

1 Association between sexes of successive 2 siblings in data from Demographic and 3 Health Survey program

4 Mikhail Monakhov (misha.monahov@gmail.com)

5

6 **Abstract**

7 Prediction of future child's sex is a question of keen public interest. Probability of having a child of
8 either sex is close to 50%, although multiple factors may slightly change this value. Some
9 demographic studies suggested that sex determination can be influenced by previous pregnancies,
10 although this hypothesis was not commonly accepted. This paper explores correlations between
11 siblings' sexes using data from Demographic and Health Survey program. In the sample of about
12 2,214,601 women (7,985,855 children), frequencies of offsprings with multiple siblings of same sex
13 were significantly higher than can be expected by chance. These deviations of frequencies were
14 driven by positive correlation between sexes of successive siblings (coefficient = 0.065, $p < 0.001$), i.e.
15 a child was more likely to be of same sex as preceding sibling. This correlation could be caused by
16 secondary sex ratio adjustment *in utero* since the effect was decreasing with length of birth-to-birth
17 interval, and birth-to-birth interval was longer for siblings with unlike sex. A variation of sex ratio
18 between families was present as well.

19 **Keywords:** sex ratio, sex determination, Lexian variation, Markov dependency, sex composition

20 Introduction

21 A family with multiple children of same sex, say, a family with ten boys and no girls, always draws
 22 attention, and speculations are made about possible reasons of such an incident. These may include a
 23 lack of either parent's ability to conceive girls, influence of witchcraft, astrology etc.

24 Surely, large unisex families may emerge simply by chance, when sexes of siblings are
 25 completely independent from each other. However, some studies indicate that sex of a child may
 26 depend statistically on sexes of other siblings (see references below). There are at least two possible
 27 sources of such dependence, that may increase frequency of unisex families. First, some parents
 28 indeed may have a predisposition to have children of particular sex. This bias can be caused by
 29 chemical compounds, infectious diseases, psychological distress (James 2000) and other conditions
 30 specific for given parents. As a result, each couple can be characterized by a probability of male birth
 31 p that may differ from proportion of boys in the population¹. This variation of p among couples is
 32 dubbed "Lexian".

33 Second, the sex of a child could be affected somehow by the sex of immediately preceding
 34 sibling, or sexes of several preceding siblings. This effect could be positive (birth of a child with
 35 either sex increases chances that next child will be of same sex) or negative. The variation of p within
 36 an offspring, dependent on previous birth, is called "Markovian". The unisex families would be in
 37 excess in a population with positive Markovian correlation.

38 Another source of variation in sex ratio is called "Poisson". It refers to random or systematic
 39 variation of p within a family, for example a decrease of p with mother's age. Markovian, Lexian and
 40 Poisson variations can be present simultaneously in same sample. Additionally, statistical dependence
 41 between siblings' sexes can be caused by parents' sex preferences, e.g. boy preference documented in
 42 many Asian cultures, or a preference for mixed-sex offspring.

¹ The probability of male birth should be distinguished from sex ratio, which is number of boys per 100 girls. But if the probability is calculated as a frequency of boys, then higher probability implies higher sex ratio and vice versa.

43 Since at least 1889² the problem was studied using demographic data. Geissler ((Geissler
44 1889), reviewed in (Gini 1951)) analyzed about 5 million births in 19th century Germany and did not
45 detect significant deviation from expected offspring frequencies. Nevertheless, he noted that in same-
46 sex offsprings probability of birth of one more child of same sex was higher than in other families (a
47 positive correlation). For mixed sex offsprings the correlation was negative (for instance, probability
48 of female birth increased after many male births).

49 Gini (Gini 1951) re-analyzed Geissler's data, together with additional datasets from Germany,
50 Italy and Netherlands, and concluded that some couples had a predisposition to have children of
51 particular sex, although this predisposition may reverse with time. For example, a family may start
52 procreation with a tendency to produce one sex, and end with a tendency to produce the opposite sex.
53 He noticed a positive association between sexes of successive siblings in a data from Italy, and
54 interpreted it as an evidence for such predispositions.

55 Turpin and Schutzenberger (Turpin and Schutzenberger 1949) (reviewed in (Gini 1951))
56 analyzed a sample from France (14,230 families) and observed positive correlation between sexes of
57 successive siblings. Unlike Gini, they hypothesized that the correlation could be explained by an
58 influence of one birth on the following birth. This idea was supported by the fact that the interval
59 between births of same sex siblings was on average shorter than for opposite sex siblings (however,
60 Gini interpreted this result in favor of his "reversal of predisposition" explanation).

61 (Bernstein 1952) analyzed a sample from USA (7,616 families). She reported positive
62 correlation between sexes of first two children in family, and an excess of unisexual three-child
63 families.

64 Malinvaud (Malinvaud 1955) (reviewed in (James 1975)) investigated nearly 4 million births
65 in France. The probability of male birth correlated positively with number of preceding boys and
66 negatively – with number of preceding girls.

² Other aspects of sex ratio variation were studied as early as in 1710, see James, W. H. (2000). "The variation of the probability of a son within and across couples." Human Reproduction **15**(5): 1184-1188.

67 Renkonen *et al* (Renkonen 1956, Renkonen, Makela et al. 1961) collected data from 31215
68 families in Finland. Their dataset was analyzed by several researchers, who reached contradictory
69 conclusions. Original authors confirmed associations reported earlier by Geissler (Renkonen, Makela
70 et al. 1962). First, probability of male birth was higher when all previous siblings were boys, and
71 lower when they were girls (positive correlation). Second, for mixed offsprings the probability of
72 male birth was decreasing when number of preceding boys increased (negative correlation). The
73 negative correlation was explained by possible immunization of mother's organism against male fetal
74 antigens. Edwards (Edwards 1961, Edwards 1962, Edwards 1966) re-analyzed the data using different
75 statistical methods and demonstrated that correlation between sexes of successive siblings was
76 positive (presumably due to some kind of *in utero* adjustment of sex ratio), although it was negative
77 when only two last siblings in a family were concerned (probably due to birth control). Beilharz
78 (Beilharz 1963) analyzed same data and dismissed previous conjectures about biological factors.
79 Instead, he concluded that there is no influence of sex of previous sibling, and apparent positive
80 correlation is caused by heritable tendency of some parents to have same-sex offsprings.

