DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 51, ARTICLE 14, PAGES 459–500
PUBLISHED 20 AUGUST 2024
https://www.demographic-research.org/Volumes/Vol51/14
DOI: 10.4054/DemRes.2024.51.14

Research Article

Transitions to adulthood in men and women in rural Malawi in the 21st century using sequence analysis: Some evidence of delay

Estelle McLean **Emma Slaymaker**
Maria Sironi **Amelia C. Crampin**
Albert Dube **Rebecca Sear**

Contents

Demographic Research: Volume 51, Article 14
Research Article

Transitions to adulthood in men and women in rural Malawi in the 21st century using sequence analysis: Some evidence of delay

Estelle McLean[1]
Maria Sironi[2]
Albert Dube[3]
Emma Slaymaker[4]
Amelia C. Crampin[5]
Rebecca Sear[6]

Abstract

BACKGROUND
Many sub-Saharan African countries have large populations of young people, and these cohorts have the potential to bring significant change. Understanding the changing lives of young people is important for ensuring individuals experience healthy and successful transitions to adulthood and for understanding how best to ensure they realise their full potential, for themselves and their communities.

METHODS
This analysis used detailed, prospectively collected data on men and women from the Karonga Health and Demographic Surveillance Site in rural Malawi between 2004–2017. Multi-channel sequence analysis was conducted for the domains of leaving school, getting married, and having children. Sequences were grouped into clusters and descriptive and multinomial logistic regression carried out to assess differences in the transition to adulthood by sex, calendar time, and sociodemographic factors.

RESULTS/CONCLUSIONS
Many participants followed a fairly early and rapid transition to adulthood; however, a sizable proportion experienced later marriage. These participants were split into those

[1] Corresponding author. Malawi Epidemiology and Intervention Research Unit, PO Box 46, Chilumba, Karonga District, Malawi and Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom. Email: info@meiru.mw.
[2] Social Research Institute, University College London, 55-59 Gordon Square, London, WC1H 0NU, United Kingdom.
[3] Malawi Epidemiology and Intervention Research Unit, PO Box 46, Chilumba, Karonga District, Malawi.
[4] Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom.
[5] Malawi Epidemiology and Intervention Research Unit, PO Box 46, Chilumba, Karonga District, Malawi and Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom.
[6] Centre for Culture and Evolution, Brunel University London, Kingston Lane, Middlesex, UB8 3PH, United Kingdom.

(mostly men) who received little education and had a long delay between school-leaving and marrying, and those who spent longer in education. Access to secondary school increased over time for both men and women, while only women experienced delays in age at marriage. Where associations between sociodemographic factors and transitions to adulthood were found, they suggested that greater access to resources was associated with a later transition out of education, and therefore to marriage.

CONTRIBUTION

Sequence analysis applied to longitudinal HDSS data proved a useful tool to identify different trajectories to adulthood in this rural Malawian context.

1. Introduction

Adolescence and young adulthood are important stages of development: events and decisions made during these periods can have far-reaching effects in later life (Bennett and Waterhouse 2018; Coall et al. 2016; Day et al. 2015; Kreniske et al. 2019). Most sub-Saharan African countries have young populations with large numbers of adolescents and young adults, and this 'demographic dividend' of young, potentially productive individuals can bring significant benefits to these countries. Many such countries are experiencing a period of change (Lutz et al. 2019): with increasing urbanisation and access to schooling and the internet, young people now may expect, and have, very different lives to their parents. Understanding the changing lives of young people is important for ensuring individuals experience healthy and successful transitions to adulthood and for understanding how best to ensure the demographic dividend delivers on its potential for good.

The transition to adulthood is typically not defined by a single event but by a series of transitions, which may differ between societies but are often related to changes in family and working life (Juárez and Gayet 2014). For example, many studies in sub-Saharan Africa have examined transitions to adulthood such as sexual debut, school-leaving, pregnancy, and marriage (Delprato et al. 2015; Odimegwu and Mkwananzi 2016; Yakubu and Salisu 2018), and a few have also looked at leaving home (Chae, Hayford, and Agadjanian 2016). Both the timing and the nature of transitions differ over time. For example, the rise of formal education and wage labour means that in many societies the shift from education to work may now be considered a key part of the transition to adulthood. In the past, and in some contemporary societies which are still predominantly subsistence-based, children and adolescents engage in productive work at young ages, meaning the transition into work is not a meaningful marker of adulthood. The formation of romantic relationships and entry into parenthood are more universal

transitions to adulthood, though education may delay the formation of formal marital or cohabiting relationships (and therefore also first births), because it is typically socially normative to complete education before entering such relationships (Amoo 2017; Shapiro and Gebreselassie 2014). Access to contraception and changing social norms may extend the transition to adulthood by allowing the entry to parenthood to happen much later than sexual debut (Reda and Lindstrom 2014; Sauvain-Dugerdil et al. 2008). These societal shifts may also lead to greater diversity and complexity in transitions to adulthood (Nahar, Xenos, and Abalos 2013). These broad generalisations hide a range of contextual factors which affect the definitions and timing of transitions to adulthood in different societies, and how they change over time.

Here, we use existing prospectively collected data from the Karonga Health and Demographic Surveillance Site (HDSS) in Northern Malawi to describe transitions to adulthood and assess whether there is evidence for change over time, for women and men. This is a context which is experiencing considerable change: Education and paid employment (often involving migration to urban centres) are increasingly important for young people, though the majority of this rural community still engages in subsistence farming. In this analysis we focus on the three transitions to adulthood for which data are available: leaving education, marriage, and having children.

Most studies of transitions to adulthood tend to assess each event singly by calculating the proportion who have experienced the events by a certain age. This kind of data can be collected in cross-sectional surveys, is relatively easy to collect, and can be useful for understanding the situation at a particular time-point. More sophisticated studies have collected retrospective data on age/date of events, and some have looked at the events in the context of each other by examining the order and timing of events (Beguy et al. 2011; Biddlecom and Bakilana 2003; Pesando et al. 2021). Such retrospective studies can be subject to recall bias (Mensch et al. 2014) and require greater skill on the part of interviewers to help respondents to report accurate data, but have the benefit of producing results quicker than prospective surveys. We are fortunate in having access to a longitudinal dataset in which data have been collected prospectively over many years and are therefore subject to less recall bias. However, such longitudinal datasets may be subject to biases due to selective loss to follow-up (which may be particularly significant in a young adult population, as young adults are often highly mobile). We discuss later the implications of loss to follow up for our results.

2. Background

2.1 The transition to adulthood in Malawi

In Malawi, marriage and child-bearing is expected and almost universal, often happening at a relatively young age, especially for women: in 2015–2016 nearly half of Malawian women were married by age 18 (National Statistical Office 2016). Marriage is one of the main traditional markers of moving from adolescence to adulthood, conferring respect and allowing individuals to take part in community meetings (Kok et al. 2021). Many young people enjoy the independence and responsibilities it grants them (Ansell et al. 2018). Young people tend to choose their own partners; however, a marriage might be encouraged by parents in the event of a pregnancy (Ansell et al. 2018; Melnikas et al. 2022), though this is not always the case if they come from a wealthier background (Poulin, Beegle, and Xu 2021).

Young people in Malawi tend to have high aspirations for their education (Zahra 2020); however, especially in rural areas they are often unable to fulfil them. Being over-age for the school grade is common in rural areas due to delayed school enrolment (Moyi 2010), or students having to pause their studies due to lack of money or repeat grades due to poor performance (Zahra 2020). Being too old for their grade leads to a higher likelihood of eventual school drop-out (Sunny et al. 2017), and for women is associated with early pregnancy and marriage (Glynn et al. 2018). This highlights the links between the different events which mark transition to adulthood in this context. Transitions out of education and into marriage and first birth are often temporally close because the decisions to leave education and to get married or give birth are not independent. Women do not tend to marry while in formal education because education provides some protection from social pressure to marry, but a pre-marital pregnancy while still at school might result in leaving education and marrying. Even without a pre-marital pregnancy, some women, especially, for example, an older student struggling to complete primary school (i.e., a student who is over-age for their grade), may feel that marriage is a preferable option to staying in education (Bertrand-Dansereau and Clark 2016).

Transitions into marriage are not necessarily permanent in Malawi: divorce is also common at all ages (Clark and Brauner-Otto 2015). This often occurs within the first three years of the union (Bertrand-Dansereau and Clark 2016; John 2022), particularly if the couple are young, have not known each other long, or if the marriage was instigated or rushed into due to a pregnancy (Bertrand-Dansereau and Clark 2016), though remarriage also tends to follow quickly (Grant and Pike 2019). Migration is common among young Malawians during the transition to adulthood, especially for women, who tend to move for marriage. Young men are less likely to move than women, but when they do they tend to move longer distances and are more likely to report moving for work

or education (Beegle and Poulin 2013; McLean et al. 2023). We consider later how migration influences our analysis.

