Family Size of Children and Women during the Demographic Transition

David Lam
University of Michigan
davidl@umich.edu

Letícia Marteleto
University of Texas
marteleto@prc.utexas.edu

Prepared for presentation at XXVII IUSSP International Population Conference
Busan, Korea
August 26-31, 2013

Date of Draft: July 14, 2013

David Lam is Professor of Economics and Research Professor in the Population Studies Center at the University of Michigan. Letícia Marteleto is Assistant Professor of Sociology and Research Associated in the Population Research Center at the University of Texas-Austin. Excellent research assistance was provided by Kendra Goostrey. Support for this research was provided by the U.S. National Institutes of Health (NICHD Grant Number R01HD045581.

ABSTRACT

This paper analyzes links between declines in the family size of women and declines in the family size of children during the demographic transition. We extend Preston’s (1976) model in two ways. First, we derive the relationship between the variance of women’s family size and children’s family size, a relationship that has important implications for inequality in children’s family size. Second, we analyze family size from the perspective of children of a given age rather than women of a given age. We apply the framework to 310 data sets from the IPUMS-International census project and the Demographic and Health Surveys, representing 101 countries. Consistent with Preston’s conjecture, we find that mean family size of children tends to fall more slowly than mean family size of women as fertility declines. The increase in resources per child is 5%-20% smaller than it would be if children’s family size decreased at the same rate as women’s family size. We show that inequality in children’s family size increases substantially as fertility declines, the result of increasing skewness in women’s family size.
Family Size of Children and Women during the Demographic Transition

INTRODUCTION

Rapid fertility decline in most of the developing world has been accompanied by rapid declines in the family size of school-aged children. These declines in children’s family size may in turn have led to increased resources for children at the household level. While a link between falling fertility and falling family size for children may seem inevitable, the actual dynamics of this link depend on the change in the variance of fertility. This fundamental demographic point, elegantly demonstrated by Preston (1976), has generally been neglected in discussions of changing family size in developing countries. The goal of this paper is to expand on Preston’s result and look empirically at the relationship between declining family size for women and declining family size for children in a number of developing countries throughout the demographic transition.

We begin with a brief review of the literature looking at links between family size and children’s schooling in developing countries. We then discuss Preston’s model, extending it in two directions. First, we derive the relationship between the variance of children’s family size and the variance of women’s family size. Second, we look at family size from the perspective of children of a given age, rather than women of a given age. As we will show, this is a more complicated problem than the one analyzed by Preston, but nonetheless leads to some simple analytical expressions that can be applied to micro-data. We then analyze the relationship between the family size of women and the family size of children in a large number of countries, using census data from the IPUMS-International project (Minnesota Population Center, 2010) and survey data from the Demographic and Health Surveys. Consistent with Preston’s conjecture, we find that the mean family size of children tends to fall more slowly than the mean
family size of women during the demographic transition. The differences in the rates of decline are relatively small, however. While mean family size of children tends to fall only slightly more slowly than the mean family size of women during the demographic transition, there is a significant divergence in the variance of family size between women and children. The standard deviation of women’s family size falls at roughly the same rate as the mean, but the standard deviation of children’s family size falls very little as fertility declines. We show that this has important implications for inequality, with inequality in children’s family size increasing substantially as fertility declines.

**RESEARCH ON FAMILY SIZE AND SCHOOLING**

Researchers have long been interested in the influence of family size on children’s schooling. The dilution of resources framework theorizes that resources are diluted within families that have more children, and therefore the larger the family, the fewer the resources available per child, and the worse the outcomes for each child (Blake 1981). The early empirical evidence on developing countries has produced mixed results, however (Lloyd, 2005). Several of the early empirical studies on the association of family size and schooling in developing countries have confirmed Blake’s prediction that children from large families attain less schooling than children with fewer siblings (Ahn, Knodel, Lam, and Friedman 1998; Knodel and Wongsith 1991; Lam and Marteleto, 2005; Parish and Willis 1993, Patrinos and Psacharopoulos 1997, Psacharopoulos and Arriagada, 1989; Pong 1997; Post and Pong 1998; Post 2002; Shavit and Pierce 1991). Other studies, however, have found a positive association between family size and education (Chernichovsky 1985; Hossain 1988; King et al 1986; Mueller 1984). This heterogeneity in the association between family size and children’s education potentially reflects the different demographic, economic and social conditions parents face when deciding family size and the amount to invest in each child’s education. It can also reflect an important issue of
endogeneity—parents who highly value children’s education may decide to have fewer children in the first place, which could explain the association found in these early studies.

A new wave of research addressing the endogeneity between family size and schooling in developing countries has also yielded mixed results. A negative effect of sibship size was found for the schooling of Chinese children (Li, Zhang and Zhu 2008). Maralani (2008) reports positive effects for older cohorts and negative effects for younger cohorts in Indonesia (2008). A recent study also reported mixed results—a positive effect of family size on schooling among pre-demographic transition adolescents that disappears for young cohorts of adolescents and in the most developed parts of Brazil (Marteleto and Souza 2012). Thus, both the early and more recent evidence suggests that the effect of family size on adolescents’ schooling is not homogenous over time or across regions or cohorts within the same country.

The purpose of this paper is not to provide any new evidence on the impact of family size on children’s outcomes, but to analyze how the family size of school-aged children changes during the demographic transition. Given the rapid declines in fertility in most developing countries in recent decades, coupled with the potential importance of family size for children’s outcomes, we find it surprising that there has not been more systematic analysis of the dynamics of family size from a child’s perspective during these fertility declines.