81 No correlation was found between successive siblings' sexes in 116,458 sibships from Utah,
82 sUSA (Greenberg and White 1967). Also, interval between births in this sample did not depend on
83 whether the siblings were of same or opposite sex.

84 In the data from 1970 USA census, with more than 230,000 families, sex ratio increased with
85 number of preceding boys, and decreased with number of preceding girls. The authors suggested that
86 their data provide evidence for both Markovian and Lexian variation (Ben-Porath and Welch 1976).

87 In another USA sample (649,366 births) the sex of last child in a family was negatively
88 correlated with sexes of preceding siblings (probability of male birth decreased with number of
89 preceding brothers) (Thomas Gualtieri, Hicks et al. 1984).

90 Mitter *et al* reported an excess of unisexual offsprings in small sample (451 families) from
91 India (Mitter and Anand 1975).

92 (Maconochie and Roman 1997) reported that sex ratio was not associated with sexes of
93 preceding siblings in 330,088 offsprings from Scotland. (Jacobsen, Moller et al. 1999) used a sample
94 of 815,891 children from Denmark. According to their analysis, neither sex of immediately preceding
95 sibling nor sexes of two or three preceding siblings had significant effect on sex ratio, when the
96 regression model was adjusted for paternal age. (Rodgers and Doughty 2001) analyzed 6,089 families
97 from USA and concluded that sexes of previous siblings had no effect on sex ratio.

98 Garenne (Garenne 2009) analyzed over 2 million births in Sub-Saharan Africa. There was no
99 correlation between sexes of successive siblings (although, only the correlation between two last
100 births was tested). The sex ratio was dependent on number of preceding boys (positively) and number
101 of preceding girls (negatively), as it was the case in Malinvaud's sample.

102 Overall the results of published studies are mixed. Many, but not all of them, report a
103 correlation between sexes of siblings in a family. When the correlation is found, researchers disagree
104 about it's possible sources, since it can be attributed to both positive Markovian and to Lexian
105 variation. The situation is complicated by a variety of statistical methods employed. Some authors
106 tested pair-wise correlation between sexes of two successive siblings. Others used a regression of
107 probability of male birth on numbers of preceding boys and girls. Methods that enable simultaneous
108 testing of Markovian and Lexian variation were rarely employed. Few studies explicitly tested for an
109 excess of families with particular composition, such as unisex families. Sample size is also an issue,
110 with many studies conducted on only few thousands of families (while others used hundreds of
111 thousands).

112 Recently large scale demographic data, including birth sequences, became available for
113 analysis from Demographic and Health Survey (DHS) program. The DHS is international
114 organization that assists in conducting demographic surveys, mostly in developing countries. I used
115 DHS data from about 7 million births (to my knowledge, largest sample so far) to test (1) if there are
116 significant deviations from expected frequencies of offsprings with various sex compositions, (2) if
117 there is a correlation between sexes of successive siblings, and (3) if this correlation can be attributed
118 to Lexian or Markovian variation.

119 **Methods**

120 The surveys were carried out in 1985-2014 mostly in developing countries of Africa, Asia and South
 121 America. Datasets were downloaded from DHS website (dhsprogram.com) in August-September
 122 2015. Surveys of following types were selected for analysis: Standard DHS, Continuous DHS, Interim
 123 DHS, Special DHS. For each survey, the “individual response” dataset was downloaded, in Stata
 124 format (.dta). The sequences of siblings sexes were taken from variables b4_01-b4_20. Other
 125 variables used in analysis were b11_01-b11_20 (intervals between births), b3_01-b3_20 (child’s date
 126 of birth), v011 (mother’s date of birth), b0_01-b0_20 (indicator of plural birth), v201 (total number of
 127 children ever born).

128 Data handling and statistical analysis were done with Stata/MP 13. Logistic regression used
 129 sex of each child (except of first child in an offspring, who doesn’t have any preceding sibling) as
 130 dependent variable, and sex of previous child – as independent variable. Pairs of successive siblings
 131 where at least one of them was from plural birth were excluded. Offsprings with reported total number
 132 of children ever born (variable v201) did not match number of children recorded in variables b4_01-
 133 b4_20 were excluded as well.

134 Maximum likelihood estimation was performed using the model from (Astolfi and Tentoni
 135 1995). Only offsprings with four or less children were analyzed. Briefly, Markovian dependency was
 136 modelled with parameters k_{im} , k_{if} (birth order $i = 2, 3, 4$). The probability of i ’th child being boy
 137 equals $k_{im}p_1$ when $i-1$ ’th child was boy and $k_{if}p_1$ when $i-1$ ’th child was girl. The probability of i ’th
 138 child being girl equals $1 - k_{im}p_1$ when $i-1$ ’th child was boy and $1 - k_{if}p_1$ when $i-1$ ’th child was girl.
 139 Other parameters reflected the shape of probability distribution of p_1 and stopping rules. Markovian
 140 variation was excluded by constraining $k_{im} = k_{if} = k_i$ ($i = 2, 3, 4$); Poisson – by constraining $k_{im} = k_m$, k_{if}
 141 $= k_f$ ($i = 2, 3, 4$); Lexian – by constraining 2nd, 3rd and 4th central moments of probability distribution
 142 to zero. Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm was used for likelihood maximization.
 143 The code for maximum likelihood estimation was written in Stata, and tested on Astolfi’s data to
 144 ensure that it generated same results as in the original publication.