In Malawi during the period under analysis (2004–2017) the Malawian government campaigned and committed to end child (under 18) marriage (Daniel 2017), to reduce adolescent pregnancy (Pot 2019), and to increase access to secondary school. Evidence from the Malawi Demographic and Health Survey shows that while the overall mean age at first marriage for women has remained quite stable, in the north (where the Karonga HDSS is located) it increased between 2010 and 2015/16 (Baruwa, Amoateng, and Biney 2019). There was also a decrease in the prevalence of teenage child-bearing in Karonga from 2010 to 2015/16, though it increased overall in Malawi, apparently driven by increases in urban areas (Chintsanya, Magadi, and Likupe 2021). During this period there was also widespread access to anti-retroviral therapy (ART), which reversed increasing adult mortality rates that were mostly a result of the HIV/AIDS pandemic (Price et al. 2017). These changes in adult mortality, along with declining infant and child mortality rates (Lungu et al. 2019) and declining fertility rates (McLean et al. 2017), have likely had an effect on family structures and formation. All these factors plus increasing access to smartphones and the internet may cause changes to transitions to adulthood in Karonga.

2.2 Sequence analysis and its use in studying the transition to adulthood

Sequence analysis uses various methods to indicate how similar or dissimilar text strings are from one standard string, or from each other. It has been used extensively in genomics for assessing stretches of DNA, and over the last few decades has been used increasingly in life course research (Liao et al. 2022). In genomics each letter in the text strings represents a DNA base pair; in life course research it represents a state that a person experiences, such as marital status. A sequence of NNNNMMMDDD would represent someone spending 4 periods never married, 3 married, and 3 divorced. The period of time that each letter represents varies across analyses; i.e., it could be yearly, quarterly, monthly, or even daily. Sequence analysis has been used to study the transition to adulthood with data from Europe (Oris and Ritschard 2014; Schwanitz 2017) and from low and middle income countries (Pesando et al. 2021).

3. Methods

3.1 Data source

Our data come from a Health and Demographic Surveillance System (HDSS) in southern Karonga district in Northern Malawi. This is a predominantly rural, subsistence farming and fishing area. The main ethnic group are Tumbuka, who since the 19th century have followed patrilineal and patrilocal customs (Vail 1989): women tend to move to their husband's village when they marry. Land is held by men, and a father will assign land to his married sons (Mutangdura 2004). Both men and women tend to remain living with their parents/guardian until they marry: though young men may decide to move out earlier to get some freedom this is not very culturally acceptable (personal communication). Marriage is traditionally negotiated between the groom's paternal uncle and the bride's father and paternal aunts and uncles. If accepted, the bride price (*lobola*) is agreed upon (Bertrand-Dansereau and Clark 2016). Anecdotally, elopement, where the bride moves with the groom without negotiation or payment of *lobola*, is becoming more common due to the increased cost of weddings and *lobola* (personal communication). Marrying and having children is the norm, and fertility rates are relatively high, although decreasing (McLean et al. 2017). Polygyny is widespread: at the end of 2016 about 15% of households in the HDSS were headed by men with multiple wives. The area has been affected by the AIDS pandemic, though not as much as in some areas: in 2009 HIV prevalence was estimated to be 9% in women and 7% in men (Floyd et al. 2013). ART became widely available in the area over the period of analysis, which contributed to reduced mortality rates and increased life expectancy (Price et al. 2017).

The Karonga HDSS was set up between 2002 and 2004 in a 150m^2 area in the south of the Karonga district (Crampin et al. 2012). All households were surveyed in the initial census and information gathered on household and individuals. The area remains under continuous follow-up: births and deaths are collected on a monthly basis and migrations in, out, and within the area on an annual basis. In-migrants who are returning to the area are consistently and reliably linked back to their original identification number, even if they left the area for a long time and/or moved to a different household. Households are asked to identify a head (who must be resident) and report if/when this changes. Regular survey rounds gather further information on each household and individual (including their own and their parents' education status), and unique household and person identifiers allow linkage of all data collected at the site. For all individuals, the identification numbers of their mother and father are recorded, even if the parent is not an HDSS member. All participants over the age of 12 are regularly asked to report their marital status and report past and current spouses, who are also linked to existing personal

identifiers (or are assigned new ones). For this analysis, data from 1 January 2004 to 31 December 2017 were used.

3.2 Data management

The HDSS data are longitudinal: each person's time in the area is arranged as episodes of when they started and ended living in a particular household. This episode dataset was converted to a quarterly dataset with 4 records per person per year. The open cohort nature of HDSSs means that information is not collected on people who leave the area, so a sequence analysis which only includes information on people with complete sequences may introduce bias, as those who stay may differ from those who remain. Work has been done to suggest ways to impute missing data on sequences and include imputed data in analyses (Gabadinho and Ritschard 2016; Halpin 2016). However, in this case the level of missing data was felt to be too high to consider that option: on average about 50% of young people left the area during the sequence period examined. Instead, we included individuals with complete data, but conducted an assessment of the validity of conducting sequence analysis excluding people with missing information due to in- and out-migration (this can be found in Appendix 1). This assessment also allowed us to determine the length of the sequences (i.e., the period of time over which people are observed) that were included in our analyses. If only individuals with complete data are included, then the sample size decreases and the potential for bias increases as the length of time over which people are observed increases. We compared 3 sequence lengths: 9 years (ages 14–22 for women, 17–25 for men), 7 years (ages 15–21 for women, 18–24 for men), and 5 years (16–20 for women and 19–23 for men) [the age ranges are different as women tend to marry earlier than men]. We decided to use the 7-year sequence length (ages 15–21 for women, 18–24 for men), as it was found to provide an appropriate balance between having the most representative sample and being able to study the transition to adulthood over a reasonably long time period. However, the assessment of representativeness suggests that the dataset we used may underestimate the proportion who do not marry for both men and women, and overestimate the proportion who marry in the mid period for men, and the early period for women – biases which need to be taken into account when interpreting results (Appendix 1).

In the final datasets, all women are included between the ages of 15 and 21 and all men between 18 and 24, meaning we include women born between 1989 and 1995 and men born between 1986 and 1992. Note that choosing these age ranges means that some transitions to adulthood may have occurred before the individual is 'observed' in the dataset, though this is only the case for a small proportion of individuals. To explore change over time, for some analyses we divide the datasets into two 'eras', one sequence

period starting between 2004 and 2007 (so ending no later than 2013), the other starting between 2008–2011 (so ending up to 2017).

Initially, we considered 4 transition-to-adulthood variables: leaving school, leaving home, marrying, and having children. However, results observed for leaving home and marrying were very similar. Married participants almost always lived in a household where the household head was listed as either the participant, their spouse, or an in-law (and not a parent). Never-married or divorced participants, on the other hand, typically reported a parent or other relative as head: only 4.2% of quarters deviated from this pattern. Thus, it was deemed unlikely that leaving home for reasons other than marriage would be detectable in this dataset, and for the sake of simplicity we dropped this variable.

The sources of data for the markers of the adulthood variables are described below.

- Schooling: participants (or close informants) were asked (usually annually) about their schooling, whether they were currently attending and what grade, and/or what was the highest grade they attended/qualification received. They were also asked the age or year they left school if they had left. Using this information, each person's schooling history was constructed, including gaps where they left school and returned. This variable contains the following categories: never attended, currently in primary, currently in secondary, left primary, left secondary.
- Marriage: participants (or close informants) were regularly asked (usually annually) to report their current marital status, the start and end dates of their current and previous marriages, and the age/year they first got married. Combining the above, a person's marital history was constructed with the following categories: never married, married (or re-married following a divorce), divorced/widowed.
- Children: mother and father identifiers have been continuously recorded for all individuals during the HDSS. Birth dates or estimates are also recorded for all participants. These were used to classify participants as having no children, 1 child, or 2 or more children.

3.3 Sequence analysis

We carried out multi-channel sequence analysis using the TraMineR package in R (Gabadinho et al. 2011). In this form of sequence analysis each participant has a separate sequence for each of the 3 domains. The resulting dissimilarity matrices are combined into one, before cluster analysis is carried out. Multichannel sequence analysis has advantages over just using one variable as it allows the observation of the inter-connectedness of life experiences. In some cases it may be used to understand how other

life processes or events interact with or impact the main outcome under investigation, or the interactions of all factors may be the goal of the analysis (Pollock 2007). There are various methods which can be used to create the dissimilarity matrix and several have been recommend for use, depending on the goal of the analysis (Studer and Ritschard 2016). However, for multi-channel sequence analysis in the R package only 4 are available: optimal matching (OM), longest common sequence (LCS), Hamming distance (HAM), and dynamic Hamming distance (DHD). With the default settings, OM and LCS are identical, so initially we used LCS, HAM, and DHD.

