Published in European Journal of Population (online first)

Reflections on the Search for Fertility Effects on Happiness

Øystein Kravdal

Department of Economics, University of Oslo

Abstract

There have been many studies of how the number of children in a family affects the parents' or the children's lives. One strand of this research focuses on the implications of fertility for the parents' level of self-reported well-being or happiness. It is argued in this paper that an overall "happiness effect" is not very informative because of the presumably large variation in individuals' perceived gains from having children. Furthermore, it is explained that such an effect would be difficult to estimate. Most importantly, the highly varying ideas about how a child will affect life quality are important for the decision about whether to *have* a child. Many of those who have few or no children have chosen this because they think their life will be best this way, and their happiness therefore tells us little about how happy their more fertile counterparts - who to a large extent have different views about the consequences of childbearing – would have been if *they* had few or no children. This estimation problem that arises when effects of a certain event (here childbearing) are heterogeneous, and the individuals who experience that event tend to be among those for whom the effects are particularly positive or negative, is acknowledged in the treatment effect literature. However, there is little consciousness about it in the fertility-happiness research. In addition, there is a more "standard" selection problem: factors with implications for childbearing desires, or for the chance of fulfilling these, may also affect or be linked to happiness for other reasons. Unfortunately, even the most advanced statistical approaches that have been used in this research area fail to handle all these problems, so reported results should be interpreted very cautiously.

Contact details:

Address: P.O. Box 1095 Blindern, 0317 Oslo, Norway

Telephone (47)48020904; Fax: (47) 22855035

Email: okravdal@econ.uio.no

1. Background

There has been a strong interest in aggregate-level consequences of high fertility and high population growth (Headey and Hodge 2009; Panayoutou 1994; O'Neill 2010), and much research has also dealt with the possible societal implications of low fertility (Blake and Mayhew 2006; Dormont et al. 2006; Rand 2004). Furthermore, many studies have taken an individual- or household-level approach and addressed the implications of low or high fertility for the children in the family or the parents. For example, much attention has been given to the potentially adverse effects of large sibship size on children's education (Black et al. 2005; Kravdal et al. 2013; Li et al. 2008), how the number of young children affects the parents' work activity (Boushey 2008; Dommermuth and Kitterød 2009), the importance of children as providers of help to elderly parents (Brandt et al. 2009; Wenger et al. 2007), and the emotional rewards from parenthood (Eiback and Mock 2011; Nelson et al. 2013; White and Dolan 2009). In recent years, there has also been a strong interest in how the various negative and positive effects of children on parents' lives sum up to an effect on measures of overall subjective well-being, including what one may refer to as "life satisfaction" or "happiness". Many authors have pointed out that "life satisfaction" and "happiness" are not the same, and therefore also may be differentially affected by fertility (e.g. Haller and Hadler 2006), but it is not important in this paper to make such a distinction. All arguments simply refer to "happiness".

Some studies have shown a positive relationship between parenthood or number of children on the one hand and happiness on the other, but there are also several that have pointed in the opposite direction, and many authors have emphasized that the relationship varies with factors such as age, sex, marital status, socioeconomic status, and welfare support systems (see Aassve et al. (2012), Fritjers et al. (2004), Keizer et al. (2010), Myrskylä and Margolis (2012), Nelson et al. (2012), Soons et al. (2009), Vannasche et al. (2013), Zimmerman and Easterlin (2006) and studies cited in a recent review by Hansen (2011)). The implications of childbearing may also depend on how early in the parents' lives the children were born, whether they are biological children or step children, the sex of the children, and whether they have had a difficult life (Umberson et al. 2010).

As in other social science areas, the underlying aim of most of this fertility-happiness research is probably to come as close as possible to identifying a causal effect, though authors may not always be so explicit about that. A causal effect may be thought of as the difference in the outcome (happiness) resulting from experiencing versus not experiencing the event, if everything that is not a consequence of the event is fixed. It is widely acknowledged that one cannot find such an effect just by comparing, for example, two-child couples with one-child couples, because some of the factors that have led a couple to have only one child – for example a difficult economic situation or poor health - may also have implications for happiness. Researchers typically try to take this into account by controlling for such factors in various kinds of regression analysis. However, there is another problem – widely acknowledged in the treatment effect literature - that comes in addition to the mentioned selection problem, and to some extent is integrated into it (thus making it more complex than it might seem): it is difficult to estimate an overall effect of a "treatment" when the

"treatment" effect varies and the factors linked to this effect heterogeneity also affect the "treatment take-up" (e.g. Angrist and Pischke 2009; Xie et al. 2012). Translated to the fertility-happiness case, the problem would be that effects of childbearing to a large extent are foreseen and taken into account, so that those who have an additional child tend to be among those experiencing the most positive effects of childbearing. In other words, it may not be a good idea to compare with one-child couples, because they may have preferred to have only one child, and not necessarily because of a difficult life situation, but perhaps because they have other preferences than those with two children and think their life will be best this way. Their level of happiness may well be much higher than it would be among the two-child couples if *they* had only one child. In fact, they may be just as happy as their two-child counterparts. Going back to a more general level again, one may say that a problem arises if one tries to estimate an effect of an event or behavior and the probability of that event or behavior is influenced by people's expectations about that very effect – expectations that vary greatly.

This problem related to heterogeneous effects linked to the probability of the event under consideration seems not to have attracted its due attention in the fertility-happiness research (and according to Xie et al. (2012) this is true also for many other empirical social science areas). Margolis and Myrskylä (2011) very briefly hinted to it, and some authors have at least pointed out that fertility decisions are based on expected gains from childbearing (e.g. Kohler et al. 2005), but on the whole there is very little consciousness in the fertility-happiness literature about this source of bias. In principle, there would be similar concerns when estimating effects of children on more specific outcomes, but there are special reasons to worry when the focus is on an overall measure of well-being, because this is so central when making important decisions such as whether to have a child (Haybron 2003).

The objective of this paper is to explain in a non-technical way - using simple examples and no equations - the challenges one meets when trying to assess the importance of childbearing for happiness. The first step (section 2) is to elaborate on the idea that many, but not all, consequences of childbearing can be foreseen and are judged differently – leading to precisely the "treatment heterogeneity" problem that is recognized in the treatment effect literature. The second step (sections 3-5) is to give an impression of the nature of that problem by showing through hypothetical examples how a gap may arise between a simple measure such as the difference in happiness between one- and two-child couples and the true average effect of childbearing on happiness. The situation with known and varying consequences of childbearing is contrasted with the situation where the consequences are rather general or unknown. It is also explained that the true average effect, if it could be estimated, may not be very interesting. In the third step (section 6), complexity is added to illustrate what one may consider a combination of the "standard" type of selection mechanism that is well recognized in the fertility-happiness literature and the problem related to heterogeneous effects. Fourthly (section 7), an example with only a very simple selection process is constructed. The motive for presenting this example is to use it as a contrast when showing (section 8) how the methods used so far in the fertility-happiness literature deal with the supposedly problematic

effect heterogeneity. The paper ends (section 9) with a summary of the methodological challenges and a few brief reflections on the way forward.

2. Various types of consequences of childbearing and their relevance for fertility desires

Childbearing has several consequences. Many of these are probably known to people, in the sense that they at least are aware about the existence of this type of consequences, and may have quite reasonable expectations about which way they are going, without knowing exactly what will happen, of course. A trivial example is that all parents probably know that they will experience periods with less sleep than they would otherwise have had, though they cannot know precisely how sleep deprived they will be and how they will feel about it. One may see the actual outcomes as a sum of expected outcomes plus various types of additional unknown influences. The latter are returned to later in the section.

The consequences that in this broad sense are known, and therefore also are likely to influence fertility desires, are of different types. Some may be rather general, while others may vary markedly between individuals – in strength and even in sign. To start with an example of consequences that vary in sign, most people surely realize that, if they have children, they will be intensely involved in care for some time. This is probably seen as positive by many (and indeed an important reason for having children), while others may see this activity as largely a burden and prefer to do as little of it is as possible (Poortman and van der Lippe 2009). A related issue is that some may expect that the co-operation with a partner about child-raising and -caring will strengthen their relationship, while others may have a type of relationship that is adversely influenced by this (Twenge et al. 2003).

Let us now turn to the known consequences that perhaps are seen as generally positive, though to varying degree. One example is that most parents – even those who do not strongly enjoy playing with children or otherwise interacting with them – probably find parenthood rewarding in the sense that they derive emotional pleasure from seeing a child growing up, feeling needed, and knowing that they have given life to another human being who is likely to survive them (Eiback and Mock 2011; Nelson et al. 2012; White and Dolan 2009). However, there are probably different views about how *much* this positive aspect of childbearing and -rearing would add to happiness. Furthermore, many may have an idea that children may be helpful in old age (Brandt et al. 2009; Wenger et al. 2007), though there may be different views about this, and the inclination to take such uncertain and long-term implications into account may vary. In developing countries (which have been poorly represented in the fertility-happiness literature and where childbearing to a lesser extent is determined by fertility desires), also younger parents derived practical advantages from children, as they often do much household work and are involved in income-generating activities. Finally, children may have high value from a religious perspective (McQuillan 2004).