145 **Results**

146 **Large same-sex offsprings were more numerous than expected**

147 Total sample included 7,985,855 children in 2,214,601 offsprings (Table 8). An offspring can be
 148 characterized, regardless of birth order, by total number of children n and number of boys k . If sex of
 149 each child is determined independently of siblings, then expected frequency of the offspring can be
 150 calculated according to binomial law: $E_{n,k} = \binom{n}{k} r_n^k (1 - r_n)^{n-k}$, where r_n is proportion of boys in all
 151 offsprings with size n . Chi-squared test showed significant difference between expected and observed
 152 frequencies (Chi2 = 994, df = 178, $p < 0.001$). The deviations remained significant when offsprings
 153 with at least one plural birth were excluded from the analysis (Chi2 = 1071, df = 176, $p < 0.001$), or
 154 when only one proportion of boys r for whole sample (instead of specific proportion for each
 155 offspring size) was used to calculate expected frequencies (Chi2 = 1700, df = 177, $p < 0.001$).

156 Next, chi-squared test was applied to test deviation from expected frequency for each type of
 157 offspring (as characterized by n and k) (Table 1).

158 **Table 1 Odds ratios demonstrating relative abundance of offsprings with various sizes n and numbers of boys k .**
 159 **Odds ratio above 1 indicates an excess of offsprings with given composition, relative to expected frequency. *p-value**
 160 **< 0.05. For 31 out of 63 chi-square tests performed the p-value was below 0.05. Offsprings with $n > 10$ are not shown.**

k	n									
	2	3	4	5	6	7	8	9	10	
0	0.94*	0.95*	0.94*	0.97	0.98	1.08	1.25*	1.27	1.39	
1	1.09*	1.01*	0.99	1.01	1.02	1.05*	1.08*	1.18*	1.14	
2	0.95*	1.05*	1.04*	1.01	1.01	1.01	1.02	1.05	1.09*	
3		0.93*	0.98*	0.98*	0.99	0.97*	0.98	0.99	1.01	
4			0.97*	1.01	0.98*	0.99	0.98	0.97*	0.99	
5				1.02	1.03*	0.99	0.97*	0.98	0.98	
6					1.07*	1.05*	1.02	0.99	0.97	
7						1.16*	1.06*	1.05*	0.99	
8							1.31*	1.15*	1.06	
9								1.08	1.31*	
10									1.63*	

161
 162 The deviations were significant for many types of offspring. Notably, offsprings with many
 163 boys were in excess among large families. For instance, among 30,393 offsprings with exactly ten
 164 children 52 were all-boys, while expected number was only 32 (OR = 1.63, $p = 0.037$). Similarly,

165 among large offsprings number of those composed mostly with girls tended to be higher than
 166 expected. At the same time among small offsprings (<5 children) mixed-sex types were more
 167 common. For example, among offsprings with only 2 children those with one boy and one girl were
 168 overrepresented (OR = 1.09, $p < 0.001$).

169

170 Sex ratio depended on the sex of preceding sibling

171 The dependence between sexes of successive siblings was tested using logistic regression (boy coded
 172 as “0” and girl as “1”). The sex of preceding sibling had small but significant positive effect on sex of
 173 a child (coef = 0.016, $p < 0.001$, $\sim 0.4\%$ ³ change of sex ratio). Assuming that this effect could depend
 174 on offspring size and birth order of a child, the regression was repeated for different offspring sizes n
 175 and birth orders i (Table 2).

176 **Table 2 Correlations between sexes of successive siblings, by offspring size n and birth order i . Regression coefficients**
 177 **from logit model are presented (sex of i 'th child as dependent variable, sex of $i-1$ 'th child – independent variable).**
 178 **Offsprings with more than ten children are not shown. *p-value < 0.05.**

i	n								
	2	3	4	5	6	7	8	9	10
2	-0.168*	0.013	0.063*	0.089*	0.075*	0.063*	0.104*	0.081*	0.046
3		-0.11*	-0.002	0.069*	0.08*	0.085*	0.078*	0.094*	0.171*
4			-0.072*	0.011	0.08*	0.077*	0.092*	0.1*	0.116*
5				-0.028*	-0.001	0.093*	0.085*	0.105*	0.093*
6					-0.037*	0.028*	0.075*	0.078*	0.073*
7						-0.016	0.068*	0.072*	0.066*
8							0.005	0.068*	0.035
9								-0.003	0.041
10									0.063*

179

180 When dependent variable was the sex of last child in an offspring (that is, $i = n$), the
 181 regression coefficients were mostly negative. Meanwhile, for most other combinations of n and i the
 182 coefficients were positive and significant. When the analysis (for all offspring sizes n combined) was
 183 performed excluding last child of each offspring (i.e. when $i \neq n$, for all n), the regression coefficient

³ As a rule of thumb in these models, logistic regression coefficient of 0.1 corresponds to 2.5% increase of probability that the i 'th child has same sex as preceding sibling.

184 was 0.067 ($p < 0.001$). Conversely, when only last children were included ($i = n$, for all n), the
185 regression coefficient was -0.083 ($p < 0.001$).

186 For confirmation of the results obtained in pooled sample the regression was performed in
187 each dataset from DHS website (one dataset per survey), and then random effects meta-analysis was
188 applied to pool estimates of regression coefficients. When all children were included (regardless of
189 birth order) the result was not significant (coef = 0.002, $p = 0.67$, $I^2 = 48.4\%$, $p < 0.001$). When last
190 child of each offspring was excluded, the effect of preceding sibling's sex was positive and significant
191 (coef = 0.065, $p < 0.001$, $I^2 = 68.6\%$, $p < 0.001$). When only last children were included the effect of
192 previous sibling's sex was negative and significant (coef = -0.121, $p < 0.001$, $I^2 = 86.2\%$, $p < 0.001$).