We then used agglomerative cluster analysis to assign the sequences to groups using the cluster R package (Maechler et al. 2023), using the Wards hierarchical algorithm. We tested other hierarchical algorithms but they produced unusable solutions (i.e., with some very small and some very large groups). We calculated several statistical measures of the quality of the different cluster solutions using the WeightCluster (Studer 2013) R package. For each dissimilarity matrix we produced 2–15 clusters and examined the cluster statistics (Appendix 2, Figure A-5 and Table A-1). Based on the cluster statistics, we displayed 3–10 clusters in sequence index graphs which allow for visual assessment of the characteristics of each cluster: these figures present each sequence within each group as a line, with different colours representing the different sequence states (Brzinsky-Fay 2014). We examined the sequence index plots from all 3 transition variables concurrently with the cluster statistics to identify the most useful solution for each dissimilarity matrix method, based on optimising the cluster statistics and whether the groups produced were useful and distinctive. From this assessment we found solutions with 6 or 7 clusters to be optimal; however, some of the clusters exhibited relatively low homogeneity, as indicated by their Average Silhouette Width (ASW) (Appendix 2 Table A-1). As other hierarchical algorithms had not worked, we used a non-hierarchical method, Partitioning Around Medoids (PAM) (Studer 2013), to try to improve the homogeneity in the 3 dissimilarity matrices for 5–7 groups.

For all 3 dissimilarity matrix methods the 6-cluster PAM solution was the most useful, both in terms of the visual inspection of the clusters and in optimising the overall group ASWs, so we chose it over the hierarchical methods described above. The solutions for all 3 dissimilarity matrix methods were similar and we selected LCS as it had marginally higher ASWs (Appendix 2 Table A-2). The overall ASW still falls in the 'weak' category according to cut-offs proposed by Kaufman and Rousseeuw (1990). On inspection of the sequence groups, this appears to be driven by two main factors. First, the 'divorced' category is a feature of almost all groups; however, it is not experienced by all members of each group. Second, many groups appear to have some level of heterogeneity with regards to the timing of leaving school and marrying (i.e., the two events happening at the same time or the school leaving event occurring at least a year previously).

3.4 Analysis

We described the sequence categories according to the timing of events, whether schooling was completed, and whether there was a delay between leaving school and marrying or between marrying and having a first child. We calculated these characteristics overall and by sex, and compared the overall proportion in each group across the two sexes and by era (sequence period starting 2004–2007 or 2008–2011).

We then explored sociodemographic correlates of group membership. We considered several factors from the participants' pre-marriage households: If participants were not married at the start of the sequence period the first complete record at the beginning of the sequence was used; if they were already married their latest 'never married' record (prior to the start of the sequence period) was used. Therefore, this analysis was only carried out on a subset of participants with available data: participants may not have had available data if they were already married at the start of the sequence and were not previously present in the HDSS. We carried out multinomial logistic regression models with sequence group as the outcome; group 2 (mid marriage, primary) was treated as the baseline as this group was a good size for both men and women, and a good proportion had sufficient data to be included in the model. All the socioeconomic variables under examination were included as covariates simultaneously, and all analyses were carried out separately by sex.

In addition, for the groups that were predominantly not married by the end of the sequence period, the proportions not married 1 and 2 years after the sequence period were calculated. This analysis could only be performed on a subset of participants who had this data available.

4. Results

4.1 Description of sequence categories

1,326 female and 1,325 male adolescents were present in the HDSS at age 15 for women and 18 for men and remained for at least 24 of the subsequent 28 quarters, so were included in the analysis. We identified 6 clusters through the multi-channel sequence analysis combining men and women, using the Longest Common Sequence dissimilarity matrix method and Partition Around the Medoids clustering algorithm. Sequence index plots separated into men and women are shown in Figure 1, and summaries of the transition variables for each group, overall and by sex, are shown in Table 2. The groups are described below.

4.1.1 Group 1: Earlier marriage (left primary)

This group is characterised by early marriage and having left school with only primary education (Figure 1), though only 40.6% (men 45.1%, women 38.0%) had completed all grades of primary school. All experienced marriage, and 20.8% (men 18.3%, women 22.2%) experienced divorce. While the majority left school and married at a similar time (the average number of quarters between the 2 events was –0.6), 23.8% (men 41.3%, women 12.5%) left school a year or more before marriage. Less than 1% had no children and the majority (95.3%) had 2 or more children by the end of the period. The majority had children quickly (the average number of quarters between marrying and having a child was 3.3 for women and 4.6 for men), with only 5.5% (men 7.5%, women 4.3%) having a delay of 2 or more years between marriage and first child (Table 1).

4.1.2 Group 2: Mid marriage (left primary)

This group also only attended primary school, but married later in the period compared to group 1 (Figure 1). The proportion that completed primary school was higher than group 1 at 57.4% (men 56.8%, women 57.8%). Women in this group tended to leave school and marry at a similar time (average number of quarters between the two events was 2.0), while a group of men left school before marrying (68.9% of men left school at least a year before marrying compared to 22.6% of women). Only 1% of this group did not experience marriage (all women), and 16.9% (men 8.7%, women 23.6%) experienced divorce. The average number of quarters between marrying and having a child is higher than in group 1 (5.5), as is the proportion with at least a 2-year gap between the two events (22.1%). Under 3% had no children, and 32.6% had at least 2 by the end of the period (Table 1).

4.1.3 Group 3: Mid marriage (left secondary)

This group married in the mid-period, similar to group 2, but attended secondary school (Figure 1); however only 19.3% (men 28.3%, women 10.2%) completed secondary school. Less than 1% did not marry, and the majority left school and married at a similar time (average number of quarters between the two events was 0.3), though 33.3% of men and 10.6% of women left school 1 year or more before marrying. 18% (men 11.8%, women 24.2%) experienced divorce. The average number of quarters between marrying and having a child was slightly lower than in group 1 (women 2.8, men 3.9), though the proportion with 2 or more years between the two events was higher (men 12.2%, women

9.5%). Less than 2% had no children and 35.7% had at least 2 by the end of the period (Table 1).

4.1.4 Group 4: No/late marriage (left primary)

This group left primary school early in the period but only married late, or did not marry before the end of the sequence period (Figure 1). Only 33.2% (men 35.2%, women 24.5%) completed primary school. 46.6% married, and 25.2% had a child by the end of the period (Table 1). The time between leaving school, marrying, and having a child is not calculated for groups 4, 5, and 6 as fewer group members experienced these transitions.

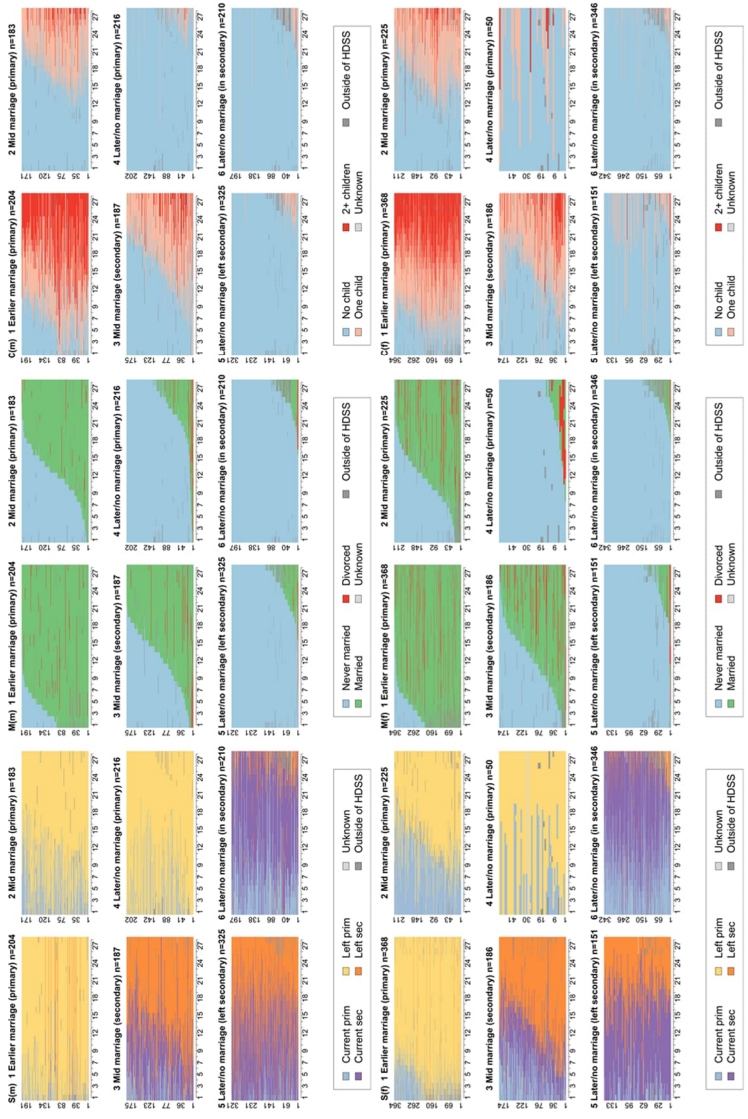
4.1.5 Group 5: Late/no marriage (left secondary)

This group attended and left secondary school, but married late in the period or did not marry. Some of the women had a child before marrying (Figure 1). Overall, 67.9% (men 67.7%, women 68.2%) completed secondary school and 1.1% were still attending school by the end of the period. 30.5% (men 32.6%, women 25.8%) married and 25.4% (men 20.6%, women 35.8%) had a child by the end of the period (Table 1).