On the more negative side there is the mentioned short-term sleep problem (Dørheim et al. 2009). Also, most parents probably realize that there may be poorer work opportunities (Boushey 2008, Dommermuth and Kitterød 2009), more economic worries (Aassve et al. 2007), worries about the children's well-being, and less time for own leisure activities (Bittman and Wajkman 2000) as a result of childbearing. In total, these burdens may lead to

mental distress (Evenson and Simon 2005). However, there may be different ideas about the chance of these outcomes – depending, for example, on whether help from grandparents can be expected - and there may be different evaluations of how such outcomes would affect life quality.

Children may exert some control of the parents' lifestyle (Joutseneemi et al. 2007; Kendig et al. 2007), and parenthood may increase the level of social integration into the community (Bühler 2008; Knoester and Eggebeen 2006; Nomaguchi and Milkie 2004). Such factors, along with some of those mentioned above, probably contribute to the better health and lower subsequent death rates observed among those who have children (Grundy and Kravdal 2010). The possibility of such social effects with long-term health implications is perhaps not widely considered, but may enter into the decision for some people.

As already indicated, reality may be better or worse than expected. In particular, consequences of a type that people do take into account may be stronger or weaker than foreseen. An example about the unpredictability of sleep problems has already been mentioned. Furthermore, some may draw fewer practical advantages from children in the long run than expected because they are less in need of such help, or because the children are not willing to provide support; and some may enjoy the interaction with children more than they thought originally. In the extreme case, what has been expected to be positive consequences of childbearing may turn out to be negative, and vice versa. We may consider this difference between actual and expected consequences as additional unknown contributions to the actual happiness. Most of these probably vary in size and sign between individuals, but it is not impossible that some of them on the whole are positive (negative), which would mean that people generally tend to be too negative (positive) in their expectations about that type of childbearing consequence.

Furthermore, there may be consequences beyond what people usually take into account. For example, there is a literature suggesting that women are influenced physiologically by pregnancies (Britt et al. 2007; Fletcher et al. 2002; Rieck and Fiander 2006; Russo and Russo 2007), with positive or negative implications for later health and thus happiness (and ultimately mortality). These findings are probably not well known to the public. There may, of course, also be other such health effects that *no one* currently is aware of, and that may be revealed in future research.

In the argumentation below, it is an underlying assumption that people are able to form an idea about whether having a child on the whole is good for them, weighing the positive consequences they expect against the negative ones. This idea presumably enters heavily into the formation of their fertility desires, which in turn are an important determinant of actual fertility (along with fecundity and the ability to use contraception efficiently). In support of this rather obvious assumption, a qualitative study has suggested an association between women's and men's desires for more children and their experiences with earlier births and pregnancies and how the first years of parenthood affected their lives for better or worse (Newman 2008). Presumably, they consider these experiences an indication of what may happen if they have another child.

The fact that some consequences of childbearing are felt soon and others only in the longer term makes it particularly difficult for an individual to form an overall conclusion. For example, if parenthood on the whole is expected to reduce happiness during young adult years and improve it in old age (as indicated by a number of studies, including Margolis and Myrskylä 2011), it is not obvious what the overall judgment would be. In the discussion below, this timing aspect is ignored. The level of happiness is considered as constant except for an immediate change after childbirth, if any. Estimation of a "happiness effect" is problematic enough even with that simplification.

3. Example 1: The least problematic type of variations in effects of childbearing

Let us assume that there are 300 couples who already have one child and consider having another. This is a highly relevant perspective, given the current below-replacement fertility in Europe. Some of the consequences of having another child that were reviewed above may be more relevant for the childless than for those who have a child (as indicated by Kohler et al. (2005), who found more positive happiness effects of first-born than later-born children). However, the basic picture is the same, and it is easier to discuss one-child couples than childless couples or individuals, among whom childbearing is more deeply intertwined with changes in partnership status, which also are important for happiness (Soons et al. 2009). The *timing* of this second birth may also be an issue the parents consider, but that is ignored for simplicity.

Let us further assume that there are 200 couples among the 300 who expect that a second child would give them (in total, as a couple) a level of happiness 5 units higher than if they remain one-child parents. Therefore, they would also be particularly likely to want a second child. (Social expectations and considerations about how a child may affect grandparents' or other people's happiness may also enter into the decisions - if not embedded in the parents' evaluations of their own happiness – and less rational ideas may play a role.) Let us call these couples Group 1. Furthermore, let us assume that the remaining 100 couples (Group 2) have the opposite view: they think life will be better without a second child than if they have a second child, the difference again being 5. Let us also assume that other contributions to expected happiness are the same for the two groups, in the sense that the couples in the largest group expect a level of 17 if they have a second child and 12 if they do not, while those in the smaller group expect a level of 12 if they have a second child and 17 if they do not. These numbers are, of course, completely arbitrary. It would seem reasonable that the intensity of the views about childbearing varies, for example that some couples expect a child to change happiness by only 3 in either direction. However, this is not essential for the arguments that follow and therefore ignored for now (but see some elaboration below). Finally, let us assume that the couples are correct in their happiness expectations. Thus, to summarize, *expectations* about effects of childbearing affect fertility desires, and the expected happiness associated with each potential fertility outcome is the same as the happiness that actually *is experienced* as a result of these outcomes (i.e. the actual happiness levels are also 12 and 17).

The symmetry in the evaluations is meant to symbolize that one fertility outcome is not considered generally inferior to the other. People simply have different ideas about what is good for them. Those who think it is best to have one child and also have one child are just as happy as those who prefer two children and have two children, and failure to attain the preferred fertility level results in the same welfare loss for everyone. This simple example should represent well enough a part of reality, and therefore be a relevant starting point.

Let us now for an illustrative purpose also make the assumption – very unrealistically - that the expectations about childbearing consequences, and thus presumably the fertility desires, have no impact on actual childbearing: the chance of having a child is 50% for everyone. In that case, 100 of the 200 couples in Group 1 (let us call them Subgroup 1a) will have two children and a happiness of 17 (as opposed to 12 if they had not had the second child), while 100 couples (Subgroup 1b) will have one child and a happiness level of 12 (as opposed to 17 if they had had the second child). In Group 2, a subgroup (2a) consisting of 50 will have two children and a happiness of 12 (as opposed to otherwise 17), and 50 will have one child and a happiness of 17 (as opposed to otherwise 12). This situation is illustrated in Table 1.

(Table 1 about here)

An intuitively reasonable definition of an "effect of childbearing on happiness" is: the difference in happiness that a couple would experience if they had a child compared to if they did *not* (everything else being the same, except any consequence of childbearing). This is also what is called a "treatment effect" in what we may refer to as the treatment-effect literature (for basic terminology and ideas, see e.g. Angrist and Pischke 2009). In the appendix, some main ideas in that literature are summarized in a more formal language, and the examples presented in this paper are specified mathematically.

Because "treatment effects" typically vary across observation units - which is indeed also the case in the examples considered here - the key intention is typically to identify a so-called "average treatment effect" (ATE). Within Group 1, the situation is simple: the "treatment effect" is 5 for everyone, and thus the average is 5 as well. Similarly, the "treatment effect" is -5 for everyone in Group 2 and the average also -5. If we instead consider the entire sample, the situation is slightly more complex: the "treatment effect" is 5 for 200 couples and -5 for 100 couples, so the "average treatment effect" in the sample is 1.67. This positive figure reflects that the second group is smaller, i.e. that there are fewer who think that having one child is the best solution than who think two children would be most advantageous. It is referred to below as the "true overall effect" to avoid leaning heavily on the terminology in the treatment effect literature.

Most of the analysis of the fertility-happiness effect has, at least until recently, been based on a simple cross-sectional design. Applied to the situation described here, the essence of such an analysis is to compare the happiness of two-child couples with that of one-child couples. Typically, this is done in a regression approach that allows control for factors affecting both fertility and happiness, but that kind of complexity is left for later sections. Given the assumptions above, how large is the difference in happiness between one- and two-

child couples? Among the former, the average happiness is 13.67, while it is 15.33 among the latter. This gives a difference of 1.67, which we recognize as the true overall effect. If the chance of having a child had been different from 50%, but the same in both main groups of couples, the conclusion would have been the same. In other words, when childbearing consequences and expectations about them vary as in this example, and fertility is unrelated to these expectations, the difference in happiness between those with two children and those with one child is the same as the true overall effect. The happiness among those with one child gives a correct impression of how life would have been for those with two children if they had not had that child, because those who have had an additional child and those who have not are the same kind of people. In both these groups there are two-thirds expecting positive consequences of childbearing (coming from Group 1) and one-third expecting negative consequences (coming from Group 2); there is no systematic variation in the "assignment to treatment".

4. How meaningful is an average effect of childbearing?

The average effect of childbearing may not be a very interesting measure if there is much variation, as in the foregoing example and as we may also see in reality. In principle, presentation of such a measure could lead to unfavourable behaviour changes, because individuals informed about the "effect" might revise their expectations about the happiness resulting from childbearing upwards, and if they were at the margin with respect to fertility desires, they might end up wanting and having a child that it would not really be in their own interest to have. Stated differently, they knew what they were doing, and information about an "effect" has not had any beneficial impact. Conversely, if those who expected adverse consequences of childbearing constituted a majority, the average "effect" as calculated above would be negative, and some of those barely wanting a child might be pushed to the other side against their own interests.