193 Why does the sign of this correlation depend on whether "dependent" child is last or not last
194 in a family? Is it possible that sex of preceding sibling affects sex of a child via two different
195 mechanisms, depending on whether the next sibling is a last child in the family or not? The negative
196 correlation for last child may be explained by deliberate family planning, namely parents' willingness
197 to have a family with mixed sex composition ("balance preference", discussed by (Gini 1951) and
198 others). Such preference may dictate specific stopping behavior. For instance, if a couple have several
199 daughters they may continue procreation until having a son, and then stop. A couple having a boy and
200 a girl may decide to stop the procreation, while a family with two girls may wish to have one more
201 child in a hope that it will be a boy.

202 However, the last child recorded in DHS data is not necessary the last child in the family. The
203 respondent woman may continue the procreation after interview date (i.e. the family could be
204 incomplete). The proportion of children who are last in complete families should increase with the
205 interval between date of last birth and date of interview. If the negative correlation between sexes of
206 two last siblings is caused by parents' stopping behavior, the correlation should be stronger when
207 proportion of complete families is higher. Indeed, in the logit model including the time since last birth
208 as a covariate, the effect of interaction term was negative and significant (Table 3, column 3). The
209 negative sign suggests that coefficient's absolute value is increasing when the time since last birth
210 increases. For instance, when last birth was less than 6 months before interview (bottom 10% of

211 distribution of interval values) the effect of previous sibling's sex was not significant (coef = -0.017, p
 212 = 0.096). When last birth was more than 167 months before interview (top 10%) the negative
 213 correlation was highly significant (coef = -0.24, $p < 0.001$), presumably because the proportion of
 214 complete families in this subsample was high.

215 **Table 3 Effects of preceding sibling's sex and time since last birth on sex of last sibling in an offspring. Logistic**
 216 **regression coefficients. p-values in italic.**

Independent variable	<i>1</i>	<i>2</i>	<i>3</i>
Sex of preceding sibling	-.0979 <i>< 0.001</i>	-.0994 <i>< 0.001</i>	-0.02 <i>< 0.001</i>
Time since last birth, months		-.0006 <i>< 0.001</i>	-0.00008 <i>< 0.001</i>
Interaction			-0.0011 <i>< 0.001</i>

217

218 Alternatively to deliberate family planning, some biological factors behind the correlation
 219 between sexes may change their direction from positive to negative, and such a change could be
 220 accompanied by the end of procreation.

221

222 **Was the correlation caused by Lexian or Markovian variation?**

223 If some parents have a tendency to produce children of particular sex, then a correlation should be
 224 observed between sexes of siblings regardless their birth order, not only between successive siblings
 225 (Edwards 1961). If the association between successive siblings has major role, then little or no
 226 correlation should be observed between sexes of siblings separated by other births. In DHS data the
 227 correlation was strong between successive siblings, less strong but still significant— between siblings
 228 separated by one birth, and negligible when siblings were separated by two or more births (Table 4,
 229 column 2).

230

231 **Table 4 Correlations between sexes of siblings separated by various numbers of births. Logistic regression**
 232 **coefficients. Offsprings with more than 10 children are omitted.**

Number of births separating the two siblings	1. All pairs of siblings			2. Excluding last children			3. Only last children		
	<i>Coef</i>	<i>P</i>	<i>N</i>	<i>coef</i>	<i>P</i>	<i>N</i>	<i>coef</i>	<i>p</i>	<i>N</i>
0	0.016	<0.001	5540403	0.067	<0.001	3819694	-0.098	<0.001	1720709
1	-0.011	<0.001	3766164	0.021	<0.001	2519485	-0.077	<0.001	1246679
2	-0.019	<0.001	2524926	0.003	0.32	1648152	-0.059	<0.001	876774
3	-0.015	<0.001	1653104	0.004	0.27	1046960	-0.049	<0.001	606144
4	-0.013	<0.001	1051994	-0.008	0.13	638503	-0.022	<0.001	413491
5	-0.014	0.004	643119	-0.006	0.37	370079	-0.026	<0.001	273040
6	-0.003	0.70	373788	0.004	0.62	201079	-0.010	0.28	172709
7	0.001	0.92	203656	-0.006	0.63	101590	0.008	0.54	102066
8	-0.001	0.96	103270	0.004	0.83	46867	-0.004	0.80	56403
9	0.004	0.82	47840	-0.015	0.59	20087	0.018	0.45	27753
10	-0.018	0.52	20606	-0.037	0.42	7770	-0.007	0.85	12836

233

234 When only last children were concerned (i.e., last sibling's sex was dependent variable) the
 235 negative coefficient remained highly significant even when siblings were separated by up to five
 236 births (Table 4, column 3). This negative correlation with sexes of several preceding siblings is not
 237 surprising if the last child's sex is influenced by parents' balance preference (parents take into
 238 consideration the sexes of several existing children).

239 When all children, regardless of birth order, were included in the model, the coefficient was
 240 positive and significant for successive births, but it was negative and significant when siblings were
 241 separated by 1-4 births (Table 4, column 1). Apparently, this result is caused by combination of two
 242 effects: positive correlations in most successive pairs, and negative correlation between sex of last
 243 child and sexes of several siblings preceding the last one.

244 More formal way to test the presence of Markovian dependency is to use statistical model
 245 where impacts of Markovian, Lexian and Poisson variations are represented simultaneously by
 246 different sets of parameters (see Methods). Likelihood ratios can be used to test significance of

247 particular parameters for model fit. Four models were fit to DHS data: a model with all parameters
 248 being estimated (full model), with Markovian variation excluded (Model 1), with Poisson variation
 249 excluded (Model 2) and with both Markovian and Lexian variation excluded (Model 3). Only
 250 offsprings with four or less siblings were considered (71% of all offsprings). The influences of all
 251 three sources of variation were significant: Markovian (LR = 39, df = 3, $p < 0.001$), Poisson (LR = 39,
 252 df = 4, $p < 0.001$) and Lexian (LR = 1182, df = 3, $p < 0.001$). If Lexian variation is mediated by beta-
 253 distribution of probability of male birth, then evaluated distribution parameters are $a = 64.1$ and $b =$
 254 60.3 (variance equals 0.00196).