4.1.6 Group 6: Late/no marriage (in secondary)

This group is similar to group 5, but members remained in secondary school for longer (Figure 1). Overall, 29.5% (men 39.5%, women 23.4%) had finished secondary school, and 62.2% (men 56.7%, women 65.6%) were still in education at the end of the period. 20.3% (men 21%, women 19.9%) married, and 16% (men 11%, women 19.1%) had a child by the end of the period (Table 1).

Figure 1: Sequence index plots for S – schooling; M – marriage; and C – having children, separated for men (m) – top row, and women (f) – bottom row



Note: Each individual is represented by 1 row in the figures, and individuals are ordered by increasing number of quarters in the 'never married' category (lowest number at the bottom of each graph).

Table 1: Characteristics of the 6 sequence groups, overall and by sex

		Overall						Male						Female					
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Total		572	408	373	266	476	556	204	183	187	216	325	210	368	225	186	50	151	346
Finished primary	n	278	239	364	101	462	517	117	102	183	83	317	193	161	137	181	18	145	324
	%	48.6%	58.6%	97.6%	38.0%	97.1%	93.0%	57.4%	55.7%	97.9%	38.4%	97.5%	91.9%	43.8%	60.9%	97.3%	36.0%	96.0%	93.6%
Finished secondary	n	3	1	80	0	300	295	3	0	55	0	213	131	0	1	25	0	87	164
	%	0.5%	0.2%	21.4%	0.0%	63.0%	53.1%	1.5%	0.0%	29.4%	0.0%	65.5%	62.4%	0.0%	0.4%	13.4%	0.0%	57.6%	47.4%
In primary	n	0	0	0	0	0	10	0	0	0	0	0	4	0	0	0	0	0	6
	%	0.0%	0.0%	0.0%	0.0%	0.0%	1.8%	0.0%	0.0%	0.0%	0.0%	0.0%	1.9%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%
In secondary	n	0	1	6	0	5	346	0	0	5	0	4	119	0	1	1	0	1	227
	%	0.0%	0.2%	1.6%	0.0%	1.1%	62.2%	0.0%	0.0%	2.7%	0.0%	1.2%	56.7%	0.0%	0.4%	0.5%	0.0%	0.7%	65.6%
Left school 1 yr+before marriage[1]	n	66	166	75				45	122	57				21	44	18			
	%	23.8%	44.6%	22.0%				41.3%	68.9%	33.3%				12.5%	22.6%	10.6%			
Q between leaving school and marrying [2]	mean	–0.6	2.0	0.3				–0.4	4.5	0.9				–0.7	0.5	–0.2			
Never married	n	0	4	2	142	331	443	0	0	0	105	219	166	0	4	2	37	112	277
	%	0.0%	1.0%	0.5%	53.4%	69.5%	79.7%	0.0%	0.0%	0.0%	48.6%	67.4%	79.0%	0.0%	1.8%	1.1%	74.0%	74.2%	80.1%
Ever divorce	n	117	69	67	23	15	11	37	16	22	17	11	4	80	53	45	6	4	7
	%	20.5%	16.9%	18.0%	8.6%	3.2%	2.0%	18.1%	8.7%	11.8%	7.9%	3.4%	1.9%	21.7%	23.6%	24.2%	12.0%	2.6%	2.0%
First child 2yr+ from marriage[3]	n	30	85	39				15	39	22				15	46	17			
	%	5.5%	22.1%	10.8%				7.5%	22.3%	12.2%				4.3%	21.9%	9.5%			
Q between marrying and 1st child[4]	mean	3.8	5.5	3.3				4.6	5.7	3.9				3.3	5.3	2.8			
No children	n	2	11	5	199	355	467	2	7	4	168	258	187	0	4	1	31	97	280
	%	0.3%	2.7%	1.3%	74.8%	74.6%	84.0%	1.0%	3.8%	2.1%	77.8%	79.4%	89.0%	0.0%	1.8%	0.5%	62.0%	64.2%	80.9%
2+ch	n	545	133	133	9	7	1	193	63	66	4	2	1	352	70	67	5	5	0
	%	95.3%	32.6%	35.7%	3.4%	1.5%	0.2%	94.6%	34.4%	35.3%	1.9%	0.6%	0.5%	95.7%	31.1%	36.0%	10.0%	3.3%	0.0%

Note: 1 = Earlier marriage (primary); 2 = Mid marriage (primary); 3 = Mid marriage (secondary); 4 = No/late marriage (primary); 5 = Late/no marriage (left secondary); 6 = Late/no marriage (in secondary); [1] % calculated from participants with marriage observed during the sequence; [2] mean calculated from participants with both leaving school and marriage observed in the sequence; [3] % calculated from participants with first child observed during the sequence; [4] mean calculated from participants with marriage and first child observed in sequence.

4.2 Distributions by sex and era

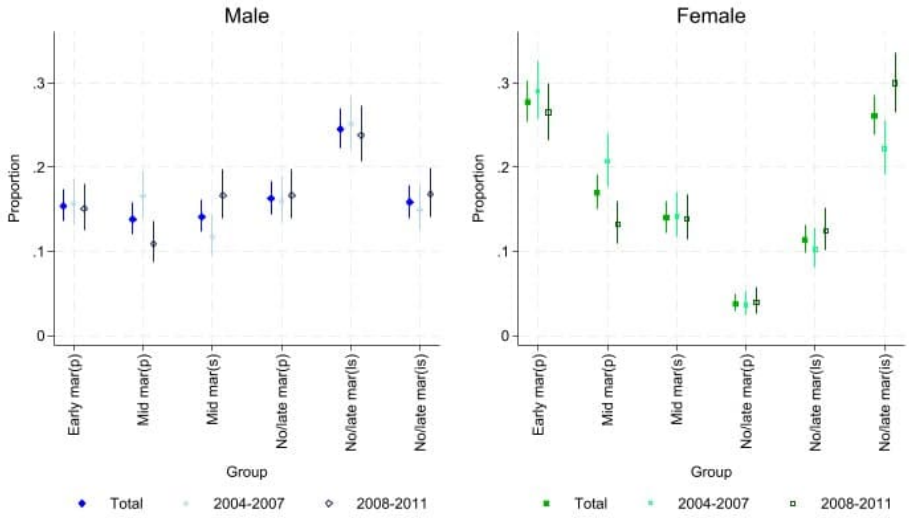
For men, the most common group was 5 (Late/no marriage (left secondary)) with 24.5% of men falling into this group; the remaining groups had around 14%–15% of men overall. Most groups had similar proportions of men in both eras, suggesting little change over time, apart from group 2 (Mid marriage (primary)) which included 16.5% of men in the earlier era and 10.9% in the later era, and group 3 (Mid marriage (secondary)) which included 11.7% and 16.6% of men respectively. For women, there was more variation in group size. The largest groups were group 1 (Earlier marriage (primary)), which accounted for 27.8%, and group 6 (Late/no marriage (in secondary)) which accounted for 26.1%, and included a higher proportion of women in the later than the earlier era (22.2% versus 30.0%). The proportions of women in group 2 (Mid marriage (primary)) also showed differences between the two eras for women: 20.7% in the earlier era and 13.3% in the later. Group 4 (No/late marriage (primary)) was very uncommon for women, accounting for only 3.8% compared to 16.3% for men (Figure 2).

4.3 Sociodemographic characteristics of sequence groups prior to marriage

Overall, 2,463 (93%) participants had information available about their pre-marriage household. The group with the lowest availability was the early marriage group, group 1 (women 66%, men 88%), but all other groups had over 90% with data for each sex, with over 99% availability for groups 4–6. For the majority (93%) the pre-marriage data were from the start of the sequence period (i.e., age 15 for women and age 18 for men). For group 1 the data was more likely to be from before the sequence period (31% of men and 39% of women), but for groups 3 and 4 the proportion of data from the pre-marriage period was under 3% and for groups 4–6 it was under 1%.

Figure 3 and Table A-3 show results from multinomial logistic regression models with all sociodemographic variables, with sequence group as the outcome and group 2 (mid-marriage (primary)) treated as the baseline. Living in parents' household, father having secondary education (but not mother), higher dwelling quality, improved water, and proximity to tarmac road in general were associated with membership of the 3 groups which attend secondary school (group 3 'mid marriage (secondary)'; group 5 'no/late marriage (left secondary)'; group 6 'no/late marriage (in secondary)'). Proximity to a tarmac road was also associated with a higher chance of membership of group 4 (no/late marriage (primary)), but there were few other differences between the 3 primary-school-attending groups.

