

**Investigating additive genetic, maternal, and paternal (co-)variation in fertility and educational level in the Netherlands. An application of the ‘animal model’.**

Jornt J. Mandemakers  
j.j.mandemakers@rug.nl

Submitted to IUSSP conference 2013 in Busan, Korea

Session 13-01 Biodemography.

Graziella Caselli, Sapienza University of Rome

Session 13-05 Biodemography of human and non-human populations.

Annette Baudisch, Max Planck Institute for Demographic Research

**Short abstract (1350 chars)**

Low levels of fertility in Western countries are often attributed to men and women pursuing careers first and babies later. Research has shown considerable similarity between parents-children and siblings in completed fertility and education. This raises the possibility that the effect of education on fertility is not causal, but has a genetic basis and/or arises due to non-genetic factors of the family of origin. Relatively little research investigates the (common) genetic origin of these outcomes. Most of the existing bio-demographic research on fertility uses the Danish Twin Study (cf. Kohler and Rodgers). It is crucial to extend this line of research to other contexts to generalize findings and gain insight into possible environmental variability. Unfortunately, twin data are not readily available for most contexts, but genetic origins can also be investigated using information from multi-generation surveys using the so-called ‘animal model’. This is a mixed model that takes advantage of information on the relatedness of all individuals in a pedigree to estimate the genetic merit of an individual. We apply such a model to the Netherlands Kinship Panel Study (8,200 families and 56,000 individuals) to investigate the role of additive genetic effects, maternal and paternal effects on (the co-variation of) fertility and education.

## Extended abstract

### **Investigating additive genetic, maternal, and paternal (co-)variation in fertility and educational level in the Netherlands. An application of the ‘animal model’.**

Low levels of fertility in western countries are often attributed to men and women pursuing careers first and babies later (Balbo, Billari, and Mills 2012; Mills, Rindfuss, McDonald et al. 2011). Educational level is usually taken as the main indicator of socio-economic status and considered as the driving force behind the postponement of childbearing and subsequent decreased completed fertility. Research has demonstrated considerable similarity between parents-children and siblings in levels of fertility (Murphy 1999; Rijken and Liefbroer 2009) and education (van Doorn, Pop, and Wolbers 2011). This raises the possibility that the link between education and fertility is in part spurious because both may have a (common) genetic origin and/or arises due to (common) non-genetic factors related to the family of origin (e.g. socialization, wealth transfers). To study these issues, information on related individuals is needed.

The relatively little research that exists on this specific issue is based on comparing dizygotic and monozygotic twins and uses the Danish Twin Study (Kohler and Rodgers 2003; Rodgers, Kohler, McGue et al. 2008). The chapter by Kohler and Rodgers (2003) finds little to no common genetic variation for number of children and completed educational level for men and women, but some common shared environmental variation for women. The later contribution of Rodgers et al. (2008) on age at first birth, cognitive ability, and educational level does not find common genetic source of variance, but does find substantial common shared environmental influences for fertility and education, which implies that the relation between age at first birth and education is indeed spurious, resulting from a common rearing environment.

These interesting results bear further study. Most research into the heritability of fertility use a twin design and are set in Denmark (cf. Kohler et al. 2003; Kohler, Rodgers, and Christensen 1999; Kohler, Rodgers, and Christensen 2002; Rodgers, Hughes, Kohler et al. 2001; Rodgers, Kohler, Kyvik et al. 2001; Rodgers et al. 2008). We know of only one other study based on Australian twins (Kirk, Blomberg, Duffy et al. 2001). These studies have established the existence of a significant heritable component to variation in human fertility, but there is a need to expand this type of research to other contexts in order to generalize these findings, gain insight into possible environmental variability, and to further use genetically-informed designs to test social science explanations.

Unfortunately, twin data are hard to come by and therefore not available for most contexts. Genetically informed designs are also possible using kinships other than twins, such as full siblings, half-siblings, cousins (see for example on a limited set of dyads with the NLSY data, Rodgers, Bard, and Miller 2007). Recent research has started to analyze large multi-generation datasets using the so-called ‘animal model’ (Pettay, Helle, Jokela et al. 2007; Pettay, Kruuk, Jokela et al. 2005; Stearns, Byars, Govindaraju et al. 2010). The ‘animal model’ takes advantage of information on the relatedness of all individuals in a pedigree to estimate the genetic merit of an individual using a mixed modeling approach. The ‘animal model’ despite its somewhat unfortunate name (it was originally developed to optimize breeding among livestock) can be applied to study humans as well. For an introduction see (Wilson, Réale, Clements et al. 2010) and for some applications (Pettay et al. 2007; Pettay et al. 2005). The animal model is a mixed model, which allows us to separate phenotypic observations into an additive genetic component and other random and fixed effects.

