The Impact of Infant and Child Mortality Risk on Fertility

Kenneth I. Wolpin

INTRODUCTION

The relationship between the infant and child mortality environment and human fertility has been of considerable interest to social scientists primarily for two reasons: (1) The fertility and mortality processes are the driving forces governing population change, so an understanding of the way they are linked is crucial for the design of policies that attempt to influence the course of population change. (2) The "demographic transition," the change from a high fertility-high infant and child mortality environment to a low fertility-low mortality environment, which has occurred in all developed countries, has been conjectured to result from the fertility response to the improved survival chances of offspring.

Fundamental to either of these motivations is an understanding of the micro foundations of fertility behavior in environments where there is significant infant and child mortality risk. My purpose in this chapter is to clarify and summarize the current state of knowledge. To that end, I survey and critically assess three decades of research that has sought to understand and quantify the impact of infant and child mortality risk on childbearing behavior. To do so requires the explication of theory, estimation methodology, and empirical findings.

I begin by posing the basic empirical (and policy relevant) question: "What would happen to a woman’s fertility (children born and their timing and spacing) if there was a once-and-for-all change in infant or child mortality risk?" Alternative behavioral formulations, encompassing static and dynamic decision-theoretic models found in the literature, answer that question and are reviewed. An illustrative three-period decision model, in which actual infant and child deaths are revealed sequentially and behavior is both anticipatory and adaptive, is developed in some detail, and the empirical counterparts for theoretical constructs derived from that model are developed and related to those found in the literature. Specifically, I demonstrate how replacement and hoarding "strategies," which are prominent hypotheses about reproductive behavior in this setting, fit explicitly into the dynamic model and how these concepts are related to the question posed above.

I review a number of empirical methods for estimating the quantitative effect of infant and child mortality risk on fertility, connecting them explicitly to the theoretical framework. I pay particular attention to the relationship between what researchers have estimated and the basic behavioral question. Finally, I present and discuss an overview of empirical results.

THEORY

Static Lifetime Formulations

The earliest formal theoretical models were static and lifetime (i.e., the family attempts to satisfy some lifetime fertility goal decided at the start of its "life"). The "target fertility" model is the simplest variety of such models. Suppose a couple desired to have three surviving children. If the mortality risk were zero, they could accomplish their goal by having three births; if instead they knew that one of every two children would die, then they would need six births. Thus, it appeared straightforward that fertility would be an increasing function of mortality risk. Also, quite obviously, the number of surviving children would be invariant to the fraction of children who survive because the number of births is exactly compensatory. The target fertility model provides the intuitive basis for the concept of "hoarding," that is, of having more births than otherwise would be optimal if mortality risk were zero.

The target fertility model ignores the fact that children are economic goods, that is, that they are costly. A number of authors have introduced a budget constraint into the optimization problem (O'Hara, 1975; Ben-Porath, 1976). Although most formulations included additional decision variables, usually following the quality-quantity trade-off literature (Becker and Lewis, 1973; Rosenzweig and Wolpin, 1980), the essential features of the mortality-fertility link can be

---

1The theoretical notion of a target level of fertility, leading to a positive association between fertility and infant and child mortality, was conceptualized at least as early as 1861 (see the quotation from J.E. Wappons in Knodel, 1978).

2More formally, if the utility function is \( U(N) \) and achieves a maximum at \( N = N^* \) (the "target" fertility level), where \( N \), the number of surviving children, is equal to \( sB \), the (actual) survival rate \( s \) times the number of births \( B \), then by maximizing utility and performing the comparative statics, it can be shown that \( dB/ds = -Bis < 0 \) (iB is constant).
demonstrated in a simpler framework in which the family maximizes a lifetime utility function only over the number of surviving children and a composite consumption good, subject to a lifetime budget constraint.\textsuperscript{3} In this model, as in the target model, there is no uncertainty; parents know exactly how many children will survive for any number of births.

In this model it is easily shown that an increase in the survival rate will reduce the number of births, as in the target model, only if fertility has an inelastic demand with respect to its price (i.e., to the cost of bearing and rearing a child).\textsuperscript{4} Thus, it could be optimal to have fewer births at a positive mortality rate than at a zero mortality rate (the opposite of hoarding).\textsuperscript{5} This result can arise because births per se are costly. At the higher mortality rate, although the number of surviving children is lower for the same number of births, increasing the number of surviving children by having additional births is costly. Depending on the properties of the utility function, the optimal response may be to reduce births. Therefore, the hoarding-type implication of the target fertility model is not robust to the addition of a resource cost to bearing a child.

These models have several shortcomings. First, although the family might be assumed to know the survival risk their children face, they cannot know with certainty the survival fraction (realized survival rate) (i.e., exactly how many children will die for any given number of births). Furthermore, if the number of surviving children is a random variable, these formulations are inconsistent with expected utility analysis unless utility is linear. Second, fertility is clearly discrete. The number of children can take only integer values. Third, fertility decision making would seem a priori to be best described as a sequential optimization problem in which one child is born at a time and in which there is, therefore, time to respond to realized deaths (Ben-Porath, 1976; Knodel, 1978; O'Hara, 1975; Williams, 1977).

Sah (1991) considered the case of an expected utility maximizing family choosing the number of discrete births to have. He showed that if there is no ex ante birth cost (a cost that is incurred regardless of whether or not the child survives), then the number of births must be a nonincreasing function of the survival risk (as is true of the previous model). Consider the case in which the choice is between having two, one, or no children. In that case, the difference in expected utilities associated with having one versus no child is the survival risk \( s \) times the difference in utilities (i.e., \( s[U(1) - U(0)] \)). Similarly, the difference in expected utilities between having two children versus having one child is \( s^2[U(2) - U(1)] - [U(1) - U(0)] \). Now suppose that for a given \( s \), it is optimal to have one child but not two, a result that requires satiation at one surviving offspring \( [U(2) - U(1) < 0] \). Clearly, at a higher \( s \), it will be optimal to have at least one child. However, at the higher value of \( s \) it will still not be optimal to have a second child, and indeed the difference in expected utilities between having two and having one cannot increase. As Sah demonstrates, the argument generalizes beyond a feasible set of two children to any discrete number of children.

This result, that increasing the mortality risk of children cannot reduce fertility (except in the neighborhood of certain mortality, \( s = 0 \)), is the obvious analog to the target fertility result. However, unlike the target fertility model, it does not imply that the number of surviving children will be invariant to the survival rate. The reason is due to the discreteness (and the uncertainty). An example may be helpful. Suppose that \( U(1) - U(0) = 2 \) and \( U(2) - U(1) = -1 \). Now, assuming \( s \) is nonzero, it will always be optimal to have at least one child, \( s[U(1) - U(0)] = s > 0 \). However, in this example, for any survival rate less than two-thirds, it will be optimal to have two children. At a survival rate just below two-thirds, the expected number of surviving children is close to 1.33, whereas at a survival rate just above two-thirds, the expected number of surviving children is close to 0.67. There is, thus, a decline in the expected number of surviving children as the survival rate increases in the neighborhood of two-thirds. However, the relationship is not monotonic; the higher the survival rate within the zero to two-thirds range, and again within the two-thirds to unity range, the more surviving children there will be on average because the number of births is constant within each range.

The example also illustrates hoarding behavior. Because utility is actually lower when there are two surviving children as opposed to one, if the survival rate were unity (zero mortality risk) only one child would be optimal. However, when survival rates are low enough, below two-thirds in the example, the couple will bear two children because there is a significant chance that they will wind up with none who survive to adulthood. Indeed, at survival rates above one-half (but below two-thirds), on average the couple will have more than one surviving birth, exceeding the optimal number of births with certain survival. The key to this result, as will be apparent in the dynamic framework considered below, is that the family’s fertility cannot react to actual infant and child deaths.

Sah (1991) demonstrates, however, that adding a cost of childbearing, as before, leads to ambiguity in the effect of the survival rate on fertility. He

---

\textsuperscript{3} If \( X \) is a composite consumption good, \( Y \) is wealth, and \( c \) the fixed cost of bearing a child, the problem is to choose the number of births \( B \) that will maximize \( U(sB, X) \) subject to \( Y = cB + X \).

\textsuperscript{4} The optimal number of births is found by setting the marginal rate of substitution between the consumption good and the number of surviving births, \( U_1/U_2 \), equal to the “real” price of a surviving birth, \( c/s \). An increase in the survival rate reduces this price. The elasticity of fertility with respect to the survival rate, \( d \ln B \ln s \), is equal to negative one plus the elasticity of fertility with respect to its cost, \( -(1 + d \ln B \ln c) \).

\textsuperscript{5} However, although the number of births may rise or fall as the survival fraction increases, the number of surviving children must increase. The elasticity of the number of surviving children \( (sB) \) with respect to \( s \) equals minus the elasticity of births with respect to \( c \). Strictly speaking, the result follows if children are not Giffen goods. The target fertility result will arise only if the elasticity of fertility with respect to \( c \) is zero (fertility is perfectly inelastic with respect to \( c \)).
develops two sets of sufficient conditions for fertility (in the general case of any finite number of children) to decline with the survival rate (for hoarding to be optimal) that depend on properties of the utility function: that the utility function is sufficiently concave (in discrete numbers of children), or that for any degree of concavity the marginal utility of the last optimally chosen birth be nonpositive, that is, that the marginal utility of the last child be nonpositive if all of the optimally chosen children were to survive. Obviously, this second condition will fail to hold if there is no target fertility level, that is, if children always have positive marginal utility. Sah shows that these conditions are weaker than those that would be required if fertility were treated as a continuous choice within the same expected utility framework, and it is in that sense that discreteness reduces ambiguity.