Figure 2: Proportion in each sequence group by sex and era

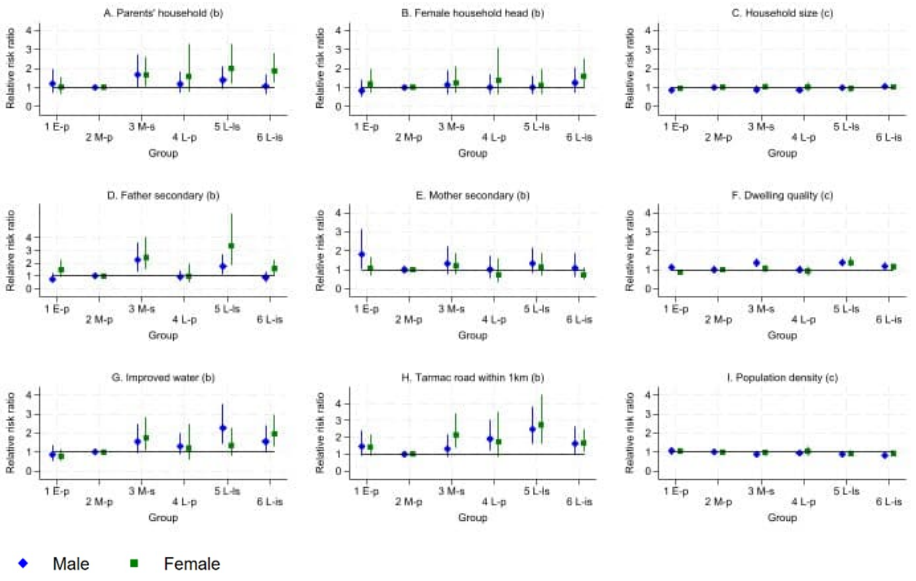


4.4 Post-sequence period

For men in group 4, 5, or 6 (groups which had mostly not yet married by the end of the period) 85% had data available for 1 year after the sequence period. Group 4 (no/late marriage (primary)) was most likely to have data available, with 9% unavailable compared to 18% in groups 5 (no/late marriage (left secondary)) and 6 (no/late marriage (in secondary)), which was mostly due to out-migration. Of those with available data, only 37% of group 4 remained never married (compared to 49% at the end of the sequence period). This was 53% for group 5 (67% at end of sequence) and 62% for group 6 (79% at end of sequence). Data availability for two years after the sequence period drops to 65% overall due to increasing numbers who were too young to have this age observed before the end of the data period. For those with available data, the proportion never married in group 4 (no/late marriage (primary)) remained at 36%, but reduced to 43% in group 5 and 52% in group 6. Data availability was lower for women than for men, especially for group 6, where 20% was unavailable 1 year later (mostly due to out-migrations) and 47% was unavailable 2 years later (split between being too young and out-migration). Of those with available data, in group 4, 67% were never married after 1 year and 55% after 2 years (compared to 74% at the end of the sequence); in group 5,

64% were never married after 1 year, reduced to 53% after 2 years (74% at end of sequence); and in group 6, 70% were never married after 1 year, reduced to 62% after 2 years (80% at end of sequence). The proportion with data available for 3 or more years after the sequence period was very low, mostly due to participants being too young to have that age observed.

Figure 3: Relative risk ratios for sociodemographic factors from multinomial logistic regression model with sequence group as outcome (group 2 as baseline) adjusted for all covariates displayed here, separate models by sex. A. living in parents' household, B. female household head, C. household size, D. father with secondary school, E. mother with secondary school, F. dwelling quality score, G. improved drinking water source, H. tarmac road within 1km, I. population density



5. Discussion

Sequence analysis proved a useful tool to investigate transitions to adulthood in the context of rural Malawi. It produced groups which were quite easy to interpret, though there was some degree of heterogeneity across all groups, especially with regards to timing of school-leaving relative to marriage. Many participants followed a fairly rapid transition to adulthood, leaving school, marrying, and having children all within a short time period (groups 1, 2, and 3 in the sequence analysis (as was seen in a cross-national study of transitions to adulthood in 69 countries in rural Africa: Pesando et al. 2021)). However, a sizable proportion experienced later marriage. These participants were divided between those (mostly men) who received little education and had a long delay between school-leaving and marrying (group 4, 'late/no marriage, left primary' in the sequence analysis) and those who were able to access more education so that both leaving school and marriage were delayed relative to other young people (group 5, 'late/no marriage, left secondary' and group 6, 'late/no marriage, in secondary'). With respect to change over calendar time, group membership looked fairly similar across time periods for most groups, which is perhaps unsurprising given that the calendar time period we are considering is relatively short. Nevertheless, we did pick up some differences over time, largely related to access to education. For men, these changes over time seem related to increased access to secondary education only (fewer participants left school with only primary education), while for women the changes over time involved increased access to secondary education and also some delay in the timing of marriage.

The rapid transition between marrying and having a child for groups 1 (early marriage (primary)) and 3 (mid marriage (secondary)) (average of less than a year between the two events) makes it likely that many of these marriages were precipitated by a pregnancy. In rural Malawi, pre-marital pregnancy is still regarded as taboo, and in some locations requires paying a fine to the traditional authority (Naphambo 2022), so parents will often pressurise a young couple to marry if there is a pregnancy or they are deemed to be at high risk of one. This strategy is common in Malawi and neighbouring countries (Kok et al. 2021; Melnikas et al. 2022); however, the young people are not without agency in these decisions. For many, especially those who cannot see that continuing to attend school is possible for them, or that it will change their prospects, a marriage or pregnancy at an early age is a choice to gain independence and respect as an adult (Baraka et al. 2022; Kidman et al. 2024; Kok et al. 2021). In our data there were a few women with observable sequences involving having a child without being married; however, this was very rare and did not constitute a sequence group in its own right. These results are similar to those observed in rural Zimbabwe, where only 5% of women and 10% of men experience having a child before entering marriage (Del Fava et al. 2016). However, there are societies in the region where childbearing outside a union is

more normative and some women are able to remain in school while having a child, such as South Africa (Bennett and Waterhouse 2018; Goldberg 2013) and in an informal settlement in Nairobi, Kenya, where 28% of female sequences involved having a child before marriage (Beguy, Ndugwa, and Kabiru 2013). Across sub-Saharan Africa, rural location, lower schooling, and lower socioeconomic status have been found to be associated with adolescent marriage (Belachew et al. 2022). In rural Malawi early pregnancy and marriage have been found to be associated with being behind at school and living in communities where marriage before age 18 is considered acceptable (Kidman et al. 2024). In our study there were not many observable differences between the group that married early with only primary education (likely due to a pregnancy) and the group that married slightly later (less likely to be due to a pregnancy).

Three groups in the present analysis only attended primary school, accounting for just under half of the whole sample for both sexes, though this proportion decreased over time. Those who only attended primary school tended to have lower pre-marriage household socioeconomic status and lived in more rural locations, further from the main road, and without an improved drinking water source. Adolescents from very poor households in Malawi work both in and outside the household to support their families, so often do not have the time or energy for their schoolwork, despite wanting to continue studying (Rock et al. 2016; Zietz, de Hoop, and Handa 2018). Most of the adolescents in the analysis belong to subsistence farming households, which are vulnerable to harvest fluctuations that may affect the choices available to them: an analysis of DHS data linked to weather data shows that experiencing drought during the growing season was associated with faster transition to marriage and child-bearing for young rural Malawian women (Andriano and Behrman 2020). To alleviate extreme poverty in Malawi the national Social Cash Transfer Programme targets vulnerable households with unconditional cash transfers, which has been found to have some positive effects for adolescents, including increased school enrolment. However, it has had less effect on actual school attendance (Lambon-Quayefio et al. 2024) and very little effect on adolescent pregnancy and co-habitation/marriage (Dake et al. 2018).

One of the groups whose members only had primary education, did not marry, or married late in the period (group 4) consists mainly of men. Apart from living closer to a road at the beginning of the period, there does not appear to be anything that distinguishes them from the group that has primary education but does marry (group 2). In group 2, however, there is heterogeneity, with a mix of people who leave primary school and marry at a similar time and people who leave school before marrying, so this may be masking any differences. While many of the men in the ‘no/late marriage (primary)’ group do go on to marry, the proportion never married did not drop between 1 and 2 years after the sequence period, suggesting that some men remain unmarried for a while. A qualitative study in rural Malawi found that adolescents living in poverty are often

isolated within the social and community networks due to stigma related to poverty and related appearance and behaviours (Rock et al. 2016), so it could be that this group of men struggle to develop social relationships. An analysis of Malawian men from Salima district (Central region) found that rural men were more likely to marry than semi-urban men, that increasing earnings made marriage more likely, and that education had no independent effect (Pike 2021). Individuals' earnings were not available to analyse in the present study, so it is possible that the men in the group that did not marry were not able to due to lack of funds. Further qualitative work is needed to understand this group: to see whether they perceive this marriage delay as negative, and, if so, what has hindered them. However, as we discuss below, leaving education early may be a strategic choice for some individuals if it brings few benefits in the context of rural Malawi.