Some might also consider a positive happiness-effect measure as justifying pronatalist policies: by subsidizing childbearing, so that more people have a child, the well-being of the population would increase (unless the resources allocated to this might have produced larger welfare gains elsewhere). However, this is not a good argument. Making childbearing less expensive could correspond to adding, say, 6 to the happiness for those who have a child. The two groups consisting of 200 and 100 couples would then not rank the outcomes as 17/12 (a short-form for 17 if the couple have two children and 12 if they have one) and 12/17, but 23/12 and 18/17. The first group of 200 couples would still want a child, which would give them an even higher level of happiness, and also the second group of 100 couples would want a child, because this would now be a marginally better outcome. They would have a level of happiness 1 higher than without the policy. Obviously, there would be a rationale for introducing such a policy even if the first main group consisted of 100 couples and the second consisted of 200 (i.e. reversed relative size), so that the overall effect had been -1.67 instead of 1.67. The only difference would be that the total happiness gain would be smaller (100 would gain 6 from the policy and 200 would gain 1, rather than the opposite). To conclude, it

is very unclear what an estimate such as 1.67 would tell us that could have any relevance for policy.

5. Example 2: The variations in effects of childbearing that cause trouble

As already pointed out, the assumption made above about fertility not being influenced by the expected consequences of childbearing is far-fetched. In real life, those who think a child makes life better are also particularly likely to want a child and actually *have* a child. To build that into the example, let us assume that couples who expect most benefit from childbearing (Group 1) have 90% probability of having that child and 10% probability of remaining one-child couples.[1] Furthermore, let us assume that 70% of those who think childbearing will make them less happy (Group 2) succeed in avoiding further childbearing, while 30% have an unintended birth. These numbers are shown in Table 2.

(Table 2 about here)

The average happiness among those with two children is, of course, higher (16.29) in this situation than in the "random-childbearing" situation described in the preceding section (15.33), since those who have a second child to a larger extent are among those who expect positive consequences of childbearing. Also the level of happiness among those who remain one-child couples is higher than in the situation with "random" childbearing (15.89 vs 13.67), for the same reason: this fertility outcome largely occurs among those who consider it the best. Even more interestingly, the difference in the average happiness between one- and two-child couples is only 0.40 and no longer the same as the true overall effect.

Among those who have a second child and thus attain an average happiness of 16.29, the happiness would only have been 12.71 if they had not had this child. The difference between these two numbers, which is 3.58, is referred to as the "average treatment effect among the treated" in the treatment effect literature. (Obviously, the happiness of 12.71 associated with this hypothetical situation is lower than the happiness of those who end up with one child, among whom a much larger proportion have preferred to have only one child.) The fact that the "average treatment effect among the treated" is much larger than the "average treatment effect" reflects, of course, that those who have been exposed to the "treatment" are also particularly likely to have much effect of it; it is the anticipation of that large effect that has driven many of them into the "treatment group".

To summarize, because of the type of variation in the chance of having a child in this example, the difference of 0.40 between one- and two-child couples is neither equal to the "average treatment effect" (1.67) nor the "average treatment effect among the treated". As explained earlier, the effect of 1.67 would be the most interesting to know about, though still rather meaningless given the large variation around this average. The "average treatment effect among the treated" (if it could be estimated) would tell us what those who do have a child gain from that, and one might think of purposes for which this could be interesting but, again, an average conceals variation. However, it would be even more meaningless to present

0.40 as the effect. This number has no straightforward interpretation. It reflects in a complex way the variations in (perceived) childbearing consequences and the degree to which these are translated into fertility desires and actual childbearing.

The assumed heterogeneity in these examples can be considered somewhat extreme, because of what was referred to above as symmetry (the happiness levels being 12/17 vs. 17/12). However, by making some changes in the examples it is easy to show that the same problem arises if there is less difference in the effects (e.g. one being 5 and the other 3), or if the effects are still 5 and -5, but without the same symmetry (e.g. 12/17 vs. 12/5).

5.1 Contrasting with the situation where there are general or unknown effects of childbearing

To elaborate on the underlying substantive issues, let us go back to the distinctions made in section 2: some effects of childbearing are unknown (therefore not taken into account in the decision-making either) and can be either rather general or vary considerably between couples; others can to a larger extent be considered as known (therefore potentially taken into account), and these can also be general or varying. The problem described above – with the observed difference of 0.40 between one- and two-child couples not being the same as the after all more interesting overall effect of 1.67 - arises only in the latter situation, i.e. when people are aware of certain consequences of childbearing and evaluate these differently and these evaluations have implications for fertility. If the consequences of childbearing vary but are *unknown*, and therefore not taken into account in the decision-making, we would essentially be in the same simple situation as in Example 1 (Table 1): the difference in happiness between one- and two-child couples would be equal to the true overall effect. (In that example it was assumed, very unrealistically, that expected consequences vary and are the same as the actual consequences, but do not affect fertility. The situation with unknown and varying consequences could be illustrated by just having “actual” happiness rather than “expected assumed equal to actual” happiness as the label of the fourth column of Table 1). Consequences that are unknown and *general* can be seen as forming a special case and would be just as unproblematic. Also consequences that are *known* and *general* cause no trouble. This can be illustrated (at the risk of being seen as overly pedantic) by assuming that there are not two groups of couples with different attitudes, but that Group 1 is the only one. Then, a happiness of 17 would be observed for those with two children and a happiness of 12 for those with one child, the difference being 5 just as the true effect.

As already mentioned, all consequences of childbearing are, in real life, not of the same type; there is a mixture. Some vary and are largely known and taken into account (thus creating the problems just described) and some are largely unknown or rather general (and would be unproblematic if they ruled the ground alone). Let us consider the implication of this for the estimation and first look at the case where there is an additional effect that is unknown and general. Assume, for example, that politicians unexpectedly decide to give a bonus to all those who have a child (e.g. by subsidizing childbearing costs). This would correspond to adding the same number, say 1, to the actual happiness for everyone who has a

child. The difference in happiness between one- and two-child couples would then be 1.40 rather than 0.40 in our example. In other words, the difference reflects the general unknown consequence plus the rather uninformative measure. Note that, when seeing an estimate such as 1.40, one would not be able to tell what the general effect is and what the other contribution is.

Most childbearing consequences that come in addition to those foreseen probably do vary between couples, however. That would be the case, for example, with the physiological effects, which we could consider stochastic. If we assume that childbearing increases the chance of certain positive health outcomes, some couples may experience a health advantage corresponding to a happiness addition of 0.5 if they have a child; others would get an advantage of 1.0 or 1.5 or 0. Such beneficial health effects might influence Group 1 and Group 2 in the same way (if unrelated to any factors that determine views about childbearing consequences). Then, the difference in happiness between one- and two-child couples would be 0.40 plus the average of this unknown effect.

The third possibility, in principle, is that there are additional consequences of childbearing that are known and general. That would correspond to adding a certain amount, for example 2, both to the expected happiness associated with having a child and to the actual happiness for those who do have a child. In a more complex example than described here, with a larger number of main groups having different views about the consequences of childbearing, such an addition could move some groups into wanting rather than not wanting a child, but this would not fundamentally change the situation: the difference in happiness between one- and two-child couples (now larger), would tell us little about the true effects.

To summarize, there are various types of childbearing consequences. Some are taken into account in people's decision-making, and the evaluation of them varies, while others are more general or unknown. It is the former that are responsible for the estimation problems, and that therefore are in focus also in the remaining discussion. Summarizing instead from a more technical perspective, we may say that the treatment effect literature describes a problem that arises if "treatment effect heterogeneity" is linked to the "take-up of treatment". A key message from this and the earlier sections is that this problem is highly relevant also for fertility-happiness research because some (but not all) of the consequences of fertility vary and are probably also to some extent known and therefore taken into account when making fertility decisions.

6. Example 3: some additional realistic features incorporated

A real population does not consist only of two groups at the same overall happiness level, but with opposite evaluations, such as in the preceding example. In particular, there are surely some who are generally less happy, and whose evaluation of the childbearing consequences may differ from that of others in the population. Let us now build this into the example (Example 3; see Table 3) by assuming that there are some couples who are poor and therefore also more likely than others not to have a very happy life. Obviously, there are differences in

the views about childbearing also within this group of poor couples, just as among the others. Some may expect a happiness of 6 if they remain one-child couples and 11 if they have a second child (i.e. an effect of 5), while another subgroup of the poor may reach the opposite conclusion (i.e. an effect of -5). Let us further assume that the poor are more likely than the non-poor to be in the group where the effect is -5: half of the 100 couples with negative views about further childbearing are poor, while only 10% of the 200 couples with positive views are poor.[2] Let us also assume that, regardless of poverty, the chance of having a child is 90% among those who are most positive to childbearing and 30% among the others. In other words, poverty affects fertility (only) by influencing the chance of expecting positive rather than negative consequences of childbearing.