Sequential Decision Making

In formulating the theoretical linkages between infant and child mortality and fertility, the early contributors to this area of research clearly had in mind sequential decision-making models under uncertainty. No biological or economic constraints would force couples to commit to a particular level of fertility that is invariant to actual mortality experience. However, as in other areas of economics, the formalization of such dynamic models of behavior, particularly in the context of estimation, awaited further development. To illustrate the informal argumentation of that time, consider the following discussion by Ben-Porath (1976:S164):

Let us distinguish between two types of reaction to child mortality: "hoarding" and "replacement." Hoarding would be the response of fertility to expected mortality of offspring; replacement would be the response to experienced mortality. . . . If children die very young and the mother can have another child, the same life cycle can be approximated by replacement. Where the age profile of deaths is such that replacement can reconstitute the family life cycle, replacement is superior to hoarding as a reaction, since the latter involves deviations from what would be the optimum family life cycle in the absence of mortality. If preferences are such that people have a rigid target of a minimum number of survivors at a given phase in the life cycle, hoarding involves a large number of births and the existence of more children than necessary who have to be supported in other phases of the life cycle. . . .

The superiority of replacement is clear, but of course it is not always possible. The risks of mortality are often quite significant beyond infancy. Parents may be afraid of a possible loss of fecundity or some health hazard that will make late replacement impossible or undesirable. The reaction to mortality which is expected to come at a late phase of either the children's or the parents' life cycle may be partly in the form of hoarding.

It is obvious from Ben-Porath's remarks that he viewed the replacement decision as a sequential process made in an environment of uncertain mortality and that the hoarding of births is a form of insurance that depends on forward-looking behavior. Furthermore, Ben-Porath postulated that the essential features of the environment that lead to hoarding behavior as an optimal response are those that make replacement impossible, namely that children may die beyond the period of infancy and that the fertile period is finite (and possibly uncertain).

Although several sequential decision-making models of fertility are discussed in the literature, which include nonnegligible infant mortality risk (Wolpin, 1984; Sah, 1991; Mira, 1995), none explicitly model sequential fertility behavior when mortality past infancy is significant (probably because of its intractability in a many-period setting). However the essential behavioral implications of sequential decision making and the intuition for them can be demonstrated in a sequential decision-making model with only three periods. Moreover, a three-period formulation is sufficient to illustrate and operationalize replacement and hoarding concepts.

Suppose that births are biologically feasible in the first two periods of a family's life cycle, but that the woman is infertile in the third (Figures 3-1 and 3-2 provide a graphical representation of the structure of the model). Each offspring may die in either of the first two periods of life, as an infant or as a child, with probabilities given by \( p_1 \) (the infant mortality rate) and \( p_2 \) (the child mortality rate, conditional on first-period survival). Within periods, deaths occur subsequent to the decision about births. Thus, an offspring born in the first period of the family's life cycle may die in its infancy (its first period of life) before the second-period fertility decision is made. However, that same offspring, having survived infancy, may instead die in its childhood (its second period of life) after the second-period fertility decision is made. Such a death cannot be replaced by a birth in the third period because the woman will be infertile. For the same reason, a birth in the second period is not replaceable even if its death occurs in that period (as an infant). It is assumed for simplicity that the survival probability to the adult period of life, conditional on surviving the first two nonadult periods, is unity. The family is assumed, for ease of exposition, to derive utility from only those offspring who survive to adulthood. This corresponds to the notion that children are investment goods as in the old age security hypothesis (e.g., Willis, 1980) that offspring provide benefits only as adults.

\[ d_2^0 = d_1^0 = 0 \]
FIGURE 3-1 Decision tree: Period 1.

Formally, let \( n_j \) indicate a birth at the beginning of period \( j = 1, 2 \) of the family’s life cycle and zero otherwise. Likewise, let \( d_j^k \) indicate the death of an offspring of age \( k, k = 0, 1 \) at the beginning of period \( j \), zero otherwise, given that a birth occurred at the beginning of period \( j - k \). By convention, an infant is age 0 (in its first period of life) and a child is age 1 (in its second period of life). Thus, letting \( N_{j+1}^{k} \) be the number of surviving offspring at the beginning of period \( j = 1, 2, 3, 4 \) of the family’s life cycle,

\[
\begin{align*}
N_0 &= 0, \\
N_1 &= n_1 (1 - d_1^0) = M_1^0, \\
N_2 &= M_1^0 (1 - d_2^0) + n_2 (1 - d_2^1) = M_2^1 + M_2^0, \\
N_3 &= M_2^1 + M_2^0 (1 - d_3^1) = M_3^2 + M_3^0.
\end{align*}
\]

where \( M_j^k \in \{0, 1\} \) indicates the existence of an offspring of age \( k = 0, 1, 2 \) at the end of period \( j \) (and beginning of \( j + 1 \)). Further, let \( c \) be the fixed exogenous cost of a birth and \( Y \) the income per period. Finally, utility in period 1 is just period 1 consumption, \( Y - cn_1 \), utility in period 2 is that period’s consumption, \( Y - cn_2 \), and period 3 utility is consumption in that period plus the utility from the number of surviving children in that period, \( Y + U(N_3) \).\(^8\) Lifetime utility is not discounted and income is normalized to zero for convenience.\(^9\)

\(^8\)Note that the couple does not care about the age distribution of children.

\(^9\)One can normalize income at zero without loss of generality because utility is linear in consumption.
Because the decision horizon is finite, the problem of optimally choosing a sequence of births so as to maximize lifetime utility is most easily solved backwards. Define $V^*_t(N_{t-1})$ to be the expected lifetime utility at time $t$ if fertility decision $n_t = 1$ or $0$ is made at the beginning of period $t$, given that there are $N_{t-1}$ surviving offspring at the end of period $t-1$. Furthermore, define $V_t(N_{t-1}) = \max\{V^*_t(N_{t-1}), V^0_t(N_{t-1})\}$ to be the maximal expected lifetime utility at period $t$ for given surviving offspring at the end of period $t-1$. Because no decision is made at the beginning of period 3, consider the lifetime expected utility functions at period 2 conditional on the number of surviving children, namely

$$V^0_2(0) = U(0),$$

$$V^0_2(1) = (1 - p_1)(1 - p_2)U(1) + [(1 - p_1)p_2 + p_1]U(0) - c,$$

$$V^1_2(1) = (1 - p_2)U(1) + p_2U(0),$$

$$V^1_2(1) = (1 - p_2)^2(1 - p_1)U(2) + (1 - p_2)[2p_2(1 - p_1) + p_1]U(1) + p_2[p_2(1 - p_1) + p_1]U(0) - c.$$

For either of the two states, $N_t = 0$ or 1, the decision of whether or not to have a birth is based on a comparison of the expected lifetime utilities of the two alternatives. If the family has no surviving offspring at the beginning of period 2, either because there was no child born in period 1 or because the infant did not survive to period 2, then from equation (2), the family will choose to have a birth in period 2 if and only if $V^1_2(0) > V^0_2(0)$, or $(1 - p_1)(1 - p_2)[U(1) - U(0)] - c > 0$. If there is a surviving offspring at the beginning of period 2, then the condition for choosing to have a birth is $V^1_2(1) > V^0_2(1)$ or $(1 - p_2)(1 - p_2)[U(2) - U(1)] + p_2[U(1) - U(0)]] - c > 0$. It is easily seen from these expressions that as long as the utility function exhibits diminishing marginal utility in the discrete stock of surviving offspring, that is, $U(2) - U(1) < U(1) - U(0)$, then for values of $p_1$ and $p_2$, the difference between expected lifetime utilities associated with having and not having a birth in period 2 is greater when there is no surviving offspring at the beginning of period 2 ($N_t = 0$) than when there is a surviving offspring ($N_t = 1$) (i.e., the gain to have a birth in period 2 is larger if an offspring born in period 1 dies as an infant than if it survives to period 2). The extent to which the gain from a birth in period 2 is increased by the death of an infant born in period 1 [$V^1_2(0) - V^0_2(0)] - [V^1_2(1) - V^0_2(1)]$ is equal to $(1 - p_2)^2(1 - p_1)[U(1) - U(0)] - [U(2) - U(1)]$. This gain is clearly larger the more rapid the decline in the marginal utility of surviving offspring and the smaller the age-specific mortality probabilities. It is this gain that represents the motivation for "replacement" behavior.

To isolate the effect of infant mortality risk on second-period fertility, suppose that the child mortality probability $p_2$ is zero. In this case, the birth decision in period 2 is governed by the sign of $(1 - p_1)[U(2) - U(1)] - c$ if there is a surviving first-period birth and by the sign of $(1 - p_1)[U(1) - U(0)] - c$ if there is not. Clearly, the family would have a second birth as insurance against the child death of the firstborn (i.e., there would be no hoarding because such a death, given the survival of infancy, is impossible by assumption). However, the absence of such a hoarding motive does not imply that there is no effect of mortality risk on fertility.

An increase in infant mortality risk $p_1$ has two effects on fertility. First, because an offspring born in period 1 is more likely to die during infancy, the family is more likely to enter the second period without a surviving offspring ($N_1 = 0$). In this case, according to the previous analysis, the gain from a birth in period 2 would be larger. Second, the value of having a second-period birth is lower in the new mortality environment regardless of the existing stock of children (assuming nonsatiation). The effect of a (unit) change in the infant mortality probability on the gain from having a second-period birth is $-\{U(N_1 + 1) - U(N_1)\}$. For expositional purposes, call this the "direct" effect of mortality risk. If at an initial level of $p_t$ it were optimal for the household to have a second birth even if the first survived infancy, $(1 - p_1)[U(2) - U(1)] - c > 0$, then increasing infant mortality risk sufficiently would make it optimal to have a second birth only if the first died in infancy. Further increases in the infant mortality rate would eventually lead to optimally having zero births (at some level of $p_t$, $(1 - p_1)[U(1) - U(0)] - c < 0$). If having a second surviving child reduces utility (satiation), then a second birth would only be optimal if the first died during infancy.

