/2061635

Müller, M. W., Hamory, J., Johnson-Hanks, J., & Miguel, E. (2022). The illusion of stable fertility preferences. *Population Studies*, *76*(2), 169-189.

Naghibi, S. A., Khazaee-Pool, M. & Moosazadeh, M. (2019). Psychometric properties of fertility desire scale (FDS) developed for Iranian parents. *BMC Public Health, 19***,** 1-11.

Oakley, A. (2016). A small sociology of maternal memory. *The Sociological Review, 64***,** 533-549.

Price, N. L., & Hawkins, K. (2007). A conceptual framework for the social analysis of reproductive health. *Journal of Health, Population, and Nutrition*, *25*(1), 24.

Rackin, H. M. & Morgan, S. P. (2018). Prospective versus retrospective measurement of unwanted fertility: Strengths, weaknesses, and inconsistencies assessed for a cohort of US women. *Demographic Research, 39***,** 61.

Rodrigues, P., Manlosa, A. O., Fischer, J., Schultner, J., Hanspach, J., Senbeta, F., & Dorresteijn, I. (2022). The role of perceptions and social norms in shaping women's fertility preferences: a case study from Ethiopia. *Sustainability Science, 17*(6), 2473-2488.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322-331.

Rutstein, S. O., & Rojas, G. (2006). Guide to DHS statistics. *Calverton, MD: ORC Macro*, *38*, 78.

Ryder, N. B. (1973). A critique of the national fertility study. *Demography*, *10*(4), 495–506. https://doi.org/10.2307/2060877

Ryder, N. B. (1980*). Components of temporal variations in American fertility*. In R.W.Hiorns (Ed), *Demographic Patterns in Developed Societies*. London, UK: Taylor & Francis.

Sennott, C., & Yeatman, S. (2012). Stability and change in fertility preferences among young women in Malawi. *International Perspectives on Sexual and Reproductive Health*, *38*(1), 34–42. https://doi.org/10.1363/3803412

Testa, M.R. (2006). Childbearing preferences and family issues in Europe. Special Eurobarometer 253/Wave 65.1. *Vienna Yearbook of Population Research*, *5*, pp.357-379.

Tinker, S. C., Gibbs, C., Strickland, M. J., Devine, O. J., Crider, K. S., Werler, M. M., ... & National Birth Defects Prevention Study. (2013). Impact of time to maternal interview on interview responses in the National Birth Defects Prevention Study. *American Journal of Epidemiology*, *177*(11), 1225-1235.

Yeatman, S., Sennott, C., & Culpepper, S. (2013). Young women's dynamic family size preferences in the context of transitioning fertility. *Demography*, *50*(5), 1715-1737.

List of tables

List of Figures

Table 1: Distribution of key variables by childlessness status and duration since first birth for women (15-49) in 37 sub-Saharan African countries, DHS 2003-2021

	All women		Women with surviving child(ren)	
	Childless*	With surviving child(ren)	0-10 years since 1st birth	10+ years since 1st birth
Desired Family Size	4.20 (2.27)	5.39 (2.75)	4.83 (2.45)	5.90 (2.91)
Years since first birth	-	12.5 (8.42)	5.14 (3.04)	19.0 (5.94)
Birth Cohort				
Before 1965	.008	.080	.001	.149
1965-1974	.030	.237	.040	.405
1975-1984	.134	.366	.340	.372
1985-1994	.476	.272	.512	.072
1995-	.351	.045	.105	-
Number of living children				
1-2		.418	.723	.147
3-4		.313	.258	.362
5+		.269	.018	.491
Marital Status				
Never In union	.785	.066	.111	.026
Currently married/live with partner	.189	.822	.811	.832
Formerly married	.026	.112	.078	.142
Area of residence				
urban	.460	.336	.352	.321
rural	.540	.664	.648	.679
Education				
None	.165	.394	.327	.454
Primary	.309	.341	.343	.340
Some Secondary or more	.526	.264	.330	.206
Household wealth index				
Poorest (q1)	.135	.213	.206	.220
Poorer (q2)	.149	.200	.198	.202
Middle (q3)	.176	.196	.193	.199
Richer (q4)	.213	.194	.195	.193
Richest (q5)	.326	.197	.209	.186
Age at interview date	20.0 (5.83)	31.7 (8.47)	25.0 (4.85)	37.7 (6.19)
N	281,215	693,059	326,316	366,743

Notes: *Childless women are those with no surviving child on the date of the interview.
For the desired family size, years since first birth, and age variables, means and standard deviations (in parenthesis) are reported; for the categorical variables, the numbers represent relative frequencies.