For both men and women there were groups who attended secondary school (though did not necessarily complete it) and did not marry within the sequence period. Combined, groups 5 (no/late marriage (left secondary)) and 6 (no/late marriage (in secondary)) make up about 40% of the total for both men and women, and the assessment of representativeness indicates that the group that married late may be underestimated in those with more complete sequences. These groups do seem to marry eventually, but at older ages compared to the other groups, and often not immediately after leaving school. The young people in these groups tended to have higher household socioeconomic status and be less likely to live in a more rural area at the beginning of the period than those who married after leaving primary school. Since the year 2000 there has been a large global increase in secondary and tertiary education attendance; however, this has not uniformly improved access to secure and adequately paid work (Juárez and Gayet 2014), so the increase in access to secondary education for both men and women in our data may not indicate an improvement in their life prospects. An ethnographic study of Lesotho, India, and Laos found that while a huge amount of global funding and effort has been dedicated to improving access to education, the systems are not necessarily designed to enable rural children to succeed, with frequent testing and barriers to progressing to subsequent grades serving to filter out pupils unsuitable for formal employment. It was found that aspirations of an unattainable future were used to encourage rural children to attend school, despite teachers recognising that the majority of their students would not achieve any qualifications. Students either left with very little schooling, or continued to strive and repeat grades, yet still found themselves unable to achieve careers in the formal sector and were given very little education that would actually help them to succeed within rural livelihoods (Ansell et al. 2020). A 2007 qualitative study that compares aspirations of transition to adulthood demonstrates the barriers to rural Malawians succeeding in school. Many of the young people struggle to overcome barriers and finish secondary school, despite their best efforts, before finally giving up, having wasted a lot of time and money for very little gain (Hajdu et al. 2024). In the present analysis the

proportion in group 5 who finished school was relatively high (around 68%), but it is likely that this was after many repeated years and they may have found that their education did not help them to succeed.

The majority of children in this rural area will grow up taking on household and farming responsibilities regardless of whether they are in school, and many will continue farming whilst not defining themselves as employed or self-employed. Thus, it is not possible to identify the transition to 'work' nor to identify whether they are actively looking for other work while continuing to farm. Other studies have shown that this transition may not always be smooth, with young rural Malawians being at risk of exploitation and struggling to find reasonably paid work. Often, those attempting to develop a livelihood beyond agriculture, which many have been led to feel should be possible with the additional secondary school education they have received, experience many set-backs and losses, ending up in a worse position than when they started (Hajdu et al. 2024). Malawian men working in agriculture have been found to be more likely to marry earlier than those working in other sectors, possibly because a wife is able to help with farming but not with other work (Pike 2021), so it may be that the men who leave secondary school and experience a delay before marrying are establishing themselves in their work.

There appears to be a group of women who are delaying marriage by remaining in school: previous work has suggested that community pressure to marry may be alleviated somewhat if the young person is in education (Bertrand-Dansereau and Clark 2016). However, there are also women who are experiencing a delay between leaving secondary school and marrying: as women tend not to be seen as 'providers' in the Karonga area and there is a strong societal expectation that they will marry relatively young (Bertrand-Dansereau and Clark 2016), the reason for this delay is less clear. Qualitative work would also help to understand this group of women, and whether they are able to be more selective in their partner, or if they have fewer options once they are older with more education. This delayed marriage group are more likely to live with their parents in a less rural location and have higher socioeconomic status than the group which marry in the mid-period with only primary education. They are also more likely to be living in a female-headed household. This may be due to the father living away from home: in rural Mozambique a father's migration has been found to be associated with delayed transition to adulthood in girls (Chae, Hayford, and Agadjanian 2016).

The experience of divorce was scattered across all groups, confirming that this is a common occurrence for young people in Malawi (Bertrand-Dansereau and Clark 2016). Women were more likely to have periods of divorce, with between 22% and 25% of the participants in groups 1 to 3 having at least one period of divorce (the proportion was similar across the 3 groups). Men were less likely to have periods of divorce, but their level of divorce was higher in the group with very early marriage (group 1), which is

more likely to be instigated by a pregnancy, which has been shown to be a risk factor for divorce (Bertrand-Dansereau and Clark 2016). The sequence index plots show that while some people experience long periods of divorce, some rapidly move back to the marriage category. This confirms previous findings for Malawi which show that in general remarriage happens very quickly after divorce (John 2022).

6. Strengths and limitations

One of the strengths of our dataset is having different markers of adulthood available, allowing for a more holistic analysis of the transitions using multi-channel sequence analysis than looking at a single transition. It has been reported that multi-channel sequencing can sometimes be dominated by one particular domain, especially if that domain is more turbulent than others (Liao et al. 2022). This did not appear to be an issue in the present analysis; however, in some of the groups it did struggle to separate the people who left school long before marrying from those who experienced these two events at the same time. The HDSS data is also very flexible, allowing us to choose an appropriate sequence length, the length of the individual sequence periods, and to use data from before and after the sequence period to further understand the groups found. We only examined data from just before or after the sequence period, though earlier life experiences have been shown to be associated with risk behaviours in adolescence (Kidman et al. 2022). Linking HDSS data to much earlier events would be possible; however, the proportion in our sample with data from early childhood is likely to be small due to in- and out-migration and because the HDSS has only be operating since 2004.

The dates of the 'having children' variable should be quite accurate as births are collected prospectively, but the 'schooling' and 'marital status' variables are collected through regular survey rounds so may not be so accurate. In addition, HDSS data are often collected from proxy respondents (to collect data from the whole population as quickly as possible), which may also introduce inaccuracies. Of the records used to construct the sequences in this analysis, about 50% were reported by proxies for women, and 35% by proxies for men. For both sexes the proportion reported by proxies reduces with age. In our analysis the marital status variable was constructed from annual reports of marital status and the marriage dates of specific spouses. In some cases there were inconsistencies; i.e., a period of 'never married' following one of 'married'. In such cases the inconsistent data was over-written with the most likely state (i.e., one report of 'married' within a period of six 'never married' reports would be over-written as 'never married'). This data cleaning may have introduced some inaccuracies.

We only included three markers of adulthood in the analysis: leaving school, marrying, and having children. Leaving home was initially considered but was found to

be almost completely correlated with the marriage variable. As indicated above, it was not possible to include a 'transition to work' variable. It was also not possible to include sexual debut, which has been examined in other studies (i.e., Del Fava et al. 2016), as this was not available for the majority of participants. Including this may have provided further insight into the transition to adulthood in this area. The feeling of being an adult is fairly subjective and varies from person to person: a study in Mozambique of children aged under 17 found that boys, those not enrolled in school, those in a romantic relationship, and those who had experienced food insecurity were likely to report subjectively feeling like an adult (Axxe, Hayford, and Eggum 2022). Thus, some of the participants in our analysis categorised as not having transitioned to adulthood may feel differently themselves.

Our assessment of representativeness indicated that the 7-year sequence dataset would produce a reasonable picture of the transition to adulthood, albeit with some caveats regarding the relative size of some of the groups. However, in the absence of data on those who leave the area, it is difficult to determine the extent of any bias that results from only including individuals with complete information. Here, we present our analysis as transparently as possible so that readers can understand our methods, and that researchers conducting similar analysis can determine whether their results might be similar to ours. We therefore believe our analysis provides useful insight into the transition to adulthood in young rural Malawians.

7. Conclusion

The results suggest that for many young people in the study the transition to adulthood is quite traditional: they leave school to marry and rapidly have children, though there are some differences in the timing of this transition (earlier or later in adolescence/young adulthood). Some adolescents (mostly men) do not follow this standard pathway: It requires qualitative work to understand whether these individuals form a potentially vulnerable group, or if they are making strategic decisions in this context – or both. Though we are only able to investigate change over a relatively short calendar time period, we do see some evidence of increasing access to secondary education for men and women, and delayed marriage for some women. This suggests that in this context young people are experiencing a rapidly changing environment, and that transitions to adulthood might be becoming more heterogenous.

References

Amoo, E.O. (2017). Trends and determinants of female age at first marriage in sub-Saharan Africa (1990–2014): What has changed? *Etude de la Population Africaine* 31(1): 3565–3577. doi:10.11564/31-1-1024.

Andriano, L. and Behrman, J. (2020). The effects of growing-season drought on young women's life course transitions in a sub-Saharan context. *Population Studies* 74(3): 331–350. doi:10.1080/00324728.2020.1819551.

Ansell, N., Froerer, P., Huijsman, R., Dungey, C., Dost, A., and Piti (2020). Educating 'surplus population': Uses and abuses of aspiration in the rural peripheries of a globalising world. *Fennia* 198(1–2): 17–38. doi:10.11143/fennia.90756.

