

Sibling similarity in family formation

Marcel Raab, *WZB Berlin Social Science Center*

Anette Fasang, *Humboldt University Berlin and WZB Berlin Social Science Center*

Aleksi Karhula & Jani Erola, *University of Turku*

Abstract

Sibling studies have been widely used to analyze the impact of family background on socioeconomic and, to a lesser extent, demographic outcomes. We contribute to this literature with a novel research design that combines sibling comparisons and sequence analysis to analyze longitudinal family formation trajectories of siblings and unrelated persons. This allows us to scrutinize in a more rigorous way, whether there is sibling similarity in family formation trajectories and if siblings' shared background characteristics, such as parental education and early childhood family structure can account for similarity in family formation. We use Finnish register data from 1987 until 2007 to construct longitudinal family formation trajectories in young adulthood for siblings and unrelated dyads (N=14,257 dyads). Findings show that siblings' family formation is moderately but significantly more similar than for unrelated dyads, also after controlling for crucial parental background characteristics. Shared parental background characteristics add surprisingly little to account for sibling similarity in family formation. Instead, gender and the respondents' own education are more decisive forces in the stratification of family formation. Yet family internal dynamics seem to reinforce this stratification, such that siblings have a higher probability to experience similar family formation patterns. Particularly patterns that go along with economic disadvantage are concentrated within families. This is in line with a growing body of research highlighting the importance of family structure in the reproduction of social inequality.

INTRODUCTION

Family of origin effects on family behavior are at the center of a controversial debate. On the one hand, growing evidence on profound social change in family formation with the proliferation of cohabitation, lone parenthood and diverse family forms questions, whether the family of origin still matters for peoples' family behavior (Elzinga and Liefbroer 2007; Fussell and Furstenberg 2005; Popenoe 1993; Shanahan 2000). On the other hand, several studies compellingly demonstrate continuing importance of the family of origin for peoples' family behavior (Liefbroer and Elzinga 2012; Murphy 2013). The broader relevance of the topic is highlighted by another line of research emphasizing the role of family structure in the reproduction of social inequality and the perpetuation of family disruption across generations (Carlson and England 2011; McLanahan and Percheski 2008).

Sibling studies have been the method of choice to study family of origin effects - everything siblings share - on socioeconomic outcomes such as education and occupational status (e.g. Conley 2008). To a lesser extent they have also been employed to examine family of origin effects on demographic outcomes such as family behavior. To date, sibling studies on family formation largely focus on isolated fertility transitions (Kuziemko 2006; Lyngstad and Prskawetz 2010). Yet, these fertility transitions are embedded in family formation trajectories that typically lead through cohabitation, marriage, then having one or more children, and possibly separation and re-partnering in between.

This paper combines the sibling approach with sequence analysis to scrutinize the impact of shared parental background characteristics on sibling similarity in holistic family formation trajectories. We address three research questions: First, are family formation trajectories of siblings more similar than among comparable unrelated persons? Second, can shared parental background characteristics account for sibling similarity in family formation? Third, in which way is siblings' family formation more similar, i.e. are siblings more likely than unrelated persons to experience specific family formation patterns?

Drawing on Finnish register data our empirical analysis employs (a) conditional assignment to build an analysis sample of sibling dyads and comparable unrelated dyads, (b) sequence analysis to measure similarity in family formation within these dyads, (c) dyadic regression analysis to assess the impact of shared parental background characteristics, and (d) cluster analysis to examine in which way siblings' family formation is similar. To our knowledge this is the first study that combines a sibling design with sequence analysis to study family of origin effects on family formation. This analytical strategy aims to contribute to the literature in three regards.

First, this is the first study to establish the simple descriptive fact that siblings are more similar to one another also in their holistic family formation trajectories and not only in isolated fertility transitions. This is possible through the combination of a sibling design with sequence analysis. We thereby recognize the importance of linked lives within families (Elder et al. 2003) as well as the interdependence of multiple family formation events across the life course and the diversity of family formation processes (Wu and Li 2005).

Second, we compare sibling dyads with a control group of unrelated dyads that are identical with regard to important parental background characteristics. Siblings naturally share parental background characteristics, such as parental education and early childhood family structure that unrelated persons do not generally share. This compositional effect alone might generate sibling similarity in family formation. Our sibling design enables us to scrutinize to what extent these shared background characteristics account for sibling similarity in family formation.

Third, by employing sequence analysis and cluster analysis we can examine qualitative patterns of family formation and the nature of siblings' similarity in family formation. Beyond directly testing the impact of parental background characteristics on sibling similarity, we thereby show in which way siblings' family formation is more similar, i.e. which substantive family formation patterns are more likely to 'run in the family'.

BACKGROUND AND THEORY

Family of origin effects on family behavior fall into two broad categories: shared parental background effects and mutual sibling influence. Sibling similarity in family formation can result from either or both of them.

Research on *parental background effects* focuses on intergenerational transmission – the degree to which parents and their children are similar in the occurrence and timing of focal family events: fertility (Barber 2001; Murphy and Knudsen 2002; Murphy 2013), marriage (Feng et al. 1999; Jennings et al. 2012; van Poppel et al. 2008), or divorce (Amato 1996; Amato and DeBoer 2001; Diekmann and Engelhardt 1999; Wolfinger 2000). Most of these studies find relatively modest but significant similarity between parents' and their children's family behavior. Two recent studies compare holistic family formation trajectories of parents and their children directly in a dyadic approach (Fasang and Raab forthcoming; Liefbroer and Elzinga 2012). Liefbroer and Elzinga (2012) conclude that in spite of profound social change in family formation over the past decades, there is persistent continuity in parents' and their children's family formation. Fasang and Raab (forthcoming) show that in addition to strong intergenerational transmission, i.e. similarity, there are other systematic patterns between parental and filial family formation: a pattern of delayed transmission, and a pattern of intergenerational contrast, where children show systematically different family behavior from their parents.

Research on *sibling similarity* in family behavior has concentrated on particular aspects of fertility. Findings include that older siblings, particularly brothers, affect younger siblings' sexual initiation (Haurin and Mott 1990; Widmer 1997), sisters influence one another in the transition to first birth but not notably in second parity transitions (Lyngstad and Prskawetz 2010), the number of siblings affects own family size preferences (Axinn et al. 1994) as well as own fertility (Murphy and Knudsen, 2002). Further, girls' teenage pregnancy

and premarital birth increase their sister's risk of experiencing these events as well (East 1998; East and Jacobson 2001; Powers 2001; Powers and Hsueh 1997). In these studies it is generally unclear both whether the sibling effect extends to the entire family formation process or only specific transitions and how important shared parental background is relative to mutual sibling influence.

This paper aims to extend prior research by examining whether there is sibling similarity in holistic family formation trajectories and what could account for it if it exists. Specifically, we ask if sibling resemblance in family formation is mainly brought about by shared background characteristics. We thereby offer a detailed analysis of one of the two broad categories of family of origin effects - parental background and mutual sibling influence - that potentially generate sibling similarity in family formation. In particular, we examine parental background in terms of parental marital history as an indicator of children's early childhood family structure and parental education as an indicator for socio-economic background. While these two indicators can certainly not capture the full "package" of shared parental background effects, they are two central factors that have been shown to correlate considerably with other observable and unobservable parental background characteristics, such as parenting styles (Chan and Koo 2010; Lareau 2003). Subsequently, we discuss empirical findings and theoretical considerations on these two parental background characteristics.

Parental marital history and early childhood family structure

Research on the intergenerational transmission of family behavior persuasively shows that the structure of the family of origin is associated with family formation in adulthood. Children from disrupted families have a higher risk of divorce than children from intact families (Amato 1996; Diekmann and Engelhardt 1999). Moreover, children of divorce have lower overall rates of marriage (Erola et al. 2012; South 2001), although they tend to have an

elevated risk of teenage marriage (Wolfinger 2003). In addition, children who experience several marital transitions or alternative living arrangements during childhood have a higher “risk of forming a first cohabitational union” (Teachman 2003). More generally, instable living arrangements during childhood promote instable family trajectories in adulthood (Carlson and England 2011).

Family structure is closely linked to economic inequalities. Children from disrupted families often grow up in single-parent households and are more likely exposed to economic hardship (Amato 1996; McLanahan and Percheski 2008), which negatively affects children’s status attainment (McLanahan and Sandefur 1994; Sigle-Rushton and McLanahan 2004). Low socioeconomic status, in turn increases marital instability (Amato 1996; Wolfinger 2005), as it is associated with a number of relationship stressors including higher incidences of poverty, substance abuse and domestic violence. Moreover, parents’ own family experiences shape parental preferences for their children’s family trajectories. Several studies report that divorced parents and their children are more tolerant toward non-traditional family forms (Axinn and Thornton 1996; Cunningham and Thornton 2006). Although, value transmission in families of separation is weaker than in intact families, parental preferences still affect their children’s attitudes (van der Valk et al. 2008).

Siblings might be more similar in their family formation than unrelated persons, simply because they grow up in similar family structures and share their parents’ marital history to a greater extent. Of course, there will be considerable variation in how two siblings experience events such as a parental divorce. They will generally experience divorce at different developmental stages and will have different coping resources available to them. However, compared to random unrelated persons, they are likely to share more of their parental marital history and early childhood family structure. If these factors are influential in generating sibling similarity in family formation, unrelated persons that share similar parental marital history and early childhood family structure experiences, should also be more similar

in their family formation. Sibling similarity in family formation compared to unrelated persons should then be smaller once parental marital history and family structure of the family of origin are factored out.

Parental education

Parental education can influence their children's family formation *directly* and *indirectly* through the transmission of parental education to children's education. Socialization and social control theories (Barber 2000; Bernardi 2003; Starrels and Holm 2000) emphasize a *direct effect* of parental education. Following this rationale, highly educated parents will increase the similarity of their offspring's family trajectories in two regards. First, highly educated parents have higher occupational aspirations for their children and are more likely to favor delayed family formation (Barber 2001; Plotnick 2007; Trent and South 1992). Through socialization parents shape their children's family plans in accordance with these preferences. Although the socialization effect is partly mediated by status transmission, several studies have shown an impact of parental education even after children's education has been taken into account. For instance, evidence from Norway shows a direct influence of parents' education on children's first union formation (Blom 1994; Wiik 2008). Second, parents with higher status are in a better position to exert social control. They likely have more educational and financial resources to influence their children to behave according to their preferences than lower educated parents (Axinn and Thornton, 1992; Barber 2001).