Table 2. Average Treatment effect on the Treated (ATT): the mean causal effect of motherhood (answering retrospective DFS question) on reported DFS of mothers vs childless women, DHS 2003-21

	Desired Family Size		
	All Women	Women (15-19)	Women (20-49)
Average effect of responding as a mother (ATT)	.619*** (.005)	-.180** (.012)	.684*** (.007)
Constant	6.59*** (.074)	5.43*** (.137)	6.61*** (.047)
Matched Obs.	974,274	212,742	761,532
Matched treated- units	693,059	37,821	655,238
Number of countries	37	37	37

Notes: The table reports the estimated impact of motherhood (i.e., answering the retrospective DFS question) on reported DFS of mothers versus childless women. The average treatment effect was estimated using 'Full matching' method. Matching weights with robust standard errors and covariate adjustments were used in the estimate.

Covariate balances are based on the following pre-treatment variables: birth cohort, marital status, area of residence, women's educational attainment, household wealth quintile, survey year fixed effect, and country fixed effects.

***, **, and * represent significance at 1%, 5% and 10% levels, respectively. Terms in parentheses are robust standard errors.

Table 3. Average Treatment effect on the Treated (ATT): the mean causal effect of motherhood (answering a retrospective DFS question) on reported DFS of mothers with 1-2 children vs childless women

	Desired Family Size	
	All mothers with 1-2 children versus childless women	Mothers with 1-2 children and duration >10 versus childless women
Average effect of being a mother (ATT)	.032*** (.005)	.378*** (.010)
Constant	6.00*** (.075)	6.26*** (.223)
Matched Obs.	570,880	335,047
Matched treated- units	289,665	53,832
Number of countries	37	37

Notes: The table reports the estimated impact of motherhood (answering the retrospective DFS question) on reported DFS of mothers with 1-2 surviving child(ren) versus childless women. The average treatment effect was estimated using 'Full matching' method. Matching weights with robust standard errors and covariate adjustments were used in the estimate.

Covariate balances are based on the following pre-treatment variables: birth cohort, marital status, area of residence, women's educational attainment, household wealth quintile, survey year fixed effect, and country fixed effects.

***, **, and * represent significance at 1%, 5% and 10% levels, respectively. Terms in parentheses are robust standard errors.

Table 4. Average Treatment effect on the Treated (ATT): the mean causal effect of duration on reported DFS of mothers with long versus short duration since first birth by parity, DHS 2003-2021

	Number of Surviving Children			
	All	1-2	3-4	5+
Mean treatment effect on the treated (ATT)	.833***	.321***	.223***	.391***
	(.005)	(.009)	(.010)	(.012)
Constant	6.96***	6.34***	7.40***	8.10***
	(.144)	(.265)	(.208)	(.132)
Duration (in years) for control vs treated women	0-10 vs 10+ years	0-10 vs 10+ years	0-15 vs 15+ years	0-20 vs 20+ years
Matched Obs.	693,059	289,665	217,240	186,154
Matched treated- units	366,743	53,832	68,004	88,043
Number of countries	37	37	37	37

Notes: The table reports the average treatment effect of duration since first birth among mothers with long durations (Treated) versus short durations (Control), cross-classified by the number of living children (parity). The average treatment effect was estimated using 'Full matching' method. Matching weights with robust standard errors and covariate adjustments were used in the estimate. Short and long durations are adjusted for each parity to accommodate the time required to achieve the respective parity.
Covariate balances are based on the following pre-treatment variables: birth cohort, marital status, area of residence, women's educational attainment, household wealth quintile, survey year fixed effect, and country fixed effects.
***, **, and * represent significance at 1%, 5% and 10% levels, respectively. Terms in parentheses are robust standard errors.