In previous work we analyzed how changes in family size are related to changes in cohort size during the demographic transition (Lam and Marteleto 2005, 2008). We point out that during part of the demographic transition there is a period in which family size is falling while cohort size is increasing. Not until fertility decline catches up with population momentum do both family size and cohort size begin to decline. In this paper we focus on another aspect of changes in family size that we believe has received inadequate attention—the relationship between the family size of women and the family size of children. As Preston pointed out more
than three decades ago, the family size of children can change at a different rate than the family size of women (Preston, 1976). Given the dramatic changes in both mean fertility and the distribution of fertility during the rapid fertility declines in developing countries, it is instructive to analyze how those changes are related to changes in the family size of children.

FAMILY SIZE OF CHILDREN AND FAMILY SIZE OF WOMEN

Preston (1976) derived expressions for the mean family size for women of a given age, say 45-49, and contrasted this with the mean family size for the children of those women. Restating his result, if \( \bar{W} \) is the mean family size for women and \( \bar{C} \) is the mean family size of their children, Preston derived the surprisingly simple result that

\[
\bar{C} = \bar{W} + \left( \frac{\sigma_W}{\bar{W}} \right),
\]

where \( \sigma_W \) is the standard deviation of family size for women.\(^1\) The proof of the result in Equation (1) uses the fact that the proportion of children with family size \( s \), \( f_c(s) \), is a simple function of the number of women with family size \( s \), \( f_w(s) \), as follows:

\[
f_c(s) = \frac{f_w(s)\bar{s}}{\int f_w(s)ds} \frac{f_w(s)s}{\bar{s}_w}, \quad (2)
\]

Using (2) to calculate \( \bar{C} \) leads directly to Equation (1). Inspection of Equation (2) also demonstrates that there is a single crossing of the densities of children’s family size and women’s family size, with the proportion of children with family size \( s \) being exactly equal to the proportion of women with family size \( s \) at the mean of women’s family size, \( \bar{W} \). For all \( s < \bar{W} \), \( f_c(s) < f_w(s) \) and for all \( s > \bar{W} \), \( f_c(s) > f_w(s) \). We illustrate this below with empirical examples from several countries.

A convenient restatement of Preston’s result in Equation (1) is the following:

---

\(^1\) As in Preston, we use “family size” to refer only to children, omitting parents. A woman’s family size is the number of children born (or surviving) to that woman. A child’s family size is the number of children with the same mother in that child’s family (also known as sibship size).
where $CV_w$ is the coefficient of variation of family size for women. Equation (3) says that the mean family size of children will be greater than the mean family size of women by a multiple that is equal to 1 plus the squared coefficient of variation in women’s family size. In other words, mean family size for children will always be greater than mean family size for women as long as there is any variation in women’s fertility.

We can simplify Equation (3) to

$$\bar{s}_c = \bar{s}_w \left(1 + \frac{\sigma_w^2}{\bar{s}_w^2}\right) = \bar{s}_w \left(1 + CV_w^2\right),$$

where $\bar{s}_c$ is the mean family size of children and $\bar{s}_w$ is the mean family size of women. Equation (3) tells us that the mean family size of children will always be greater than the mean family size of women by a multiple that is equal to 1 plus the squared coefficient of variation in women’s family size. In other words, mean family size for children will always be greater than mean family size for women as long as there is any variation in women’s fertility.

We can simplify Equation (3) to $\bar{s}_c = \bar{s}_w k$, where $k = 1 + CV_w^2$ can be interpreted as the multiplier that translates mean family size for women into the mean family size of their children. Below we will use $k$, the ratio of mean children’s family size to mean women’s family size, as one main focus of our empirical analysis. To examine changes over time it will be useful to look at proportional changes by taking the natural logarithm, $\ln(\bar{s}_c) = \ln(\bar{s}_w) + \ln(k)$ and taking the derivative with respect to time:

$$\frac{\partial \ln \bar{s}_c}{\partial t} = \frac{\partial \ln \bar{s}_w}{\partial t} + \frac{\partial \ln k}{\partial t}$$

Equation (4) tells us, for example, that a 10% reduction in mean fertility ($\bar{s}_w$) will lead to a 10% reduction in children’s mean family size if and only if $k$ remains constant. Changes in $k$, resulting from changes in the coefficient of variation of women’s family size, will amplify or dampen the impact of the change in women’s family size. If, for example, $k$ were to increase by 10% at the same time that fertility fell by 10%, it would completely offset the reduction in women’s family size, leaving mean family size of children constant.

Looking at historical data from the United States, Preston observed that the variance in fertility fell more rapidly or less rapidly than mean fertility during different stages of fertility decline (1976). In terms of Equation (4) this means that $k$ was not constant, causing movements
in children’s family size that sometimes differed substantially from movements in fertility. For example, Preston found that the average family size of women fell by 53% between 1890 and 1950, while the average family size of children only fell by 37% over the same period. The reason for the discrepancy is that the standard deviation in women’s fertility fell more slowly than mean fertility over this period. The coefficient of variation, and therefore the multiplier $k$, increased, partially offsetting the decline in women’s mean family size. Preston suggested that this might be a typical pattern during fertility decline, with fertility falling faster among some groups than others. The resulting increase in dispersion would cause children’s mean family size to fall more slowly than women’s mean family size during the demographic transition. As Preston put it, “These patterns are a disconcerting precedent for those concerned with issues of population quality in less developed countries; the pace of reductions in family size for children can be expected to lag behind that for women in the process of fertility transition” (1976: 108).

Surprisingly, there have been very few studies applying Preston’s result to developing countries. One of the few is an application to Cambodia by Neupert (2005), who shows that the mean family size of children was about 25% higher than the mean family size of women aged 45-49 in 1998. We will see below that this 25% difference is quite typical of countries with high levels of fertility. We will also see below that Preston’s conjecture appears to be consistent with the trend we see in most countries during the demographic transition. The mean family size of children tends to fall at a somewhat slower rate than the mean family size of women. While this pattern is not universal, the evidence from comparing countries and from looking at changes in countries over time is that mean children’s family size falls 10-20% more slowly than mean women’s family size.