To illustrate the effect on second-period fertility of increasing the probability of death in the second period of life, assume that the increase occurs from an initial state in which there is no mortality risk in either period of life, $p_1 = p_2 = 0$. It is useful to contrast that effect relative to the effect of increasing the first-period mortality risk from the same state. Furthermore, assume that in the zero mortality environment it is optimal to have only one surviving child (i.e., $U(2) - U(1)|\{U(1) - U(0)| - c < 0$ and $[U(1) - U(0)] - c > 0$). Then, taking derivatives of the relevant expressions in equation (2) evaluated at zero mortality risk yields

$$\frac{d[V^1_2(1) - V^0_2(1)]}{dp_1}\bigg|_{p_1 = p_2 = 0} = -(U(2) - U(1)),$$

$$\frac{d[V^1_2(1) - V^0_2(1)]}{dp_2}\bigg|_{p_1 = p_2 = 0} = -(U(2) - U(1)) + \{U(1) - U(0)\},$$

and

$$\frac{d[V^1_2(1) - V^0_2(1)]}{dp_1}\bigg|_{p_1 = p_2 = 0} = -(U(2) - U(1)),$$

$$\frac{d[V^1_2(1) - V^0_2(1)]}{dp_2}\bigg|_{p_1 = p_2 = 0} = -(U(2) - U(1)) + \{U(1) - U(0)\}.$$
The effect of a change in the "infant" mortality rate is, as previously derived, the direct effect, which is negative if there is no satiation at one surviving offspring. The effect of a change in child mortality risk is the negative direct effect plus an additional non-negative term whose magnitude depends on the degree of concavity of the utility function. As with Sah’s result, this positive offset arises because survival of the first offspring to adulthood is now uncertain and the decision about the second birth must be made before that realization. The hoarding effect generalizes to any levels of mortality risk in the sense that concavity is a necessary condition for its existence. Both replacement and hoarding behavior depend on the curvature of the utility function.

The analysis of the second-period decision, taking the first-period birth decision as given, does not provide a complete picture of the effect of infant and child mortality risk on the family’s fertility profile. To see how the decision to have a birth in the first period varies with mortality risk, it is necessary to consider the relevant expected lifetime utilities in period 1, namely,

\[
V_1^0 = \max[V_2^0(N_1 = 0), V_2^1(N_1 = 0)] = V_2(0),
\]

\[
V_1^1 = (1 - p_1) \max[V_2^0(N_1 = 1), V_2^1(N_1 = 1)] + p_1 \max[V_2^0(N_1 = 0), V_2^1(N_1 = 0)] - c,
\]

\[
= (1 - p_1)V_2(1) + p_1V_2(0) - c.
\]

The value (expected lifetime utility) of forgoing a first-period birth is simply the maximum of the values attached to entering period 2 without a surviving offspring. The value attached to having a first-period birth depends on the probability that the infant will survive. If the offspring survives infancy, the family receives the maximum of the values associated with entering the second period with an offspring and choosing either to have or not to have a birth in that period [see equation (2)]. If the offspring does not survive, the family receives the maximum of the values associated with entering the second period without a surviving offspring [see equation (2)]. The couple has a birth in period 1 if \( V_1^1 > V_1^0 \).

To characterize the decision rules in period 1, consider the types of behavior that would be optimal in period 2 under each of the two regimes, having or not having a surviving offspring at the beginning of period 2. There are three scenarios to consider:

1. It is optimal to have a birth in period 2 regardless of the value of \( N_1 \) [\( V_2^1(0) \geq V_2^3(0), V_2^1(1) > V_2^3(1) \)].

2. It is optimal to have a birth in period 2 only if \( N_1 = 0 \) (i.e., if there are no surviving offspring [\( V_2^1(0) > V_2^3(0), V_2^1(1) < V_2^3(1) \)]).

3. It is not optimal to have a birth in period 2 regardless of the value of \( N_1 \), [\( V_2^1(0) < V_2^3(0), V_2^1(1) < V_2^3(1) \)].

Without providing the details, which are straightforward, the optimal behavior in period 1 is as follows:

1. If it is optimal to have a birth in period 2 when there is already a surviving offspring, then it will be optimal to have a birth in period 1.

2. If it is optimal to have a birth in period 2 only if there are no surviving offspring, then it will be optimal to have a birth in period 1.

3. If it is not optimal to have a birth in period 2 regardless of whether there are surviving offspring, then it will not be optimal to have a birth in period 1.

Together, these results imply that increasing infant mortality risk cannot increase fertility. At very low mortality risk, it will be optimal (also assuming that the birth cost is low enough) to have a second birth independent of whether there is a surviving first birth, implying that case (1) holds. As infant mortality risk increases, it will eventually become optimal to have a second birth only if there is not a surviving first birth, implying that case (2) holds. Finally, at some higher level of mortality risk, it will not be optimal to have a second-period birth regardless of whether there is a surviving first birth, implying that case (3) holds.

The effect of increasing child mortality risk is more complex. If we assume that infant mortality risk (when child mortality risk is zero) is such that case (2) holds, then increasing child mortality risk will at some point produce case (3) (having two children regardless of the mortality experience of the first child, i.e., hoarding behavior as in Sah). Further increases in child mortality risk will lead eventually to a decline in births back to case (2) and then to case (1).

Table 3-1 provides a numerical example. It presents two scenarios, one in which the infant mortality probability is varied, holding the child mortality probability fixed at zero, the other allowing for the opposite (although the infant mortality probability is positive). In both cases, the expected number of births and the expected number of survivors to adulthood are calculated as the relevant mortality risk is varied. In the example, the values of the utility differences \( U(2) - U(1) \) and \( U(1) - U(1) \) are fixed and exhibit concavity and nonsatiation, as is the cost of a birth, \( c \).

With respect to \( p_1 \), the expected number of births is piecewise linear. Between \( p_1 = 0 \) and the value of \( p_1 \) given by the solution to \( V_2^3(1) = V_2^3(1) \) (i.e., the

---

\([\text{These results are special to the assumption that offspring do not yield contemporaneous utility flows.}]\)
value of \( p_1 \) at which there is indifference between having a birth and not having a birth when there is a surviving offspring, 0.1125 in the example), the family will have a birth in both periods regardless of the survival outcome of the first birth (the expected number of births is two). Between the above solution and the value of \( p_1 \) that makes the family indifferent between having a birth or not when there is no surviving child, the value that solves \( V_1^d(0) = V_2^d(0) \) (0.606 in the example), the family will have a birth in period 1, but only have a birth in period 2 if the first offspring dies. The expected number of births is \( 1 + p_1 \) over this range. At higher levels of infant mortality (greater than 0.606), it will not be optimal to have any births (the expected number of births and survivors are zero). Thus, there is a positive relationship between the number of births and the infant mortality risk over a considerable range of values of the infant mortality rate (from 0.1125 to 0.606 in the example). But rather than indicating hoarding behavior, the positive relationship arises because over this range of increasing infant mortality risk it is optimal to replace infants who die. Notice, however, that the expected number of surviving children (as seen in Table 3-1) never declines (increases) with decreasing (increasing) infant mortality risk.

With respect to \( p_2 \), there may be as many as five possible piecewise linear segments of the expected birth function over the \( p_2 \) domain. In the first segment, beginning from zero (and ending at 0.044 in the example), it is optimal to have a birth in both periods only if the first offspring dies (the expected number of births is \( 1 + p_1 \)). The expected number of surviving children (to adulthood) declines in this segment. In the second segment, it is optimal to have a birth in both periods, and the utility gain from the second birth is an increasing function of \( p_2 \) (between \( p_2 = 0.044 \) and 0.10 in the example). The third segment repeats the second except that the gain is now declining in \( p_2 \) (between \( p_2 = 0.10 \) and 0.157 in the example). The fourth segment repeats the first with the expected number of births equal to \( 1 + p_1 \) \( (p_2 = 0.157 \) to 0.552), while in the last segment the family has no births (between \( p_2 = 0.552 \) and 1.0). As seen in the table, expected births first increase and then decrease as the child mortality rate increases. Notice that the jump in fertility between the first and second segments (between \( p_2 \) equal to 0.025 and 0.05) is exactly the hoarding response. Moreover, it is accompanied by an increase in the expected number of surviving children. Thus, increased child survival induces, over this range, a decrease in the expected number of surviving children.

As Table 3-1 makes clear, knowledge of the values of utility differences and of the cost of a birth are sufficient for the calculation of fertility behavior under any hypothetical mortality environment. Thus, if the model provides an accurate representation, or at least approximation, of fertility behavior, it would provide a tool for assessing the effect of health-related policy interventions and thus of the experiment postulated at the beginning of this chapter. Of course, the three-period model is intended to be only illustrative. For example, in a model with a longer fertile period and in which there are non-negligible age-specific mortality

<table>
<thead>
<tr>
<th>( p_2 )</th>
<th>0.044</th>
<th>0.10</th>
<th>0.157</th>
<th>0.552</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_1^d(0) )</td>
<td>0.606</td>
<td>0.606</td>
<td>0.606</td>
<td>0.606</td>
</tr>
<tr>
<td>( V_2^d(0) )</td>
<td>0.606</td>
<td>0.606</td>
<td>0.606</td>
<td>0.606</td>
</tr>
</tbody>
</table>

NOTES: \( 0.2(1 - 0.1) = 0.8(1 - 0.1) = 0.9 \), and \( c = 0.71 \).
risks over an extended range of childhood ages, there would be a considerably more complex fertility response to a change in the mortality environment.