Indirect effects of parental education through intergenerational status transmission are rather obvious given the tight link between educational trajectories and family formation in developed societies. Parental education will then matter to the extent that it predicts children's own education. The prolongation of education during the last decades is seen as a major driver of postponement in the transition to adulthood (Billari and Liefbroer 2010; Fussell and Furstenberg 2005). Transitions like marrying and childbearing usually take place after a

person left the educational system and reached some degree of financial independence. Although higher levels of education are often linked to individualization, educational institutions also standardize early life courses to be more uniform (Brückner and Mayer 2005; Mayer and Schoepflin 1989; Shanahan 2000). The standardizing effect of education continues into later adulthood for highly educated more than for lower educated persons whose lives unfold outside the grip of educational institutions at much younger ages.

Beyond structural effects, education is also related to children's own attitudes and preferences toward family formation. High education is associated with post-material values of self-realization and with liberal attitudes toward family formation (Inglehart and Baker 2000; Lesthaeghe and Neidert 2006; Trent and South 1992). Adolescents in higher educational tracks expect and desire marriage and parenthood at older ages and see themselves at a lower risk of childbearing out of wedlock (Plotnick 2007).

Attitudinal differences by education are also important in structuring gender differences in family formation. To explain educational differences in teenage childbearing and lone motherhood, McLanahan and Percheski (2008) propose that motherhood plays a different role as a source of identity for higher and lower educated women. Lower educated women expect motherhood as psychologically more rewarding and view it as a more essential role in their life course. In contrast, highly educated women "have other possible sources of identity from which they can derive meaning and fulfillment." (McLanahan and Percheski 2008: 262). In addition, the opportunity costs of parenthood are higher for highly educated women. Consequently, they more often postpone the transition to parenthood or remain childless than their lower educated counterparts.

For men on the other hand, both high and low education likely delays family formation compared to medium education, albeit for quite different reasons. Highly educated men delay family formation until they completed extensive education and are situated in an economically secure job, similar to highly educated women. In contrast, in view of the

average increase in women's education, lower educated men with few resources appear as less attractive marriage partners. Struggling to find a partner they postpone marriage and are more likely to live with their parents until later ages (Carlson and England 2011; McLanahan and Percheski, 2008).

Both indirect and direct effects of parental education suggest that siblings will be more similar in their family formation simply because they share parental education and are more likely to have similar educational trajectories themselves. If indirect and direct effects of parental education are important mechanisms generating sibling similarity in family formation, unrelated persons who share the same parental educational background will also be similar to one another in their family formation. Part of the sibling similarity in family formation should then vanish once parental education and own education of siblings and unrelated dyads are taken into account. Subsequently, we empirically test to what extent two crucial parental background characteristics - parental marital history and parental education – account for sibling similarity in family formation.

DATA AND METHODS

The Finnish register data

The empirical analysis uses Finnish register data that consists of a random sample of 1 percent of the population in 1970 (Österbacka 2004; Statistics Finland 1996). All subsequent family members were included in the sample. We use yearly panel data that was collected between 1987 and 2007 for the cohorts born 1969 to 1977 for whom we can construct family formation sequences from age 18 to age 30. Data requirements for the analysis are high: we need to follow individuals over several years to reconstruct their family formation process, and establish intergenerational links to parents and between siblings. Only people who have at least one sibling are included. This allows us to separate the impact of having any siblings from the specific impact of siblings and thus ensures greater rigor in the sibling design.

People are identified as siblings if they have the same mother. Parents are identified through living in the same household as the newborn child. However, until 1987 data was only collected every 5 years. Accordingly, the children were aged 0-4 years when the information was available for the first time. The same applies to parental education, which we defined as the educational level of the highest educated parent in the household. In most cases (71.1 percent) both parents have the same level of education. For 16.8 percent father's education and for 10.6 percent mother's education was higher. For a remaining 1.5 percent data was only available on mother's level of education. We decided to use the earliest available education information for two reasons. First, this strategy ensures consistency in the measurement of parental education because it usually refers to the biological parents. Second, compared to other developmental stages socio-economic conditions during early childhood have been shown to matter most for predicting offspring outcomes in adulthood (Duncan and Brooks-Gunn 2000; Heckman 2006).

For assessing the effect of parental marital history we draw on children's family experiences up to age 18. Since siblings were defined by having the same mother we only refer to the marriage biographies of mothers. We decided against age-specific effects of childhood family structure because they were not supported by recent studies examining the effects of childhood living arrangements on the transition to adulthood (Musick and Meier 2010; Teachman 2003). Moreover, age-specific effects would prohibit creating a robust sample of sibling dyads and unrelated dyads with fixed family background by introducing too much complexity to the matching procedure described below.

Our starting sample comprised 11,803 individuals. We excluded persons with missing data or inconsistencies in the family trajectories. After this selection we only kept families with complete data on all siblings. These restrictions removed 25 percent of the starting sample and reduced the sample size to 9,266 individuals. From these data we extracted 4,994 unique sibling dyads and generated 9,263 unrelated dyads in a procedure detailed in the

following. This adds up to a total of 14,257 dyads for which we can construct family formation trajectories from age 18 to age 30.

The timing and sequencing of family formation in the “third decade of life” (Fussell and Furstenberg 2005) is of particular interest as a “demographically dense phase” with multiple important life course transitions (Rindfuss 1991). In countries like Finland that are considered forerunners of the Second Demographic Transition (van de Kaa 1987; Surkyn and Lesthaeghe 2004) this is particularly true for two markers of the transition to adulthood: leaving the parental home and entry into a first union. Compared to other developed countries children start independent living early in Finland (Billari and Liefbroer 2010). By the age of 20 over half of the children left the parental home. Similarly, Finnish young adults experience a relatively early transition to a first union. The median age of entering a first union is approximately 22 for women and 24 for men born between 1969 and 1981 (Jalovaara 2012). Less than ten percent of all first unions are formal marriages (Finnäs 1995). After ten years 88 percent of cohabiting couples have separated or married (Jalovaara 2013). By the age of 30, almost half of the population has children, two thirds are living in a coresidential union or are married, and about 20 percent live as singles (Statistics Finland 2013). Moreover, Finland is one of the forerunners in female labor force participation; 61 % in 1970 and 70 % as early as 1980 (OECD 2013). This lowers women’s threshold to leave a partner since they are less bound by economic dependence (Jalovaara 2013). In sum, these stylized facts show that family formation sequences until age 30 in Finland provide plenty of variation to meaningfully examine differences between siblings and unrelated persons.

Moreover, the Finnish case can be considered a conservative test of the impact of shared parental background characteristics. Finland, along with other Scandinavian countries, represents a fairly egalitarian social democratic welfare state regime. In this context family of origin effects on offspring’s family formation can be assumed relatively small compared to liberal and conservative welfare regimes as far as they are a by-product of status transmission.

Stratification research has shown that the relationship of family background and socioeconomic status is comparatively low in Finland (Österbacka 2001; Solon 2002). In terms of educational and class mobility, the family of origin effect has even become somewhat weaker during the last decades (Kivinen, Hedman & Kaipainen 2007; Erola 2009). However, the modest intergenerational mobility effect does not imply that social status is irrelevant in Scandinavian countries. Evidence from Norway, for example, shows that parental education directly affects the entry into a first union (Wiik 2008). In addition, recent studies analyzing Finnish data demonstrate an educational gradient in family formation: persons with low education are more likely to experience early parenthood (Jalovaara and Miettinen 2013) and have a higher risk of partnership dissolution (Jalovaara 2013).

Analytical strategy

To address our research questions we constructed two analysis samples. We employed sequence analysis to quantify the difference between two family formation sequences in a dyad (RQ1), used dyadic regression analysis to identify which factors account for this difference (RQ2), and applied cluster analysis to explore substantive family formation patterns and sibling similarity within them (RQ3). In the following section we explain in detail how we created the analysis samples and introduce the methods used.

Analysis sample

As illustrated in the upper panel of figure 1, siblings naturally share the same parental background characteristics, whereas this is generally not the case for unrelated dyads. As a result there is heterogeneity in parental background characteristics for most unrelated dyads but not for sibling dyads. This has two implications: First, it obscures the comparison of siblings and unrelated dyads, such that we are at risk of overestimating the difference between them. Second, it would lead to ambiguous reference categories in the dyadic regression for

parental background characteristics: more combinations of parental background characteristics are possible for unrelated dyads than for sibling dyads. For example, siblings always share the same level of parental education, whereas in unrelated dyads one person might have highly educated parents but the other person not. We therefore construct two analysis samples: First, randomly assigning each sibling an unrelated person to generate dyads of unrelated persons (upper panel of figure 1). Second, to equalize the variation of parental background characteristics between sibling dyads and unrelated dyads, we perform a conditional assignment to generate unrelated dyads by assigning two individuals to one another conditional on sharing the same parental background characteristics (lower panel of figure 1).

- Figure 1: Random and conditional assignment of unrelated dyads -

The analysis samples are constructed as follows. To generate sibling dyads, each of the individual focal children in our sample is matched with a sibling. For two-child families there naturally is only one possible sibling match. In families with more than two siblings (10.5% of all families), we randomly chose a sibling. To ensure that the sample only contains unique sibling dyads we exclude doublets, i.e. a match of the same two siblings once treating sibling 1 as the first dyad member and once treating sibling 2 as the first dyad member. Below we describe the construction of analysis sample one and two using these sibling dyads. Note that the case numbers between analysis sample one and two marginally deviate, due to the exclusion of a few more doublet sibling dyads that are generated during random assignment but not in conditional assignment.

For the first analysis sample, we randomly assign unrelated persons to form 9,261 unrelated dyads and can compare them with 4,981 sibling dyads. To generate the second conditionally assigned analysis sample of unrelated dyads, each focal child is matched with an

unrelated child conditional on sharing the same combination of three parental background characteristics: parental education (low, medium, high)¹, mothers age at first marriage (lowest 25 percent, middle 50 percent and top 25 percent of age distribution) and a dichotomous variable, whether the parents divorced. This yields 18 ($3*3*2$) possible combinations. Based on these conditions we are able to generate 9,263 unrelated dyads, i.e. there is a suitable unique unrelated match for virtually each individual in the sample. They can be compared to 4,994 sibling dyads. This conditional assignment simply equalizes the possible variance in observed parental background characteristics for sibling dyads and unrelated dyads. We later also include these parental background characteristics in the dyadic regression model to directly estimate their impact on similarity in family formation.

Methods

Sequence analysis is used to measure similarity between family formation sequences. We define ten family formation states combining residential situation, relationship status and the number of children: “single, parental home, no child” (SPH), “single, own home, no child” (SNC), “single, own home, 1+ child” (S1C), “cohabiting, no child” (CNC), “cohabiting, 1 child” (C1C), “cohabiting, 2+ children” (C2C), “married, no child” (MNC), “married, 1 child” (M1C), “married, 2 children” (M2C), “married, 3+ children” (M3C). Those who separate from a cohabiting relationship or divorce are considered “single” again, to prioritize their residential situation over their legal marital status. Divorce before age 30 is rare in Finland and cohabitation as a substitute for marriage is relatively common. “Single” generally refers to not being in a cohabiting or married relationship, but might include other relationship forms that we are unable to identify, such as living apart together.