Another of Preston’s results using U.S. data was a difference in the distribution of family size for whites versus nonwhites. Preston found that the difference between the family size of
women and the family size of children was much greater for nonwhites than for whites, a result of the fact that non-white women had higher mean-adjusted dispersion in fertility than white women. In terms of Equation (3), the coefficient of variation in women’s family size in 1970 was 1.0 for nonwhites and 0.76 for whites. An implication of this difference is that the mean family size of nonwhite children was 50% higher than the mean family size of white children, even though the mean family size of nonwhite women was only 19% higher than the mean family size of white women. This could have important implications for developing countries when we compare differences by region, education, ethnicity, or income. We might expect to see, for example, that the rural-urban gap in family size of children is larger than the rural-urban gap in family size of women, potentially exacerbating the rural-urban schooling gap. We leave the issue of comparing subgroups within a population to future work, but note that a similar phenomenon can occur when comparing countries. As we will show below, two countries may have similar mean family size for women but have quite different mean family size for children. In particular, a country with higher inequality in fertility will have a higher mean family size of children than a country with the same mean fertility but with lower inequality in fertility.

*The variance in family size of women and family size of children*

If family size has a significant impact on children’s outcomes then we may be interested in what happens to the variance in children’s family size as well as the mean. One important question would be whether the variance in children’s family size (or some mean-adjusted measure of dispersion) tends to increase or decrease during the demographic transition. We might also be interested in how the variance in children’s family size differs across population subgroups or across countries. Building on Preston’s simple result for mean family size, it is interesting to consider whether there is some analogous relationship between the variance of family size for women and the variance of family size for children. Not surprisingly, just as the
relationship between the mean of children’s family size and the mean of women’s family size
depends on the variance (and thus the second moment) of women’s family size, the relationship
between the variance of children’s family size and the variance of women’s family size depends
on the skewness (and thus the third moment) of women’s family size. Specifically,
\[
\sigma_c^2 = \sigma_w^2 \left[ 1 + CV_w (S_w - CV_w) \right],
\]
where \(S_w\) is the skewness in women’s family size.\(^2\)

The result in Equation (5) is somewhat less intuitive than Preston’s result for the mean, but
leads to some important points. We will see below that skewness tends to be slightly negative
when fertility is high, becoming increasingly positive as fertility declines. This implies that the
variance in children’s family size is smaller than the variance in women’s family size in early
stages of the demographic transition, a pattern we document below. While the mean of
children’s family size must be greater than the mean of women’s family size, this is not the case
for the variance (or standard deviation). When fertility declines the skewness of women’s family
size tends to increase, eventually surpassing the coefficient of variation (which, as we will see
below, increases slightly as fertility declines). This tends to cause the variance in children’s
family size to exceed the variance in women’s family size in low fertility populations.

Since we will show trends in both the mean and standard deviation of children’s family size,
we will also be able to make statements about what happens to mean-adjusted inequality of
children’s family size, as measured by the coefficient of variation and by other measures. We
will see that inequality in children’s family size tends to rise as fertility declines, a potentially
important factor in understanding inequality in children’s outcomes.

\(^2\) For a random variable \(x\), skewness is defined as
\[
S = \sum f(x) \left( x - \bar{x} \right)^3 / \sigma^3.
\]
\(S=0\) implies a symmetric
distribution. \(S>0\) implies that the right tail falls off less rapidly than the left tail. \(S<0\) implies that the left
tail falls off less rapidly than the right tail. For women’s family size we tend to observe \(S>0\).
We take the square root of Equation 4, and denote the square root of the term in brackets by $m$, so that $m$ is the ratio of the standard deviation in children’s family size to the standard deviation in women’s family size. Taking logs and taking the derivative with respect to time, 

$$\frac{\partial \ln \sigma_c}{\partial t} = \frac{\partial \ln \sigma_w}{\partial t} + \frac{\partial \ln m}{\partial t}. \quad (6)$$

As in the case of the mean, this provides a useful way to think about the change in the standard deviation of children’s family size. We will see that while the standard deviation in women’s family size falls along with the mean, the multiplier $m$ tends to increase when fertility declines, the result of increasing skewness in the distribution of women’s family size. The net result is that the variance in children’s family size falls very little as fertility falls, even though the mean of children’s family size is falling.

Now consider the change in the coefficient of variation (CV) of children’s family size. Since the CV is a standard mean adjusted measure of inequality, this will provide some insights about how inequality in children’s family size changes as fertility declines. Combining our results from above:

$$\frac{\partial \ln CV_c}{\partial t} = \left(\frac{\partial \ln \sigma_w}{\partial t} + \frac{\partial \ln m}{\partial t}\right) - \left(\frac{\partial \ln \bar{w}}{\partial t} + \frac{\partial \ln k}{\partial t}\right) \quad (7)$$

$$= \frac{\partial \ln CV_w}{\partial t} + \frac{\partial \ln m}{\partial t} - \frac{\partial \ln k}{\partial t}$$

As noted, we will see below that the coefficient of variation for women, $CV_w$, tends to increase as fertility declines. We will also see that both $k$ and $m$ tend to rise, potentially offsetting each other, but with $m$ rising considerably faster than $k$. The result from Equation (7) is that the coefficient of variation for children’s family size increases substantially as fertility declines. We will show empirically that this and several other measures of inequality in children’s family size tend to increase as fertility declines.
Family size of school-aged children

The previous results, like the original analysis of Preston, focuses on the family size of women in a given age range (Preston used 45-49), comparing their average family size to the average family size of their children. If we are interested in the family size of school-aged children during the demographic transition, we will want to consider a somewhat different version of the question addressed by Preston. The children of women aged 45-49 cover a broad age range extending from roughly 5 to 30. We are also interested in focusing on children in a narrower age range, say 9-11, and looking at the relationship between the family size of those children and the family size of some corresponding group of women who represent the potential mothers of those children. As we will see, this is a somewhat more complicated problem, although the main insights of Preston’s result still apply.