Table 5. Sensitivity tests: estimated impacts of motherhood/duration using alternative matching methods

Method	All women (Mothers vs childless women)		Mothers with 1-2 children (Duration 10+ vs 0-10 years)	
	ATT	Treated vs (control) units	ATT	Treated vs (control) units
Nearest Neighbor with replacement	.707*** (.003)	693,059 (13,006)	.537*** (.014)	53,832 (1,451)
Exact Matching	.597*** (.005)	110,073 (87,338)	.290*** (.025)	53,832 (194,482)
Full Matching	.619*** (.005)	693,059 (281,215)	.321*** (.009)	53,832 (235,833)

Table 6. Robustness test: estimated robustness values for the ATT using the Cinelli and Hazlett (2020) method

Treatment	*Outcome: Desired Family Size*		
	Estimate (SE)	$RV_{q=1}$	$RV_{q=1,\alpha=0.05}$
Mothers vs childless women	.639*** (.005)	11.17 %	11.15 %
Duration since first birth (0-10 vs 10+ years)	.820*** (.005)	15.6 %	15.8 %

Appendix Table 1: Number of women included in the study by country and childlessness status

Country	DHS survey year(s)	Childless women	Has Surviving child(ren)
Angola	2015-16	3,363	11,015
Benin	2001, 2006, 2011-12, 2017	12,123	35,894
Burkina Faso	2003, 2010	6,933	21,356
Burundi	2010-11, 2016-17	9,390	16,392
Cameroon	2004, 2011, 2018-19	11,340	24,344
Chad	2004, 2014-15	4,390	14,524
Comoros	2012	2,222	2,527
Congo	2011-12	4016	16286
Cote d'Ivoire	2011-12	2,522	6,703
DR Congo	2007, 2013-14	7,526	19,320
Eswatini	2006-07	1594	3358
Ethiopia	2005, 2011, 2016	15,283	26,143
Gabon	2012	2029	5,887
Gambia	2013, 2019-20	7034	13290
Ghana	2003, 2008, 2014	6,293	13,341
Guinea	2005	6664	17,722
Kenya	2003, 2008, 2014	8207	21,684
Lesotho	2004-05, 2009-10, 2014	7098	14188
Liberia	2006-07, 2013, 2019-20	4863	18,081
Madagascar	2003-04, 2008-09	6451	17,279
Malawi	2004-05, 2010, 2015	13448	44,607
Mali	2006, 2012, 2018	7127	24,406
Mauritania	2019, 2020-21	4950	7789
Mozambique	2003-04, 2011	6407	19,430
Namibia	2006-07, 2013	6012	12,751
Niger	2006, 2012	3871	14,375
Nigeria	2003, 2008, 2013, 2018	34,904	77,944
Rwanda	2005, 2010, 2014, 2019	19732	32,483
Senegal	2005, 2010	8613	15,195
Sierra Leone	2008, 2013, 2019	10398	27878
South Africa	2016	2456	6,029
Sao Tome & P	2008-09	596	1978
Tanzania	2004, 2009, 2015	9182	23,460
Togo	2013-14	2563	6,659
Uganda	2006, 2011	4450	12,198
Zambia	2007, 2013-14, 2018-19	9217	26,702
Zimbabwe	2005-06, 2010-11, 2015	7948	19,841
N		281,215	693,059

Appendix Table 2: Summary of covariate balance before and after Full matching exercise for selected covariates