(Table 3 about here)

This situation may give associations to a "standard" selection problem, in the sense that there is a factor (poverty) affecting both fertility desires and happiness. However, the example should rather be seen as illustrating a combination of a "standard" selection problem and the problem due to differences in evaluations of childbearing consequences that is dealt with in preceding sections: people make widely different judgments, but the degree to which they land in one camp rather than the other depends on certain factors that also are important for their general happiness level. There is much realism in this example, as effects of various factors on fertility would often result from different views about the consequences of childbearing, which in turn affect fertility desires. Another influence on fertility could be through the chance of *realizing* the fertility desires. Such a simpler selection problem that is *not* in the same way mixed with the problem related to the differences in the expected consequences of childbearing is illustrated in Example 4 below.

With a situation such as described in Example 3, the overall effect of childbearing on happiness ("average treatment effect") is still 1.67, but the difference in happiness between one- and two-child couples is 1.92. The reason why that difference is much more positive than in Example 2 (0.40) is, of course, the very low happiness of those with one child, who to a large extent come from the subgroup who are poor, in which the level of happiness is generally low.

7. Example 4: A simpler selection problem

It was assumed in Example 3 that the chance of actually having a child, given the ideas about childbearing consequences and thus childbearing desires, is independent of poverty. That is not necessarily reasonable. Poverty, which may be linked with low education and poor health, could be associated with infecundity or inadequate use of contraception. Rather than adding this feature to the already quite complex Example 3, let us set up an example (Example 4, Table 4) where there is no heterogeneity in the evaluation of childbearing consequences.

More specifically, assume that 200 couples fully agree that having a second child reduces their happiness by 5, so in that sense they resemble the 100 couples in Group 2 in the

earlier examples. However, while 150 are non-poor and expect a happiness level of 17 if they avoid having another child and 12 if they have another child, there are 50 poor couples among whom the corresponding happiness levels are (as in Example 3) generally lower, 11 and 6. Let us further assume that these poor couples also are less able to avoid having the unwanted second child: 50% have a second child, as opposed to 10% among the non-poor. Then, the difference in happiness between one- and two-child couples is -7.81, which is clearly different from the true overall effect of -5. As shown below, this particular type of bias (if alone) is not difficult to get rid of.

(Table 4 about here)

8. Procedures that have been used in studies of the fertility-happiness effects

8.1 Cross-sectional multivariable regression

At least until quite recently, most studies have essentially estimated the fertility-happiness effect by comparing the happiness level of persons who have n_1 number of children with the happiness level of persons who have n_2 number of children. The models usually control for some other factors, and many authors have additionally stratified their analysis by, for example, age or socioeconomic status or alternatively estimated the corresponding interaction effects (see e.g. Margolis and Myrskylä 2011).

One may think of such "standard" cross-sectional multivariable regression analysis as producing effects of childbearing for each level of each covariate that is controlled for and then averaging these – with certain weights - up to one effect. There will, of course, be less variation in the evaluation of childbearing consequences within a specific covariate level than in an entire national population, but there will surely be some, so the concerns that are mentioned are still valid.

Let us see what happens if reality is as described in Example 3 and a linear regression model is estimated.[3] If having a second child is the only independent variable in the model, the corresponding coefficient is, of course, the same as the difference between one- and two-child couples mentioned earlier: 1.92. If also poverty is included the estimate is 0.29. Put differently, the difference between one- and two-child couples, which may be seen as a very naïve estimate of the effect of childbearing, increases (from 0.40 to 1.92) when the difference between poor and non-poor is added to the example, while the true overall effect of childbearing remains the same, and since inclusion of poverty in the model reduces the estimate, we could say that this control to some extent takes away the additional bias produced by the poor/non-poor difference. There isn't much value in this, however, as we are no closer to knowing what the true effect is.

What would happen in the very hypothetical situation where we could observe the expectations about childbearing consequences that influence people's decisions, and there are no other factors that both affect fertility and are linked to the actual consequences of childbearing? This would essentially mean that we could include a Group 1 vs. Group 2

dichotomous variable in the model, i.e. control for the group affiliation. Returning to the simpler Example 2, the estimate of the effect of having a second child would no longer be 0.40, but -0.38, if a group variable is added. One might think that a regression of this type would be like estimating one effect for Group 1 (the effect being 5) and another for Group 2 (the effect being -5) and then getting the true population average 1.67 as the resulting estimate, but the weighting is different. For example, while Group 1 includes 200 couples who would have a happiness of 17 if they had a child and 12 if they did not, we observe 180 in that group who actually have a child (with happiness 17) and 20 who do not (happiness 12). If we performed a weighted regression, with weights set so that there were 200 observations of each of these two types and 100 corresponding observations of each type in Group 2, the estimate would be 1.67. Using the more complex Example 3, we also get an estimate of -0.38 if a group dummy is added, alone or in interaction with poverty.

In other words, even with rather unrealistic data on the key underlying heterogeneity, a simple regression would not give us the correct answer. This also means that moving a step in this direction by adding the "ideal number of children" or similar measures into the models (Stanca 2012) may not help much. That said, if there really *were* data that allowed observation of Group 1 and Group 2, one could analyse these groups separately and then calculate an overall effect from the estimates 5 and -5 using the appropriate population weights (200 vs. 100). More importantly, one could take yet another step and report these two effect estimates and the size of the groups. As argued above, this would be much more informative and less misleading than reporting the overall effect.

As a contrast to the situation with varying consequences of childbearing, let us consider the case with a very simple type of selection, illustrated in Example 4. An effect of -7.81 appears, of course, in a regression model if having a second child is the only variable. If poverty is added, however, the correct effect of -5 is estimated. In other words, this simple selection is dealt with adequately.

If there are differences between poor and non-poor as in these examples, but no poverty variable available in the data – or if there are other such unobserved differentiating factors – one may resort to other techniques. Twin fixed effects can be used to control for unobserved factors shared between twins, or a longitudinal (within-individual or within-couple) analysis can be carried out to control for individual unobserved variables that are constant over time. These alternatives are addressed below.

8.2 Twin fixed-effects analysis

If there is not information about poverty in the data available to the researcher, and it can be assumed that poverty reflects family background factors and genetic traits, it would make sense to compare among twins, and preferably monozygotic rather than dizygotic. The idea (admittedly not very convincing) is then that twins are at the same poverty level also when they are adults. Besides, they are similar - presumably even more so - with respect to a

number of other characteristics that may affect both fertility and happiness and that it might be hard to get information about.

Such an analysis of monozygotic twins was carried out by Kohler et al. (2005), who concluded that many of the results were quite different from those obtained with "standard" regression. Since many probably do not have a clear intuition of the approach, it may be worth presenting it in some detail. It was assumed in Example 3 that 200 couples think (rightly) that having an additional child would increase their happiness by 5, either from 12 to 17 or (for the few who are poor) from 6 to 11. There are also 100 who expect childbearing to reduce their happiness by 5, from 17 to 12 or (for the much larger group of poor) from 11 to 6. Assume now that these four groups of couples are representative of one-child women in a larger population, that childbearing attitudes and poverty reflect family background factors that are shared between twins, and that the twinning rate is independent of such factors. Then, at a certain level of the twinning rate, there will be 1800 pairs of non-poor twins who agree that the happiness outcomes resulting from having/not-having a second child are 17/12, 200 pairs of poor twins from whom the outcomes are 11/6, and 500 pairs of non-poor and 500 pairs of poor twins who have the opposite view of childbearing (12/17 among the non-poor and 6/11 among the poor). We have now turned from considering couples to considering women, but that does not change the essence of the arguments. Everything discussed so far could have been based on women – or men – as the unit rather than couples without undermining the fundamental logic. The expansion from 300 couples to 3000 twin pairs is done for a very trivial reason: to avoid decimals further down in the discussion.

Let us now consider the first group of 1800 non-poor twin pairs in more detail. These women have positive attitudes to childbearing, and we assume as earlier that they have 90% probability of also having a second child (regardless of whether the twin actually has a child or not). Then, there will be 18 twin pairs (1%) consisting of two women with one child, 324 pairs (18%=$2 \cdot p \cdot (1-p)$, where p is the individual 90% probability of having a child) consisting of one woman with one child and one woman with two children, and 1458 (81%) twin pairs consisting of two women with two children. In total, 3240 (90%) of the women have two children and 360 (10%) have one child. These numbers, which are shown in Table 5, are 20 times larger than the corresponding numbers in Example 3 (because 10 times as many twin pairs as couples are considered and because there are two women within a twin pair).

(Table 5 about here)

Similarly, among the poor there are 360 who have a second child and 40 who remain one-child mothers – again 20 times the corresponding numbers in Example 3. In the other main group, consisting of women who are not so positive to childbearing and who have only 30% chance of actually having a child, there are 300 women with two children and 700 with one child both among the poor and the non-poor (9% of the twin pairs consist of two women with two children, 42% consist of exactly one woman with one child, and 49% consist of two women with one child).