Looking at the problem from the standpoint of children of a given age would be simple if women had all of their children at some particular age $\mu$, with some women having more children than others. In that very unrealistic case, Equation (3) would provide an exact characterization of the relationship between the family size of children born in a given year and the family size of women who give birth in that year (women age $\mu$). More realistically, we must recognize that the children born in a given year have mothers who span a wide age range. If we are interested in tracking something like the mean family size of school-age children, the problem becomes more complicated than in Preston’s results. If we are interested in the mean family size of children age 10 in 1980, the mothers of those children could have ranged in age from roughly 25 to 49, representing the experience of a wide range of cohorts. Assume that we can define a group of women who could be considered the potential mothers of children who are age 10 in 1980. All women age 25 to 49 in 1980 might be considered to have been at risk of having a 10-year-old child in 1980 (although women in the middle of the age range are much
more likely to have done so than those on the extremes), so this is one logical candidate for the relevant group of women. We can then analyze how the mean family size of women age 25-49 in 1980 compares to the mean family size of children age 10 in 1980.

Note that the family size of the mothers of children aged 10 will correspond almost exactly to the family size of the children themselves. When we take children of a single age we do not observe the phenomenon discussed by Preston. Preston’s result is driven by the fact that a broad age group of children (such as children of women aged 45-49) will over-represent children born in large families. Mothers with eight children will have eight times as many children included in the calculation of children’s mean family size as mothers with one child. This will not happen in a sample of 10-year-old children. With the minor exception of multiple births, each mother of a 10-year old will have only one 10-year-old child, so the family size of 10-year-old children will be the same as the family size of the mothers of 10-year-old children. These mothers, however, are not a random sample of all women. They over-represent women with large numbers of children, since those women are more likely to have children of any given age3. If we want to understand how changes in the family size of 10-year-old children compare to changes in fertility, we cannot look simply at the mothers of those children, but must look at the fertility of women of a broader age range. If we consider an approximate childbearing span of age 15 to 40, then women aged 25 to 49 would be an approximate representation of the potential mothers of 10-year-old children in a given year. We can then compare how changes in the fertility of women aged 25-49 compares to changes in the family size of children aged 10.

As we will see below, empirically it turns out to be the case that the relationship between the

3 In the extreme case, women with no children are never represented among mothers of children of any given age. More generally, the mothers of children of a given age will be more likely to include mothers with large numbers of children. This generates the point commonly demonstrated in statistics and demography courses that “the average child is not from the average family” (Jenkens and Tuten, 1992).
mean family size of 10-year-old children and the mean family size of women aged 25-49 can be roughly approximated by the following variation on Equation (2):

\[
\bar{x}_{C(10)} \approx \bar{x}_{W(25-49)} \left[ 1 + CV_{W(25-49)}^2 \right].
\] (8)

Equation (8) states that the mean surviving family size of children aged 10 is roughly equal to the product of the mean surviving family size of women aged 25-49 and a term equal to 1 plus the squared coefficient of variation in family size for women aged 25-49. Unlike Preston’s result, or its restatement in Equation (3), this is not an exact equality since children aged 10 do not represent all children born to women aged 25-49. They are just one single-year age group drawn out of the children of these women. The intuition behind the expression is in some ways the inverse of the logic behind Preston’s result. When we look at children of a single year of age, their mothers are not a random sample of all women, but are overly representative of women who have large numbers of children. The larger the variance in women’s fertility, the larger the gap between the average family size of children and the average family size for all women.

Although Equation (8) is not an exact equality, it turns out empirically to be a simple and surprisingly accurate approximation, and provides a convenient way of summarizing the relationship between trends in fertility and trends in children’s family size for one particular age of children.

**DATA AND METHODS**

In order to look at changes in family size we require micro data from censuses or surveys at multiple points during the demographic transition in a given country. Our analysis uses 310 data sets representing 101 different countries. We use 144 census samples taken from 59 countries in the Integrated Public Use Microsamples – International (IPUMS-I) project of the University of Minnesota (Minnesota Population Center 2011), listed in Appendix 1. We also use 166 Demographic and Health Surveys (DHS) representing 68 countries (some of which are also in
the IPUMS-I data sets). As seen in the list of countries in Appendix1, these countries represent all regions of the world and cover a wide range of demographic experiences. In addition to the large number of developing countries covered by DHS and IPUMS-International, we include all of the high-income countries in the IPUMS-International database that provide data on fertility. While our focus is on the demographic transition, comparisons to high-income countries are interesting in providing a complete picture of the relationship between women’s family size and children’s family size.

We will begin by looking at family size from the perspective of women aged 45-49 and the children of those women, following the approach of Preston. We will also look at family size from the perspective of school-aged children in a narrow age range. Age 10 is quite an interesting age to examine because it represents an age at which most children would be expected to be in school. We use the 9-11 age group rather than age 10 alone in order to reduce problems that might result from age misreporting or small cell sizes. We prefer 9-11 over a broader group such as age 7-14 in order to focus on something that is closer to a single birth cohort, providing a better match to the model summarized in Equation (8).