**ESTIMATION ISSUES**

It is possible within this theoretical framework to describe all of the methods that demographers and economists have used to estimate the responsiveness of fertility to changes in mortality risk. In the three-period model, the family decides deterministically in each of the two fecund periods whether or not to have a birth. A population of homogeneous families with respect to preferences (the two utility differences) and constraints (the cost of childbearing and the infant and child mortality probabilities) will all act identically. Moreover, because knowledge of preferences and constraints by the researcher would imply perfect prediction of behavior by the model (assuming for this purpose that the model is literally true), there would be no statistical estimation problem. But suppose the population differs randomly in the cost of childbearing, $c$. Furthermore, assume that different values of $c$ for periods 1 and 2, $c_1$ and $c_2$, are drawn independently from the same distribution, $F(c; \Theta)$, at the beginning of their respective periods. Thus, the decision in period 1 is conditioned on knowledge of $c_1$ only. Neither period birth cost is known to the researcher, so that, although each family’s behavior is still deterministic (each family either decides to have a birth or not in each period), the researcher can determine behavior only probabilistically.

What can be learned from estimation obviously depends on the data that are available. In this regard, it is useful to divide the discussion of estimation issues into two cases corresponding to whether the population is homogeneous or heterogeneous with respect to mortality risk.

**Population-Invariant Mortality Risk**

Consider a sample of families for whom we observe fertility and infant and child mortality histories. By assumption, they have the same utility function, draw their period-specific birth costs from the same distribution, and face the same mortality schedule for their offspring. In the context of the three-period model, it is sufficient for our purpose that we have information on $n_1$, $n_2$, and $N_I$, that is, that we know about all births and at least whether the firstborn died in infancy. It is not necessary to know the mortality experience in period 2 because there can be no subsequent fertility response in period 3. In addition, let us suppose that the researcher also knows the mortality schedule faced by the sample, $p_1$ and $p_2$, and in addition, for identification purposes, knows the form and parameter values that describe the birth cost distribution $F(c; \Theta)$.

Then the likelihood function for this sample is

$$L(\omega | \text{data}) = \prod_{i=1}^I \Pr(n_{i1}, n_{i2} | d_{i1}, p_1, p_2, F),$$

which consists of the two utility differences. There are $I$ families in the sample, with the lack of an $i$ subscript implying constancy in the population.

Now, the probability (from the researcher’s perspective) that a randomly drawn family will be observed to have a birth in period 2 conditional on the two possible values of the beginning period stock of surviving offspring is

$$\Pr(n_2 = 1 | N_I = 1) = \Pr(c_2 < (1 - p_2)(1 - p_1)(1 - p_2)[U(2) - U(1)] + p_2[U(1) - U(0)]),$$

$$\Pr(n_2 = 1 | N_I = 0) = \Pr(c_2 < (1 - p_2)(1 - p_1)[U(1) - U(0)]).$$

Similarly, the probability of a first-period birth is given by

$$\Pr(n_1 = 1) = \Pr(c_1 < (1 - p_1)[E_1[V_2(1)] - E_1[V_2(0)]].$$

where $E_1$ is the expectations operator given the information set at period 1 and is taken over the distribution of infant and child mortality and birth costs, and the value functions in the integral of equation (7) are given by equation (4). There are three sample proportions, corresponding to the theoretical probabilities in equations (6) and (7), from which the utility differences can be recovered.

---

11Alternatively, utility differences could have a random component.

12The assumption of imperfect foresight with respect to the future cost of childbearing is not consistent with the solution of the model presented for the first-period fertility decision. In deciding on first-period fertility, the family would have to take into account the possible future actions that would be optimal for all possible values of the randomly drawn second-period cost of childbearing. Changing the informational structure in this way simplifies the estimation problem and its exposition.

13This is a strong assumption. In the two-period decision context, where there are only two utility differences, adopting a specific functional form for the utility function would not reduce the number of parameters. For the longer-horizon model, that would realistically apply, the reader can think of placing parametric restrictions on the form of the utility function and on the distribution function for birth costs. Such assumptions would be sufficient for identification of the distribution function parameters.

14It might appear that we could have relaxed the assumption that all of the parameters of the birth cost distribution are known, given that there are more sample proportions than unknown parameters. However, this turns out not to be the case given the structure of the model. As an example, if $F$ is assumed normal and we normalize the variance to unity, the mean of the cost distribution does not enter the decision rule in a way that allows it to be identified separately from the utility differences.
Given these estimates, the response of fertility behavior in the population to variations in mortality risk can be obtained by solving the behavioral optimization problem. More specifically, given the estimates of the utility differences, it is possible to forecast the impact of changing age-specific mortality risk on the expected number of births and surviving children exactly as in Table 3-1.

Notice that in this example there is no other way to estimate policy responses to mortality risk variation because mortality risk is assumed not to vary in the population. However, the sample proportions of births, the data analogs of equations (6) and (7) used in the structural estimation, obviously provide information about behavior. Indeed, taking the difference in the probabilities in equation (6) provides a natural way to define a measure of replacement behavior. Specifically, the definition of the replacement rate \( r \) is given by

\[
r = \Pr(n_2 = 11n_1 = 1, d_i = 1) - \Pr(n_2 = 11n_1 = 1, d_i = 0),
\]

where the conditioning event in equation (8) is restricted to a first-period birth with and without its death. (This detail is unnecessary given the construction of the model, because the second-period birth probability is the same regardless of why the stock of children is zero.) Analogous to the three-period dynamic model, if there is diminishing marginal utility, then the probability of a birth is larger when there is no surviving offspring than when there is, and \( r > 0 \). According to equation (8), full replacement, \( r = 1 \), would require that the probability of having a second birth be unity when the first birth did not survive and zero otherwise. On the other hand, the replacement rate will be zero if the probability of a birth is independent of the number of surviving offspring, a result that requires [see equation (6)] that the marginal utility of surviving offspring be constant, \( U(2) - U(1) = U(1) - U(0) \).

What do we learn from this transformation of the underlying probabilities (i.e., from calculating \( r \))? Or alternatively, for what policy experiment would calculation of \( r \) be relevant? Implicit in the original policy experiment is the notion that the effect of the change in mortality risk of infants and children is known to families or becomes immediately obvious from its (population) impact (for example, as seems to have been the case when the polio vaccine was introduced in the United States). However, suppose that, although effective, families did not alter their beliefs about mortality risks (or did so only very slowly, as might be the case with generalized improvements in nutrition). Then we would observe families responding only to the reduced number of infant deaths as they are experienced and not to the decline in the infant and child mortality risk per se. In this case, the replacement rate would measure the full response to the program.

Now the replacement rate times the number of first-period infant deaths in a population yields the number of extra births that arise in that population because of the infant deaths. This result follows from the fact that the definition of \( r \) in equation (8) is equivalent to the expected number of births given a birth and infant death in the first period, \( 1 + \Pr(n_2 = 11n_1 = 1, d_i = 1) \) minus the expected number of births given a birth and no infant death, \( 1 + \Pr(n_2 = 11n_1 = 1, d_i = 0) \). So suppose, for example, that the government institutes a health program that will reduce the infant mortality risk by 0.05. Following its introduction, there will be a reduction in the number of first-period deaths of 0.05 \( \times \Pr(n_1 = 1) \times \) (number of families). If \( r = 1 \) at the preprogram, and perceived to be postprogram, mortality risk, then all of the second-period births that had resulted from the replacement of first-period deaths will not occur and the number of second-period births will thus decline by the same amount as the number of first-period deaths. Furthermore, the number of surviving children would be approximately unchanged. Alternatively, if \( r = 0 \) so that no first-period deaths were replaced, the number of second-period births will be unaffected by the reduction in the number of deaths, and the number of surviving children will increase by the number of averted deaths.

If the fertile stage is extended beyond two periods, the replacement rate would have to account for births, arising from a death, that occur in later periods. If there are \( T \) fecund periods, then the replacement rate for the specific case of a birth in the first period and its subsequent infant death is given by

\[
r = \sum_{j=2}^{T} \left[ \Pr(n_j = 11n_1 = 1, d_i = 1) - \Pr(n_j = 11n_1 = 1, d_i = 0) \right].
\]

Notice that to calculate equation (9) knowledge of the probabilities of all future birth sequences conditional on \( N_t \) is required. In the \( T \)-period case, a replacement rate can be calculated at any period for any given birth and death sequence. For example, in a setting in which the entire birth and death sequence determined the decision rule, there would be, for example, a seven-period replacement rate given the death of a 3 year old in period 6, and it would be conditional on other births and deaths as of the end of period 6. There is thus potentially a very large set of replacement rates, all of which are determined by the parameters of the underlying behavioral model.

Although perhaps less transparent than was the case in the three-period

---

15It should be noted that the behavioral formulation assumes that the policy is itself a surprise (i.e., that families attach a zero probability to its occurrence). Otherwise, in the dynamic model, one would have to allow for a distribution over future infant and child mortality risk (conditional on current information). Introducing additional parameters would require a reconsideration of identifi-

16Replacement births could be postponed in this model due to random fluctuations in birth costs. In richer models there could be additional reasons.

17Thus, replacement rates are not restricted to infant deaths.