	Pre-matching			Post-matching		
Covariate	Mean of treated units*	Mean of control units	Standardized mean differences	Mean of treated units*	Mean of control units	Standardized mean differences
Birth Cohort						
before 1965	.0801	.0082	.2648	.0801	.0804	-.001
1965-74	.2374	.0306	.4861	.2374	0.237	.001
1975-84	.3658	.1341	.4811	.3658	.3658	-.0001
1985-94	.2716	.4762	-.4601	.2716	.2717	-.0002
1995-	.0452	.3510	-1.4728	.0452	.0452	-.0002
Marital Status						
Never in union	.0661	.7845	-2.892	.0661	.0662	.0000
Currently in union	.8221	.1892	1.654	.8221	.8222	-.0002
Formerly married	.1118	.0263	.2714	.1118	.1118	.0002
Area of residence						
Urban	.3360	.460	-.264	.3360	.3358	-.0003
Rural	.6644	.5396	.2643	.6644	.6642	.0003
Education						
None	.3944	.1652	.4691	.3944	.3952	-.0017
Primary	.3412	.3092	.0675	.3412	.3413	-.0001
Secondary or more	.2644	.5256	-.5924	.2644	.2644	-.0000
Household Wealth						
Poorest	.2133	.1353	.1904	.2133	.2131	.0005
Poorer	.2000	.1491	.1271	.2133	.2131	.0005
Middle	.1961	.1758	.0511	.1961	.1959	.0005
Richer	.1940	.2134	-.0491	.2133	.2131	.0005
Richest	.1966	.3263	-.3264	.194	.1941	-.0004
Survey Year						
2003-05	.0907	.0888	.0064	.0907	.0904	.0008
2005-09	.2185	.2249	-.0155	.2185	.2187	-.0006
2010-14	.4082	.3931	.0307	.4082	.4082	-.0001
2015-21	.2826	.2932	-.0234	.0907	.0904	.0008

*Treated units = mothers and Control units = childless women.

Figure 1. Estimated effect of motherhood (answering a retrospective DFS question) on desired family size of women with surviving child(ren) vs childless women by country, DHS 2003-2021