EMPIRICAL RESULTS

*Distribution of family size of women aged 45-49*

Following Preston (1976), it is useful to begin by looking at examples of how the distribution of family size of women and children has changed as fertility has declined. Figure 1 shows the cases of Brazil, Thailand, Ghana, and Malawi, using multiple censuses from IPUMS-I. Each panel shows the density for the distribution of children ever born for women aged 45-49 in a single census year, along with the distribution of family size for the children of those women.
As in Preston, the distribution of family size of children is calculated using Equation 2, and is not based on actual data for the children.\(^4\)

The first row shows Brazil in 1970 and 2000. Brazil had a rapid fertility decline, with mean children ever born to women 45-49 falling from 5.6 to 3.4 over this period. The mean family size of children of women 45-49 in 1970 was 8.8, 1.6 times greater than women’s family size. We will see below that this is a relatively large difference between women’s family size and children’s family size. This is because of the high coefficient of variation in women’s family size in Brazil, presumably a reflection of the generally high level of socioeconomic inequality in Brazil. As must be true by Equation (2), the density for children’s family size crosses the density for women’s family size at the mean of women’s family size. Comparing 1970 with 2000 demonstrates the increasing skewness in women’s family size as fertility declines.

The second row of Figure 1 shows Thailand in 1970 and 2000. Thailand has an even faster fertility decline than Brazil, with an extremely skewed distribution of children ever born in 2000. Mean family size of children was 1.25 times larger than mean women’s family size in 1970, a substantially lower ratio than in Brazil. Mean children’s family size was 0.6 lower in Thailand than in Brazil in 1970, even though mean women’s family size was almost a full child higher in Thailand. This reflects the much lower coefficient of variation in women’s family size in Thailand. Mean family size of women fell by 61\%, from 6.5 to 2.5, while mean family size of children falls by 58\%, from 8.2 to 3.5. The difference in the rates of change is fairly small, though it goes in the direction predicted by Preston.

The bottom row shows two African countries, Ghana in 2000 and Malawi in 2008. Note that the mean family size of women in Ghana in 2000 was 5.6, almost identical to Brazil’s mean

\(^4\) Given our interest in resources available to school-aged children, surviving children is arguably a more appropriate measure than children ever born. We begin with children ever born because it is available in more censuses and is the measure used by Preston. Below we will also look at surviving children.
in 1970. The Ghanaian distributions look quite different than the Brazilian distributions, however, with important implications for children’s family size. Mean family size of children in Ghana in 2000 was 7.1, which is 1.8 children fewer than Brazil 1970, in spite of the fact that mean fertility was identical in the two cases. Ghana’s much lower dispersion in fertility causes the distribution of children’s family size to be only slightly different than the distribution of women’s family size, while the two distributions are much more different in Brazil. Malawi’s pattern also shows relatively low dispersion in fertility. The striking result here is that children’s mean family size in Malawi in 2008 was lower than children’s mean family size in Brazil in 1970, even though women’s mean family size was almost one birth higher in Malawi.

In this paper we will look at both changes over time and differences across countries in the relationship between women’s family size and children’s family size. In general we will find that cross-country differences are much larger, and are much more consistent with Preston’s prediction, than are changes over time in a given country.

**Patterns across all countries**

We now look at women’s family size and children’s family size for all of our IPUMS-I and DHS data sets. We continue focusing on children ever born to women aged 45-49, the same measure and same age group used by Preston. Figure 2 looks at two summary statistics that are motivated by Equations (3) and (5) that will drive all of our results. The figure shows scatterplots of the coefficient of variation and the skewness in children ever born, both plotted against the mean, using all of the 310 data sets shown in Appendix 1. This combines cross-section differences across countries with time-series variation within countries, an issue we discuss further below. The patterns in Figure 2 are very striking. Looking at the top panel, the coefficient of variation (CV) has a strong negative relationship to the mean, implying that the CV tends to increase as fertility declines. The OLS regression line implies that a decrease of one in
mean CEB is associated with a 0.05 increase in CV, about 10% of the overall mean. We also report the fixed effect regression coefficient when we include 101 country fixed effects. These estimates use only the variation over time within a given country. The fixed effect (FE) estimate is -0.26, roughly half the OLS estimate. We will report both the OLS and the FE estimate for all of the relationships we show in the paper. The bottom panel shows that skewness is also negatively associated with the mean. Skewness is slightly negative when mean CEB is above about 6.5, becoming highly positive when fertility is low. The relationship between skewness and mean CEB is very similar in the OLS and FE estimates.

The patterns in Figure 2 have important implications for what happens to the distribution of children’s family size as fertility declines. Using our estimates for the mean and coefficient of variation of women’s family size (the two numbers plotted in the top panel of Figure 2), we use the formula in Equation (3) to calculate the mean family size of children. Figure 3 is a scatterplot of the mean family size of children plotted against the mean family size of women for women aged 45-49. As must be the case by Equation (3), the mean family size of children is always larger than the mean family size of women. The surprising pattern in Figure 2 is that the absolute difference between the family of children and the family size of women is about the same at low levels of fertility as it is at high levels of fertility. The constant for the regression line in Figure 2 is 1.4 (s.e.=0.097), while the slope is 1.04 (s.e.=0.017), implying that as we reduce fertility by one birth the mean family size of children also falls by about one birth. If $k$, the multiplier in Equation 2, were constant (implying that the CV is constant), then mean children’s family size would decline at the same proportional rate as mean women’s family size, not at the same absolute rate.
The ratio of children’s family size to women’s family size

In the top panel of Figure 4 we plot $k$, the ratio of children’s mean family size to women’s mean family size, against the mean family size of women, along with the least squares regression line. Several features are apparent in the top panel of Figure 4. The most important feature is the negative relationship between $k$ and women’s mean family size. We know this has to happen given Equation (3) and the pattern in Figure 2, which shows that the CV of women’s family size is a negative function of mean women’s family size. Put another way, the standard deviation in women’s family size tends to fall more slowly than the mean as fertility declines, causing the multiplier $k$ to increase.