To measure similarity in family formation optimal matching analysis, the most common form of sequence analysis in the social sciences, calculates pairwise distances

¹ Levels of education are defined as follows: low = elementary and lower secondary; medium = secondary and lower university/polytechnic; high = higher university and higher polytechnic.

between all sequences using two transformation operations – substitution and insertion/deletion of a state – to turn one sequence into another. The transformation operations are associated with specific costs and the distance between two sequences is given as the sum of these costs for aligning two sequences. For a comprehensive introduction see MacIndoe and Abbott (2004). Because the sequences are censored at age 30 and the timing and spacing of events is crucial in family formation, we chose an algorithm that emphasizes the timing of events in determining sequence similarity but nonetheless accounts for the order of family formation states as well. This can be achieved by using only substitution operations and no insertion/deletion operations in optimal matching (Lesnard 2010; MacIndoe and Abbott 2004).

The substitution costs are specified to reflect both the substantive closeness of family formation states and the timing of transitions between them. Clearly, “married, 3+ children” is more similar to “married, 2 children”, than to “parental home, single, no children”. Therefore we specify a substitution cost matrix based on the substantive closeness of family formation states (appendix Table A1) where for instance the first substitution above comes at a cost of only 1 and the latter costs 11. To appropriately account for similarity in terms of timing, we combine this substitution cost matrix with the time point specific transition rates between family formation states (probability to transition from one state to another), such that the substitution of states between which transitions occur very frequently are less costly and generate less distance (Lesnard 2010).²

This results in a separate substitution cost matrix for each time point (year) by summing up the substantive substitution cost matrix and the time point specific transition probabilities for each year. The approach can be considered a modified version of the Dynamic Hamming measure (Lesnard 2010) that accounts for the timing and the order of

² For example, the relationship between age and the probability of transitioning from the parental home (SPH) to a single household without children (SNC) follows the form of an inverted-U shape: the substitution costs decrease from age 18 to 24 and increase again afterwards.

family formation events at the same time. As a robustness check we conducted all analyses for three additional cost specifications in the sequence analysis: only Dynamic Hamming Distance, only substantive substitution costs in table A1, and Elzinga's Longest Common Subsequence metric (Elzinga and Liefbroer 2007). Although the regression results slightly varied across cost specifications, the substantive findings were robust. The cluster analysis based on our proposed combined cost specification produced the best cluster solutions according to several cluster cut-off criteria (available upon request) and the highest explanatory power in the regression models and therefore was best suited to capture the variability in our data.

The output of the sequence analysis is a pairwise distance matrix that contains sequence distances for each possible combination of siblings and unrelated persons. Each of the 9,266 individual family formation sequences is compared to all other sequences in a pairwise comparison, which yields $\frac{N(N-1)}{2}$ comparisons = 42,924,745, i.e. cells in the distance matrix. This matrix is the basis for three sets of analyses that speak to our three research questions. First, to address the descriptive question, whether siblings' family formation is more similar than for unrelated dyads, we simply extract the distance values for the sibling dyads and the randomly assigned unrelated dyads (upper panel of figure 1) from the larger distance matrix. The distribution of the sequence distances among siblings is then compared to that among randomly assigned unrelated dyads.

Second, to test whether parental background characteristics account for sibling similarity in family formation we use the dyadic distances for the siblings and the unrelated dyads as the dependent variable in a dyadic regression. The independent variables include sibling status of the dyad, gender constellation, age difference, education, parental education, mother's age of first marriage and parental divorce. Table A2 in the appendix shows descriptive statistics on all independent variables and the dyadic outcome measure used in the regressions. Sibling and unrelated dyads do not differ in parental background characteristics

due to the conditional assignment of dyads but there are small differences for some of the remaining variables. Given the rarity of twins individuals in unrelated dyads have a higher probability of being born in the same year than siblings (11 vs. 3 percent). Siblings in turn have a higher chance of obtaining the same level of education than unrelated persons (52 vs. 44 percent).

We calculate three models. Model 1 (M1) is a baseline model using the first analysis sample of randomly assigned unrelated dyads and including only sibling status. Model 2 (M2) is the same as M1 but using the second analysis sample of conditionally assigned unrelated dyads that are equalized on parental background characteristics to resemble the sibling dyads. Model 3 (M3) is a fully interacted model by sibling status to identify sibling specific effects including the full set of covariates. M3 uses the conditionally assigned unrelated dyads that equalize variation on parental background characteristics and thus allows specifying the same reference categories for siblings and unrelated dyads. The dyads in our analysis samples are unique but by design the same individual can occur more than once in a sibling and an unrelated dyad. We therefore calculate robust standard errors.

Third, to examine whether siblings are concentrated in specific substantive family formation patterns, we apply Ward cluster analysis using the entire distance matrix. Several cluster cut-off criteria, including Point Biserial Correlation, Average Silhouette Width, and Hubert's C, support a five clusters solution (Hennig and Liao 2010; Kaufman and Rousseeuw 1990). We then examine whether siblings have a higher probability to be in the same family formation cluster than unrelated dyads, and whether this is particularly the case for specific substantive family formation patterns. The sequence analysis and the calculation of different cluster cut-off criteria were conducted using the TraMineR and the WeightedCluster packages in R (Gabadinho et al. 2011; Studer 2013).

RESULTS

Is siblings' family formation more similar (RQ 1) and what role do shared parental background characteristics play for sibling similarity (RQ 2)?

We begin by establishing that sibling's family formation is significantly more similar than for unrelated dyads, particularly for same sex siblings. Figure 2 shows the distribution of the dyadic sequence distances for siblings (light grey line) and unrelated dyads (dark grey line) separately for the complete sample, opposite sex dyads, female dyads and male dyads for analysis sample one using the randomly assigned unrelated dyads. The distances are normalized between zero and 100, where zero indicates two identical sequences and 100 indicates the most dissimilar family formation trajectories. The width of the curves reflects 90 percent asymptotic confidence intervals. For all subgroups the two curves for siblings and unrelated dyads largely do not overlap indicating that in these locations of the distribution the difference between them is statistically significant. This difference is particularly pronounced for same sex siblings, where the gap between the two curves is largest (bottom graphs in figure 2).

- Figure 2: Distribution of sequence distances among sibling dyads and randomly assigned unrelated dyads (analysis sample one) -

Family formation sequences are most different for unrelated women followed by sisters, unrelated men and brothers. Note that sisters' family formation is still more different (42) than for unrelated men (38.9) up to the age of 30. Nonetheless, the difference between siblings and unrelated persons is slightly larger for sisters at $47.8 - 42 = 5.8$ compared to $38.9 - 33.9 = 5$ for brothers.

These findings underline the primacy of gender in structuring family formation in two regards. First, same sex siblings are much more similar to one another than opposite sex

siblings. Second, the higher average distances for women might partly result from more complexity within individual family formation sequences. There might simply be more going on in women's trajectories with more family states and more frequent changes between them. A sequence complexity measure (Elzinga and Liefbroer 2007; Elzinga 2010) that captures the number and duration variation of different family formation states within individual sequences confirms significantly higher sequence complexity for women than for men (available upon request). If most men's family formation events are simply delayed, expanding the observation window to age 40 might somewhat equalize this difference between men and women. However the significantly higher sequence similarity for brothers than for unrelated men also until the age of 30, suggests that the censoring does not majorly distort the analysis of sibling similarity in family formation. Our findings further highlight that women are much more actively engaged in family formation in the crucial years between age 18 and 30 that coincide with the accumulation of educational credentials and labor market entry and thus set the stage for future life-time labor market success.

Overall, the sibling effect, calculated as the difference in mean distance between siblings and randomly assigned unrelated dyads, is $44.1 - 40.1 = 4$. Sibling similarity after conditional assignment (table A2) is slightly reduced to $43.7 - 40.2 = 3.5$. The conditional assignment on parental background characteristics thus reduces sibling similarity by 12.5 percent, but siblings remain significantly more similar in their family formation also after conditional assignment.

The dyadic regression presented in table 1 further supports this finding. Negative coefficients indicate less distance and thus more similar family formation in a dyad. Model 1 (M1) shows that siblings are on average significantly more similar to one another than randomly assigned unrelated dyads by -3.94 on the distance measure that ranges from zero to 100. Except for rounding errors this corresponds almost exactly to the mean difference of 4.00 established in figure 2. Model 2 (M2) shows that this effect remains similar at a difference of

-3.52 for siblings and conditionally assigned unrelated dyads corresponding to the 12.5 percent reduction in sibling similarity established above.

Even though many of the covariates in M3 significantly account for the variance in the dyadic distance measure and the adjusted R-square increases from 0.01 in M1 to 0.12 in M3, they add little to explaining sibling similarity. The sibling main effect remains significant when including the full set of covariates (M3), but is reduced to -2.69. Note that in M3 the coefficients for the main effects in the left column refer to the effects for unrelated dyads. The interaction effects in the right column capture the additional effects for siblings. We find only three significant sibling interactions: for sisters, twins and high education. Overall, the results support that sibling similarity in family formation is only moderately generated by the compositional effects of shared parental background characteristics in terms of early childhood family structure and parental education (RQ2). They affect similarity in family formation in the direction we would expect - but as can be seen from the interaction effects in M3 mostly for siblings and unrelated dyads alike.

- Table 1: Ordinary Least Squares Regression predicting dyadic distance -

The results from M3 substantiate the descriptive gender findings (figure 2): women are overall most different in their family formation but sisters are relatively more similar than brothers compared to conditionally assigned unrelated dyads of the respective sex. This is visible in the positive main effect for female dyads and the significant negative interaction for sisters. In contrast, unrelated men and brothers both are more similar to one another by 6.91 distance points than opposite sex dyads.

An age difference of more than three years increases sequence distance slightly by 0.89 for siblings and unrelated dyads - possibly a “mini cohort effect” for successive birth cohorts. There is a strong twin effect of -7.2 measured as siblings who were born in the same year. A

closer examination of the sequences revealed that this effect is mainly driven by the joint experience of home leaving. The probability to leave the parental home in the same year is 27 percent for twins compared to only 13 percent for siblings, 13 percent for unrelated young adults who were born in the same year, and 11 percent for unrelated young adults who were not born in the same year.