255

256 **Effects of birth-to-birth interval on correlation between sexes**

257 If previous sibling's sex exerts it's effect by leaving some kind of temporary chemical mark in
 258 mother's organism, then this effect could be stronger for shorter intervals between births (Edwards
 259 1961). The interaction between birth-to-birth (BTB) interval and sex of previous sibling was tested
 260 using logistic regression (Table 5).

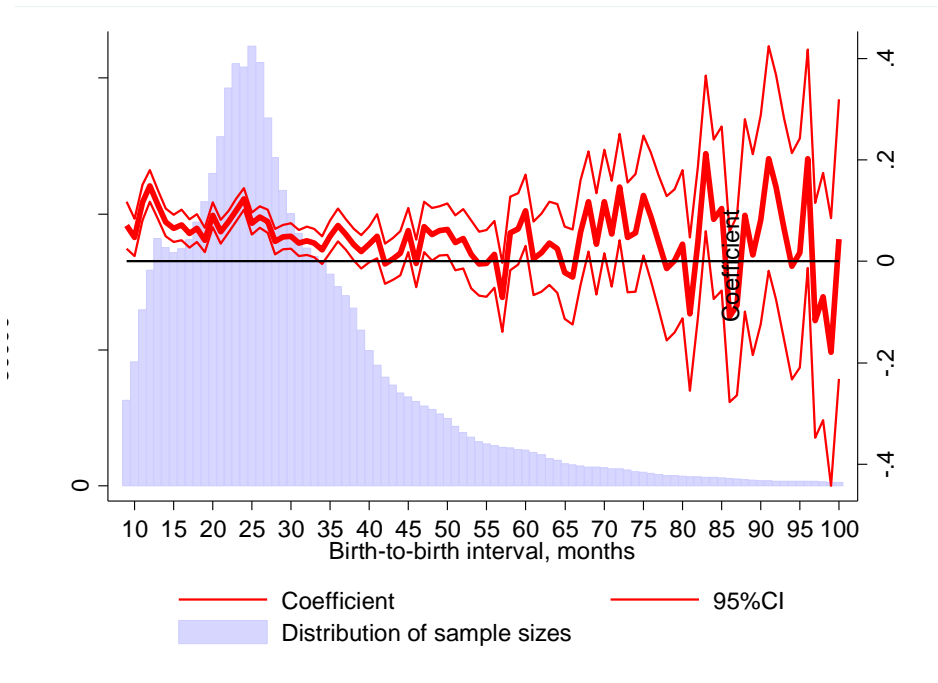
261 **Table 5 Interaction of preceding sibling's sex and birth-to-birth (BTB) interval. Logistic regression coefficients are**
 262 **shown (p-values in italic).**

Independent variable	Excluding last children			Only last children		
	1	2	3	4	5	6
Sex of preceding sibling	0.062 <i><0.001</i>	0.063 <i><0.001</i>	0.087 <i><0.001</i>	-0.065 <i><0.001</i>	-0.082 <i><0.001</i>	-0.084 <i><0.001</i>
BTB interval, months		.00051 <i><0.001</i>	0.00089 <i><0.001</i>		.00048 <i><0.001</i>	.000489 <i><0.001</i>
Interaction term			-.00077 <i><0.001</i>			.0000325 <i>0.8</i>

263

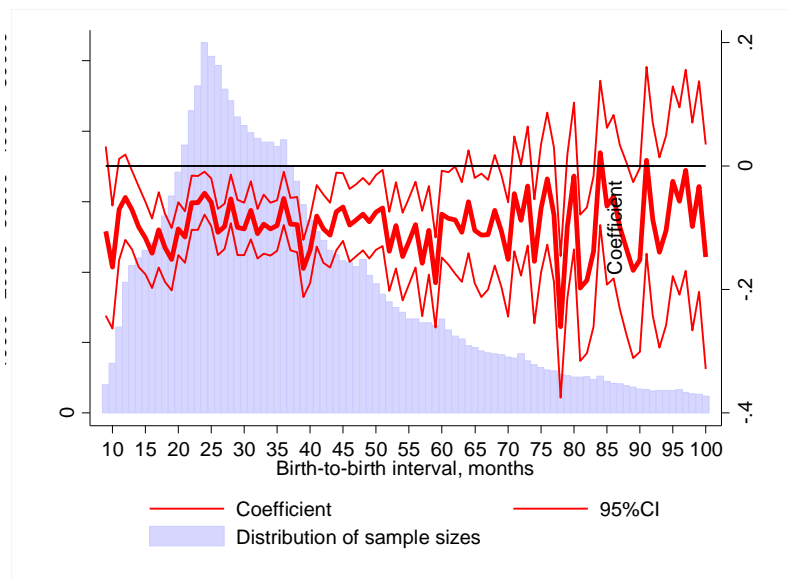
264 The effect of BTB interval alone was significant, regardless of whether last children were
 265 excluded or not: every additional month since last birth increased chances to have a girl by
 266 approximately 0.01%. The interaction term was significant only in the model excluding last children

267 (coef = -0.00077, $p < 0.001$). The negative sign of the coefficient suggests that the effect of preceding
 268 sibling's sex is vanishing with time. This interaction can be seen on Figure 1 (the value of regression
 269 coefficient is decreasing from 9 to 40-45 months of BTB).



270
 271 **Figure 1 Logistic regression coefficients (effect of preceding sibling's sex) by BTB intervals, with last child of each**
 272 **offspring excluded from analysis. Distribution of sample sizes over BTB values is shown in the background.**

273 When the analysis was limited to last children, the interaction was not significant (coef = -
 274 0.00003, $p = 0.8$), Table 3, Figure 2.



