

Similarity of Siblings' Family Formation

*Extended abstract prepared for the 2013 XVII IUSSP International Population Conference
in Busan, Republic of Korea*

Jani Erola

University of Turku, Finland

(jani.erola@utu.fi)

Anette Eva Fasang

Humboldt-University Berlin & Social Science Research Center (WZB), Germany

(anette.fasang@hu-berlin.de)

Aleksi Karhula

University of Turku, Finland

(aleksi.karhula@utu.fi)

Marcel Raab

Bamberg University & Social Science Research Center (WZB), Germany

(raab@wzb.eu)

Extended abstract

Decisions about cohabitation, marriage, and fertility are made in the context of social interaction with significant others. This includes intergenerational transmission of family behavior from parents to their children (Liefbroer and Elzinga, 2006; Booth and Kee, 2006, Barber, 2000), as well as contagion among siblings (Lyngstad and Prskawetz 2010, Kuziemko 2006) and within peer networks (Keim et al. 2009). To date, studies on sibling effects in family formation focus on fertility transitions to first and higher order births (Lyngstad and Prskawetz 2010, Kuziemko 2006). Yet, fertility transitions are embedded in family formation trajectories that typically lead through cohabitation, possibly marriage, then having one or more children, and for some, union dissolution and re-partnering in between. Individuals plan and make decisions about fertility in the biographical context of these holistic family formation trajectories.

The main analytical contribution of this study is to combine the sibling approach with sequence analysis to disentangle the mechanisms that link family background to family formation trajectories. We use Finnish register data to compare holistic family formation sequences of siblings. We use data from 1987 until 2007 to construct family formation sequences. The empirical analysis consists of three parts and follows both a descriptive and an explanatory objective. First, we apply sequence analysis and calculate pair wise sequence distances to determine whether siblings' family formation trajectories are more similar to one another than those of non-sibling dyads. This allows us to explore, whether siblings have a higher than random probability to follow the same family formation pattern.

Second, we extend the sibling correlation approach applied in the multilevel regression framework to the analysis of the sequence distance data. We find that siblings' holistic family formation sequences are significantly more similar to one another than those of non-siblings.

Third, to disentangle the mechanisms that link family background to family formation and thereby generate sibling similarity, we use pair-wise sequence distances for both sibling and non-sibling dyads as the dependent variable in a regression analysis. In the scope of this extended abstract we are not able to go to these results in depth.

To our knowledge this is the first study to combine the sibling approach with sequence analysis (see MacIndoe & Abbott 2004, and Aisenbrey & Fasang, 2010 for an introduction and overview of sequence analysis).

The methods used and the intervals of the data forced some restrictions upon us. First we restricted our analysis to people born 1969 – 1977. This was done to assure that we could follow everyone from the age of 18 to the age of 30 (1987 – 2007). We also restricted the sample to include only people who had at least one sibling born in the time frame. Our sample

includes 9581 people, 4872 men and 4754 women. This results in 9581 sibling dyads in the whole sample, 5038 opposite sex sibling dyads, 2244 sister dyads and 2299 brother dyads.

At the first stage of the study we analyze the dynamic Hamming distances between siblings compared to those among randomly chosen individuals and compute the sibling correlations based on this data. In a regression analysis framework using data arranged according to siblings the models provide information on unobserved family-level heterogeneity, *the family level variation*, which does not vary between the children of the same family, and unobserved *individual level variation* affecting the modeled outcome that does not originate from the shared background. The unobserved family component includes all unobserved sources of background variation, whether this includes shared social environment, genetic background, cognitive or non-cognitive traits, but controlling for something shared by the siblings reduces this variation.

The total variance can be understood as the sum of the variances of both unobserved components, $Var_{family} + Var_{individual}$. Thus when applying the multilevel regression approach we can calculate the share of the variance in the modeled outcome that stems from family background with the following:

$$\rho_{regression} = \frac{Var_{family}}{Var_{family} + Var_{individual}} \quad (2)$$

which equals the correlation between two randomly drawn pairs of siblings. The bigger the ρ , the stronger the influence of the effects shared by the siblings.