The results support the paramount importance of education in structuring early life courses and family formation. The higher the dyads combined education, the more similar is their family formation for both siblings and unrelated dyads. This underlines the greater heterogeneity of family formation among the lower educated whose early life courses unfold outside of educational institutions from a much younger age. The only sibling-specific effect is that two highly educated siblings are significantly more similar in their family formation than two highly educated unrelated persons, net of controlling for parental education. Additional analyses (available upon request) on the interaction of gender and education show that education has a much stronger effect for female than for male dyads substantiating that women's family formation is more tightly interrelated with their educational and employment trajectories during the third decade of life.

High parental education generates more similar family formation for siblings and unrelated dyads alike, net of their own education. Apparently parental education has a direct impact beyond indirect effects through intergenerational status transmission as posited in socialization and social control theories. Possibly, highly educated parents make use of greater resources to guide their children's family formation, including financial support for a longer time into early adulthood. In contrast, children of lower educated parents might leave the radar of parental influence and establish independence at younger ages.

In line with our expectations, parental marital history as an indicator of early childhood family structure equally impacts similarity in family formation. Like education, the effects do not vary for siblings and unrelated dyads. Two members of a dyad are more similar to one

another if their mother married late, likely by also delaying their own family formation. They are less similar to one another if their mother married young. This dissimilarity might be driven by intergenerational behavioral transmission: if they start family formation early themselves there is more going on early on in the sequences and the risk of future union disruption and instable family formation increases. Additional analyses substantiated that sequence complexity is indeed significantly higher for children whose mother's married young than for those whose mothers married late.

In line with previous research (McLanahan and Bumpass 1988; Wolfinger 2000), we find a similar pattern for divorce: calculating the complexity measure showed that children whose parents' divorced have more instable family formation themselves and are less similar to one another in their family formation, again within sibling and unrelated dyads alike. Parental divorce is a very different experience even for siblings in the same family depending on the age of the child, custodial arrangements and the level of parental conflict prior, during and after the divorce (Amato, 2000). Our results underline this heterogeneity in divorce experiences and refute a uniform impact of parental divorce on siblings' family formation.

We conclude that parental background characteristics impact similarity in family behavior much in line with our expectations, but contribute surprisingly little to explaining sibling similarity in family formation. To gain a deeper insight into the nature of this sibling similarity, we next directly examine whether siblings have a higher probability to experience the same substantive family formation patterns than unrelated dyads.

In which way is siblings' family formation similar (RQ 3)?

Figure 3 shows relative frequency sequence and dissimilarity-to-medoid plots (Fasang and Liao 2013) for the five family formation patterns derived with cluster analysis. The clusters are ordered ascending according to average education in each group starting with the most educated group in the top panel of figure 3. Within each cluster the sequences are sorted

by the age of leaving the parental home for the first time. Each horizontal line in the left panel of Figure 3 represents one individual sequence, with different colors indicating different family formation states. It is impossible to plot all 9,266 sequences due to over-plotting. The left panels of Figure 3 therefore show a selection of representative sequences for each cluster that comprise 2 percent of the sample (186 representative sequences). They are chosen as the medoid sequence (Aassve et al. 2007), the sequence with the minimum distance to all other sequences within equal sized frequency groups across the clusters (see Fasang and Liao 2013 for details). The dissimilarity-to-medoid plots in the right panels of Figure 3 indicate the degree of sequence variation within each cluster and thereby are informative about how homogeneous each family formation pattern is. They are box-plots of the mean and interquartile range of the pairwise sequence distances between the chosen representative medoid sequences displayed in the left panels of Figure 3 and the other sequences that this medoid represents. The size of the clusters is proportional to their relative frequency in the population. Table 2 shows descriptive statistics for the clusters. We subsequently describe the five family formation patterns referring to Figure 3 and Table 2.

- Figure 3: Relative Frequency Sequence Plots of five family formation clusters -

- Table 2: Description of five family formation clusters -

The highest educated cluster (66 percent medium or high education, table 2) at the top of figure 3 follows a “traditional” pathway of leaving the parental home, cohabitating, marrying in their late twenties and having one or two children by age 30. Only 26 percent of young adults in this group do not experience the transition to parenthood until age 30.

In the middle of the educational distribution we find two groups that are characterized by childlessness (roughly 85 percent childless): the first is called “home stayers” because they leave the parental home at the latest age with a median home leaving age of 25 compared to

age 22 in the total sample (table 2). The left panel of Figure 3 shows that this is the most homogenous group with low dissimilarities to the representative sequences displayed in the right panel. Extended living in the parental home is primarily a male experience (66 percent men). The second group with medium education, called “living alone/transitional cohabitation”, shows a pattern of establishing an independent household in the early twenties (median age 21) and moving in and out of transitional periods of cohabitation throughout their twenties. Finally, we find two smaller but highly distinct family formation patterns at the bottom of the educational distribution: “early marriage, high fertility” and “parenthood out of wedlock”. In these groups, roughly 70 percent have the lowest level of education. The same is true for their parents. Compared to the population average their family formation is characterized by early parenthood (median ages 23 and 24) and high parities (2.46 and 1.72) either within marriage or out of wedlock in cohabiting unions and as single parents. Moreover, figure 3 illustrates that virtually all persons in these two patterns move directly into a co-residential union when leaving the parental home. The high distances to the medoid in the right panels of Figure 3 shows that these are particularly heterogeneous groups. The corresponding family formation experiences are primarily made by women (70 and 69 percent).

The clusters underline the stratification of family formation by education and gender visible in the dyadic regression (table 1): low education goes along with more eventful and variable family formation trajectories of high parities and parenthood out of wedlock. High education is associated with a “traditional” timing and sequencing of family formation and establishing a one- or two-child family secured in marriage by age 30. Another pattern that is characteristic for high and medium education is a relatively uniform delay of family formation (table 2). The two postponement clusters are the most prevalent patterns and represent more than half of the study population.

- Figure 4: Probability of two dyad members to be in the same cluster -

After establishing these main patterns of family formation among our cohorts of young adults in Finland, we can address research question three, whether siblings are concentrated in any of these patterns. Figure 4 shows the probability of the focal person's dyad partners – siblings and conditionally assigned unrelated persons – to be in the same cluster as the focal persons. If two members of a sibling dyad have a higher probability to be in the same cluster than two members of an unrelated dyad this indicates that specific family formation patterns are concentrated within families and offers insights into the nature of sibling similarity in family formation. The line in the middle of figure 4 and the corresponding percentages show the overall probability to be in the same cluster as the focal person conditional on the focal person's cluster membership. The graph sets this line to zero to illustrate deviations from the mean for siblings and unrelated persons. First, across all clusters siblings have a higher probability to be in the same group as the focal person (solid lines on the right of figure 4) than unrelated persons (dashed lines on the left). For instance, the mean probability to be in the “extended cohabitation, parenthood out of wedlock” group – given that the focal person is in this cluster – is 22 percent for siblings but only 14 percent for unrelated persons. This corresponds to a 29 percent deviation from the mean, i.e. a 29 percent higher probability for a sibling to be in the same cluster as the focal person than for the population average. The clusters in figure 4 are ordered descending according to the degree of sibling concentration within each group. Siblings particularly have a higher probability to be in the same cluster in the family formation patterns that are associated with educational disadvantage (see table 2): “Extended cohabitation, parenthood out of wedlock” and “early marriage, high fertility”.

DISCUSSION

In this paper we propose a novel research design using sibling comparisons and sequence analysis to study family of origin effects on holistic family formation trajectories. Findings show that siblings are moderately but significantly more similar to one another in holistic family formation trajectories than unrelated persons. This is particularly the case for same sex siblings. Equalizing siblings' and unrelated dyads' parental background characteristics decreases the sibling similarity moderately by about 13 percent. Moreover, sibling similarity is particularly pronounced for sisters in family formation patterns of 'extended cohabitation, parenthood out of wedlock' and 'early marriage, high fertility'. These family formation patterns are at the same time associated with economic disadvantage.

In addition to the moderate decrease of sibling similarity through parental background, the dyadic regression showed that parental background largely affects similarity in family formation for siblings and unrelated dyads in the same way. However, we want to highlight two notable sibling-specific interactions, which reduce sibling's dissimilarity in family formation. First, twins, measured as siblings born in the same year, are much more similar in their family formation than other siblings. On the one hand, shared genetic background may account for this twin effect (Kohler et al. 1999, 2002). On the other hand, twins also share more of their environment and socialization experiences than siblings. There might be a culture or norm for twins to jointly step through 'rites of passage' in the transition to adulthood, such as leaving the parental home together or orchestrating a double marriage. As noted above, our results indeed support a clustering of joined home leaving for twins. Small case numbers (133 twins) and the relatively crude twin indicator unfortunately prohibited further investigation of the twin effect in this study. Second, family formation is significantly more similar for highly educated siblings than for two highly educated unrelated persons, net of parental education. This corresponds with Conley and Glauber's (2008) finding that sibling correlations in education and earnings are strongest among siblings from advantaged families.

Our study suggests that this might also be the case for similarities in family formation. Arguably there is particularly strong mutual reinforcement of specific family formation patterns among highly educated siblings.

Overall, we conclude that the observed parental background characteristics and sibling interactions only play a relatively marginal role in generating sibling similarity in family formation. At the same time, the strong impact of respondents' own educational level and gender supports continuing importance of the family of origin for shaping people's family behavior. These effects might be even larger in countries with closed stratification systems in which intergenerational status transmission is stronger than in Finland. The same is true for the size of the sibling effect. Cross-national comparisons show relatively low sibling correlations in earnings for Nordic countries (e.g., Björklund et al. 2002). Systematic comparisons with countries representing other social mobility and welfare state regimes would be useful to test whether this also holds for sibling correlations in family formation. Nonetheless, our study supports that even in Finland siblings are notably and significantly more similar in family formation than unrelated persons.

At first glance, the explanatory power of the sibling effect might seem small, explaining only one percent of the variance as indicated by the \mathcal{R}^2 in table 1. Generally, sequence distances are a highly complex outcome where small effect sizes and relatively low \mathcal{R}^2 are common (see Liefbroer and Elzinga 2012). Further, a comparison with the explanatory power of education puts the explanatory power of the sibling effect into perspective. An additional regression model including only education on the conditionally matched sample accounts for five percent of the variance indicated by \mathcal{R}^2 (available upon request). This means that the sibling indicator accounts for as much as 20 percent of the variance in the sequence distances as education accounts for. Given the uncontested crucial relevance of education for demographic behavior, a sibling effect of this size, even if comparatively moderate, warrants future research. Family internal dynamics, such as parent-child relationship quality and

mutual sibling influence, might generate the remaining unexplained sibling similarity. The moderate size of the sibling effect can also be interpreted in the context of these family internal dynamics. Although siblings share the same family of origin they also have many experiences that they do not share, such as differential parental treatment or sibling position (East and Jacobson 2000; Whiteman et al. 2011). Our findings support that overall, the mechanisms that generate sibling similarity in family formation outweigh such differentiating factors.