18It is also possible for the replacement rate to be larger than one if children die at older ages. A death close to the end of the fertile period might induce hoarding behavior.
model, the expression for the replacement rate in equation (9) is equivalent to the
difference in the expected number of births given the birth and death of a child in
period 1 and the expected number of births given the birth and survival of a child
in period 1. Therefore, equation (9) measures exactly the excess births that arise
from an infant death. Analogous replacement rates would measure the excess
births that would arise from the death of a child of any age given any birth and
death history.

The value in estimating replacement effects for policy analysis rests on an
assumption about the extent to which effective programs alter families’ perceptions
about mortality risk. Eventually, one would expect that the new mortality
environment would become known, in which case replacement rates would pro-
vide an inaccurate picture of the fertility response. The value of structural estima-
tion of the behavioral model does not depend on assumptions about learning,
because identifying the fundamental parameters allows either policy experiment
(or any combination) to be simulated (although predicting the effect of the change
would depend on those assumptions). For the rest of the discussion I will
assume, as has the literature, that replacement rates correspond to experiments of
interest.

Longitudinal or retrospective data on birth and death histories are not always
available. Often, only information on total births and deaths is reported for a
cross section of households (i.e., in the three-period model, \(n_1 + n_2\) and \(d_1\)).
Mortality risk is the same for all families, it would seem natural to estimate a
regression of the number of births on the actual number of infant and child
deaths. Such a regression would determine the additional births that arise from
one additional death (i.e., the replacement rate in equation (8) or equation (9) in
the \(T\)-period case). Thus, for the regression

\[
B_j = b_0 + rD_j + \nu,
\]

where \(B_j\) is the number of births in family \(j\), \(D_j\) is the number of deaths, and \(\nu\) is
a stochastic element, the regression coefficient \(r\) is the replacement rate in equation
(8). In the three-period model, \(B\) can equal only 0, 1, or 2, whereas \(D\) can be
only 0 or 1, and the only relevant deaths are those of infants. Again, second-
period deaths are ignored because they cannot influence fertility.

From the birth probabilities given in equation (6), it is possible to derive the

\[
\begin{align*}
\hat{r} &= \frac{E(B | D = 1) - E(B | D = 0)}{g_1 + g_2} = \frac{g_1 + g_2}{g_1 + g_2},
\end{align*}
\]

where \(g_1 = \Pr(n_1 = 0)\) and \(g_2 = \Pr(d_1 = 1 | n_1 = 1)\). It is easy to see from
equation (11) that the ordinary least-squares regression coefficient of births on
deaths in general will overstate the replacement effect.²¹

There have been several attempts to provide estimates of replacement effects
that correct for the “spurious” correlation between births and deaths.²² Olsen
(1980) considers the following joint stochastic representation of total births and
deaths:

\[
\begin{align*}
B_j &= \overline{B} + r(D_j - \overline{D}) + \nu_j, \\
D_j &= pB_j + e_j,
\end{align*}
\]

where the overbars indicate means. Substituting the first equation in (12) into the
second yields

\[
D_j = p\overline{B} + \frac{p\nu_j}{1 - pr} + \frac{e_j}{1 - pr}.
\]

As Olsen argues, because the number of deaths is not statistically independent of
\(\nu_j\) as seen in equation (13), the regression coefficient estimator of \(r\) from the \(B_j\)
equation will be biased and inconsistent. Specifically, the probability limit of the
regression estimator is

²¹ Notice that the replacement rate can be estimated correctly by restricting the estimation sample
to couples with at least one birth (\(g_1 = 0\) in that sample). However, this is an artifact of the three-
period model. In general, the number of deaths must rise with the number of births for a constant
mortality rate, and the resulting positive correlation between births and deaths is built into the
estimated replacement effect.

²² Notice that in the three-period model, knowledge of \(g_1\), the proportion of families not having a
child in period 1, and of \(g_2\), the proportion of families who have both a birth in the first period and
for whom the infant dies, is sufficient to solve equation (11) for the true replacement effect. Of
course, if one had this information it would not be necessary to estimate the replacement effect from
equation (11), as equation (6) could be computed directly as discussed above. The use of equation
(11), as noted, is predicated on the lack of such event history data.
\[
\text{plim}(\hat{r}) = r + \frac{\text{cov}(D_j, \nu_j)}{\text{var}(D_j)} > r. \tag{14}
\]

The ordinary least-squares regression estimator overstates the true replacement rate because deaths and births are positively correlated independently of the existence of replacement; families with more births experience more deaths simply because their "sample" size is larger. Deriving expressions for the moments in equation (14) under the assumption that \(D_j\) and \(B_j\) are binomial random variables, Olsen further shows that

\[
\text{plim}(\hat{r}) = r + \left(1 - pr\right)\left[\frac{p}{\text{var}(B)} \right]^{-1}. \tag{15}
\]

The probability limit of the ordinary least-squares estimator can deviate substantially from the true replacement rate. Assuming \(r = 0\), Olsen reports that the ordinary least-squares regression estimate of \(r\) ranges between 0.9 and 1.7 for five different populations. But what is important is that equation (15) can be used to "correct" the replacement effect estimate using the observed mortality rate \(p\) (along with the mean and variance of births). Given the ordinary least-squares estimate (\(\hat{r}\)), the sample mortality rate (\(p\)), and the first two moments of the sample birth distribution \([\bar{B}, \text{var}(B)]\), equation (15) can be explicitly solved for \(r\).

Mauskopf and Wallace (1984) present a somewhat different procedure for estimating the replacement rate that solves explicitly for the death distribution that is consistent with replacement behavior.\(^{23}\) To outline their method, define \(B^*\) to be the number of births that would occur if the family experienced no child deaths. In the three-period model, \(B^*\) is a well-defined entity obtained by solving the sequential model for \(n_1\) and \(n_2\) conditional on there being no deaths. Given that \(c\) varies randomly in the population, \(B^*\) is a random variable with expectation in the three-period model given by \(1 \times \text{Pr}(B = 1 \mid D = 0) + 2 \times \text{Pr}(B = 2 \mid D = 0)\). In general, one can write \(B^* = E(B^*) + u\), where \(E(u) = 0\), \(E(u^2) = \sigma^2\), and where \(E(B^*)\) is determined by the exact optimizing model that is adopted.

Because replacement children can themselves die and be replaced, Mauskopf and Wallace (1984) conceptualize the process as sequential (although they conceive of the actual decision process as being static), consisting of separate rounds of deaths, replacement births, deaths of the replacement births, replacement births of the replacement birth deaths, etc. So in the first round, \(d_0 = pB^* + e_0 = pE(B^*)\)

\[+ \omega_0,\text{ where } \omega_0 = pu + e_0 \text{ and } p \text{ is the (non-age-specific) mortality rate of children. First round deaths, } d_0, \text{ conditional on } B^*, \text{ are assumed to be binomially distributed. Assuming independence between } u \text{ and } e_0, \text{ the variance of } \omega_0 \text{ is } \text{var}(\omega_0) = p^2 \sigma^2 + p(1 - p)E(B^*). \text{ First-round replacements are } r_0 = rd_0 + \eta_0 = rpE(B^*) + \xi_0, \text{ where } \xi_0 = ro_0 + \eta_0 \text{ and } r \text{ is the replacement rate. The variance of } \xi_0 \text{ given the independence between } \eta_0 \text{ and the other stochastic elements is } \text{var}(\xi_0) = r^2 \text{var}(\omega_0) + r(1 - r)pE(B^*). \text{ In general, in the } i\text{th round, } d_i = pE(r_{i-1}) + \omega_i \text{ and } r_i = rE(d_i) + \xi_r. \text{ Summing over all rounds, } i = 0, 1, \ldots \text{, they derive the following expressions for the first two moments of total deaths and total births:}

\[E(D) = \frac{pE(B^*)}{1 - pr}, \]

\[E(B) = \frac{E(B^*)}{1 - pr}, \]

\[\text{var}(D) = \frac{\left[p^2 \sigma^2 + pE(B^*)(1 + pr - p)\right]}{(1 - pr)^2}, \]

\[\text{var}(B) = \frac{\left[p^2 \sigma^2 + pE(B^*)\right]}{(1 - pr)^2}, \]

\[\text{cov}(D, B) = \frac{\left[p \sigma^2 + prE(B^*)\right]}{(1 - pr)^2}. \]

These five equations have four unknowns, \(p\), \(r\), \(E(B^*)\), and \(\sigma^2\), satisfying the necessary condition for identification. Mauskopf and Wallace estimate the parameters by matching the theoretical moments to the observed moments in the data.

In the \(T\)-period model, the replacement rate estimated from the total birth and total death relationship using either Olsen's correction or the Mauskopf and Wallace method would be an "average" of replacement rates specific to the actual birth and death sequences in the sample. Such "average" replacement rates could be quite different for samples that differ, for example, only in the age distribution of the families (women), but with the same underlying infant mortality risk, preferences, and birth costs.

**Population-variant Mortality Risk**

**Observable Heterogeneity**

Assume now that population variation in the mortality environment is observable to the researcher (e.g., geographic variation in mortality rates). Struc-

\(^{23}\)Mauskopf and Wallace (1984) argue that Olsen's correction is only strictly valid at a zero replacement rate. Olsen assumed that the distribution of deaths conditional on births was binomial (i.e., \(E(D) = pB\)). However, if deaths are in part replaced, the binomial assumption cannot be valid as the number of deaths will depend on the replacement rate and on the number of births that would occur if there were no deaths.
tural estimation could proceed with the pooled (over geographic areas) data if utility differences (the structural parameters) were cross-sectionally invariant or separately by geographic area if not.