If a model is estimated for all women in these 3000 twin pairs, and only the number of children is included, the effect of this variable is, of course, 1.92 as in the example above. If

instead a fixed-effects model is estimated by adding a dummy for each twin pair, the estimated effect of having a second child is -0.38 (just as in the "standard" regression if "Group" could be controlled for; see section 8.1). If the poor twins had been exactly like the non-poor (i.e. a situation as in Example 2), a "standard" regression based on the women in the twin pairs would have given 0.40 (just as the difference between one- and two-child couples in Example 2), while a fixed-effects model again would have given -0.38. In other words, one could consider the fixed-effects approach as succeeding to some extent in dealing with the additional bias that comes from the selection due to (in this example unobserved) poverty, in the sense that the estimate is brought down from a high positive level and is the same regardless of whether there is a group of poor or not. However, there are remaining problems: the estimate (-0.38) is still different from the true effect of 1.67.[4]

In the unrealistic situation where the chance of having a child (p) is the same among all women, regardless of their ideas about the consequences of childbearing, the difference in happiness between the woman with two children and her twin with one child is 5 for the $2000 \cdot 2 \cdot p \cdot (p-1)$ twin pairs that are positive to childbearing (recall that $2 \cdot p \cdot (p-1)$ is the proportion of the twin pairs among whom the two women have different outcomes). It is -5 for the $1000 \cdot 2 \cdot p \cdot (p-1)$ twin pairs that are negative to childbearing. With this 2:1 relationship between the main groups, we again end up with 1.67 as the overall effect. In other words, it is the (trivial) fact that the differences in expected gains from childbearing have effects on actual childbearing that creates the problem, just as observed above with a simpler analysis.

An additional complicating factor when using such an approach is that there is an aspect of reality that is not reflected in the presentation so far: if the two twins have the same ideas about childbearing consequences, the difference in actual fertility is a result of one of them having had one more child or one less child than wanted. That might be associated with characteristics (woman-specific and not shared between twins) that also are important for the later happiness. For example, an unwanted child could be due in part to modest socioeconomic resources, which could reduce happiness. Faced with such a reality, one would get a more positive estimate. Yet another source of bias is that twins may not necessarily share the childbearing preferences, contrary to the assumption above. For example, some pairs may include one woman who evaluates the consequences as 17/12 and another who evaluates them as 12/17.

It should also be noted that a fixed-effects analysis, not surprisingly, works well (i.e. gives the correct estimate of -5) if the situation is as in Example 4 and the same assumptions as above with respect to twinning rates and sharing of characteristics between twins are made. It should be unnecessary to spell the argument out in detail. In other words, when there is homogeneity in the expected value of childbearing (or an heterogeneity that is not linked to the chance of having a child) and there are unobserved factors shared between twins that influence the chance of actually having a child as well as the level of happiness – just as poverty in Example 4 - a twin fixed-effects analysis can be applied. (If the factors were observed, it would be easier to control for them in a "standard" regression, and if there are factors not shared between siblings that affect fertility and happiness, there would be a bias also with a twin fixed-effects approach.)

8.3 Longitudinal within-individual or within-couple analysis

An alternative and increasingly common approach is to focus on those who have a child and examine how their happiness has changed over time before and after birth (see, for example, Angeles 2010; Clark 2008; Fritjers et al. 2004; Zimmermann and Easterlin 2006; Soons et al. 2009; Keizer et al. 2010; Myrskylä and Margolis 2012). As mentioned earlier, however, possible differences between short- and long-term effects of childbearing are for simplicity ignored in this paper, so let us just say that there is one measurement before and one after having had a (second, in our case) child. The idea behind a comparison of these measures is that there may be something "special" with those who have a second child that would make them relatively happy (or unhappy) anyway, and that the time-invariant parts of these factors are controlled for, while they would contaminate a comparison across individuals. Additional factors varying over time and suspected to be determinants of or co-determined with fertility may be controlled for, as well as any general drift in happiness over age or period. However, to the extent that there is variation in effects of childbearing among those who have and those who do not have a child, one could only hope to get a measure of the effect of childbearing among the former (i.e. "average treatment effect for the treated"). Also, an important limitation is that such longitudinal data typically cover too few years to allow estimation of implications of childbearing in the long run, should there be an interest in those (though more than 20 years were included in the recent study by Myrskylä and Margolis (2012)).

As a simple illustration of how the method works, let us assume that happiness is measured twice, that all couples have one child at the first measurement, and that it is possible to have a second child before the second measurement. Let us also assume that everything is as in Example 3 except that happiness refers to the time of second measurement and that the happiness at that time if they had not had a child (which they may not have had either) is the same as they felt at the time of the first measurement. Then, 162 non-poor couples would have a child and experience an increase from a happiness of 12 to a level of 17; 18 non-poor couples would not have a child and remain at 12; 18 poor couples would have a child and see an increase from 6 to 11; 2 poor couples would not have a child and remain at 6; 15 non-poor couples would have a child and reduce their happiness from 17 to 12; 35 non-poor couples would not have a child and remain at 17, 15 poor couples would have a child and reduce their happiness from 11 to 6; and 35 poor couples would not have a child and remain at 11. When we focus on those who have had a second child, the difference in the happiness between the two measurements is, by definition, equal to the "average treatment effect on the treated" (even now when there are both poor and non-poor, with different happiness levels), and we see (as in Example 2) that it is 3.57.

To understand better what has been achieved, consider instead an analysis where all measurements – two for each couple – are pooled together and form the basis for a regression on a second-child indicator. Its coefficient would then be 2.50 and wouldn't tell us much of interest. (In a simpler situation where the poor are exactly like the non-poor, the corresponding estimate would be 2.68.) However, by taking the longitudinal approach and

considering the difference over time among those having a child (using the entire sample and adding couple dummies would actually give the same) we get at least a result that has an interpretation: the "average treatment effect among the treated". The difference between 3.57 and the true overall effect of 1.67 reflects, of course, that if the 35 + 35 couples who do not expect to benefit much from a second child and remain one-child couples *did* have a second child, they would have contributed negatively (each with -5) to the estimate. There are also couples who have not had the additional child they would prefer, and who would have contributed positively had they had this child, but this group is smaller (18+2 couples).

Again, Example 4 can stand as a contrast. A similar longitudinal approach built on that example would give four groups of couples – one not having a child and remaining at a happiness of 17, one having a child and reducing the happiness from 17 to 12, one not having a child and remaining at 11, and one having a child and reducing the happiness from 11 to 6. The effect would be correctly estimated as -5. In contrast, a "standard" regression based on two observations for each couple would give -7.50, but if poverty could be observed in the data, and therefore controlled for, also this set-up would give the correct result.

9. Conclusions

9.1 Summarizing the problems

Childbearing is quite obviously a matter of taste to a large extent. Some expect that an additional child will give them much happiness, and may be right about that, while others – perhaps in a rather similar life situation but with other preferences – may have the opposite attitude. When the (perceived) implications of childbearing differ widely, an "effect" – defined as the population average of the difference between the level of happiness a couple would attain if they had a child and the level they would attain if they did not have a child (in econometric jargon the "average treatment effect") - is not very informative. For example, a weakly positive effect may conceal the fact that some people may derive large advantages from having a child while a smaller group may gain just as much from *not* having a child.

Another problem is that it in practice would be hard to estimate such an overall effect. One reason is precisely this variation in the ideas about the consequences of childbearing, coupled with the trivial fact that these ideas are taken into account when making decisions. (If the consequences of childbearing were unknown, and therefore not taken into account in the decision-making, or if they did not vary, there would not be similar estimation problems.) The core issue is that most of those who have one child may have wanted this fertility outcome and are more happy with it than the two-child couples would have been had *they* had only one child (and perhaps just as happy as the two-child couples who have wanted the second child).

Another reason why estimation is difficult is that factors of importance for happiness also affect people's views about whether childbearing is good or bad for them, and that they affect the chance of having a child given these views about childbearing (through fecundity or quality of the contraceptive use). If only the second of these problems existed, it could be

solved by controlling for the relevant confounding variables, or in the absence of information about them, one could control for some of them through a twin fixed-effects analysis or a within-individual fixed-effects analysis. However, these two more advanced approaches, which are popular tools in social science these days, may be considered to deal only to some extent with the first kind of selection, since it is so intertwined with the problem related to differences in the expected consequences of childbearing. Moreover, even if such an additional source of bias could be isolated, it would be of little value as the other problem remains.

9.2 Better methods or data?

The main problem described in this paper has long been recognized in the treatment effect literature, and several examples have been suggested in economics and other fields, such as the possibility that effects of education differ between individuals, are known to these individuals, and are taken into account by them when they make their decisions about schooling (Blundell et al. 2005). One might consider using an instrumental-variable approach, which means that the analysis is based on one or more variables that affect the key independent variable (in our case having a second child) but not the outcome variable (in our case happiness) more directly. The latter is the so-called exclusion criterion. However, if such a variable exists, one would only (and even based on certain assumptions) be able to estimate an average effect for the observation units (in our case couples) for whom the exposure to the "treatment" (birth of the second child) can be induced by the instrument; one would not identify an effect for those who would never experience the "treatment", or who would experience it regardless of the instrument (Blundell et al. 2005). This is the so-called "local average treatment effect" (Imbens and Angrist, 1994). Blundell et al. (2005) made the point that an overall effect ("average treatment effect") in principle can be estimated by using a so-called control function method, which assumes a common unobserved factor behind the "treatment" (second birth), the outcome (happiness) and the effect of the "treatment" on the outcome. However, this approach would also require an exclusion restriction.