To gain a deeper insight into the nature of sibling similarity in family formation, we subsequently showed that substantive family formation patterns that are associated with economic disadvantage are concentrated within families. Sisters are concentrated in family formation patterns of ‘extended cohabitation, parenthood out of wedlock’ and ‘early marriage, high fertility’. That is, specific gendered family formation patterns and economic disadvantage are encapsulated within families and thus contribute to the reproduction of inequality (McLanahan and Percheski 2008). These results suggest that siblings reinforce one another in following family formation patterns stratified by education and gender. Further, unobserved parental effects could generate particularly strong sibling clustering in certain family formation patterns. East and Jacobson (2003: 395), for instance, point out that “the disproportionately high teenage birth rates among the sisters of childbearing teens” could be partly explained by the strained parenting of their mothers, which in turn was caused by the increased financial and emotional burdens following the first teenage motherhood in the family. In response to the negative parental treatment the younger sisters of teenage mothers are more likely to engage in problem behavior and are at a high risk of early pregnancy as well.

To further improve our understanding of sibling similarity in family formation, future research should address several issues. First, lacking data on the relationship quality between parents and their children as well as among siblings, we were unable to account for possible

moderating effects of these psychological characteristics. Close emotional relationships can function as ‘transmission belts’ (Schönpflug 2001) and could be one explanation for the remaining unexplained sibling similarity we observe. Second, more detailed indicators of parents’ and siblings’ social and occupational status would be desirable to assess family background effects in greater detail. In this study, we were able to rigorously scrutinize the impact of two crucial background factors that have been shown to correlate with many other potentially relevant parental background characteristics, such as parenting styles (Chan and Koo 2010; Lareau 2003). Third, we had no information on interactions between siblings and could not directly measure mutual sibling influence in family behavior - an important alternative explanation for the sibling similarity in family formation. Finally, the analysis of longer family formation trajectories until the age of 40 and beyond, would allow for studying potential variation of the degree of sibling similarity across the life course. Previous research suggests that family of origin effects are more pronounced for early transitions and diminish as people age. For instance Lyngstad and Prskawetz (2010), show that sibling’s birth events influence the transition to first parenthood but lose relevance for higher order parities. The literature on older sisters’ out-of-wedlock or teenage childbearing and their younger sister’s risk of experiencing these events as well points into a similar direction (e.g., East and Jacobson 2001; Powers and Hsueh, 1997). Since men experience family formation events later than women an extension of the observation period might weaken the sibling-gender interaction we found for women and strengthen it for men.

In sum, our findings show that there is significant sibling similarity in family formation and that shared parental education and parental marital history add little to explaining this similarity. Instead education and gender stand out as major stratifying forces of family formation for both siblings and unrelated dyads. Yet, sibling status seems to reinforce the impact of education and gender. As a result, compared to unrelated dyads siblings are particularly overrepresented in family formation patterns that go along with

specific constellations of educational levels and gender. As our cluster results show, scrutinizing the nature and driving forces of sibling similarity in family formation can potentially improve our insight into the link between family behavior and the reproduction of social inequality. Combining a sibling design and sequence analysis offers insights both into the amount of similarity in the family formation trajectories of siblings and unrelated dyads and into the substantive content of this similarity in the family formation clusters. This approach is in principle easily transferable and might yield promising results when applied to other research questions, such as sibling similarity in educational and employment trajectories, or health trajectories.

REFERENCES

- Aassve, A., Billari, F. C., & Piccarreta, R. (2007). Strings of Adulthood: A Sequence Analysis of Young British Women's Work-Family Trajectories. *European Journal of Population*, 23, 369–388. doi:10.1007/s10680-007-9134-6
- Amato, P. R. (1996). Explaining the Intergenerational Transmission of Divorce. *Journal of Marriage and Family*, 58, 628–640.
- Amato, P. R., & DeBoer, D. D. (2001). The Transmission of Marital Instability across Generations: Relationship Skills or Commitment to Marriage? *Journal of Marriage and Family*, 63, 1038–1051.
- Axinn, W. G., Clarkberg, M. E., & Thornton, A. (1994). Family Influences on Family Size Preferences. *Demography*, 31, 65–79.
- Axinn, W. G., & Thornton, A. (1992). The influence of parental resources on the timing of the transition to marriage. *Social Science Research*, 21, 261–285.
- Axinn, W. G., & Thornton, A. (1996). The Influence of Parents' Marital Dissolutions on Children's Attitudes Toward Family Formation. *Demography*, 33, 66-81. doi:10.2307/2061714
- Barber, J. S. (2000). Intergenerational Influences on the Entry into Parenthood: Mothers' Preferences for Family and Nonfamily Behavior. *Social Forces*, 79, 319–348. doi:10.1093/sf/79.1.319
- Barber, J. S. (2001). The Intergenerational Transmission of Age at First Birth among Married and Unmarried Men and Women. *Social Science Research*, 30, 219–247.
- Bernardi, L. (2003). Channels of Social Influence on Reproduction. *Population Research and Policy Review*, 22(5/6), 527–555. doi:10.1023/B:POPU.0000020892.15221.44
- Billari, F. C., & Liefbroer, A. C. (2010). Towards a new pattern of transition to adulthood? *Advances in Life Course Research*, 15, 59–75. doi:10.1016/j.alcr.2010.10.003
- Björklund, A., Eriksson, T., Jäntti, M., Raaum, O., & Österbacka, E. (2002). Brother correlations in earnings in Denmark, Finland, Norway and Sweden compared to the United States. *Journal of Population Economics*, 15, 757–772. doi:10.1007/s001480100095
- Blom, S. (1994). Marriage and Cohabitation in a Changing Society: Experience of Norwegian Men and Women Born in 1945 and 1960. *European Journal of Population*, 10, 143–73.
- Brückner, H., & Mayer, K. U. (2005). De-Standardization of the Life Course: What it Might Mean? And if it Means Anything, Whether it Actually Took Place? *Advances in Life Course Research*, 9, 27–53. doi:10.1016/S1040-2608(04)09002-1
- Carlson, M. J., & England, P. (Eds.). (2011). *Social Class and Changing Families in an Unequal America*. Stanford University Press.

- Chan, T. W., & Koo, A. (2010). Parenting Style and Youth Outcomes in the UK. *European Sociological Review*, 27, 385–399. doi:10.1093/esr/jcq013
- Conley, D. (2008). Bringing Sibling Differences In: Enlarging Our Understanding of the Transmission of Advantage in Families. In A. Lareau & D. Conley (Eds.), *Social Class: How Does It Work?* (pp. 179–200). New York: Russell Sage Foundation.
- Conley, D., & Glauber, R. (2008). All in the family? Family composition, resources, and sibling similarity in socioeconomic status. *Research in Social Stratification and Mobility*, 26, 297–306. doi:10.1016/j.rssm.2008.08.003
- Cunningham, M., & Thornton, A. (2006). The Influence of Parents' Marital Quality on Adult Children's Attitudes Toward Marriage and Its Alternatives: Main and Moderating Effects. *Demography*, 43, 659–672. doi:10.1353/dem.2006.0031
- Diekmann, A., & Engelhardt, H. (1999). The Social Inheritance of Divorce: Effects of Parent's Family Type in Postwar Germany. *American Sociological Review*, 64, 783–793.
- Duncan, G. J., & Brooks-Gunn, J. (2000). Family Poverty, Welfare Reform, and Child Development. *Child Development*, 71, 188–196. doi:10.1111/1467-8624.00133
- East, P. L. (1998). Impact of Adolescent Childbearing on Families and Younger Sibling: Effects that Increase Younger Siblings' Risk for Early Pregnancy. *Applied Developmental Science*, 2, 62–74. doi:10.1207/s1532480xads0202_1
- East, P. L., & Jacobson, L. J. (2000). Adolescent Childbearing, Poverty, and Siblings: Taking New Direction From the New Literature. *Family Relations*, 49, 287–292. doi:10.1111/j.1741-3729.2000.00287.x
- East, P. L., & Jacobson, L. J. (2001). The Younger Siblings of Teenage Mothers: A Follow-Up of Their Pregnancy Risk. *Developmental Psychology*, 37, 254–264.
- East, P. L., & Jacobson, L. J. (2003). Mothers' differential treatment of their adolescent childbearing and nonchildbearing children: contrasts between and within families. *Journal of Family Psychology*, 17, 384–396. doi:10.1037/0893-3200.17.3.384
- Elder, G. H., Johnson, M. K., & Crosnoe, R. (2003). The Emergence and Development of Life Course Theory. In J. T. Mortimer & M. J. Shanahan (Eds.), *Handbook of the Life Course* (pp. 3–19). Springer.
- Elzinga, C. H. (2010). Complexity of Categorical Time Series. *Sociological Methods & Research*, 38, 463–481. doi:10.1177/0049124109357535
- Elzinga, Cees H., & Liefbroer, A. C. (2007). De-standardization of Family-Life Trajectories of Young Adults: A Cross-National Comparison Using Sequence Analysis. *European Journal of Population*, 23, 225–250. doi:10.1007/s10680-007-9133-7
- Erola, J. (2009). Social Mobility and Education of Finnish Cohorts Born 1936--75: Succeeding While Failing in Equality of Opportunity? *Acta Sociologica*, 52, 307–327. doi:10.1177/0001699309348701

- Erola, Jani, Härkönen, J., & Dronkers, J. (2012). More Careful or Less Marriageable? Parental Divorce, Spouse Selection and Entry into Marriage. *Social Forces*, *90*, 1323–1345. doi:10.1093/sf/sos073
- Fasang, A. E., & Liao, T. F. (2013). Visualizing Sequences in the Social Sciences: Relative Frequency Sequence Plots. *Sociological Methods & Research*. doi:10.1177/0049124113506563
- Feng, D., Giarrusso, R., Bengtson, V. L., & Frye, N. (1999). Intergenerational Transmission of Marital Quality and Marital Instability. *Journal of Marriage and Family*, *61*, 451–463.
- Finnäs, F. (1995). Entry into Consensual Unions and Marriages Among Finnish Women Born Between 1938 and 1967. *Population Studies*, *49*, 57–70. doi:10.1080/0032472031000148246
- Fussell, E., & Furstenberg, F. F. (2005). The Transition to Adulthood During the Twentieth Century. In R. A. Settersten, F. F. Furstenberg, & R. G. Rumbaut (Eds.), *On the Frontier of Adulthood* (pp. 29–75). Chicago & London: The University of Chicago Press.
- Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, *40*, 1–37.
- Haurin, R. J., & Mott, F. L. (1990). Adolescent sexual activity in the family context: the impact of older siblings. *Demography*, *27*, 537–57.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–2. doi:10.1126/science.1128898
- Hennig, C., & Liao, T. F. (2010). Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification. London. Retrieved from <http://www.ucl.ac.uk/statistics/research/pdfs/rr308.pdf>
- Inglehart, R., & Baker, W. E. (2000). Modernization, Cultural Change, and the Persistence of Traditional Values. *American Sociological Review*, *65*, 19–51. doi:10.2307/2657288
- Jalovaara, M. (2012). Socio-economic resources and first-union formation in Finland, cohorts born 1969-81. *Population studies*, *66*, 69–85. doi:10.1080/00324728.2011.641720
- Jalovaara, M. (2013). Socioeconomic Resources and the Dissolution of Cohabitations and Marriages. *European Journal of Population*, *29*, 167–193. doi:10.1007/s10680-012-9280-3
- Jalovaara, M., & Miettinen, A. (2013). Does his paycheck also matter? *Demographic Research*, *28*, 881–916. doi:10.4054/DemRes.2013.28.31
- Jennings, E. A., Axinn, W. G., & Ghimire, D. J. (2012). The Effect of Parents' Attitudes on Sons' Marriage Timing. *American Sociological Review*, *77*, 923–945. doi:10.1177/0003122412464041

- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data. An Introduction to Cluster Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9780470316801
- Kivinen, O., Hedman, J., & Kaipainen, P. (2007). From Elite University to Mass Higher Education: Educational Expansion, Equality of Opportunity and Returns to University Education. *Acta Sociologica*, *50*, 231–247. doi:10.1177/0001699307080929
- Kohler, H.-P., Rodgers, J. L., & Christensen, K. (2002). Between nurture and nature: the shifting determinants of female fertility in Danish twin cohorts. *Social Biology*, *49*, 218–248.
- Kohler, H.-P., Rodgers, J. L., Christensen, K., & Christense, K. (1999). Is Fertility Behavior in Our Genes? Findings from a Danish Twin Study. *Population and Development Review*, *25*, 253–288. doi:10.1111/j.1728-4457.1999.00253.x
- Kuziemko, I. (2006). Is Having Babies Contagious? Estimating Fertility Peer Effects Between Siblings.
- Lareau, A. (2003). *Unequal Childhoods. Class, Race, and Family Life*. Berkeley: University of California Press.
- Lesnard, L. (2010). Setting Cost in Optimal Matching to Uncover Contemporaneous Socio-Temporal Patterns. *Sociological Methods & Research*, *38*, 389–419. doi:10.1177/0049124110362526
- Lesthaeghe, R. J., & Neidert, L. (2006). The Second Demographic Transition in the United States: Exception or Textbook Example? *Population and Development Review*, *32*, 669–698. doi:10.1111/j.1728-4457.2006.00146.x
- Liefbroer, A. C., & Elzinga, C. H. (2012). Intergenerational transmission of behavioural patterns: How similar are parents' and children's demographic trajectories? *Advances in Life Course Research*, *17*, 1–10. doi:10.1016/j.alcr.2012.01.002
- Lyngstad, T. H., & Prskawetz, A. (2010). Do Siblings' Fertility Decisions Influence Each Other? *Demography*, *47*, 923–934.
- MacIndoe, H., & Abbott, A. (2004). Sequence Analysis and Optimal Matching Techniques for Social Science Data. In A. Bryman & M. Hardy (Eds.), *Handbook of Data Analysis* (pp. 387–406). London: SAGE Publications.
- Mayer, K. U., & Schoepflin, U. (1989). The State And The Life Course. *Annual Review of Sociology*, *15*, 187–209.
- McLanahan, S., & Bumpass, L. (1988). Intergenerational Consequences of Family Disruption. *The American Journal of Sociology*, *94*, 130–152.
- McLanahan, S. S., & Percheski, C. (2008). Family Structure and the Reproduction of Inequalities. *Annual Review of Sociology*, *34*, 257–276. doi:10.1146/annurev.soc.34.040507.134549

- McLanahan, S. S., & Sandefur, G. (1994). *Growing Up with a Single Parent: What Hurts, What Helps?* Cambridge, MA: Harvard Univ. Press.
- Murphy, M. (2013). Cross-national patterns of intergenerational continuities in childbearing in developed countries. *Biodemography and social biology*, *59*, 101–26. doi:10.1080/19485565.2013.833779
- Murphy, M., & Knudsen, L. B. (2002). The Intergenerational Transmission of Fertility in Contemporary Denmark: The Effects of Number of Siblings (Full and Half), Birth Order, and Whether Male or Female. *Population Studies*, *56*, 235–248.
- Musick, K., & Meier, A. (2010). Are both parents always better than one? Parental conflict and young adult well-being. *Social science research*, *39*, 814–30. doi:10.1016/j.ssresearch.2010.03.002
- OECD. (2013). Civilian labour force, females % of population 15–64. AFLS Summary Tables. Retrieved from <http://www.stats.oecd.org>
- Österbacka, E. (2001). Family Background and Economic Status in Finland. *Scandinavian Journal of Economics*, *103*, 467–484. doi:10.1111/1467-9442.00255
- Österbacka, E. (2004). *It Runs in the Family. Empirical Analyses of Family Background and Economic Status*. Åbo: Åbo Akademis förlag.
- Plotnick, R. D. (2007). Adolescent expectations and desires about marriage and parenthood. *Journal of Adolescence*, *30*, 943–63. doi:10.1016/j.adolescence.2007.01.003
- Popenoe, D. (1993). American Family Decline, 1960-1990: A Review and Appraisal. *Journal of Marriage and Family*, *55*, 527–542.
- Powers, D. A. (2001). Unobserved Family Effects on the Risk of a First Premarital Birth. *Social Science Research*, *30*, 1–24. doi:10.1006/ssre.2000.0682
- Powers, D. A., & Hsueh, J. C.-T. (1997). Sibling models of socioeconomic effects on the timing of first premarital birth. *Demography*, *34*, 493–511. doi:10.2307_3038305
- Rindfuss, R. R. (1991). The Young Adult Years: Diversity, Structural Change, and Fertility. *Demography*, *28*, 493. doi:10.2307/2061419
- Schönpflug, U. (2001). Intergenerational Transmission of Values: The Role of Transmission Belts. *Journal of Cross-Cultural Psychology*, *32*, 174–185. doi:10.1177/0022022101032002005
- Shanahan, M. J. (2000). Pathways to Adulthood in Changing Societies: Variability and Mechanisms in Life Course Perspective. *Annual Review of Sociology*, *26*, 667–692.
- Sigle-Ruston, W., & McLanahan, S. S. (2004). Father Absence and Child Wellbeing: A Critical Review. In D. P. Moynihan, L. Rainwater, & T. Smeeding (Eds.), *The Future of the Family* (pp. 116–155). New York: Russel Sage Foundation.

- Solon, G. (2002). Cross-Country Differences in Intergenerational Earnings Mobility. *Journal of Economic Perspectives*, *16*, 59–66. doi:10.1257/089533002760278712
- South, S. J. (2001). The Variable Effects of Family Background on the Timing of First Marriage: United States, 1969–1993. *Social Science Research*, *30*, 606–626. doi:10.1006/ssre.2001.0714
- Starrels, M. E., & Holm, K. E. (2000). Adolescents' Plans for Family Formation: Is Parental Socialization Important? *Journal of Marriage and Family*, *62*, 416–429.
- Statistics Finland. (1996). *The Finnish Longitudinal Census Data File*. Helsinki: Statistics Finland.
- Statistics Finland. (2013). Population according to area, gender, age, year and family type. Retrieved from <http://tilastokeskus.fi/meta/til/>
- Studer, M. (2013). WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. *LIVES Working Papers*, *24*.
- Surkyn, J., & Lesthaeghe, R. (2004). Value Orientations and the Second Demographic Transition (SDT) in Northern, Western and Southern Europe: An Update. *Demographic Research, Special 3*, 45–86. doi:10.4054/DemRes.2004.S3.3
- Teachman, J. D. (2003). Childhood Living Arrangements and the Formation of Coresidential Unions. *Journal of Marriage and Family*, *65*, 507–524. doi:10.1111/j.1741-3737.2003.00507.x
- Trent, K., & South, S. J. (1992). Sociodemographic Status, Parental Background, Childhood Family Structure, and Attitudes toward Family Formation. *Journal of Marriage and the Family*, *54*, 427. doi:10.2307/353074
- Van De Kaa, D. J. (1987). Europe's second demographic transition. *Population bulletin*, *42*, 1–59.
- Van Der Valk, I., Spruijt, E., De Goede, M., Larsen, H., & Meeus, W. (2008). Family Traditionalism and Family Structure. *European Psychologist*, *13*, 83–95. doi:10.1027/1016-9040.13.2.83
- Van Poppel, F., Monden, C., & Mandemakers, K. (2008). Marriage Timing over the Generations. *Human Nature*, *19*, 7–22. doi:10.1007/s12110-008-9033-5
- Whiteman, S. D., McHale, S. M., & Soli, A. (2011). Theoretical Perspectives on Sibling Relationships. *Journal of Family Theory & Review*, *3*, 124–139. doi:10.1111/j.1756-2589.2011.00087.x
- Widmer, E. D. (1997). Influence of Older Siblings on Initiation of Sexual Intercourse. *Journal of Marriage and Family*, *59*, 928–938.
- Wiik, K. A. (2008). “You’d Better Wait!”--Socio-economic Background and Timing of First Marriage versus First Cohabitation. *European Sociological Review*, *25*, 139–153. doi:10.1093/esr/jcn045

- Wolfinger, N. H. (2000). Beyond the Intergenerational Transmission of Divorce: Do People Replicate the Patterns of Marital Instability They Grew Up With? *Journal of Family Issues*, 21, 1061–1086. doi:10.1177/019251300021008006
- Wolfinger, N. H. (2003). Parental Divorce and Offspring Marriage: Early or Late? *Social Forces*, 82, 337–353. doi:10.1353/sof.2003.0108
- Wolfinger, N. H. (2005). *Understanding the Divorce Cycle: The Children of Divorce in Their Own Marriages*. New York: Cambridge University Press.
- Wu, L. L., & Li, J.-C. A. (2005). Historical Roots of Family Diversity. Marital and Childbearing Trajectories of American Women. In R. A. Settersten, F. F. Furstenberg, & R. G. Rumbaut (Eds.), *On the Frontier of Adulthood* (pp. 110–149). Chicago & London: The University of Chicago Press.