In the sequence analysis context we can consider the mean squared distances as being equivalent to variances in the means in regression analysis (e.g. Grower & Krzanowski 1999). We contrast the mean distances between brothers, sisters and mixed-sex siblings with the mean distances between the randomly drawn same- and different-sex pairs in the data. We assume that the distances among the siblings should always be smaller than the mean among the random pairs, if different at all. A sibling correlation for the family background effect equivalent to a sibling correlation acquired with the regression approach would then be the difference between the mean squared distances of the siblings divided by the mean squared distances of the randomly drawn pairs:

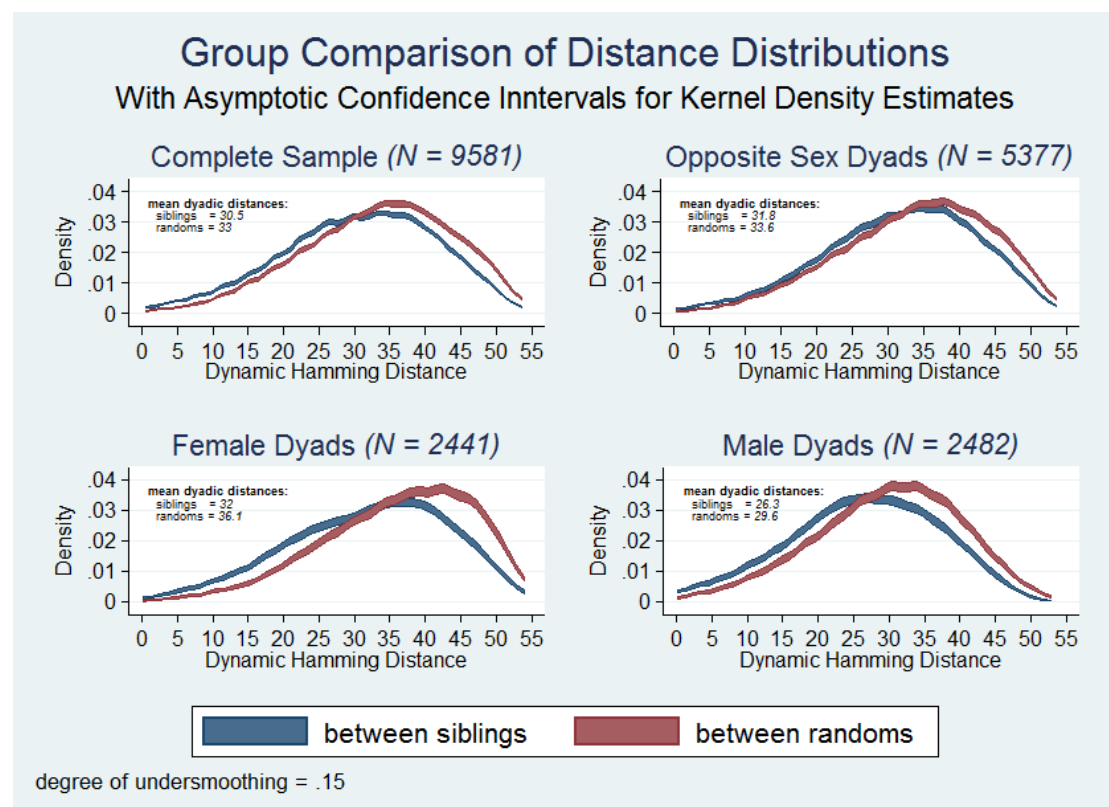
$$\rho_{sequence} = \frac{mean(dist^2)_{family}}{mean(dist^2)_{family} + mean(dist^2)_{individual}} \quad (3)$$

Using Equation 3, we can address the first research question in this analysis and quantify the similarity of siblings' family formation relative to the similarity of family formation of

randomly selected non-sibling dyads. Based on this, we can specify the total family background effect on similarity of family formation trajectories, presented as sibling correlations.

To compare how similar siblings' family formation is are compared to randomly chosen non-siblings dyads Figure 1 shows the distribution of dynamic Hamming distances for all dyads, opposite sex dyads, and same sex dyads. In each graph in figure 1, the red lines show the distribution of distances for random non-sibling dyads, and the blue lines show the distribution of distances for sibling dyads. Low distances indicate similar family formation trajectories for a dyad, whereas high distances indicate very distinct family formation trajectories within a dyad. As expected, siblings are more similar to one another in their family formation than randomly drawn pairs.