Unfortunately, it is difficult to find a variable satisfying the exclusion criterion. Fertility is influenced by the social and economic resources of the involved individuals, their health, their general values, and similar factors at the community level, and all these are also likely to be important for happiness. Some studies of various types of effects of childbearing have relied on twin births (Black et al. 2005; de Haan 2010;), miscarriages Maralani 2008), or sex composition of older siblings (Angrist et al. 2010; Conley and Glauber 2006; Goux and Maurin 2005) as instruments, but these are all problematic, as one can easily think of a number of ways in which they could influence the welfare of the parents. For example, twins often have low birth weight and may therefore, especially in poor settings, be weaker; miscarriages may be psychologically burdening or indicate health problems; and children's sex could affect parents' lifestyle and resources, with wide-reaching consequences. (For a critique of these instruments, see e.g. Rosenzweig and Zhang 2009 and Åslund and Grönqvist 2010).

Xie et al. (2012) argued that it could be helpful to check whether the "treatment" effect varies by estimating the interaction between the "treatment" variable and the predicted propensity to "take the treatment". This would not solve the problem, but give an idea about how serious it could be - though only to some extent, as the propensity score is calculated from observed variables exclusively (i.e. given the propensity score, there may still be unobserved variations in expectations about the treatment effect, which could be linked to the actual effects and to the chance of taking up the treatment).

In principle, an alternative and to some extent complementary strategy would be to collect richer data. Since the core problem is that the happiness of those who have only one child is a poor measure of how a two-child couple would have fared if *they* had only one child, and vice versa, one might consider asking all couples how they think their life would have been if they had had fewer or more children. However, it is not obvious that they would be able to give an answer that comes closer to the truth than an assessment based on the actual observations of happiness; evaluating consequences of hypothetical outcomes is indeed difficult. Another approach could be to take a prospective perspective and ask younger people how they value further childbearing compared to not having more children, but that also would be problematic. While we would learn about the perceived consequences of childbearing that do influence their actual decision-making, they may be poorly informed, in the sense that what they assume to be consequences of the various outcomes might never have happened.

A more optimistic view is that, even though such data may not themselves give us better estimates, they may shed some light on the amount of variation in the expected consequences of childbearing and the strength of the association between these views and fertility. This can be used in simulation experiments to get an idea about the magnitude of the bias produced by the currently used methods.

9.3 The broader relevance of the problems discussed

The problems discussed in this paper have relevance also for studies of more specific consequences of fertility. For example, one may be interested in how childbearing affects the parents' work activity or health, or how the children's socialization (Downey and Condron 2004) or schooling careers (Black et al. 2005; Kravdal et al. 2013; Li et al. 2008) are influenced by the number of siblings. To elaborate on the latter, this kind of consequence typically varies between families: possibilities for educating children vary, and people may have different ideas about how important education is for the children's and their own well-being (the former probably also being a crucial factor in the fertility decision-making, perhaps partly by being accommodated into the parents' evaluation of their own well-being). Assuming general agreement about education being important, families with many children may to a particular extent include those who think they can afford to educate so many children, and who may well be right about that. Thus, if those with a more moderate number of children had an unwanted child, there might be somewhat more adverse effects of that on the children's education than the difference between high- and medium-fertility groups would

suggest. Whereas all research on how sibsize affects education takes into account - with varying degree of sophistication - that many individual and community factors may lead to high wanted or unwanted fertility as well as low education, this additional problem due to subjective evaluations of how childbearing may affect education and the link between these and the fertility desires is not considered.

However, there is one important difference between such studies of more specific outcomes and analyses of the fertility-happiness effect: while the more specific consequences of childbearing may vary between individuals and may be taken into account, they are presumably not as strongly linked to variations in childbearing desires, and thus actual childbearing, as the expected consequences for happiness. Each of these consequences, such as the impact on children's education, is only one among several that contribute to the overall happiness that probably is a key factor in the decision-making.

9.4 The challenge in a nutshell

Social scientists are trained to be generally careful when presenting and interpreting their results because of various types of selection that are not adequately controlled for, but particular challenges arise when analysing effects of fertility on happiness. We definitely need to think more carefully about how to do such analysis. Conditions under which the existing methods work reasonably well should be better identified and alternative approaches developed, both of which may include collection of other types of data than currently used. Given the limitations of the types of studies that are now carried out, we should never describe results with words suggesting that we may know something about how childbearing affects the parents' happiness in general. The estimate may be far from what is the correct average, and the average would not be so interesting anyway because of the possibly large variation. Similar concerns can be raised with respect to analyses of other micro-level effects of fertility, though these problems are likely smaller.

10. Acknowledgements

The very helpful and insightful comments from Wendy Sigle-Rushton, Torkild Lyngstad, Mikko Myrskylä and two reviewers are greatly appreciated. The study is part of a project on consequences of high fertility funded by the Norwegian Research Council and the Hewlett Foundation.

References

Aassve, A., Mazzuco, S., & Mencarini, L. (2006). An empirical investigation into the effect of childbearing on economic wellbeing in Europe. *Statistical Methods and Applications*,15, 209-227.

Aassve, A., Goisis, A., & Sironi, M. (2012). Happiness and childbearing across Europe. *Social Indicators Research*, 108, 65-86.

Angeles L. (2010). Children and life satisfaction. *Journal of Happiness Studies*, 11, 523-538.

Angrist, J.D., Lavy, V., & Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28, 773-824.

Angrist J.D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An epiricists' companion*. Princeton: Princeton University Press..

Åslund, O., & Grönqvist, H. (2010). Family size and child outcomes: Is there really no trade-off? *Labour Economics*,17, 130-139.

Bittman, M., & Wajcman, J. (2000). The rush hour: The character of leisure time and gender equity. *Social Forces*,79, 165-189.

Black, S.E., Devereux, P.J., & Salvanes, K.G. (2005). The more the merrier? The effects of family size and birth order on children's education. *The Quarterly Journal of Economics*, 669-700.

Blake, D., & Mayhew, L. (2006). On the sustainability of the UK state pension system in the light of the population ageing and declining fertility. *The Economic Journal*,116, F286-F305.

Blundell, R., Dearden, L., & Sianesi, B. (2005). Evaluating the effect of education on earnings: models, methods, and results from the National Child Development Survey. *Journal of the Royal Statistical Society Series A*, 168, 473-512.

Boushey, H. (2008). "Opting out?" The effect of children on women's employment in the United States. *Feminist Economics*,14, 1-36.

Brandt, M., Haberkern, K., & Szydlik, M. (2009). Intergenerational help and care in Europe. *European Sociological Review*, 25, 585-601.

Britt, K., Ashworth, A., & Smalley, M. (2007). Pregnancy and the risk of breast cancer. *Review of Endocrine-Related Cancer*, 14, 907−933.

Bühler, C. (2008). On the structural value of children and its implication on intended fertility in Bulgaria. *Demographic Research*, 18, 569−610.

Clark, A. E., Diener, E., Georgellis, Y., & Lucas, R.E. (2008). Lags and leads in life satisfaction: A test of the baseline hypothesis. *The Economic Journal*, 118, F222–F243.

Conley, D., & Glauber, R. (2006). "Parental educational investment and children's academic risk: Estimates of the impact of sibship size and birth order from exogenous variation in fertility. *Journal of Human Resources*, 41, 722-737.

De Haan, M. (2010). Birth order, family size and educational attainment. *Economics of Education Review*, 29, 576-588.

Dormont, B., Grignon, M., & Huber, H. (2006). Health expenditure growth: reassessing the threat of ageing. *Health Economics*, 15, 947-963.

Dommermuth, L., & Kitterød, H.R. (2009). Fathers' employment in a father-friendly wewlfare state: does fatherhood affect men's working hours? *Community, Work & Family*,12, 417-436.

Downey, D.B., & Condron, D.J. (2004). Playing well with others in kindergarten: The benefit of siblings at home. *Journal of Marriage and Family*, 66, 333-350.

Dørheim, S.K., Bondevik, G.T., Eberhard-Gran, M., & Bjorvatn, B. (2009). Sleep and depression in postpartum women: A population-based study. *Sleep*, 32, 847-855.

Eibach, R.P., & Mock, S.E. (2011). Idealizing parenthood to rationalize parental investments. *Psychological Science*, 22, 203-208.

Evenson, R. J., & Simon, R.W. (2005). Clarifying the relationship between parenthood and depression. *Journal of Health and Social Behavior*, 46, 341−358.

Fletcher, B., Gulanick, M., & Lamendola, C. (2002). Risk factors for type 2 diabetes mellitus. *Journal of Cardiovascular Nursing*, 16, 17−23.

Goux, D., & Maurin, E. (2005). The effect of overcrowded housing on children's performance at school. *Journal of Public Economics*, 89, 797-819.