Figure 4, then, provides evidence consistent with Preston’s prediction that children’s mean family size falls more slowly than women’s mean family size as fertility declines. Using the combination of cross-section and time series variation in women’s family size in the scatterplot, the regression implies that the ratio of children’s family size to women’s family size would increase by 0.3, from roughly 1.2 to 1.5, when mean CEB of women 45-49 falls from 7 to 2.

Another interesting feature in the top panel of Figure 4 is the levels of the ratio. At high fertility levels we see that mean family size is about 1.2 times the mean family size of women. This is considerably lower than the values Preston estimated for the United States. Preston estimated that children’s mean family size was 1.6 times greater than women’s family size in the 1890 U.S. census, when women’s family size was 5.0. The ratio rose to 2.1 in 1950 and returned to 1.6 in 1970, when women’s family size was 2.7 (our estimates for 1970 are almost identical and are one of the data points in Figure 4). The ratio in Figure 4 increases to about 1.5 when women’s family size is around 2-3. In some respects the ratio stays in a surprisingly small band. If you guessed that children’s mean family size was 30% greater than the completed fertility of
women aged 45-49, you would usually be fairly close. On the other hand, the difference between 1.5 and 1.2 may be substantively important, as can be shown with a simple example.

To think about the importance of the increase in the relative mean family size for children, suppose mother’s time (or some other family resource) must be divided by the number of children.\(^5\) Consider a reduction in women’s family size from 8 to 2, the range covered in Figure 2, between period 1 and period 2. If children’s mean family size also fell to 25% of its original value (implying that children’s mean family size is a constant multiple of women’s mean family size), then mother’s time per child would be 4 times greater in period 2. If, on the other hand, children’s family size falls more slowly than women’s family size, so that \(\bar{s}_{c1} = 1.2\bar{s}_{w1}\) and \(\bar{s}_{c2} = 1.5\bar{s}_{w2}\) (values similar to Figure 4), then mother’s time per child would only increase to 3.2 times its period 1 value \((9.6/3.0)\), 20% less than would have occurred if children’s family size fell at the same rate as women’s family size. One bottom line summary of the pattern in the top panel of Figure 2, then, is that the increase in mean resources per child associated with a decrease in fertility is about 20% lower than it would have been based on fertility decline alone, the result of the fact that children’s family size fell more slowly than women’s family size.

It is important to point out that the estimated relationship in the top panel of Figure 4 falls by roughly half to -0.028 when we include country fixed effects. In other words, the actual changes observed in countries over time are not as steep as would be predicted by the cross-sectional relationship across countries. We will discuss this further below.

The bottom panel of Figure 4 looks at the ratio of the standard deviation in family size for children to the standard deviation in family size for women. The vertical axis is the ratio of the

---

\(^5\) While it is unrealistic to assume that parental time (or money) must be strictly divided by the number of children, with no joint production, this provides a simple way of translating changes in family size into changes in resource per child, and might be thought of as an upper bound on changes in resource dilution.
two standard deviations. The horizontal axis shows the mean family size of women, identical to the top panel. There is a strong negative relationship between mean women’s family size and the ratio of standard deviations. The ratio tends to be well below 1.0 at high levels of fertility, then rises above 1.0 at lower levels of fertility. The slope of the OLS regression line is very similar to the estimate using country fixed effects. The explanation of why the standard deviation of children’s family size does not fall at the same rate as the standard deviation of women’s family size is the increased skewness in women’s family size as fertility declines. This has important implications for inequality in children’s family size, as we discuss below.

**Surviving family size**

In developing countries it is probably more interesting to look at surviving family size rather than children ever born. Most of the developing countries in the IPUMS-I and DHS data sets collect data on both the number of children ever born and the number of surviving children. Most high-income countries only collect data on children ever born, so we lose those countries when we look at surviving children. Figure 5 repeats the scatterplots from Figure 4 using surviving family size instead of children ever born. The levels of women’s family size are naturally lower in Figure 5 than in Figure 4, but the overall patterns are very similar to those seen in Figure 4. The ratio of children’s mean family size to women’s mean family size increases from about 1.2 at family size of 8 to 1.4 at family size of 3. The slope of the regression line in the top panel of Figure 5 is very similar to the slope in Figure 4.

The bottom panel of Figure 5 is quite similar to the bottom panel of Figure 4. Whether we use children ever born or children surviving, the standard deviation in children’s family size falls at a much slower rate than the standard deviation in women’s family size as mean fertility declines. We discuss the implications of this for inequality in children’s family size below.
Family size of school-aged children

These results demonstrate that children’s family size can vary in important ways from women’s family size, whether comparing countries at one point in time or one country at different points in time. This can be substantively important if we are interested in the links between fertility rates and resources available to children. To take one extreme difference in Figure 5, Brazil in 1980 and China in 1982 both had mean surviving children for women aged 45-49 of 4.5. The mean surviving family size for children of these women was quite different, however – 6.8 in Brazil versus 5.2 in China (these are the two extreme observations at mean family size of 4.5 in Figure 5). Brazilian children had an average of 1.3 extra siblings, even though the average fertility of their mothers was identical to the average fertility of the Chinese mothers. The standard deviation of children’s family size was twice as high in Brazil as in China, 3.0 versus 1.5, meaning that Brazilian children had both a higher mean family size and greater inequality in family size than Chinese children.

Linking these results to child outcomes in any particular period is complicated by the fact that the children of women aged 45-49 span a wide range of ages, roughly from age 5 to age 35. If we are interested in the impact of family size on an outcome like schooling at a given point in time, it probably makes more sense to look at family size for children of a given age, rather than looking at family size of all children born to mothers of a given age. As discussed above, this is a more complicated problem than the one originally analyzed by Preston (1976). Equation (8) provides a useful approximation that links the family size of potential mothers of 9-11 year-olds to the family size of children aged 9-11. To demonstrate the accuracy of the approximation in Equation (8) we calculate the mean surviving family size for children of women aged 25-49 (using Equation 3) and compare it to the mean family size of children aged 9-11 calculated by taking the mean number of children ever born to their mothers. Figure 6 shows this comparison
graphically, demonstrating that the two means are very similar. One way to think of this is that while 9-11 year-olds represent only one small set of the children born to women aged 25-49, they are roughly a representative sample of all children born to women aged 25-49, at least in terms of family size.