A possible alternative estimation procedure is to approximate the decision rules of the optimization problem as a general function of the state variables. Birth probabilities could be approximated by

\[
\begin{align*}
\Pr(n_2 = 1 | N_1) &= \Pr[h_2(N_1, p_1, p_2, c_2) > 0], \\
\Pr(n_1 = 1) &= \Pr[h_1(p_1, p_2, c_1) > 0],
\end{align*}
\]  

(17)

where, as before, the c's are random variables. Given distributional and functional form assumptions, the impact of a change in mortality risk can be calculated from the h functions estimated with likelihood function shown in equation (5). For example, linearizing the h function in its arguments and assuming normality of the cost distribution leads to a standard (bivariate) probit estimation problem.\(^{24}\) Although the estimates could be used to assess the effect of policy interventions that reduce mortality risks, the ability to extrapolate from equation (17) so as to assess large policy changes would depend on the global properties of the approximation of the h function. And clearly this estimation method is unavailable without population variation in mortality risk.

Nonparametric estimation of replacement rates (equation (8) or (9)) could be obtained for each geographic area, recalling that replacement rates depend on the level of infant mortality risk. Because Olsen's correction factor (equation (15)) depends on the mortality rate, Olsen's analysis could be conducted within areas, recognizing as well the direct dependence of the replacement rate on mortality risk. Similarly, Manski and Wallace's analysis could be conducted within geographic areas, obtaining separate location-specific estimates of replacement rates.

Unobservable Heterogeneity

In this case, families differ in their underlying mortality risk to an extent not fully observed by the researcher (see Rosenzweig and Schultz, 1983, or Olsen and Wolpin, 1983, for evidence of the existence of permanent unobserved heterogeneity in mortality risk). Structural estimation, using maximum likelihood as before, would require either that couple-specific mortality schedules be treated as estimable parameters or that they be assumed to have a known distribution whose parameters would be estimated. Suppose, for example, that \(p_2 = 0\) (i.e., that there is no mortality risk beyond infancy). Then, because realized deaths provide information on family-specific mortality risk, the likelihood function would incorporate both fertility and mortality events, namely

\[
L = \prod_{i=1}^{n} \Pr(n_{i1}, n_{i2}, d_{i1}^{0}, d_{i2}^{0} | p_{i1})
\]

(18)

where \(d_{i1}^{0}\) and \(d_{i2}^{0}\) indicate a death of a first- or second-period birth in infancy as in the prior notation, and where conditioning on \(p_{i1}\) indicates that the infant mortality rate is now family specific. It is important to note that, to implement the estimation procedure (maximize equation (18)), the optimization problem of the family would have to be solved for each family separately. If instead we had assumed a specific parametric distribution for \(p_1\) or a nonparametric distribution having a fixed number of discrete values (a fixed number of family types), the likelihood function would contain an integration or discrete mixture over the possible values of \(p_1\) that each family could be exposed to. The optimization problem would, in this case, have to be solved for each possible value of mortality risk. Notice that there are cross-equation restrictions implied by the behavioral model. Not only are birth probabilities in the two periods connected by the same fundamental parameters (the last two components of the likelihood function in the second line of equation (18)), but also those probabilities are functions of the mortality risk that also enters the determination of actual deaths in the first component of the likelihood function. Thus, the estimates of mortality risk would be influenced not only by observed mortality rates, but also by the fertility response to mortality risk.

Approximate decision rules (equation (17)) can be estimated using likelihood function (equation (18)). Heckman (1982) deals with this class of models in the multinomial probit case when there is also unobserved heterogeneity. Estimating approximate decision rules differs from structural estimation in that cross-equation restrictions are ignored. Estimation must take into account that the unobserved heterogeneity is correlated with the existing stock of children in each period of the family's life cycle (see, for example, Mroz and Weir, 1989). Subject to the usual caveat about the inconsistency of fixed-effects estimates in "short" panels, one would recover an (unbiased) estimate of each family's permanent mortality risk and also obtain an estimate of the effect of mortality risk on birth probabilities. However, this procedure is equivalent to a two-step procedure of estimating the family-specific mortality risk from realized family-specific mortality rates and "regressing" measures of fertility on them. Because the realized rates measure the true risk with error, these policy-relevant estimates would be biased downward. Replacement effects estimated from the effect of a death on birth probabilities or total births, holding the estimated mortality risk
constant, would be upwardly biased (since deaths are positively correlated with the family's true mortality risk).

Replacement rates estimated nonparametrically using sample birth probabilities would not correspond to the replacement rate for any particular family. The replacement rate in equation (8) or (9) holds mortality risk constant, whereas the sample birth proportions reflect the unobserved variation in mortality risk. The number of excess births calculated this way would also not provide the correct population effect. That calculation would have to be performed for each family separately and then summed over families to obtain the correct estimate. Nonparametric replacement estimates based on birth and death histories can be used for policy analysis only if one can assume that mortality risk is homogeneous (or otherwise held constant).

Olsen suggests estimating the replacement effect from the regression (equation (10)) using realized sample (infant) mortality rates among families, reflecting in part variation in underlying mortality schedules and in part "luck," as an instrumental variable for total deaths (to correct for the spurious correlation between total births and total deaths). Although such a procedure would be valid in the case in which mortality risk was constant in the population (and also unnecessary), when mortality risk varies but is observed, regression equation (10) is misspecified; the dynamic model, for example, implies that the expected number of births will vary with the true infant mortality risk (recall the direct effect of mortality risk demonstrated in the three-period model). If it is not included as a regressor because it is unobserved, then it enters through the regression error, and the realized mortality rate, being correlated with it, cannot be a valid instrument for the number of deaths. It is important to stress that the problem with this procedure exists even if there is no child mortality risk; it has nothing to do with hoarding behavior. However, if there was significant child mortality and the risk varied in the population, child mortality rates would also not be a valid instrument for estimating replacement effects because they affect fertility independently (through the direct effect and the hoarding effect).

Both Olsen and Mauskopf and Wallace extended their methodologies to the case in which population heterogeneity in mortality risk is unobservable. It is sufficient to consider the Mauskopf and Wallace paper because the problem with the method, shared by both, is more easily demonstrated. Assume that child (but, not infant) mortality risk is zero. As Mauskopf and Wallace note, one can view the moment equations (16) as conditional on a particular value of \( p \) (\( p_i \) under the above assumption). To obtain the unconditional (population) moments, Mauskopf and Wallace integrate the conditional moments over \( p \), assuming that \( p \) comes from the two-parameter beta distribution. They then estimate the beta distribution parameters instead of \( p \). However, this procedure is inconsistent with the dynamic model presented above; both the replacement rate \( r \) and \( E(B^p) \), which appear in the moment equations, depend on \( p \). Integrating over \( p \) requires that one solve for their optimal values as \( p \) changes. Ignoring this dependence leads, in principle, to incorrect estimates of the replacement effect.

Additional Mortality-Fertility Links

There are two biological links between mortality and fertility, recognized in the literature, that create special problems for the estimation and interpretation of empirical relationships (for more details see Wolpin, in press).

Breastfeeding. There is a strong presumption that lactation reduces the propensity to conceive. Thus, in societies where breastfeeding is customarily practiced, the death of an infant may hasten the birth of another child. What may appear to be a behavioral replacement response may be due to the premature cessation of breastfeeding that accompanies the infant death (see, for example, Preston, 1978).

Endogenous Mortality Risk. To the extent that the timing and spacing of children itself influences mortality risk and there is unobserved heterogeneity in mortality risk, estimating the effect of mortality risk on fertility is problematic. In addition to biological links, a further difficulty arises if mortality risk is unknown to the decision maker.

Learning. Families may not know their own infant and child mortality risk. Deaths may provide useful information about that risk. Replacement effects, for example, will then reflect not only the direct effect of a death but also on its signal about the families' underlying propensity to experience deaths (see Mira, 1995).

A BRIEF SURVEY OF EMPIRICAL RESULTS

This section presents results of representative studies that have had as their main goal either the estimation of replacement effects or the effect of mortality risk on fertility. Estimates of replacement provide information about the effect of a hypothetical policy intervention that reduces mortality risk but for which the change in mortality risk is not perceived by families. Studies that estimate the effect of mortality risk on fertility provide information about the effect of an intervention when the change in mortality risk it induces is known to families.

Replacement Effects

Estimates Based on Birth and Death History Data

The definition of replacement given in equation (8) or (9) has rarely been adopted as the statistical measure in practice when complete fertility and mortality histories have been available. Rather, two other measures, having their roots
in the demographic literature, have been more prominent: the parity progression ratio and the mean closed interval to the next birth. They are both easily defined in terms of birth probabilities. If we let \( L \) equal the duration to the next birth, then conditional on the first-period state, the probability that \( L = t \) is

\[
Pr(L = t | n_1, d_1) = Pr(n_{i+1} = 1, n_1 = 0, ..., n_2 = 0 | n_1, d_1).
\] (19)

These probabilities define the duration density function, say \( g(L) \). There is actually a different density for each period and a different longest duration. If there is a longest feasible duration, \( L^* \), then for \( g(L) \) to be a proper density it must include the probability of having no more children, for example,

\[
g(\infty) = 1 - \sum_{t=1}^{L^*} Pr(L = t | n_1, d_1).
\] (20)

(In the three-period model, \( L = 1 \) is the only feasible duration.) The parity progression ratio (PPR) is simply the probability that the couple will have (at least) one more birth, the next parity. A measure of replacement behavior is the difference between the PPR when there is a death and the PPR when there is not, that is, in the case in which there is a birth in period 1 (as in equation (6)):

\[
\Delta \text{PPR} = G(L^* | n_1 = 1, d_i = 1) - G(L^* | n_1 = 1, d_i = 0),
\] (21)

where \( G \) is the cumulative duration distribution function. PPRs can be calculated analogously for any birth and death history and are usually calculated by conditioning on a particular parity (rather than on age). Conditioning on parity rather than age leads to a different quantitative measure for replacement because it combines the different age-specific responses (i.e., experiencing one infant death out of three births will induce a different replacement response depending on the number of periods left until the end of the fecund horizon). However, calculating the excess number of births due to a reduction in the mortality risk for a population with homogeneous mortality risk would be identical using PPRs to that calculated from equation (8) or (9). For a population with heterogeneous mortality risk, the excess birth calculation using PPRs would be incorrect, as was the case with using equation (8) or (9).