Grundy, E., & Kravdal, Ø. (2010). Fertility history and cause-specific mortality: a register-based analysis of complete cohorts of Norwegian women and men. *Social Science & Medicine*, 70, 1847−1857.

Haller, M., & Hadler, M. (2006). How social relations and structures can produce happiness and unhappiness: An international comparative analysis. *Social Indicators Research*, 75, 169–216.

Hansen, T. (2012). Parenthood and happiness: A review of folk theories versus empirical evidence. *Social Indicators Research*, 108, 29-64.

Haybron, D.M. (2003). What do we want from a theory of happiness? *Metaphilosophy*, 34, 305-329.

Headey, D.K., & Hodge, A. (2009). The effect of population growth on economic growth: A meta-regression analysis of the macroeconomic literature. *Population and Development Review*, 35, 221-248.

Imbens, G.W., & Angrist, J.D. (1994). Identification and estimation of local average treatment efefcts. *Econometrica*, 62, 467-475.

Joutsenniemi, K., Martelin T., Kestilä, L. Martikainen, P., Pirkola, S., & Koskinen, S. (2007). Living arrangements, heavy drinking and alcohol dependence. *Alcohol and Alcoholism*, 42, 480-491.

Kendig, H., Dykstra, P.A., van Gaalen, R.I., & Melkas, T. (2007). Health of aging parents and childless individuals. *Journal of Family Issues*, 28, 1457−1486.

Knoester, C., & Eggebeen, D.J. (2006). The effects of the transition to parenthood and subsequent children on men's well-being and social participation. *Journal of Family Issues*,27, 1532−1560.

Keizer, R., Dykstra, P.A., & Poortman, A.R. (2010). The transition to parenthood and well-being: The impact of partner status and work hour transitions. *Journal of Family Psychology*,4, 429-438.

Kohler, H.P., Behrman, J.R., & Skytthe, A. (2005). Partner + Children = Happiness? The effects of partnerships and fertility on well-being. *Population and Development Review*, 31, 407-445.

Kravdal, Ø., Kodzi, I., & Sigle-Rushton, W. (2013). Education in sub-Saharan Africa: A new look at the effects of number of siblings. *Studies in Family Planning,* 44, 275-297.

Li, H., Zhang, J., & Zhu, Y. (2008). The quantity-quality trade-off of children in a developing country: Identification using Chinese twins. *Demography*, 45, 223-243.

Margolis, R., & Myrskylä, M. (2011). A global perspective on happiness and fertility. *Population and Development Review*, 37, 29−56.

Maralani, V. (2008). The changing relationship between family size and educational attainment over the course of socioeconomic development: Evidence from Indonesia. *Demography,* 45, 693-717.

McLanahan, S., & Adams, J. (1989). The effects of children on adults' psychological well-being: 1957–1976. *Social Forces*, 68, 124–146.

McQuillan, K. (2004). When does religion influence fertility? *Population and Development Review*, 30, 25-56.

Myrskylä, M., & Margolis, R. (2012). Happiness: before and after the kids. MPIDR Working Paper WP 2012-013. Max Planck Institute for Demographic Research, Rostock.

Nelson, S.K., Kushlev, K., English, T., Dunn, E.W., & Lyubomirsky, S. (2013). In defense of parenthood : Children are associated with more joy than misery. *Psyhcological Science*,24, 3-10.

Newman, L. (2008). How parenthood experiences influence desire for more children in Australia: A qualitative study. *Journal of Population Research* 25, 1-27.

Nomaguchi, K.N., & Milkie, M.A. (2004). Costs and rewards of children: The effects of becoming a parent on adults' lives. *Journal of Marriage and Family*, 65, 356-374.

O'Neill, B.C, Dalton, M., Fuchs, R., Jiang, L., Pachauri, S., & Zigova, K. . (2010). Global demographic trends and future carbon emissions. *Proceedings of the National Academy of Sciences*., 107: 17521-17526.

Panayoutou, T. (1994). Population, environment, and development nexus. In R. Cassen (Ed.), *Population and development: Old debates, new conclusions* (pp. 149-180). TNew Brunswick: Transaction Publishers.

Poortman, A.R., & Van Der Lippe, T. (2009). Attitudes toward housework and child care and the gendered division of labor. *Journal of Marriage and the Family*, 71, 526-541.

Rand (2004). Low fertility and population ageing. Causes, consequences and policy options. Prepared for the European commission. Santa Monica CA.

Rieck, G., & Fiander, A. (2006). The effect of lifestyle factors on gynaecological Cancer. *Best Practice and Research: Clinical Obstetrics and Gynaecology*, 20, 227−251.

Rosenzweig, M. R., & Zhang, J. (2009). Do population control policies induce more human capital investments? Twins, birth weight and China's "one-child" policy. *Review of Economic Studies*, 76, 1149-1174.

Russo, I.H.,& Russo, J. (2007). Primary prevention of breast cancer by hormone-induced differentiation. *Recent Results in Cancer Research*,174, 111-30.

Sobotka, T., Skirbekk, V., & Philipov, D. (2011). Economic recession and fertility in the developed world. *Population and Development Review*, 37, 267-306.

Soons, J.P.M., Liefbroer, A.C., & Kalmijn, M. (2009). The long-term consequences of relationship formation for subjective well-being. *Journal of Marriage and Family*, 71, 1254-1270.

Stanca, L. (2012). Suffer the little children: Measuring the effects of parenthood on well-being worldwide. *Journal of Economic Behavior & Organization*, 81, 742-750.

Twenge, J.M., Campbell, W.K., & Foster C.A. (2003). Parenthood and marital satisfaction: A meta-analytic review. *Journal of Marriage and Family*, 65, 574-583.

Umberson, D., Pudrovska, T., & Reczek, C. 2010. Parenthood, childlessness, and well-being: A life course perspective. *Journal of Marriage and Family*, 72, 612−629.

Vanassche, S., Swicegood, G., & Matthijs, K. (2013). Marriage and children as a key to happiness? Cross-national differences in the effects of marital status and children on well-being. *Journal of Happiness Studies*, 14, 501-524.

Wenger, G.C., Dykstra, P.A., Melkas, T., & Knipscheer, K.C.M. (2007). Social embeddedness and late-life parenthood – community activity, close ties, and support networks. *Journal of Family Issues*, 28, 1419-1456.

White M.P., & Dolan, P. (2009). Accounting for the richness of daily activities. *Psychological Science*, 20, 1000-1008.

Zimmermann, A., & Easterlin, R. A. (2006). Happily ever after? cohabitation, marriage, divorce and happiness in Germany. *Population and Development Review*, 32, 511–528.

Xie, Y., Brand, J.E., & Jann, B. (2012). Estimating heterogenous treatment effects with observational data. *Sociological Methodology*, 42, 314-347.

Appendix. A more formal presentation of main ideas

Following standard terminology and ideas in the treatment effect literature (e.g. Angrist and Pischke 2009) and applying them to the fertility-happiness case, we can let h_i^0 be the potential happiness of a couple i if they do not have an additional child, while h_i^1 is their potential happiness if they do. The "treatment effect" is the difference

$d_i = h_i^0 - h_i^1$.

What one can observe is h_i^0 if the couple do not have a child (i.e. if $c_i=0$) and h_i^1 if they do (i.e. $c_i=1$).

A measure of potential interest is the "average treatment effect" ATE:

$E(d_i)=E(h_i^1-h_i^0)$.

In a situation where people are selected at random to have a child, this would be the same as the average happiness among those having a child minus the average happiness of those not having a child, i.e.

$ATE = E(h_i^1 \mid c_i=1) - E(h_i^0 \mid c_i=0)$

However, with real non-experimental data, the relationship is typically more complex (see e.g. Xie et al. 2012): .

$ATE = E(h_i^1 \mid c_i=1) - E(h_i^0 \mid c_i=0) – (E(h_i^0 \mid c_i=1) - E(h_i^0 \mid c_i=0)) – (ATT-ATU)\ q$

where ATT is the "average treatment effect among the treated" (i.e. among those having another child), ATU is the "average treatment effect among the untreated" (i.e. among those not having another child), and q is the proportion not having another child. Mathematically, $ATT=E(h_i^1-h_i^0 \mid c_i=1)$ and $ATU= E(h_i^1-h_i^0 \mid c_i=0)$. To simplify the discussion below, we can write the equation as

$ATE = Obsdiff + Term1 + Term2,$

where $Obsdiff = E(h_i^1 \mid c_i=1) - E(h_i^0 \mid c_i=0)$, $Term1 = - (E(h_i^0 \mid c_i=1) - E(h_i^0 \mid c_i=0))$ and $Term2 = - (ATT-ATU)\ q$.

Term1 and Term2 have the following simple interpretations: if there are two groups in the population, both with the same happiness if they do not have a child, but with different effects of childbearing on happiness and (therefore) also different probabilities of having a child, Term1=0 and Term2≠0. If effects of childbearing instead are the same in the two groups, but one group has generally lower level of happiness and the probabilities of having a child differ, Term1≠0 and Term2=0.