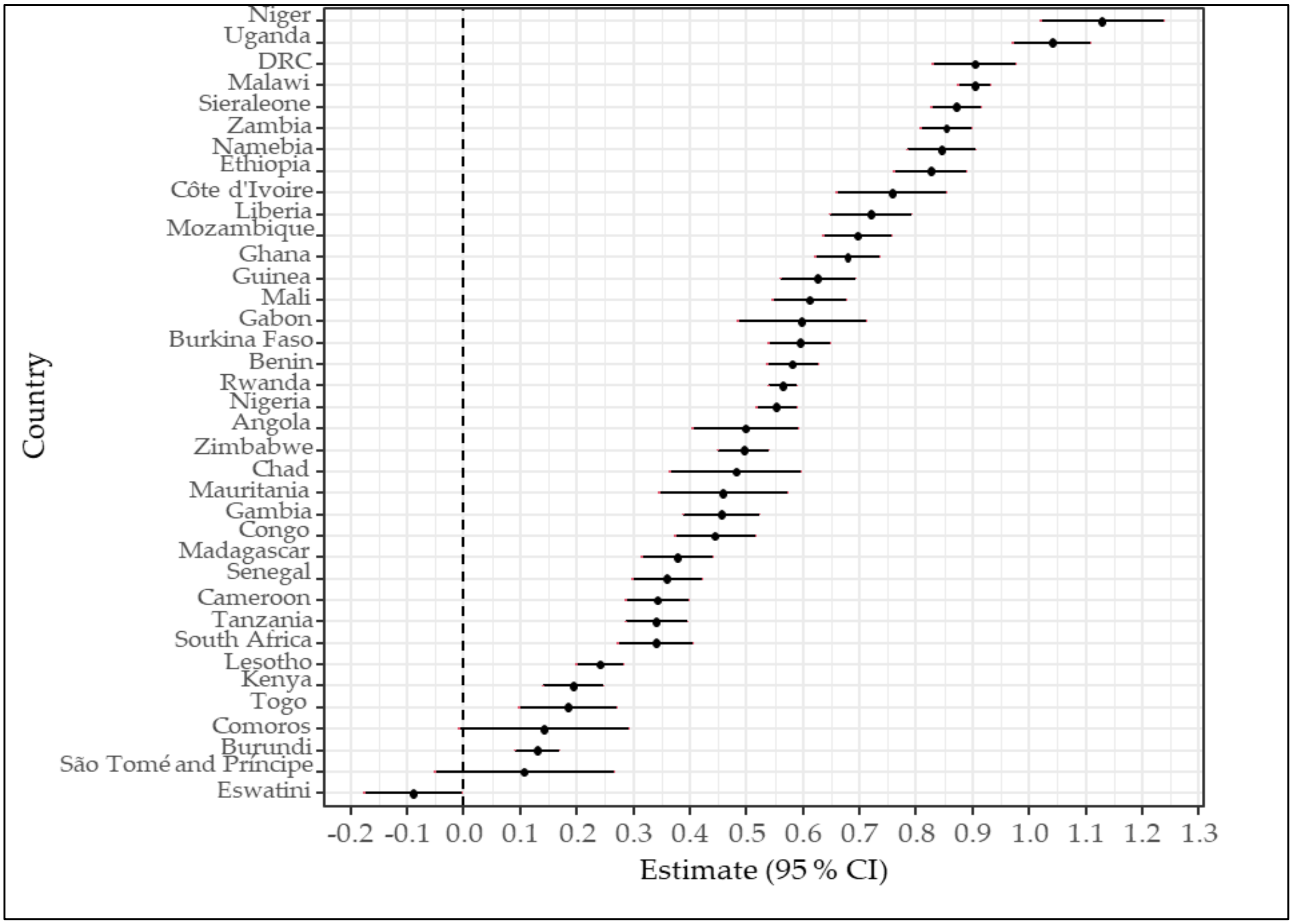


Figure 2. Robustness of the estimated average treatment effect of motherhood under the presence of unobserved confounder (mothers vs childless women)

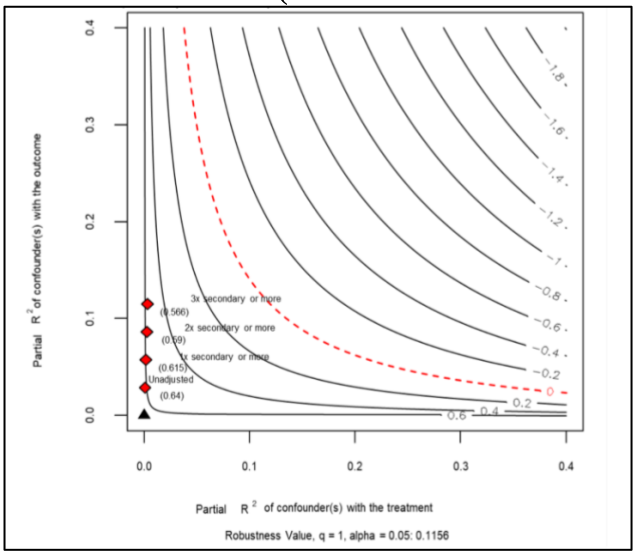
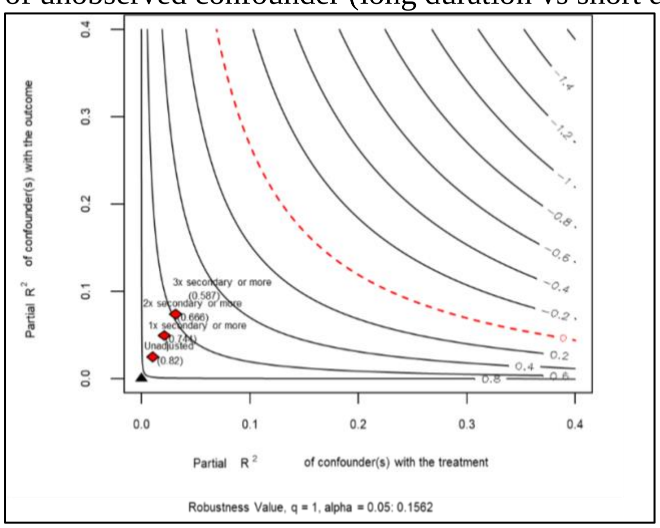


Figure 3. Robustness of the estimated average treatment effect of duration since first birth under the presence of unobserved confounder (long duration vs short duration mothers)



Legend for Figures 2 & 3:

▲ Our estimates where no unobserved confounder was assumed.

◆ Hypothetical estimates in the presence of unobserved confounder that is 1x, 2x, and 3x stronger than the effect of female education.

— Contours of partial R^2 of confounder with treatment and outcome variables when the unobserved confounder is 1x, 2x, and 3 times stronger than the effect of female education on DFS.

---- The point at which our estimated treatment effect become zero in the presence of an unobserved variable

Appendix Figure 1. Estimated effect of childbearing on DFS (Desired Family Size) of women with surviving child(ren) by age of women at survey year for each country