275

276 **Figure 2 The effect of preceding sibling's sex doesn't change substantially with BTB interval, when only last children**
 277 **are considered. Logistic regression coefficients (effect of preceding sibling's sex) by BTB intervals. Distribution of**
 278 **sample sizes over interval values is shown in the background.**

279

280 Edwards (Edwards 1961) suggested also that the correlation between successive siblings'
 281 sexes could be mediated by secondary adjustment of sex ratio, namely by lower chances of fetus
 282 survival when it's sex is opposite to preceding sibling. In this case the BTB interval could be longer
 283 when successive siblings were of unlike sex.

284 BTB interval between same sex births was significantly shorter than between opposite sex
 285 births when excluding last children (linear regression, pair of same sex births coded as "1", opposite
 286 sex – "0", coef = -0.135, $p < 0.001$). Results of nonparametric rank sum test were significant as well
 287 ($p < 0.001$).

288 (Greenberg and White 1967) noted that such analysis of intervals should control for birth
 289 order or family size (because if their data the interval decreased with family size when birth order was
 290 constant). When offspring size and birth order were included into regression model as covariates, the
 291 interval still was significantly longer for opposite sex pairs (coef = -0.085, $p < 0.001$). Additionally,
 292 the analysis was repeated for each combination of offspring size n and birth order i (Table 6).

293 **Table 6 Differences between BTB intervals for siblings of same and opposite sex. Linear regression coefficients are**
 294 **shown (the value corresponds to difference of BTB between pairs of opposite sex siblings and pairs of same sex**
 295 **siblings). * $p < 0.05$. Offsprings with more than 10 children are not shown.**

<i>i</i>	<i>N</i>							
	3	4	5	6	7	8	9	10
2	0.337*	0.03	-0.142	-0.008	-0.263*	-0.124	-0.167	-0.047
3		0.257*	-0.055	-0.127	-0.309*	-0.331*	-0.071	-0.284*
4			0.019	-0.072	-0.11	-0.339*	-0.137	-0.108
5				-0.108	-0.268*	-0.333*	-0.164	-0.24
6					-0.308*	-0.015	-0.278*	-0.367*
7						-0.167	-0.405*	-0.115
8							-0.176	-0.273
9								-0.372*

296

297 The difference of BTB intervals was significant for multiple, but not for all, combinations of

298 *n* and *i*. Notably, all coefficients were negative for offsprings with six or more children.

299

300 **Additional datasets**

301 Detailed information about siblings sexes and birth order was presented in several published papers

302 ((Renkonen, Makela et al. 1961, Greenberg and White 1967, Maconochie and Roman 1997, Jacobsen,

303 Moller et al. 1999, Rodgers and Doughty 2001)). Logistic regression was applied to these datasets to

304 investigate the reproducibility of results obtained in DHS data (Table 7).

305

306 Table 7 Logistic regressions of child's sex on preceding siblings' sex in datasets from published papers.

Publication		All observations	Excluding last child	Limited to last child
(Renkonen, Makela et al. 1961) (as presented in (Edwards 1962)) ¹				
<i>All data</i>				
	<i>coef</i>	.053	.104	-0.017
	<i>p</i>	< 0.001	< 0.001	0.32
	<i>N</i>	123445	71577	51868
< 5 siblings				
	<i>coef</i>	.036	.158	-.049
	<i>p</i>	0.021	< 0.001	0.018
	<i>N</i>	62797	26091	36706
(Greenberg and White 1967) ²				
<i>All data</i>				
	<i>coef</i>	-.008	.005	-0.044
	<i>p</i>	0.195	0.46	< 0.001
	<i>N</i>	438788	322399	116389
< 7 siblings				
	<i>coef</i>	-.024	.004	-0.073
	<i>p</i>	0.005	0.71	< 0.001
	<i>N</i>	220154	140204	79950
(Maconochie and Roman 1997)				
	<i>coef</i>	-.0009	.0051	-.002
	<i>p</i>	0.7	0.3	0.3
	<i>N</i>	217960	46119	171841
(Jacobsen, Moller et al. 1999) ³				
	<i>coef</i>		0.078	
	<i>p</i>		< 0.001	
	<i>N</i>		51016	
(Rodgers and Doughty 2001) ⁴				
<i>All data</i>				
	<i>coef</i>	-.064	.076	-0.14
	<i>p</i>	0.196	0.372	0.023
	<i>N</i>	6435	2227	4208
< 4 siblings				
	<i>coef</i>	-.153	-.036	-.194
	<i>p</i>	0.007	0.74	0.003
	<i>N</i>	5046	1301	3745

307 ¹ in original report families with five or more children were pooled together. Thus, the analysis limited to last children was
308 informative only for offsprings with four siblings or less. ² in original report families with seven or more children were
309 pooled. Thus, the analysis limited to last children was informative only for offsprings with six siblings or less. ³ the data were
310 taken from Table V of original report, that lists birth sequences of first three siblings for families with at least four children
311 (sex of fourth child was not shown in original paper). Thus, only the analysis excluding last children (and only in offsprings
312 with three or more siblings) was possible. ⁴ in original report families with four or more children were pooled. Thus, the
313 analysis limited to last children was informative only for offsprings with three siblings or less.

314

315 In data from Renkonen (Renkonen, Makela et al. 1961) the effect of previous sibling's sex
316 was positive when last children were excluded (coef = 0.158, $p < 0.001$) and negative when only last
317 children were included (coef = -0.049, $p = 0.018$), as in the DHS data. In Greenberg's data (Greenberg
318 and White 1967) positive correlation, after exclusion of last children, was not significant, while there
319 was strong negative correlation between sexes of last two siblings in a family. Similar situation was
320 observed in Rodgers data (Rodgers and Doughty 2001) (excluding last children – not significant
321 effect; including only last children, coef = -0.19, $p = 0.003$). In Maconochie's data (Maconochie and
322 Roman 1997) both kinds of correlation were not significant. In Jacobsen's data (Jacobsen, Moller et al.
323 1999) the correlation was positive and significant in the model excluding last children (coef = 0.078, p
324 < 0.001).