Ansell, N., Hajdu, F., van Blerk, L., and Robson, E. (2018). 'My happiest time' or 'my saddest time'? The spatial and generational construction of marriage among youth in rural Malawi and Lesotho. *Transactions of the Institute of British Geographers* 43(2): 184–199. doi:10.1111/tran.12211.

Axxe, E., Hayford, S.R., and Eggum, N.D. (2022). Youth's family and non-family roles as predictors of subjective adulthood in three low-income agricultural settings. *Journal of Research on Adolescence* 32(4): 1546–1565. doi:10.1111/jora.12731.

Baraka, J., Lawson, D.W., Schaffnit, S.B., Wamoyi, J., and Urassa, M. (2022). Why marry early? Parental influence, agency and gendered conflict in Tanzanian marriages. *Evolutionary Human Sciences* 4: e49. doi:10.1017/ehs.2022.46.

Baruwa, O.J., Amoateng, A.Y., and Biney, E. (2019). Socio-demographic changes in age at first marriage in Malawi: Evidence from Malawi Demographic and Health Survey data, 1992–2016. *Journal of Biosocial Science* 52(6): 832–845. doi:10.1017/S0021932019000816.

Beegle, K. and Poulin, M. (2013). Migration and the transition to adulthood in contemporary Malawi. *The ANNALS of the American Academy of Political and Social Science* 648(1): 38–51. doi:10.1177/0002716213481329.

Beguy, D., Kabiru, C.W., Zulu, E.M., and Ezeh, A.C. (2011). Timing and sequencing of events marking the transition to adulthood in two informal settlements in Nairobi, Kenya. *Journal of Urban Health: Bbulletin of the New York Academy of Medicine* 88 Suppl 2(Suppl 2): S318–S340. doi:10.1007/s11524-011-9547-8.

Beguy, D., Ndugwa, R., and Kabiru, C.W. (2013). Entry into motherhood among adolescent girls in two informal settlements in Nairobi, Kenya. *Journal of Biosocial Science* 45(6): 721–742. doi:10.1017/S0021932013000199.

Belachew, T.B., Negash, W.D., Kefale, G.T., Tafere, T.Z., and Asmamaw, D.B. (2022). Determinants of early marriage among married women in nine high fertility sub-Saharan African countries: a multilevel analysis of recent demographic and health surveys. *BMC Public Health* 22(1): 1–8. doi:10.1186/s12889-022-14840-z.

Bennett, R. and Waterhouse, P. (2018). Work and family transitions and the self-rated health of young women in South Africa. *Social Science and Medicine* 203: 9–18. doi:10.1016/j.socscimed.2018.03.001.

Bertrand-Dansereau, A. and Clark, S. (2016). Pragmatic tradition or romantic aspiration? The causes of impulsive marriage and early divorce among women in rural Malawi. *Demographic Research* 35(3): 47–80. doi:10.4054/DemRes.2016.35.3.

Biddlecom, A. and Bakilana, A. (2003). Transitions into sex, parenthood and unions among adolescents and young adults in South Africa. Bethesda, MD: National Institute of Child Health and Human Development. https://www.worldcat.org/title/transitions-into-sex-parenthood-and-unions-among-adolescents-and-young-adults-in-south-africa/oclc/870110249.

Brzinsky-Fay, C. (2014). Graphical representation of transitions and sequences. In: Blanchard, P., Bühlmann, F., and Gauthier, J.A. (eds). *Advances in sequence analysis: Theory, method, applications. (Life course research and social policies, vol 2)*. Cham: Springer Science and Business Media: 265–284. doi:10.1007/978-3-319-04969-4_14.

Chae, S., Hayford, S.R., and Agadjanian, V. (2016). father's migration and leaving the parental home in rural Mozambique. *Journal of Marriage and the Family* 78(4): 1047–1062. doi:10.1111/jomf.12295.

Chintsanya, J., Magadi, M., and Likupe, G. (2021). A multilevel analysis of risk and protective factors for adolescent childbearing in Malawi. *Social Sciences* 10(8): 303. doi:10.3390/socsci10080303.

Clark, S. and Brauner-Otto, S. (2015). Divorce in sub-Saharan Africa: Are unions becoming less stable? *Population and Development Review* 41(4): 583–605. doi:10.1111/j.1728-4457.2015.00086.x.

Coall, D.A., Tickner, M., McAllister, L.S., and Sheppard, P. (2016). Developmental influences on fertility decisions by women: An evolutionary perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371(1692). doi:10.1098/rstb.2015.0146.

Crampin, A.C., Dube, A., Mboma, S., Price, A., Chihana, M., Jahn, A., Baschieri, A., Molesworth, A., Mwaiyeghele, E., Branson, K., Floyd, S., McGrath, N., Fine, P.E.M., French, N., Glynn, J.R., and Zaba, B. (2012). Profile: the Karonga Health and Demographic Surveillance System. *International Journal of Epidemiology* 41(3): 676–685. doi:10.1093/ije/dys088.

Dake, F., Natali, L., Angeles, G., de Hoop, J., Handa, S., and Peterman, A. (2018). Cash Transfers, Early Marriage, and Fertility in Malawi and Zambia. *Studies in Family Planning* 49(4): 295–317. doi:10.1111/sifp.12073.

Daniel, A. (2017). Malawi amends constitution to remove child marriage loophole | Human Rights Watch. https://www.hrw.org/news/2017/02/23/malawi-amends-constitution-remove-child-marriage-loophole.

Day, F.R., Elks, C.E., Murray, A., Ong, K.K., and Perry, J.R.B. (2015). Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: The UK Biobank study. *Scientific Reports* 5(1): 11208. doi:10.1038/srep11208.

Delprato, M., Akyeampong, K., Sabates, R., and Hernandez-Fernandez, J. (2015). On the impact of early marriage on schooling outcomes in Sub-Saharan Africa and South West Asia. *International Journal of Educational Development* 44: 42–55. doi:10.1016/j.ijedudev.2015.06.001.

Del Fava, E., Piccarreta, R., Gregson, S., and Melegaro, A. (2016). Transition to parenthood and HIV infection in rural Zimbabwe. *PLOS ONE* 11(9): e0163730. doi:10.1371/journal.pone.0163730.

Floyd, S., Molesworth, A., Dube, A., Crampin, A.C., Houben, R., Chihana, M., Price, A., Kayuni, N., Saul, J., French, N., and Glynn, J.R. (2013). Underestimation of HIV prevalence in surveys when some people already know their status, and ways to reduce the bias. *AIDS* 27(2): 233–242. doi:10.1097/QAD.0b013e32835848ab.

Gabadinho, A. and Ritschard, G. (2016). Analyzing state sequences with probabilistic suffix trees: The PST R package. *Journal of Statistical Software* 72: 1–39. doi:10.18637/jss.v072.i03.

Gabadinho, A., Ritschard, G., Muller, N., and Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software* 40(4): 1–37. doi:10.18637/jss.v040.i04.

Glynn, J.R., Sunny, B.S., DeStavola, B., Dube, A., Chihana, M., Price, A.J., and Crampin, A.C. (2018). Early school failure predicts teenage pregnancy and marriage: A large population-based cohort study in northern Malawi. *PLOS ONE* 13(5): e0196041. doi:10.1371/journal.pone.0196041.

Goldberg, R.E. (2013). Family instability and pathways to adulthood in urban South Africa. *Population and Development Review* 39(2): 231–256. doi:10.1111/j.1728-4457.2013.00590.x.

Grant, M.J. and Pike, I. (2019). Divorce, living arrangements, and material well-being during the transition to adulthood in rural Malawi. *Population Studies* 73(2): 261–275. doi:10.1080/00324728.2018.1545919.

Hajdu, F., Blerk, L. van, Ansell, N., Hemsteede, R., Mwathunga, E., Hlabana, T., and Robson, E. (2024). Livelihood trajectories of rural young people in southern Africa: Stuck in loops? *Development and Change* 55(3): 465–492. doi:10.1111/dech.12826.

Halpin, B. (2016). Multiple imputation for categorical time series. *Stata Journal* 16(3): 590–612. doi:10.1177/1536867X1601600303.

John, B.M. (2022). Marital life courses in sub-Saharan Africa: All cause union dissolution, its timing, and time spent outside marriage. (MPIDR Working Paper WP-2022-017). Rostock: MPIDR. doi:10.4054/MPIDR-WP-2022-017.

Juárez, F. and Gayet, C. (2014). Transitions to adulthood in developing countries. *Annual Review of Sociology* 40(1): 521–538. doi:10.1146/annurev-soc-052914-085540.

Kaufman, L. and Rousseeuw, P.J. (1990). *Finding groups in data: An introduction to cluster analysis*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley. doi:10.1002/9780470316801.

Kidman, R., Breton, E., Behrman, J., and Kohler, H.-P. (2022). A prospective study on adverse childhood experiences and HIV-related risk among adolescents in Malawi. *AIDS* 36(15): 2181–2189. doi:10.1097/QAD.0000000000003377.