Figure 1. Group comparisons of distance distributions with asymptotic confidence intervals for kernel density intervals



The differences are smaller among males than females. This suggests that men's family formation trajectories are generally simpler and that there is overall more variation in women's family formation, most likely because women start family formation at earlier ages than men.

Using the distances between the dyads we can now compute the unadjusted sibling correlations based on the variances in the distances using Equation 3. We show these in Table 1. The ρ as an indicator for sibling correlation in family formation across the entire sample is .12. For opposite sex dyads the ρ is slightly lower at .09, and for the same sex dyads a bit higher, at .17 for men and .16 for women.

Table 1. Sibling correlations according to gender and education

	All	Brothers	Sisters	Opposite sexes
All educational groups	0,12	0,17	0,16	0,09
Higher tertiary (both people)	0,23	0,30	0,27	0,17
Other educational combinations	0,11	0,16	0,14	0,08

In a nutshell, table 1 shows that the siblings resemble each other more in their family formation when they share high educational attainment. This suggests that similarity in status attainment processes is one mechanism through which family background effects are transmitted in family formation. The results show that the sibling correlations range between .09-.17 for all sibling dyads, but are higher between .17-.30 for siblings who both have higher tertiary education. Thus, people with higher tertiary qualifications show roughly twice the influence of the total family background effect on family formation.

Our analysis of sibling correlations for the demographic behavior of family formation yields similar effect sizes as sibling correlations for socioeconomic outcomes (e.g. Björklund et al 2002; Conley & Glauber 2008; Björklund and Salvanes 2011).

Our preliminary regression results support the finding that sibling similarity is higher among the highly educated compared to the non-siblings. They also indicate that siblings spatial and temporal proximity increase the similarity in family formation. For dyads of non-siblings proximity does not matter. This suggests that the spatial effect is not driven by any structural traits of the particular region (e.g. high unemployment rate) but rather by social learning and contagion effects among siblings (see Kuziemko 2006). In general, the regression results show that similarities in family formation are driven by the same determinants for sibling and random dyads. For siblings, however, the “similarity effects” are stronger when both siblings are highly educated (see ρ s) or of the same sex, when they live in the same region, and – to a lesser extent – when they both experienced parental divorce until age 18.

References

- Aisenbrey, S., and A. E Fasang. 2010. "New life for old ideas: The 'second wave' of sequence analysis bringing the 'course' back into the life course." *Sociological Methods & Research* 38(3):420-462.
- Björklund, Anders, and Kjell G. Salvanes. 2011. "Education and Family Background: Mechanisms and Policies." Pp. 201-247 in *Handbooks in Economics, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. The Netherlands: North-Holland.
- Björklund, Anders, Tor Eriksson, Markus Jäntti, Oddbjorn Raaum, and Eva Österbacka. 2002. "Brother Correlations in Earnings in Denmark, Finland, Norway and Sweden Compared to the United States." *Journal of Population Economics* 15(4):757-772.
- Booth, Alison L, and Hiau Joo Kee. 2009. "Intergenerational Transmission of Fertility Patterns." *Oxford Bulletin of Economics and Statistics* 71(2):183-208.
- Conley, Dalton S., and Rebecca Glauber. 2008. "All in the family?: Family composition, resources, and sibling similarity in socioeconomic status." *Research in Social Stratification and Mobility* 26(4):297-306.
- Grower, J. C, and W. J Krzanowski. 1999. "Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 48(4):505-519.
- Kuziemko, I. 2006. "Is Having Babies Contagious? Estimating Fertility Peer Effects Between Siblings." *Princeton University manuscript*.
- Lyngstad, Torkild Hovde, and Alexia Prskawetz. 2010. "Do siblings' fertility decisions influence each other?" *Demography* 47(4):923-934.
- MacIndoe, H., and A. Abbott. 2004. "Sequence analysis and optimal matching techniques for social science data." Pp. 387-406 in *Handbook of data analysis*, edited by Melissa Hardy and Alan Bryman. London: Sage.