We apply such a model to the Netherlands Kinship Panel Study (NKPS) (8,200 families and 56,000 individuals) to investigate the role of genes and the shared environment on fertility and educational level (Dykstra, Kalmijn, Knijn et al. 2005). At the moment we aim to split the variation into three components: additive genetic, maternal and paternal effects; the last two capture the shared

environment. Previous research using the NKPS has demonstrated inter-generational transmission of fertility behavior between parents and children (Rijken et al. 2009), but has not employed a genetically-informed design and been limited to study parents and children.

To summarize, this paper aims to contribute to the literature in the following three ways: First, we examine genetic, maternal, and paternal variation in fertility and educational level for the Netherlands. So far, no studies have examined genetic variation in either of these outcomes in the Netherlands. Second, we test whether the co-variation of fertility and educational can be explained by common genetic, maternal, and/or paternal origin. Research based on Danish twins suggests that there is no common genetic origin of the link between education and fertility, but it is unclear whether this is a general finding or limited to Denmark. And third, we demonstrate the use of the “animal model” in demographic research. The application of such mixed modeling techniques holds great promise to study bio-demographic questions using already well-known large scale multi-generational studies.

Table 1. Overview of the family relationships in the NKPS

	<i>N</i>
Families (primary respondents)	8,161
Individuals	56,290
Dyads	80,258
<i>Composition of dyads:</i>	
Parent – Child*	33,158
Grandparent – Grandchild*	15,146
Uncle/Aunt – Nephew/Niece*	10,827
Siblings (Full)	20,000
Siblings (Half)	916
Siblings (Adopted)	86
Siblings (Monozygotic twins)	49
Siblings (Dizygotic twins)	76

\* Includes relationships of varying relatedness (in case of twins, adoption).

## Data

This study uses data from the first 3 waves of the NKPS. The NKPS provides information on 8,161 primary respondents and their families. The primary respondents were drawn from a random sample of private addresses in the Netherlands in 2002 – 2004. The overall response rate at the first wave of the primary respondents was 45%, which is comparable to other family surveys in the Netherlands (Dykstra et al. 2005). Subsequent waves were carried out in 2006 – 2007 and in 2010 – 2011.

At each wave, the primary respondents provided extensive information on their close family members. For our purposes, primary respondents provided information on year of birth, the number of children ever born, and attained educational level for the following family members: both parents, the current partner, ex-partner(s), two randomly selected siblings, two randomly selected children, and both parents-in-law of the current partner. In addition, at the first wave self-completion questionnaires were sent to one randomly selected parent, one of the two selected siblings, the current partner, and the two selected children. The family members were re-contacted for the second wave. The additional source of information provided by the family members’ questionnaires allows us to potentially broaden the scope of the study to include additional covariates and additional

outcome variables (e.g., age at first birth) and to assess measurement accuracy of the reports of the primary respondent on his/her family members.

The design of the NKPS allows constructing a database with information on 56,290 individuals, which are nested in 8,161 families (the primary respondents), and on 80,258 dyads of related individuals (see Table 1). The number of children ever born and the highest attained educational level are the two variables of interest. People are included in the analyses if they have finished their reproductive career (45 for women, 55 for men), for educational level if they are at least 25. The variables are constructed based on the most recent report of the primary respondent. In case information from the family member is available this is used. A comparison of reports by the primary respondents' to that of the family members revealed high correlations for number of children ever born ( $r$  ranges between .91 and .97) and somewhat lower but still high correlations for educational level ( $r$  ranges between .77 and .87).

### Analytical strategy

We model the data using the so-called 'animal model', the animal model is a mixed model, which allows us to separate phenotypic observations into an additive genetic component and other random and fixed effects. All models will be estimated using the program VCE. The general form of the model looks like the following:

$$y = Xb + Z_1a + Z_2r + e$$

where  $y$  is a vector of phenotypic values;  $b$  is a vector of fixed effects;  $a$  is a vector of random effects of the additive genetic merit (also known as breeding value) of each individual;  $r$  is a vector containing other random effects;  $e$  is a vector of residuals; and  $X$ ,  $Z_1$ , and  $Z_2$  are design matrices relating an observation to its corresponding fixed or random effects. The variance-covariance structure for the additive genetic effects  $a$  is determined by the pedigree of the population, and the residual terms  $e$  are assumed to be normally distributed. The model exploits the covariance between pairs of relatives of all types of relatedness (e.g., between grand-mothers and grand-daughters, between uncles and nephews, etc.), making optimal use of the available data in the NKPS. The models are estimated using a restricted maximum-likelihood procedure.