An advantage of looking at a narrow age group of school-aged children is that we can simply estimate their family size directly using their mother’s reported fertility, as in Figure 6, and compare it to some standard measure such as completed family size of women aged 45-49. These women will only represent a small fraction of mothers of children of some narrow age group, say 9-11, and only some of these women will be mothers of children aged 9-11. But since completed fertility will fairly closely track fertility at younger ages, we will tend to find strong systematic relationships between the family size of children aged 9-11 and the family size of women aged 45-49. We have broken the clean mathematical link shown by Preston, but we have an empirical relationship that is easy to estimate and will be useful in looking at how changes in fertility are associated with changes in the family size of school-aged children.

We look at the ratio of mean surviving family size of children aged 9-11 (estimated by linking them to their mothers and using mother’s surviving family size) to mean surviving family size of women aged 45-49. Figure 7 shows this ratio plotted against mean surviving family size of women aged 45-49. The level of the ratio in Figure 7 is not particularly interesting, since the numerator is measured when women’s fertility is incomplete while the denominator is a measure of women’s completed fertility. More interesting is the negative slope, which is once again a reflection of the fact that children’s family size does not fall as fast as women’s family size as fertility declines. The OLS regression line implies that a decline in mean completed surviving fertility from 6 to 2 is associated with a decline in mean surviving family size of 9-11 year-olds from 5.8 to 2.4, a 20% smaller decline than the decline in completed fertility. The fixed effect
slope is considerably smaller, however, and is not statistically significant.

**Inequality in children’s family size**

As shown in Figure 2, skewness in women’s family size increases substantially when fertility declines. This explains the pattern shown in the bottom panels of Figure 4 and 5 in which the standard deviation of children’s family size falls at a much slower rate than the standard deviation of women’s family size as fertility declines. The effect of increasing skewness in women’s family size is so large that in many countries the standard deviation in children’s family size stays roughly constant even as mean children’s family size is experiencing significant declines. With little or no decline in the standard deviation in spite of the falling mean, mean-adjusted inequality in children’s family size tends to increase as fertility declines.

Figure 8 shows three measures of inequality in children’s family size as a function of mean women’s family size. The top panel shows that the coefficient of variation of children’s family size is substantially higher at low fertility levels than it is at high fertility levels. The regression line implies that the CV of children’s family size is 30% higher when mean completed surviving fertility is 6 than when it is 3. In this case the fixed effect estimate is somewhat larger than the OLS estimate. The bottom line shows two measures of inequality based on percentiles of the cumulative distribution. These use only the IPUMS-I countries, since we need to use the tails of the full distribution of family size and need the much larger census samples to get meaningful estimates. The P90/P10 ratio is the ratio of children’s family size at the 90\textsuperscript{th} percentile to children’s family size at the 10\textsuperscript{th} percentile.\(^6\) The P80/P20 ratio is defined analogously.

Looking at the P80/P20 measure, when women’s mean surviving fertility is around 6,

\(^6\) Given the small number of integer values of family size, the cumulative percentiles are rarely at an integer value. We take the approach of interpolating between integer values in the cumulative distribution to estimate P90 and P10. For example, if P85 is exactly 7 and P95 is exactly 8, we set P90 at 7.5.
children at the 80th percentile have families about four times larger than children at the 20th percentile. When surviving fertility is around 2, children at the 80th percentile have families about six times larger than children at the 20th percentile. If we held total family resources constant, this implies that the difference in resources per child between small families and large families is four times when fertility is high, but six times when fertility is low. As with the coefficient of variation, the fixed effect estimates for the percentile measures in Figure 8 are very similar to the OLS estimates, meaning that the changes we observe within countries are consistent with the estimates based on the cross-section pattern across countries.

The increasing inequality in children’s family size may be more important than the fact that mean family size of children falls more slowly than women’s family size. Assuming that family size affects children’s resources, there could be a substantial increase in inequality in resources as fertility declines. The empirical explanation for the increased inequality is the increase in the skewness of fertility as fertility declines, shown in Figure 2. The small fraction of women who continue to have large families have a substantial impact on the distribution of children’s family size. Some examples from the cumulative distributions of family size (not shown) are instructive. In Brazil in 2000 only 7% of women aged 25-49 had more than five surviving children, but 20% of children aged 9-11 were in families with over five surviving children. The mathematics driving the relationship between women’s family size and children’s family size means that discrepancies between the two can be very large in the tails of the distribution. In Ecuador in 2001, for example, the proportion of children aged 9-11 in a family with 8 or more surviving children (10.6%) was three times larger than the proportion of women aged 25-49 with 8 or more surviving children (3.2%). Substantial proportions of children continue to have large families even when the proportion of women having small families has become quite small. This can be very important in understanding inequality in resources per child.
CONCLUSIONS

Preston (1976) provided an elegant and insightful way of understanding the links between the family size of women and the family size of children. In this paper we have extended his mathematical results in two directions. First, we derived results that link the standard deviation of women’s family size with the standard deviation of children’s family size. The result, which depends on the skewness of women’s family size, is useful in understanding how inequality in children’s family size changes with fertility decline. Second, we analyzed family size from the perspective of children in a narrow age range, rather than looking at all children born to women in some given age range. This is useful in understanding how the family size of school-aged children evolves during the demographic transition.