325 Thus, positive correlation between sexes of successive siblings (excluding last children) was
326 found in two out of five additional datasets. It should be noted, however, that Jacobsen and Rodgers
327 do not report statistics of birth sequences after excluding plural births. Thus the positive values of
328 coefficients could be overestimated. One should note also that unlike other datasets, Renkonen data
329 don't include children from a previous marriage of each woman.

330 **Discussion**

331 Current results demonstrate an excess of large offsprings with high (many boys) and with low (many
332 girls) sex ratios, relative to binomial expectation. The overrepresentation of such offsprings was
333 accompanied by correlation between sexes of siblings. Notably, positive correlation was observed
334 between sexes of successive siblings and siblings separated by one birth, but not for those separated
335 by two or more births. This effect is consistent with hypothesis of Markovian variation, but not with
336 Lexian one. However, formal modelling indicated that both Markovian and Lexian variation were
337 present in the sample.

338 Makela (Makela 1963) noted that in the presence of Lexian variation, lack of correlation for
339 siblings separated by several births could be observed due to other factors, for example an
340 immunization against male fetal antigens. But in this case it is not clear why such factors do not make
341 correlation between successive siblings negative as well.

342 Additional evidence for Markovian variation comes from analysis of birth-to-birth intervals.
343 Namely, the correlation between sexes was stronger when births were separated by shorter intervals.

344 Several published studies reported an absence of Markovian variation, but negative report
345 doesn't necessarily mean that the correlation is absent. Perhaps it is absent in a subsample selected for
346 testing, or can't be detected using specific methods employed. For example, Garenne (Garenne 2009)
347 performed a test of association between sexes of successive siblings and found it insignificant.
348 However, this test was conducted not on all children of the sample. Instead, from each offspring only
349 two siblings were selected: the last one and second to last in birth sequence. Moreover, the analysis
350 was restricted to women at least 35 years old.

351 Jacobsen (Jacobsen, Moller et al. 1999) reported that correlation between sexes was not
352 significant, but no distinction was made between last birth in an offspring and all other births. When
353 logistic regression was applied only to non-last births, significant positive association was seen
354 between successive siblings (Table 7).

355 Regarding samples used in (Greenberg and White 1967) and (Maconochie and Roman 1997)
356 samples, there seems to be no significant Markovian correlation even when non-last births were tested
357 separately. The lack of correlation in these (and possibly other) samples might be explained by
358 between-population heterogeneity. Both samples were collected in developed countries (USA and
359 Scotland, respectively), while most of DHS data come from developing countries. Notably, Greenberg
360 and White confirmed that the correlation was present in Renkonen's data.

361 It is unknown what biological mechanisms could underlie positive Markovian variation. Gini
362 (Gini 1951) stated that “[the] hypothesis may be considered excluded by the mechanism of sex
363 determination in a male heterogametic species like the humans”. Other researchers also noted that
364 there is no plausible mechanism for influence of one siblings' sex on the next one (Ben-Porath and
365 Welch 1976, James 2000).

366 However, the casual influence is not biologically impossible. Potential mechanism could be
367 based, for example, on fetal microchimerism. During pregnancy fetal cells penetrate into mother's
368 organism and remain there for years. These cells migrate into heart, skin, thyroid and adrenal glands,
369 liver, kidney, lung, and spleen (Bayes-Genis, Roura et al.). Thus, they may migrate into reproductive
370 system as well. They can differentiate into myocardial cells (Bayes-Genis, Roura et al.) and probably
371 into other cell types (Khosrotehrani and Bianchi 2005). In principle, such cells, remaining in the
372 reproductive system after previous pregnancy, may induce an adjustment of sex ratio by influencing
373 viscosity of cervical mucus or embryo mortality. Alternatively, fetal cells may alter activity of adrenal
374 glands (and levels of cortisol), pituitary (and levels of ACTH) and other glands related to sex
375 determination.

376 A study on Mongolian gerbils demonstrated that females, developed between male embryos
377 while *in utero*, had higher sex ratio in their offsprings later on (Clark and Galef 1995). This effect in
378 gerbils could be related to an acquisition of cells by female fetus from adjacent male fetuses. Perhaps
379 this effect is based on same mechanism as the Markovian correlation in humans.

380 One study reported that concentration of fat in human milk depends on child's sex (Fujita,
381 Roth et al. 2012). The authors noted also “near-significant negative influence of the number of sons

382 living at home on milk fat concentrations”. Biological mechanism of these effects is unknown, but it
383 is likely that sex of the child causes persistent alterations in mother’s physiology, that can be related
384 to both milk production and adjustment of sex ratio.

385 **Future directions.** Among many published studies of sex ratios only few present raw data.
386 Most papers contain only a summary of data, that isn’t sufficient for independent analysis. Presenting
387 full data is essential for replication. In case of sex sequences the data can be made quite compact: it is
388 enough to indicate birth sequence (e.g., Boy-Girl-Boy) and number of offsprings of each type. The
389 data used in current paper are available on DHS website (after an approval from DHS).

390 Biological mechanisms of Markovian variation could be investigated using model organisms.
391 To the best of my knowledge, among animals the Markovian dependence (controlling for Lexian
392 variation) was explicitly tested only in dairy cattle (Astolfi and Tentoni 1995). The variation was not
393 significant, although authors noted weak effect in expected direction: “after a male calf the probability
394 of a male birth was slightly higher than the probability of a female birth”. It would be interesting to
395 see if Markovian variation is present in mice, rats or other species amenable to experimentation.