The other prominent measure of replacement behavior, based on birth and death histories, found in the literature is the differenced mean closed interval (DMCI). The DMCI is the difference in mean birth durations, conditional on having an additional birth before the end of the fecund stage, under alternative mortality experiences. The DMCI is given by

\[
\text{DMCI} = E(L | n_i = 1, d_i = 1, L \leq L^*) - E(L | n_i = 1, d_i = 0, L \leq L^*)
\]

\[
= \sum_{i=1}^{L^*} \left[ \frac{tg(t | n_i = 1, d_i = 1)}{G(L^* | n_1 = 1, d_i = 1)} - \frac{tg(t | n_i = 1, d_i = 0)}{G(L^* | n_1 = 1, d_i = 0)} \right]
\] (22)

Because the DMCI is conditional on having a birth subsequent to the death, its relationship to the other replacement measures is not straightforward. For example, in the last fertile period the expected duration to the next birth, conditional on there being a birth, must be one period independent of the prior mortality experience; thus, in this case the DMCI will be zero even though the probability of an additional birth would be responsive to the mortality history (\( r \) in equation (8) is not zero). The DMCI should be more closely related to the other replacement measures in populations of younger families and higher fertility. This measure is clearly the most problematic in calculating excess births.

Knodel (1978) uses reconstituted birth and death information from three German villages for women who married between 1840 and 1890 and whose marriages were intact at age 45. Mean closed intervals and PPRs for the three villages are shown in Table 3-2. All of the villages have high fertility rates; total fertility is over 6. The women in the village of Mömmlingen are known to have breastfed their children over extended periods, whereas in Schönberg and Anhausen breastfeeding was rarely practiced at all. This fact is consistent with the longer average birth interval in Mömmlingen, although the prolongation of postpartum sterility due to lactation may not be the only factor.

The differences between the mean closed intervals for women who did and did not experience child deaths are clearly largest in Mömmlingen, the village where breastfeeding was normally practiced. For example, over all birth intervals, the mean closed interval was more than 10 months shorter if a woman residing in Mömmlingen had experienced at least one death, but only two months shorter for women in Schönberg and less than one month shorter for women in Anhausen. The differences in PPRs by infant mortality experience, however, seem to be largest in Anhausen and are therefore more suggestive of replacement behavior. But PPRs do not uniformly rise with additional deaths; indeed, the likelihood of a woman moving from a third to a fourth birth declines with the number of infant deaths. One possible explanation of this phenomenon would be that women with more deaths learn that they have a higher infant mortality rate, which reduces subsequent fertility (the direct effect). In an attempt to net out the lactation effect of an infant death, Knodel looks at the effect of the death of the firstborn on the mean closed interval between the second and third births. The differences are now largest in Anhausen, 5.8 months, and smallest in Schönberg, 2.8 months.

Vallin and Lery (1978) use a subsample of 92,000 French women who were born between 1892 and 1916 and were surveyed in 1962. As reported in
Table 3-2, Parity Progression Ratios and Mean Closed Intervals Based on Nineteenth Century Bavarian Village Data

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Mömmlingen</th>
<th>Schönberg</th>
<th>Anhausen</th>
</tr>
</thead>
<tbody>
<tr>
<td>All birth intervals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No infant deaths</td>
<td>30.0</td>
<td>22.0</td>
<td>19.9</td>
</tr>
<tr>
<td>One or more infant deaths</td>
<td>19.4</td>
<td>20.0</td>
<td>19.2</td>
</tr>
<tr>
<td>Second to third child</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First child survives</td>
<td>29.0</td>
<td>23.9</td>
<td>23.4</td>
</tr>
<tr>
<td>First child dies</td>
<td>25.4</td>
<td>21.1</td>
<td>17.6</td>
</tr>
<tr>
<td>Parity Progression Ratios (percent)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mömmlingen</td>
<td>Schönberg</td>
<td>Anhausen</td>
</tr>
<tr>
<td>Second to third child</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No infant deaths</td>
<td>96.3</td>
<td>97.4</td>
<td>84.0</td>
</tr>
<tr>
<td>One or more infant deaths</td>
<td>100.0</td>
<td>97.6</td>
<td>90.5</td>
</tr>
<tr>
<td>Third to fourth child</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No infant deaths</td>
<td>93.0</td>
<td>93.1</td>
<td>81.3</td>
</tr>
<tr>
<td>One infant death</td>
<td>94.5</td>
<td>87.9</td>
<td>92.9</td>
</tr>
<tr>
<td>Two or more infant deaths</td>
<td>85.7</td>
<td>76.5</td>
<td>90.0</td>
</tr>
</tbody>
</table>


Table 3-3, Parity Progression Ratios and Mean Closed Intervals Based on 1962 French Survey of Family Structure

<table>
<thead>
<tr>
<th>Parity</th>
<th>Mean Closed Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Infant Death</td>
</tr>
<tr>
<td>Total fertility</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>4.14</td>
</tr>
<tr>
<td>Three</td>
<td></td>
</tr>
<tr>
<td>First birth</td>
<td>3.35</td>
</tr>
<tr>
<td>Second birth</td>
<td>4.19</td>
</tr>
<tr>
<td>Four</td>
<td></td>
</tr>
<tr>
<td>First birth</td>
<td>2.77</td>
</tr>
<tr>
<td>Second birth</td>
<td>3.43</td>
</tr>
<tr>
<td>Third birth</td>
<td>4.06</td>
</tr>
<tr>
<td>Parity Progression Ratios (n to n + 1)</td>
<td></td>
</tr>
<tr>
<td>No Infant Death (birth n)</td>
<td></td>
</tr>
<tr>
<td>Infant Death (birth n)</td>
<td></td>
</tr>
<tr>
<td>First to second child</td>
<td>68.5</td>
</tr>
<tr>
<td>Second to third child</td>
<td>57.6</td>
</tr>
<tr>
<td>Third to fourth child</td>
<td>56.4</td>
</tr>
<tr>
<td>Fourth to fifth child</td>
<td>57.3</td>
</tr>
<tr>
<td>Fifth to sixth child</td>
<td>59.2</td>
</tr>
</tbody>
</table>


Estimates Based on Total Births and Deaths

Tables 3-4 and 3-5 report estimates of replacement effects based on the use of total births and total deaths. Table 3-4 shows replacement effects obtained by Olsen and Table 3-5 those by Mauskopf and Wallace. Olsen uses data from the 1973 Columbia Census Public Use Sample and reports his results for different age and residential location groups. Only the oldest age group, women who were age 45-49 in 1973, are shown. The uncorrected estimates, that is, the regression coefficient on total deaths, imply a replacement rate of over one for both urban and rural women, regardless of whether controls are added. The corrected estimate that assumes a homogeneous mortality rate in the population is negative, implying that there are actually fewer births when there is an infant or child death. This result is consistent with the negative "direct" effect of higher infant mortality. The replacement effect obtained under the assumption that the mortality rate varies in the population (independently from births) yields point estimates of around 0.2.25 Olsen also estimates a replacement effect when the mortality

25The independence assumption is inconsistent with optimizing behavior.
TABLE 3-5 Replacement Effects from Total Births Regressors: Mausskopf and Wallace Method

<table>
<thead>
<tr>
<th>Mortality Rate</th>
<th>All Women</th>
<th>None</th>
<th>1-4</th>
<th>5+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>0.601</td>
<td>0.348</td>
<td>0.592</td>
<td>0.964</td>
</tr>
<tr>
<td>Random</td>
<td>0.593</td>
<td>0.437</td>
<td>0.613</td>
<td>0.978</td>
</tr>
</tbody>
</table>

*Standard errors in parentheses.


rate is correlated with births. Those estimates vary between 0.13 and 0.22 depending on the joint distributional assumption for the mortality rate and total fertility.26

The estimates based on the method developed by Mausskopf and Wallace are presented in Table 3-5. Mausskopf and Wallace use data from the 1970 Brazilian census, restricting attention to women who were between 40 and 50 years old at the time of the survey. The replacement rate, assuming the mortality risk to be fixed in the population, was estimated to be 0.6 for the total sample. It was 0.35 for those with zero schooling, 0.6 for women with 1-4 years of schooling, and almost unity for women with 5 or more years of schooling. Allowing the mortality rate to differ in the population, using the method described above, only changed the estimate significantly for the lowest education group.