Several interesting patterns emerge from our empirical analysis of census and DHS data from 310 data sets representing 101 countries. First, we find support for Preston’s prediction that children’s family size falls more slowly than women’s family size as fertility declines. Looking at women aged 45-49, mean family size of their children tends to be in a range of 1.2 to 1.5 times women’s mean family size. This multiplier, which is equal to one plus the squared coefficient of variation of women’s family size, tends to increase from about 1.2 at high levels of fertility to 1.5 at low levels of fertility. This implies that resources per child increase at a slower rate than would be expected based on fertility decline. We estimate the degree of “slippage” to be about 5% to 20%, with smaller estimates when we estimate fixed effects regressions using only the variation within countries over time.

We find that the gap between children’s family size and women’s family size is also important in explaining cross-section differences across countries. Two countries with the same mean completed family size can have very large differences in the mean family size of children. To take one of the most extreme cases, Brazil and China had almost identical mean completed
fertility around 1980, but mean family size of children in Brazil was 1.3 children higher than in China.  Looking beyond the means, we find that inequality in children’s family size tends to rise substantially as fertility declines.  We show that three standard inequality measures increase when fertility declines, looking both across and within countries.  The empirical reason for this increase is the large increase in the skewness in women’s fertility as fertility declines.  We show mathematically that the standard deviation in children’s family size is a positive function of the skewness in women’s fertility, and we show empirically that this skewness rises substantially as mean fertility declines.

These results have important implications for trends in resources per child during the demographic transition.  The result for the mean implies that resources per child increase more slowly than would have been expected as fertility declines.  A decline in women’s completed fertility from 6 to 2 will not lead to a tripling of resources per child, but to 5-20% less than that.  This is important to keep in mind in studying changes in children’s outcomes in response to fertility decline.  This is a fairly modest effect, however, in comparison to our results for inequality.  Inequality in children’s family size rises dramatically as fertility declines.  The coefficient of variation increases 30% when women’s mean completed family size falls from 6 to 3.  This is a very regular pattern in the data, as large when measured for one country over time as it is across countries.  This result implies that the gap in resources per child between small families and large families increases substantially as fertility declines.  This “denominator” effect exacerbates differences in resources due to the fact that large families are likely to be poorer than small families to begin with.  This large increase in inequality in family size during the demographic transition is an important effect that has received little attention in previous research.  It deserves careful attention in future research on disparities in child outcomes in developing countries.
REFERENCES


Figure 1. Distribution of family size, women 45-49 and their children

- **Brazil 1970**
  - Women (mean=5.58)
  - Children (mean=6.81)

- **Brazil 2000**
  - Women (mean=3.38)
  - Children (mean=5.44)

- **Thailand 1970**
  - Women (mean=6.53)
  - Children (mean=8.21)

- **Thailand 2000**
  - Women (mean=2.55)
  - Children (mean=3.47)

- **Ghana 2000**
  - Women (mean=5.59)
  - Children (mean=7.10)

- **Malawi 2008**
  - Women (mean=6.43)
  - Children (mean=7.85)
Figure 2.
Coefficient of variation and skewness by mean, children ever born to women 45-49, Countries and years in Appendix Table 1

Note: $\beta$ is OLS regression coefficient using all 310 country/year observations. $\beta$(FE) is regression coefficient including 101 country fixed effects. Standard errors in parentheses. Significance levels: **=0.01, *=0.05.
Figure 3.
Mean and std. dev. of children's family size by mean women's family size,
children ever born to women 45-49,
Countries and years in Appendix Table 1

\[ \beta = -1.041(0.017)** \]
\[ \beta (FE) = -1.191(0.029)** \]

\[ \beta = -0.059(0.022)** \]
\[ \beta (FE) = -0.031(0.036) \]

Note: \( \beta \) is OLS regression coefficient using all 310 country/year observations. \( \beta (FE) \) is regression coefficient including 101 country fixed effects. Standard errors in parentheses. Significance levels: **=0.01, *=0.05.
Figure 4. Mean and std. dev. of children's family size relative to women's family size, children ever born to women 45-49, Countries and years in Appendix Table 1

Note: $\beta$ is OLS regression coefficient using all 310 country/year observations. $\beta(FE)$ is regression coefficient including 101 country fixed effects. Standard errors in parentheses. Significance levels: **=0.01, *=0.05.
Figure 5.
Mean and std. dev. of children's family size relative to women's family size,
children surviving to women 45-49,
Countries and years in Appendix Table 1

\[ k = \frac{\text{Ratio of children's mean}}{\text{Mean family size of women}} \]
\[ m = \frac{\text{Ratio of children's s.d.}}{\text{Mean family size of women}} \]

\[ \beta = -0.059(0.006)^{**} \]
\[ \beta (\text{FE}) = -0.040(0.005)^{**} \]

Note: \( \beta \) is OLS regression coefficient using all 266 country/year observations. \( \beta (\text{FE}) \) is regression coefficient including 87 country fixed effects. Standard errors in parentheses. Significance levels: **=0.01, *=0.05.
Figure 6.
Mean surviving family size of children 9-11
by mean family size of children of women 25-49
Countries and years in Appendix Table 1

\[ \beta = 1.027(0.011)^{**} \]
\[ \beta (FE) = 1.075(0.015)^{**} \]
Figure 7.
Ratio of mean surviving family size of children 9-11 to mean surviving family size of women 45-49
Countries and years in Appendix Table 1

\[ \beta = -0.057(0.007)^{**} \]
\[ \beta (FE) = -0.018(0.012) \]
Figure 8.
Inequality in surviving family size of children 9-11
by mean of children surviving to women 45-49,
Countries and years in Appendix Table 1

Note: $\beta$ is OLS regression coefficient using all country/year observations. $\beta$(FE) is regression coefficient including country fixed effects. Standard errors in parentheses. Significance levels: **=0.01, *=0.05.