Table 1: OLS Regression predicting combined (Dynamic Hamming + own subcosts) dyadic distance

	M1 (random assignment)	M2 (conditional assignment)	M3 (conditional assignment)	
			<i>Main Effects</i>	<i>Interaction Effects</i>
Sibling indicator (ref.: Unrelated dyad)				
Sibling Dyad	-3.94*** (0.27)	-3.52*** (0.27)	-2.69*** (0.79)	
Gender Constellation (ref.: Opposite sex)				
Both female			2.74*** (0.41)	-2.39*** (0.70)
Both male			-6.91*** (0.39)	-1.19 (0.66)
Age Difference (ref.: 1 to 3 years)				
Born in same year			0.16 (0.53)	-7.20*** (1.70)
More than 3 years			0.89* (0.36)	0.23 (0.60)
Educational level (ref.: both low)				
Low-high			-4.74*** (0.65)	-0.48 (1.26)
Medium-low			-2.98*** (0.46)	-0.20 (0.79)
Both medium			-6.88*** (0.56)	0.13 (0.88)
Medium-high			-8.04*** (0.63)	-1.11 (1.02)
High-high			-9.12*** (0.93)	-2.94* (1.40)
Parental educational level (ref.: Low)				
Medium			-1.48*** (0.40)	0.68 (0.62)
High			-3.63*** (0.65)	1.93 (1.00)
Mother's age at marriage (ref.: Average age)				
Late marriage			-1.93*** (0.42)	0.19 (0.68)
Early marriage			2.23*** (0.40)	0.04 (0.64)
Experienced parental divorce				
			2.28*** (0.44)	0.07 (0.70)
Constant	44.06*** (0.17)	43.70*** (0.17)	48.53*** (0.48)	
Observations	14,242	14,257	14,257	
Adjusted R-squared	0.01	0.01	0.12	

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 2: Description of five family formation clusters

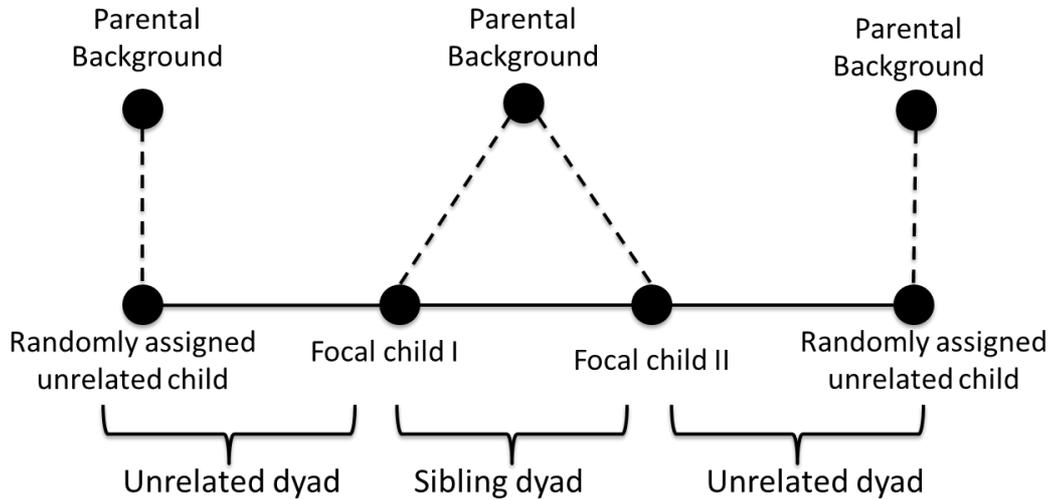
	Marriage, moderate fertility	Home stayer	Living alone, transitional cohabitation	Early marriage, high fertility	Extended cohabitation, parenthood out of wedlock	Total
Parental background						
<i>Parental Education</i>						
Low	0.56	0.59	0.59	0.72	0.75	0.62
Medium	0.32	0.31	0.30	0.25	0.22	0.29
High	0.12	0.09	0.11	0.03	0.03	0.09
<i>Mother's marriage age</i>						
Early marriage	0.27	0.23	0.29	0.39	0.36	0.28
Marriage at average age	0.51	0.49	0.48	0.46	0.47	0.49
Late marriage	0.23	0.28	0.23	0.15	0.17	0.23
<i>Experienced parental divorce</i>	0.16	0.16	0.23	0.19	0.25	0.19
Children's characteristics						
<i>Gender: Female</i>	0.54	0.34	0.49	0.70	0.69	0.50
<i>Educational level</i>						
Low	0.34	0.42	0.45	0.56	0.70	0.46
Medium	0.44	0.41	0.41	0.36	0.28	0.40
High	0.22	0.17	0.14	0.08	0.02	0.15
Children's demographic behavior						
<i>Median age at</i>						
Leaving parental home	22	25	21	20	20	22
First Cohabitation	23	26	22	20	21	23
First Marriage	26	-	-	23	-	-
First Birth	27	-	-	23	24	-
<i>Number of kids at age 30</i>						
Childless	0.26	0.85	0.86	0.02	0.06	0.56
Average number	1.15	0.19	0.16	2.46	1.72	0.78
(Standard Deviation)	(0.84)	(0.51)	(0.44)	(0.82)	(0.89)	(1.02)
Total	0.21	0.32	0.26	0.09	0.12	1.00
(Number of observations)	(1983)	(2939)	(2395)	(845)	(1104)	(9266)

Random Assignment

Parental education & marital history is

the same for sibling dyads

usually different for unrelated dyads



Conditional Assignment

Parental education & marital history is the same for sibling and unrelated dyads

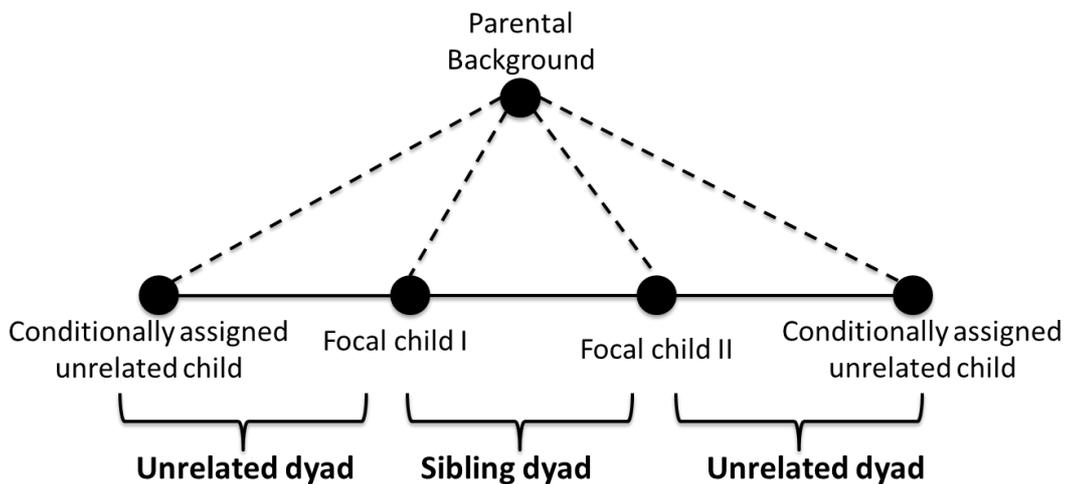
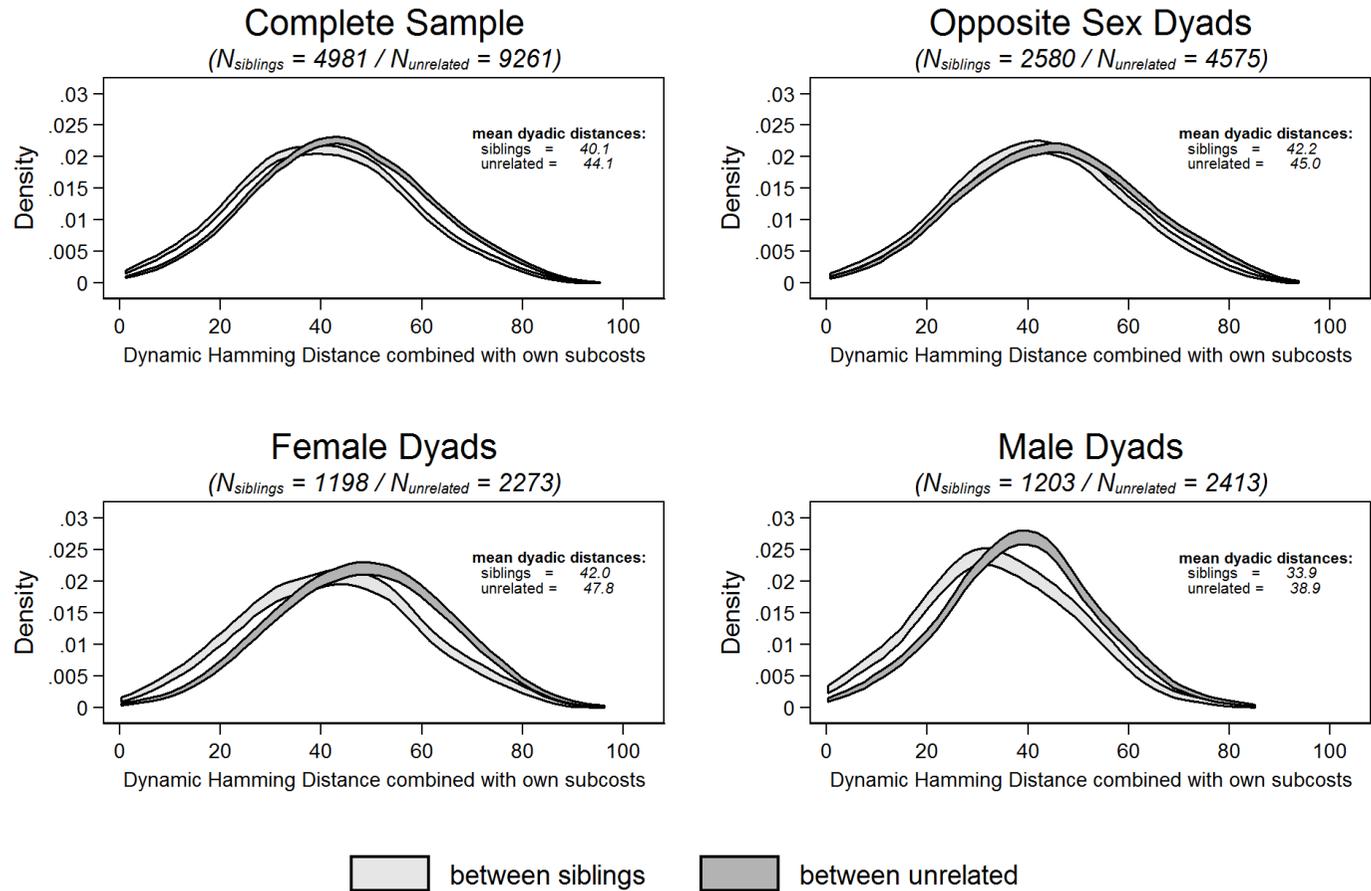


Figure 1: Random and conditional assignment of unrelated dyads



Note: Fixed-bandwidth kernel density with asymptotic confidence intervals. Case numbers marginally deviate from the conditionally assigned dyads in analysis sample two.