In this paper, there is one example of a situation where ATE is equal to the difference in happiness between one- and two-child couples (Obsdiff), and some examples where this is not the case (Term1 or Term2 being non-zero). It is also discussed what kind of substantive mechanisms that give rise to such gaps between ATE and Obsdiff. In a simple regression

analysis including only a child variable, the corresponding coefficient will be Obsdiff. In the second-last section of the paper, it is discussed whether addition of control variables or use of different types of models can give estimates closer to ATE.

Specification of examples:

In Example 1 in this paper, two groups are defined: Group 1 (g=1) and Group2 (g=2). They consist of 200 and 100 couples, respectively. The potential happiness outcomes for the two groups are defined as $h_i^x{}_g=12+5x+(g-1)(5-10x)$, where x=1 in the potential situation where they have a child and x=0 otherwise. The probabilities of having a child are assumed to be $p_1=0.50$ if g=1 and $p_2=0.50$ if g=2. Then,

$$E(h_i^1 \mid c_i=1) = (E(h_i^1{}_1 \mid c_i=1)\cdot 200\cdot p_1 + E(h_i^1{}_2 \mid ci=1)\cdot 100\cdot p_2)/(\ 200\cdot p_1+100\cdot p_2\ )$$

$$= (E(h_i^1{}_1 \mid c_i=1)\cdot 200 + E(h_i^1{}_2 \mid c_i=1)\cdot 100)/300$$

$$= (17\cdot 200+12\cdot 100)/300 = 15.33$$

when $p_1=p_2$, regardless of whether they are 50% as assumed or lower or higher.

Similarly, one can calculate $E(h_i^0 \mid c_i=0)$, $E(h_i^0 \mid c_i=1)$, ATE, ATT etc. and it is not difficult to show that Obsdiff= ATE=ATT=ATU. (Term1=0 in spite of the two groups having different happiness if they remain one-child couples, because the chance of childbearing is the same in the two groups. Also Term2=0 as a result of the non-varying fertility.)

Example 2 is similar except that the probabilities of having a child are assumed to be different in the two groups: $p_1=0.90$ and $p_2=0.30$. Then, Term1≠0 (reflecting that the happiness associated with having one child is not the same for the two groups); ATT ≠ ATU, so that Term2≠0; and Obsdiff is not equal to ATE (and not equal to ATT or ATU either).

In Example 3, the happiness function is the same as in Example 1 and Example 2 except that a term that differs between poor (r=1) and non-poor (r=0) is added:

$h_i^x{}_{gr}=12+5x+(g-1)(5-10x)-6r$

Furthermore, poverty is assumed to affect the chance of having a child through the chance of belonging to one of the two main groups g=1 and g=2: there are 70 poor and 230 non-poor; the poor have a chance of 20/70 of being in group g=1, while the non-poor have a chance of 180/230 of being in group g=1. The group-specific chances of having a child are as before $p_1=0.90$ and $p_2=0.30$. Thus, one may say that the difference compared to Example 2 is that the happiness levels in Group 2 (averaged over poor/non-poor) are no longer reversed compared to those in Group 1 (17 and 12 vs. 12 and 17), but generally lower because of a larger proportion of poor in Group 2 than in Group 1. It can be shown that Term2 is the same as in Example 2, but Term 1 is different. As in Example 2, Obsdiff ≠ ATE (and not equal to ATT or ATU either).

Example 4 is simpler in that poverty affects fertility (p) more directly, not via the distribution over the two main groups (which would symbolize differences in expected

consequences of fertility and thus fertility desires). There is supposed to be only one group (g=2 above, i.e. those less keen on childbearing), so happiness is given by

$h_i^{x}{}_r = 17-5x-6r$

The non-poor are assumed to have a 10% chance of having a child and the poor a 50% chance. Furthermore, 50 of the 200 are supposed to be poor. In this case, Term1≠0, Term2=0, and Obsdiff ≠ ATE (but ATE=ATT=ATU).

Table 1: Summary of Example 1

Group	Subgroup	Number of couples	Expected happiness (assumed equal to actual happiness) if Not having 2nd child	Expected happiness (assumed equal to actual happiness) if Having 2nd child	Having 2nd child? (proportion of group)	Actual Happiness
1	a	100	12	17	Yes (50%)	17
1	b	100	12	17	No (50%)	12
2	a	50	17	12	Yes (50%)	12
2	b	50	17	12	No (50%)	17

Average actual happiness among those with
2children (i.e. 1a+2a; 150 couples in total): 15.33
1child (i,e. 1b+2b; 150 couples in total): 13.67
Difference: 1.67

Table 2: Summary of Example 2.

Group	Subgroup	Number of couples	Expected happiness (assumed equal to actual happiness) if Not having 2nd child	Expected happiness (assumed equal to actual happiness) if Having 2nd child	Having 2nd child? (proportion of group)	Actual Happiness
1	a	180	12	17	Yes (90%)	17
1	b	20	12	17	No (10%)	12
2	a	30	17	12	Yes (30%)	12
2	b	70	17	12	No (70%)	17

Average actual happiness among those with
2children (i.e. 1a+2a; 210 couples in total): 16.29
1child (i.e. 1b+2b; 90 couples in total): 15.89
Difference: 0.40

Table 3: Summary of Example 3

Group	Subgroup	Number of couples	Expected happiness (assumed equal to actual happiness) if Not having 2nd child	Expected happiness (assumed equal to actual happiness) if Having 2nd child	Having 2nd child? (proportion of group)	Actual Happiness
1 non-poor	a	162	12	17	Yes (90%)	17
1 non-poor	b	18	12	17	No (10%)	12
1 poor	a	18	6	11	Yes (90%)	11
1 poor	b	2	6	11	No (10%)	6
2 non-poor	a	15	17	12	Yes (30%)	12
2 non-poor	b	35	17	12	No (70%)	17
2 poor	a	15	11	6	Yes (30%)	6
2 poor	b	35	11	6	No (70%)	11

Average actual happiness among those with		
2children	(i.e. 1a+2a, poor and non-poor; 210 couples in total):	15.34
1child	(i.e. 1b+2b, poor and non-poor; 90 couples in total):	13.42
	Difference:	1.92

Table 4: Summary of Example 4

Group	Subgroup	Number of couples	Expected happiness (assumed equal to actual happiness) if Not having 2nd child	Expected happiness (assumed equal to actual happiness) if Having 2nd child	Having 2nd child? (proportion of group)	Actual Happiness
non-poor	a	15	17	12	Yes (10%)	12
non-poor	b	135	17	12	No (90%)	17
poor	a	25	11	6	Yes (50%)	6
poor	b	25	11	6	No (50%)	11

Average actual happiness among those with		
2children	(i.e. a, poor and non-poor; 40 couples in total):	8.25
1child	(i.e. b, poor and non-poor; 160 couples in total):	16.06
	Difference:	-7.81

Table 5: Illustration of the twin fixed-effects approach

Group	Subgroup	Number of women	Expected happiness (assumed equal to actual happiness) if Not having 2nd child	Expected happiness (assumed equal to actual happiness) if Having 2nd child	Having 2nd child?	Actual Happiness
1 non-poor	a	3240	12	17	Yes	17
1 non-poor	b	360	12	17	No	12
		(1800 twin pars, of which				
		18 with 1+1 child,				
		324 with 1+2 and				
		1458 with 2+2)				
1 poor	a	360	6	11	Yes	11
1 poor	b	40	6	11	No	6
		(200 twin pars, of which				
		2 with 1+1 child,				
		36 with 1+2 and				
		162 with 2+2)				
2 non-poor	a	300	17	12	Yes	12
2 non-poor	b	700	17	12	No	17
		(500 twin pars, of which				
		245 with 1+1 child,				
		210 with 1+2 and				
		45 with 2+2)				
2 poor	a	300	11	6	Yes	6
2 poor	b	700	11	6	No	11
		(500 twin pars, of which				
		245 with 1+1 child,				
		210 with 1+2 and				
		45 with 2+2)				

Notes

[1] Let us also assume that the latter outcome is a result of fecundity problems and not a downward revision of fertility desires because of, for example divorce or unexpected economic hardship (i.e. a change from a situation where everyone ranks the outcomes as 17/11 to a situation where a 10% subgroup rank them as, say, 8/10). The arguments are complex enough without such heterogeneity in attitudes to childbearing developing over time within the group.

[2] Admittedly, the literature does not provide clear evidence of an income effect on fertility desires, but low income is often linked to fear about later income decline due to for example unemployment, which is more likely to have an adverse effect (Sobotka et al. 2011).

[3] The sample is set up as in Example 2 without any additional random term, and the OLS regression module in the SAS software is used. The interest lies in the point estimates.

[4] It is not difficult to see how a negative effect (-0.38) arises. Only the twin pairs consisting of women with different fertility contribute in the estimation. The majority of these pairs come from the twin pairs with negative attitudes to childbearing, because although the number of such twins in the population is smaller, the chance that the women in such a twin pair end up with different fertility is relatively high given the failure rate of 30% as opposed to only 10% among the others. Rather than having one group with a happiness difference of 5 and another group that is half as large and has a happiness difference of -5, as in reality (and which gives 1.67), the latter group with a difference of -5 is larger among the twin pairs who contribute in the fixed-effects analysis.