Approximate Decision Rules

Mroz and Weir (1989) developed a discrete-time statistical representation of the timing of births that can be viewed as an approximation to the decision rules that arise from a dynamic sequential utility maximizing model. Three stochastic processes are specified as (1) the process generating the probability of resuming ovulation after a birth, (2) the process generating the probability of conception,

---

26Olsen (1983) adds an estimate of innate mortality risk to the regression of total births and total deaths in combination with his correction method as an attempt to separate replacement and hoarding behavior. However, the effect of early age mortality risk on fertility cannot be called a hoarding response, as hoarding would not exist in an environment without significant mortality risk among older children. Controlling for innate infant frailty, however, would provide an estimate of the replacement rate that is uncontaminated by unobserved mortality risk. Olsen estimates a replacement rate of 0.17 using this method.
and (3) the process generating the onset of secondary sterility. The waiting time to a birth is the convolution of the waiting time to the resumption of ovulation and the waiting time to a conception, conditional on the resumption of ovulation and conditional on not becoming infecund. The probability of observing a woman with a particular sequence of births up to any given age is specified in terms of these three stochastic processes. Mroz and Weir allow for unobserved heterogeneity in each of the three waiting times; women may differ biologically in the postovulatory and fecund processes, and they may differ biologically and behaviorally in the conception process. However, there is neither observed nor unobserved heterogeneity in mortality risk (cross sectionally or temporally).

Monthly probabilities are modeled as logistic functions. The fecund hazard at any month depends on duration since the start of the interval, age, age at marriage, parity attained by that month (dummy variables for each attained parity), dummy variables for whether the particular month is the first month of risk of conception in the interval, a dummy for the first month of marriage, and the number of surviving children during that month. Heterogeneity shifts the monthly probability proportionately and is assumed to take on a small number of discrete values (Heckman and Singer, 1984). Identification in this model is achieved by a combination of functional form assumptions, assumptions about biological processes (for example, exactly 9 months gestation) and a clever use of data (using the timing of an infant death to tie down the beginning of the fecund period given the cessation of breastfeeding). The reader is referred to their paper for the exact details.

The model is estimated using reconstituted data between 1740 and 1819 from 39 French villages based on birth and death histories for women who were married at age 20-24. The results provide evidence on the importance of unobserved heterogeneity (in the fertility process) in the estimation of replacement effects. Mroz and Weir report that simulations conducted prior to estimation, omitting controls for unobserved heterogeneity in the fecund hazard rate and recognizing that they accounted for the cessation of lactation due to an infant death, resulted in the probability of a birth increasing in the number of surviving children (conditional on parity, age, duration, and age at marriage). Controlling for heterogeneity in estimation, however, resulted in a negative effect as is consistent with a behavioral replacement effect. Quantitatively, Mroz and Weir found that births increase by 13 percent due to the cessation of lactation alone following an infant’s death and by 17 percent overall. Given an average of about seven births, the absolute behavioral replacement effect is 0.28. Mroz and Weir essentially assume that mortality risk does not vary in the population (given covariates).

The Impact of Infant and Child Mortality Risk on Fertility

Structural Estimation

Wolpin (1984) illustrates structural estimation. The model has the following characteristics: (1) Per-period utility is quadratic in the number of surviving children in that period and in a composite consumption good, (2) fertility control is costless and perfect, (3) there is a fixed cost of bearing a child and a cost of maintaining a child in its first period of life (if it survives infancy), (4) children can die in only their first period of life subject to an exogenous time-varying (and perfectly forecasted) infant mortality rate, (5) the household has stochastic income and consumption net of the cost of children that is equal to income in each period, and (6) the household’s marginal utility of surviving children varies stochastically over time according to a known (to the household) probability distribution. Given this framework, the household chooses in each period whether or not to have a child.27

For the purpose of estimation, Wolpin assumes that the time-varying preference parameter is drawn independently over both time and across households from a normal distribution. The mortality rate faced by the household is assumed known to the researcher, measured by the state-level mortality rate in each period, and the researcher is assumed to forecast future mortality rates exactly as the household is assumed to do, namely based on the extrapolated trend in the mortality rate at the state level. Future income is forecasted from the time series of observed household income, again under the assumption that the household uses the same forecasting method.

The data are drawn from the 1976 Malaysian Family Life Survey that contains a retrospective life history on marriages, births, child deaths, household income, etc., of each woman in the sample. Wolpin used a subsample of 188 Malay women who were over age 30 in 1976, currently married, and married only once. The period length was chosen to be 18 months, the initial period was set at age 15 (or age of marriage if it occurred first), and the final decision period was assumed to terminate at age 45. Thus, there were 20 decision periods. In the implementation, the cost of a birth is allowed to be age varying as a way of capturing age variation in fecundity and in marriage rates. In addition, the woman’s schooling is allowed to affect the marginal utility of surviving children.

Parameter estimates are obtained by maximum likelihood. As already alluded to, the procedure involves solving the dynamic programming problem for each household (given their income and mortality risk profiles) and calculating

27Mira (1995) recently extended that model to the case in which families learn about the innate mortality risk they face through their realized mortality experience.
the probability of the observed birth sequence. Because the woman’s fertility is observed from what is assumed to be an exogenous initial decision period (either age 15 or age at marriage, whichever occurs first), the likelihood function is conditioned on the initial zero stock of children, which is the same for all women. The birth probability sequences that form the likelihood function can be written as products of single-period birth probabilities conditional on that period’s stock of surviving children, the output of the dynamic programming solution.

Given the parameter estimates, the replacement effect is calculated in each period and for each number of surviving children for a representative couple. The replacement effect is estimated to be small, ranging between 0.01 and 0.015 additional children ever born per additional infant death. The reason for the negligible replacement effect is that the actual fertility behavior is best fit in the context of this optimization model with utility parameters that imply essentially a constant marginal utility of children. Wolpin also calculated that an increase in the infant mortality rate by 0.05 would lead to a reduction in the number of births by about 25 percent. (Note that this effect includes the potential replacement of the increased number of infant deaths, which is in this case negligible given the very small estimated replacement effects.) Thus, the impact of a policy that altered infant mortality risk would depend quite heavily on how quickly that policy change was perceived to have been effective.28

Nonstructural Estimation

A number of studies have attempted to estimate the effect of mortality risk on fertility using nonstructural estimation methods. As already discussed, obtaining correct estimates is particularly challenging if there is unobserved mortality risk variation, and more so when mortality risk is endogenous, as when fertility spacing affects mortality risk as discussed above. Mortality risk can also be endogenous if it is affected by behaviors that are subject to choice and if, in addition, there is population heterogeneity in preferences. Several studies, having recognized this problem, have attempted to estimate the effect of innate family-specific mortality risk on fertility. To do so requires that one estimate the production function for child survival, accounting for all behavioral and biological determinants.29 Although the credibility of the estimates of the fertility-frailty relationship depends in part on the way frailty estimates are obtained, let us consider the findings of studies that estimate its effect on fertility behavior assuming the frailty estimates to be credible.

Rosenzweig and Schultz (1983), using data from the 1967, 1968, and 1969 National Natality Followback Surveys (U.S. Department of Health, Education and Welfare), find that the expected number of children ever born per woman would be 0.17 greater for an infant mortality risk of 0.1 as opposed to zero. Given that in their sample the infant mortality rate is less than 3 percent, this experiment may be within sample variation. At the sample average of 2.5 births per woman, an additional 0.25 deaths per woman leads to 0.17 more births and therefore to 0.08 fewer surviving children. Such a finding, it should be noted, must arise from replacement behavior to be consistent with the dynamic model presented above; in that model an increase in infant mortality risk cannot increase births for the same number of infant deaths.

Olsen and Wolpin (1983), also using the 1976 Malaysian Family Life Survey, estimate that a couple faced with a 1 percent higher monthly probability of death within the first 24 months of life will have their first birth approximately 2 weeks earlier. This effect is rather small given that the average interval between births is 30 months. Although seemingly inconsistent with the simple three-period model, that model is not rich enough to capture more complicated behaviors that might explain this result. For example, it is possible that greater mortality risk induces an earlier first birth so as to increase the time over which to respond to actual mortality.

CONCLUSIONS

The original question posed in the introduction to this chapter was intended to focus attention on the micro foundations of fertility behavior as a necessary prerequisite to informed population policies. It is fair to ask whether after several decades of empirical research we can confidently report to policy makers the quantitative estimates of the effects of changing infant and child mortality risk on fertility at the individual level. The answer, in my view, is unfortunately no. That assessment does not rest simply on the fact that estimates vary widely or that the empirical approaches are methodologically flawed. Rather it rests more fundamentally on the fact that we do not have a deep enough understanding of behavior to know how to generalize our results beyond the setting within which we obtain estimates. To ultimately accomplish that goal requires that we establish tighter links between theory (behavioral decision rules) and empirical methods (what is estimated).

ACKNOWLEDGMENTS

Support from National Science Foundation grant SES-9109607 is gratefully acknowledged. This chapter is in part a summarization and condensation of the
THE IMPACT OF INFANT AND CHILD MORTALITY RISK ON FERTILITY

paper "Determinants and Consequences of the Mortality and Health of Infants and Children," which will appear in the Handbook of Population and Family Economics, M. Rozenzweig and O. Stark, eds. I have received useful comments from Barney Cohen, Mark Montgomery, and several anonymous reviewers.

REFERENCES

Becker, G.G., and H.G. Lewis

Ben-Porath, Y.

Heckman, J.J.

Heckman, J.J., and B. Singer

Knodel, J.

Mauskopf, J., and T.D. Wallace

Mira, P.

Mroz, T.A., and D. Weir

O'Hara, D.J.

Olsen, R.J.


Olsen, R.J., and K.I. Wolpin

Preston, S.H.

Rozenzweig, M.R., and T.P. Schultz

KENNETH I. WOLPIN

Rosenzweig, M.R., and K.I. Wolpin

Sah, R.K.

Vallin, J., and A. Lery

Williams, A.D.

Willis, R.J.

Wolpin, K.I.