Figure 2: Distribution of sequence distances among sibling dyads and randomly assigned unrelated dyads (analysis sample one).

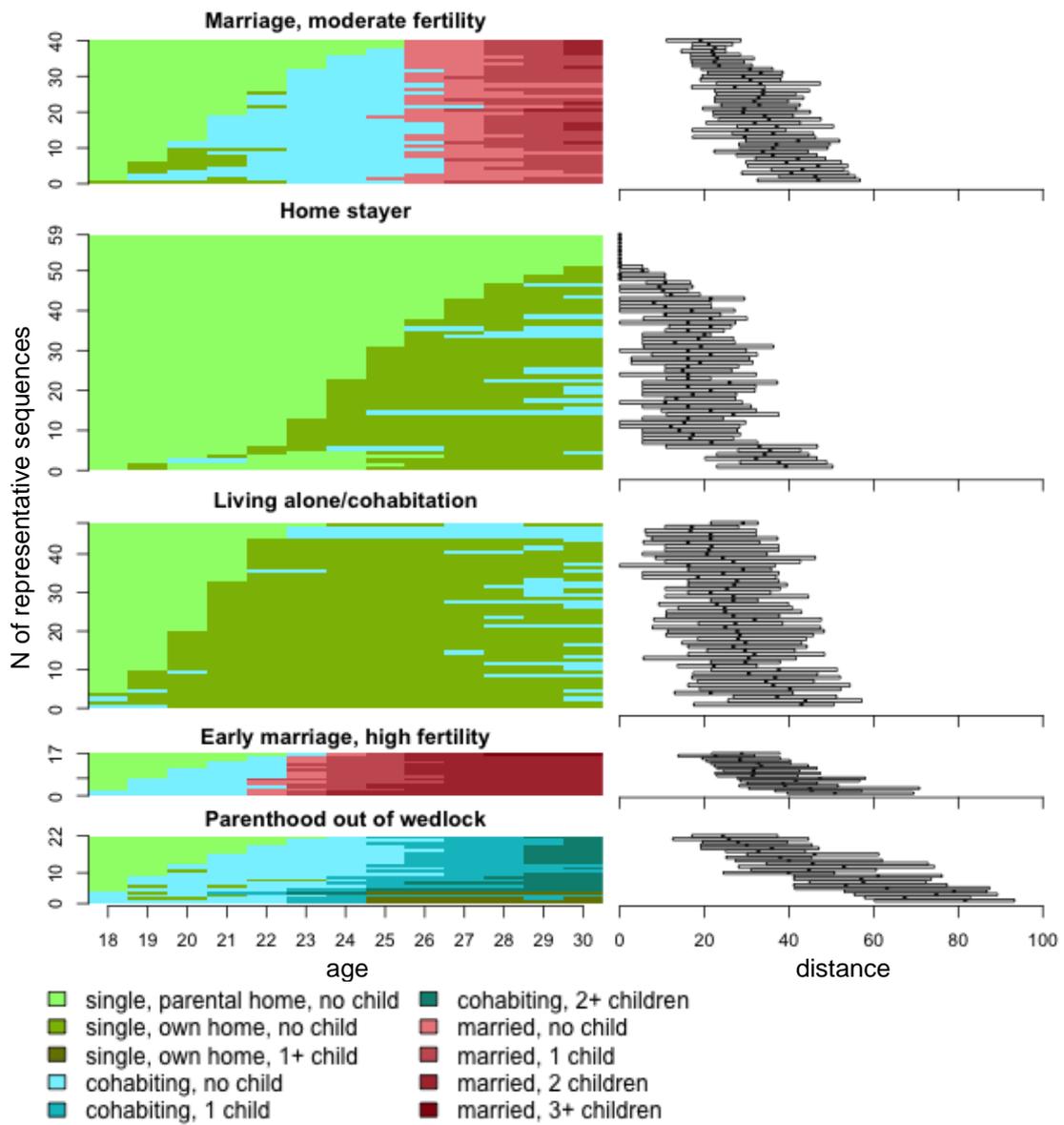


Figure 3: Relative Frequency Sequence Plot of the five family formation clusters with clusters sorted ascending according to highest educational level and sequences sorted according to age of first leaving the parental home (view in color). Size of clusters reflects proportion in population.

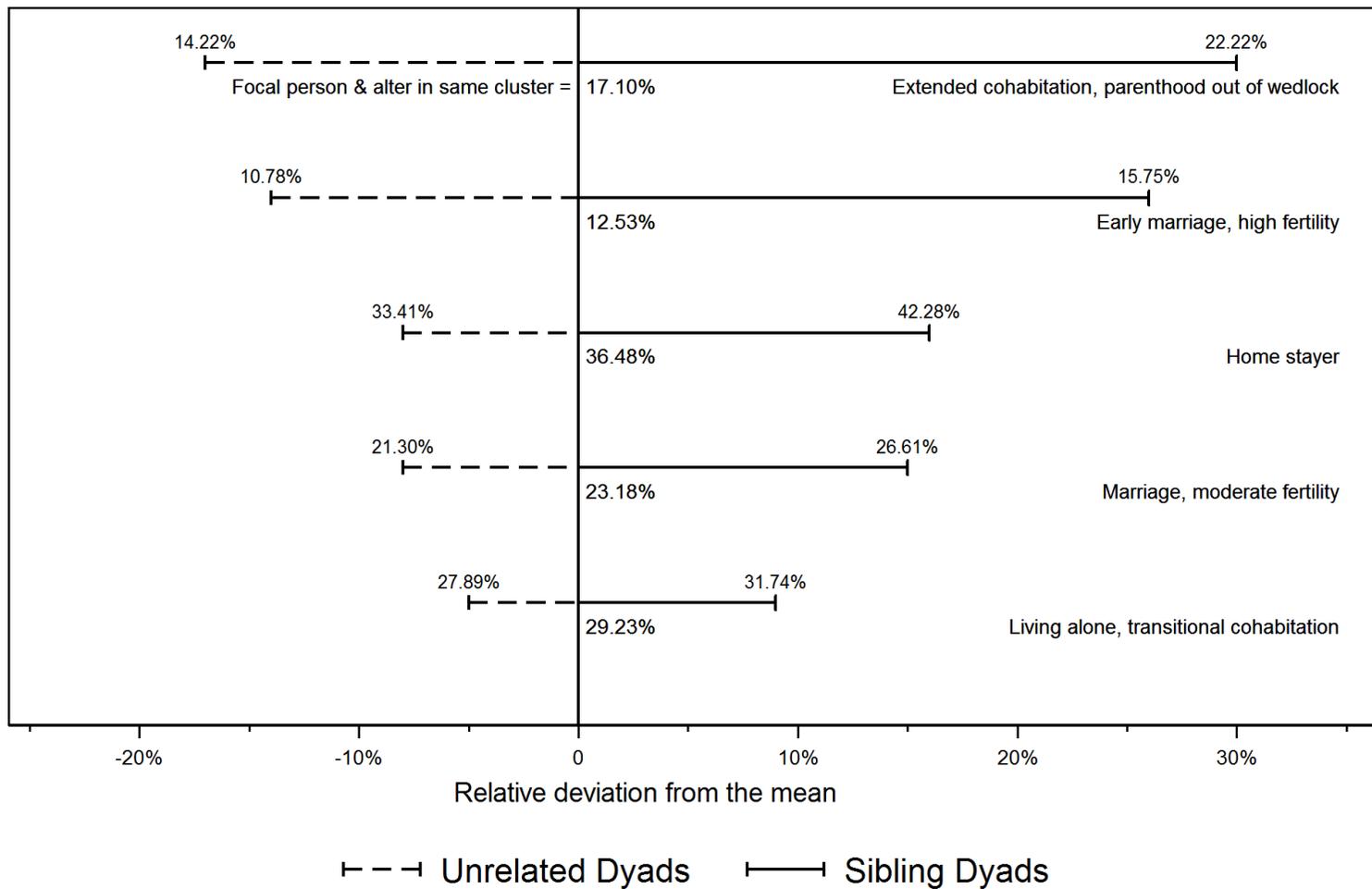


Figure 4: Focal person & alter in the same cluster: Conditional probabilities and relative deviations from overall mean

Appendix

Table A1: Substitution cost matrix

	SPH	SNC	CNC	MNC	SWC	C1C	M1C	C2C	M2C	M3C
SPH	0									
SNC	2	0								
CNC	4	2	0							
MNC	6	4	2	0						
SWC	8	6	5	5	0					
C1C	7	5	3	4	4	0				
M1C	8	6	4	2	5	2	0			
C2C	9	7	5	6	6	2	4	0		
M2C	10	8	6	4	7	4	2	2	0	
M3C	11	9	7	5	8	5	3	3	1	0

Note: SPH = “single, parental home, no child”, SNC = “single, own home no child”, S1C = “single, own home, 1+ child”, CNC = “cohabiting, no child”, C1C = “cohabiting, 1 child”, C2C = “cohabiting, 2+ children”, MNC = “married, no child”, M1C = “married, 1 child”, M2C = “married, 2 children”, M3C = “married, 3+ children”

Table A2: Descriptive Results

	Sibling dyads	Unrelated dyads	Total
Parental background			
<i>(fixed for sibling and unrelated dyads)</i>			
<i>Parental Education</i>			
Low			0.62
Medium			0.29
High			0.09
<i>Mother's marriage age</i>			
Early marriage			0.28
Marriage at average age			0.49
Late marriage			0.23
<i>Experienced parental divorce</i>			0.19
Children's characteristics			
<i>Gender constellation</i>			
Opposite sex	0.52	0.49	0.50
Both female	0.24	0.25	0.24
Both male	0.24	0.26	0.25
<i>Age difference</i>			
Born in same year	0.03	0.11	0.08
One to three years	0.62	0.54	0.57
More than three years	0.35	0.35	0.35
<i>Educational level</i>			
Low-high	0.06	0.10	0.08
Both Low	0.28	0.24	0.26
Low-medium	0.30	0.34	0.33
Both medium	0.18	0.16	0.17
Medium-high	0.12	0.12	0.12
Both high	0.06	0.04	0.04
<i>Dyadic Distance</i>	40.19	43.70	42.47
<i>(SD in brackets)</i>	(17.27)	(16.77)	(17.03)
Number of dyads	4994	9263	14257