Children acquire both their genes and an environment from their parents; we will include maternal and paternal random effects to separate the additive genetic from the shared environmental effects. The separation of maternal and paternal effects will depend on the identification of extra-pair maternities and paternities (i.e. half-siblings), which are present in the NKPS (see Table 1) but not in a significant amount, so we will see whether it is possible to fit both maternal and paternal effects. Otherwise, we might fit a 'nuclear family' effect. In addition, we will fit as a fixed effect birth year to take into account temporal environmental variation in the studied period (e.g. introduction of anti-contraceptives, legalization of abortion, educational expansion, etc.). To summarize we will fit the following models, separately for men and women:

$$y = Xb + Z_1a + Z_2m + Z_3p + e$$

Therefore we can partition the total phenotypic variance into the following four components:  $V_A$ ,  $V_M$ ,  $V_p$ , and  $V_r$ . where  $V_A$  is the additive genetic variance,  $V_m$  is the maternal effect variance,  $V_p$  is the paternal effect variance, and  $V_r$  is the residual variance. Note that  $V_r$  will include environmental effects not captured by the maternal and paternal effects, non-additive genetic effects (dominance and epistatic), and error variance.

*Multi-variate model.* To investigate the genetic correlation between educational level and number ever born we will estimate a multivariate version of the animal model in the second step. The genetic covariance between two outcomes can be estimated using relatives of differing degrees (e.g. educational level of the mother and number ever born of the daughter; and educational level of an uncle and number ever born in a nephew, etc.).

## Results

We are currently analyzing the data and unable to present meaningful results at the time of submitting this abstract. Preliminary results, however, are promising.

## References:

- Balbo, Billari, and Mills. 2012. "Fertility in Advanced Societies: A Review of Research." *European Journal of Population / Revue européenne de Démographie*:1-38.
- Dykstra, Kalmijn, Knijn, Komter, Liefbroer, and Mulder. 2005. *Codebook of the Netherlands Kinship Panel Study, a multi-actor, multi-method panel study on solidarity in family relationships, Wave 1*. NKPS. The Hague: Netherlands: Interdisciplinary Demographic Institute (NIDI).
- Kirk, Blomberg, Duffy, Heath, Owens, and Martin. 2001. "Natural Selection and Quantitative Genetics of Life-History Traits in Western Women: A Twin Study." *Evolution* 55:423-435.
- Kohler and Rodgers. 2003. "Education, Fertility, and Heritability: Explaining a Paradox." Pp. 46-90 in *Offspring: Human Fertility Behavior in Biodemographic Perspective*, edited by Wachter and Bulatao: The National Academies Press.
- Kohler, Rodgers, and Christensen. 1999. "Is Fertility Behavior in Our Genes? Findings from a Danish Twin Study." *Population and Development Review* 25:253-288.
- Kohler, Rodgers, and Christensen. 2002. "Between nurture and nature: The shifting determinants of female fertility in Danish twin cohorts." *Biodemography and Social Biology* 49:218-248.
- Mills, Rindfuss, McDonald, and te Velde. 2011. "Why do people postpone parenthood? Reasons and social policy incentives." *Human Reproduction Update* 17:848-860.
- Murphy. 1999. "Is the relationship between fertility of parents and children really weak?" *Biodemography and Social Biology* 46:122-145.
- Pettay, Helle, Jokela, and Lummaa. 2007. "Natural Selection on Female Life-History Traits in Relation to Socio-Economic Class in Pre-Industrial Human Populations." *PLoS ONE* 2:e606.
- Pettay, Kruuk, Jokela, and Lummaa. 2005. "Heritability and genetic constraints of life-history trait evolution in preindustrial humans." *Proceedings of the National Academy of Sciences of the United States of America* 102:2838-2843.
- Rijken and Liefbroer. 2009. "Influences of the family of origin on the timing and quantum of fertility in the Netherlands." *Population Studies* 63:71-85.
- Rodgers, Bard, and Miller. 2007. "Multivariate Cholesky Models of Human Female Fertility Patterns in the NLSY." *Behavior Genetics* 37:345-361.
- Rodgers, Hughes, Kohler, Christensen, Doughty, Rowe, and Miller. 2001. "Genetic Influence Helps Explain Variation in Human Fertility: Evidence From Recent Behavioral and Molecular Genetic Studies." *Current Directions in Psychological Science* 10:184-188.
- Rodgers, Kohler, Kyvik, and Christensen. 2001. "Behavior genetic modeling of human fertility: findings from a contemporary danish twin study." *Demography* 38:29-42.
- Rodgers, Kohler, McGue, Behrman, Petersen, Bingley, and Christensen. 2008. "Education and Cognitive Ability as Direct, Mediating, or Spurious Influences on Female Age at First Birth: Behavior Genetic Models Fit to Danish Twin Data." *American Journal of Sociology* 114:S202-S232.
- Stearns, Byars, Govindaraju, and Ewbank. 2010. "Measuring selection in contemporary human populations." *Nat Rev Genet* 11:611-622.
- van Doorn, Pop, and Wolbers. 2011. "Intergenerational Transmission of Education Across European Countries and Cohorts." *European Societies* 13:93-117.
- Wilson, Réale, Clements, Morrissey, Postma, Walling, Kruuk, and Nussey. 2010. "An ecologist's guide to the animal model." *Journal of Animal Ecology* 79:13-26.