396

397 **Acknowledgements**

398 I’m grateful to all participants of DHS program for collection of data and making them publicly
399 available.

400 **Supplementary information**401 **Table 8 Summary statistics**

Offspring size (number of children)	Number of offsprings	Percent	Proportion of boys in the offsprings with given size
1	438,612	19.81	0.517
2	467,637	21.12	0.524
3	387,089	17.48	0.520
4	283,323	12.79	0.512
5	202,076	9.12	0.508
6	147,510	6.66	0.506
7	105,567	4.77	0.505
8	74,526	3.37	0.507
9	48,161	2.17	0.507
10	30,393	1.37	0.505
11	15,964	0.72	0.506
12	8,477	0.38	0.510
13	3,110	0.14	0.510
14	1,266	0.06	0.516
15	527	0.02	0.510
16	239	0.01	0.502
17	67	<0.01	0.519
18	30	<0.01	0.509
19	16	<0.01	0.493
20	11	<0.01	0.527
All	2,214,601		0.512

402

403

404 **References**

- 405 Astolfi, P. and S. Tentoni (1995). "Sources of variation of the cattle secondary sex ratio." Genetics
406 Selection Evolution **27**(1): 3-14.
- 407 Bayes-Genis, A., S. Roura, C. Prat-Vidal, J. Farre, C. Soler-Botija, A. B. de Luna and J. Cinca
408 "Chimerism and microchimerism of the human heart: evidence for cardiac regeneration." Nat Clin
409 Pract Cardiovasc Med.
- 410 Beilharz, R. G. (1963). "A factorial analysis of sex-ratio data. A comment on two papers by
411 Edwards." Ann Hum Genet **26**: 355-358.
- 412 Ben-Porath, Y. and F. Welch (1976). "Do Sex Preferences Really Matter?" The Quarterly Journal of
413 Economics **90**(2): 285-307.
- 414 Bernstein, M. E. (1952). "Studies in the human sex ratio: 2. The proportion of unisexual sibships."
415 Human Biology **24**(1): 35-43.
- 416 Clark, M. M. and B. G. Galef, Jr. (1995). "A gerbil dam's fetal intrauterine position affects the sex
417 ratios of litters she gestates." Physiol Behav **57**(2): 297-299.
- 418 Edwards, A. W. (1961). "A factorial analysis of sex ratio data." Ann Hum Genet **25**: 117-121.
- 419 Edwards, A. W. (1962). "A factorial analysis of sex-ratio data: a correction to the article in Vol. 25,
420 117, including some new data and a note on computation." Ann Hum Genet **25**: 343-346.
- 421 Edwards, A. W. (1966). "Sex-ratio data analysed independently of family limitation." Ann Hum
422 Genet **29**(4): 337-347.
- 423 Fujita, M., E. Roth, Y. J. Lo, C. Hurst, J. Vollner and A. Kendell (2012). "In poor families, mothers'
424 milk is richer for daughters than sons: a test of Trivers-Willard hypothesis in agropastoral settlements
425 in Northern Kenya." Am J Phys Anthropol **149**(1): 52-59.
- 426 Garenne, M. (2009). "Sex ratio at birth and family composition in sub-saharan africa: Inter-couple
427 variations." Journal of Biosocial Science **41**(3): 399-407.
- 428 Geissler, A. (1889). "Beitrage zur Frage des Geschlechterverhältnisses der Geborenen." Zeitschrift
429 des Königlich-S ächsischen Statistischen Bureaus **35**: 1-24.
- 430 Gini, C. (1951). "Combinations and sequences of sexes in human families and mammal litters."
431 Human Heredity **2**(3): 220-244.
- 432 Greenberg, R. A. and C. White (1967). "The sexes of consecutive sibs in human sibships." Human
433 Biology **39**(4): 374-404.
- 434 Jacobsen, R., H. Moller and A. Mouritsen (1999). "Natural variation in the human sex ratio." Hum
435 Reprod **14**(12): 3120-3125.
- 436 James, W. H. (1975). "Sex ratio and the sex composition of the existing sibs." Annals of Human
437 Genetics **38**(3): 371-378.
- 438 James, W. H. (2000). "The variation of the probability of a son within and across couples." Human
439 Reproduction **15**(5): 1184-1188.

- 440 Khosrotehrani, K. and D. W. Bianchi (2005). "Multi-lineage potential of fetal cells in maternal tissue:
441 a legacy in reverse." Journal of Cell Science **118**(8): 1559-1563.
- 442 Maconochie, N. and E. Roman (1997). "Sex ratios: are there natural variations within the human
443 population?" Br J Obstet Gynaecol **104**(9): 1050-1053.
- 444 Makela, O. (1963). "Analysis of some sex ratio data." Ann Hum Genet **26**: 333-334.
- 445 Malinvaud, E. (1955). "Relations entre la composition des familles et le taux demasculinit é" Journal
446 de la soci é statistique de Paris **96**: 49-64.
- 447 Mitter, N. S. and S. Anand (1975). "Trend Towards Unisexuality In Human Sibships : A Study In
448 Two Generations." Indian Anthropologist **5**(2): 12-20.
- 449 Renkonen, K. O. (1956). "Is the sex ratio between boys and girls correlated to the sex of precedent
450 children?" Ann Med Exp Biol Fenn **34**(4): 447-451.
- 451 Renkonen, K. O., O. Makela and R. Lehtovaara (1961). "Factors affecting the human sex ratio." Ann
452 Med Exp Biol Fenn **39**: 173-184.
- 453 Renkonen, K. O., O. Makela and R. Lehtovaara (1962). "Factors affecting the human sex ratio."
454 Nature **194**: 308-309.
- 455 Rodgers, J. L. and D. Doughty (2001). "Does Having Boys or Girls Run in the Family?" CHANCE
456 **14**(4): 8-13.
- 457 Thomas Gualtieri, C., R. E. Hicks and J. P. Mayo (1984). "Influence of sex of antecedent siblings on
458 the human sex ratio." Life Sciences **34**(19): 1791-1794.
- 459 Turpin, R. and M. P. Schutzenberger (1949). "Sur la distribution de sexes chez l'homme." La Semaine
460 des Hospitaux de Paris, La Semaine Medicale(25 annee, n. 60): 2544-2545.