Kidman, R., Breton, E., Mwera, J., Zulu, A., Behrman, J., and Kohler, H.-P. (2024). Drivers of child marriages for girls: A prospective study in a low-income African setting. *Global Public Health* 19(1). doi:10.1080/17441692.2024.2335356.

Kok, M.C., van Eldik, Z., Kakal, T., Munthali, A., Menon, J.A., Pires, P., Baatsen, P., and van der Kwaak, A. (2021). Being dragged into adulthood? Young people's agency concerning sex, relationships and marriage in Malawi, Mozambique and

Zambia. *Culture, Health and Sexuality* 24(6): 1–26. doi:10.1080/13691058.2021.1881618.

Kreniske, P., Grilo, S., Nakyanjo, N., Nalugoda, F., Wolfe, J., and Santelli, J.S. (2019). Narrating the transition to adulthood for youth in Uganda: Leaving school, mobility, risky occupations, and HIV. *Health Education and Behavior* 46(4): 550–558. doi:10.1177/1090198119829197.

Lambon-Quayefio, M., Peterman, A., Handa, S., Molotsky, A., Otchere, F., Mvula, P., Tsoka, M., de Hoop, J., Angeles, G., Kilburn, K., and Milazzo, A. (2024). Unconditional cash transfers and safe transitions to adulthood in Malawi. *World Development* 175: 106483. doi:10.1016/j.worlddev.2023.106483.

Liao, T.F., Bolano, D., Brzinsky-Fay, C., Cornwell, B., Fasang, A.E., Helske, S., Piccarreta, R., Raab, M., Ritschard, G., Struffolino, E., and Studer, M. (2022). Sequence analysis: Its past, present, and future. *Social Science Research* 107: 102772. doi:10.1016/j.ssresearch.2022.102772.

Lungu, E.A., Biesma, R., Chirwa, M., and Darker, C. (2019). Is the urban child health advantage declining in Malawi?: Evidence from Demographic and Health Surveys and Multiple Indicator Cluster Surveys. *Journal of Urban Health* 96(1): 131–143. doi:10.1007/s11524-018-0270-6.

Lutz, W., Cuaresma, J.C., Kebede, E., Prskawetz, A., Sanderson, W.C., and Striessnig, E. (2019). Education rather than age structure brings demographic dividend. *Proceedings of the National Academy of Sciences of the United States of America* 116(26): 12798–12803. doi:10.1073/pnas.1820362116.

Maechler, M., Rousseeuw, P.J., Struyf, A., Hubert, M., Hornik, K., Studer, M., Roudier, P., Gonzalez, J., Kozlowski, K., Schubert, E., and Murphy, K. (2023). Package 'cluster'. https://cran.r-project.org/web/packages/cluster/cluster.pdf.

McLean, E., Dube, A., Kalobekamo, F., Slaymaker, E., Crampin, A.C., and Sear, R. (2023). Local and long-distance migration among young people in rural Malawi: Importance of age, sex and family. *Wellcome Open Research* 8: 211. doi:10.12688/wellcomeopenres.19309.1.

McLean, E., Price, A., Chihana, M., Kayuni, N., Marston, M., Koole, O., Zaba, B., and Crampin, A. (2017). Changes in fertility at the population level in the era of ART in Rural Malawi. *Journal of Acquired Immune Deficiency Syndromes* 75(4): 391–398. doi:10.1097/QAI.0000000000001395.

Melnikas, A.J., Mulauzi, N., Mkandawire, J., Saul, G., and Amin, S. (2022). 'Then I will say that we have to marry each other': A qualitative view of premarital pregnancy as a driver of child marriage in Malawi. *African Journal of Reproductive Health* 26(9): 55–63. https://www.ajol.info/index.php/ajrh/article/view/237546.

Mensch, B.S., Soler-Hampejsek, E., Kelly, C.A., Hewett, P.C., and Grant, M.J. (2014). Challenges in measuring the sequencing of life events among adolescents in Malawi: A cautionary note. *Demography* 51(1): 277–285. doi:10.1007/s13524-013-0269-2.

Moyi, P. (2010). Household characteristics and delayed school enrollment in Malawi. *International Journal of Educational Development* 30(3): 236–242. doi:10.1016/j.ijedudev.2009.11.008.

Mutangdura, G. (2004). Women and land tenure rights in southern Africa: A human rights-based approach. In: Quan, J., Tan, S.F., and Toulmin, C. (eds.). *Land in Africa: Market asset or secure livelihood. Proceedings and summary of conclusions from the Land in Africa Conference held in London November 8–9, 2004*. London. https://pubs.iied.org/pdfs/G00173.pdf.

Nahar, Q., Xenos, P., and Abalos, J. (2013). The changing transitions to adulthood across Southeast Asia: A census approach to cross-national comparisons. *Annals of the American Academy of Political and Social Science* 646(1): 42–68. doi:10.1177/0002716212469921.

Naphambo, E.K. (2022). A vexing relationship between chiefship and girls' sexuality: Insights from rural Malawi. *Journal of Adolescent Health* 70(3): S36–S42. doi:10.1016/j.jadohealth.2021.12.005.

National Statistical Office (2016). National Statistical Office, Zomba. Malawi Demographic and Health Survey 2015–16 Key Indicators Report. Zomba: NSO.

Odimegwu, C. and Mkwananzi, S. (2016). Factors associated with teen pregnancy in sub-Saharan Africa: A multi-country cross-sectional study. *African Journal of Reproductive Health* 20(3): 94–107. doi:10.29063/ajrh2016/v20i3.14.

Oris, M. and Ritschard, G. (2014). Sequence analysis and transition to adulthood: An exploration of the access to reproduction in nineteenth-century east Belgium. In: Blanchard, P., Bühlmann, F., and Gauthier, J.A. (eds). *Advances in sequence analysis: Theory, method, applications. (Life course research and social policies, vol 2)*. Cham: Springer Science and Business Media: 151–167. doi:10.1007/978-3-319-04969-4_8.

Pesando, L.M., Barban, N., Sironi, M., and Furstenberg, F.F. (2021). A sequence-analysis approach to the study of the transition to adulthood in low- and middle-income countries. *Population and Development Review* 47(3): 719–747. doi:10.1111/padr.12425.

Pike, I. (2021). Men's economic status and marriage timing in rural and semi-urban Malawi. *Social Forces* 100(1): 113–136. doi:10.1093/sf/soaa092.

Pollock, G. (2007). Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis. *Statistics in Society* 170(1): 167–183. doi:10.1111/j.1467-985X.2006.00450.x.

Pot, H. (2019). Cross-cultural studies in health and illness INGO behavior change projects: Culturalism and teenage pregnancies in Malawi. *Medical Anthropology* 38(4): 327–341. doi:10.1080/01459740.2019.1570187.

Poulin, M., Beegle, K., and Xu, H. (2021). Premarital fertility and marital timing in Malawi. *Studies in Family Planning* 52(2): 195–216. doi:10.1111/sifp.12158.

Price, A.J., Glynn, J., Chihana, M., Kayuni, N., Floyd, S., Slaymaker, E., Reniers, G., Zaba, B., McLean, E., Kalobekamo, F., Koole, O., Nyirenda, M., and Crampin, A.C. (2017). Sustained 10-year gain in adult life expectancy following antiretroviral therapy roll-out in rural Malawi: July 2005 to June 2014. *International Journal of Epidemiology* 46(2): 479–491. doi:10.1093/ije/dyw208.

Reda, A.A. and Lindstrom, D. (2014). Recent trends in the timing of first sex and marriage among young women in Ethiopia. *Etude de la Population Africaine* 28(2): 1157–1170. doi:10.11564/28-0-564.

Rock, A., Barrington, C., Abdoulayi, S., Tsoka, M., Mvula, P., and Handa, S. (2016). Social networks, social participation, and health among youth living in extreme poverty in rural Malawi. *Social Science and Medicine* 170: 55–62. doi:10.1016/j.socscimed.2016.10.005.

Sauvain-Dugerdil, C., Gakou, B., Berthé, F., Dieng, A.W., Ritschard, G., and Lerch, M. (2008). The start of the sexual transition in Mali: Risks and opportunities. *Studies in Family Planning* 39(4): 263–280. doi:10.1111/j.1728-4465.2008.00174.x.

Schwanitz, K. (2017). The transition to adulthood and pathways out of the parental home: A cross-national analysis. *Advances in Life Course Research* 32: 21–34. doi:10.1016/j.alcr.2017.03.001.

Shapiro, D. and Gebreselassie, T. (2014). Marriage in sub-Saharan Africa: Trends, determinants, and consequences. *Population Research and Policy Review* 33(2): 229–255. doi:10.1007/s11113-013-9287-4.

Studer, M. (2013). WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. (LIVES Working paper 24). Lausanne: LIVES. doi:10.12682/lives.2296-1658.2013.24.

Studer, M. and Ritschard, G. (2016). What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society Series A: Statistics in Society* 179(2): 481–511. doi:10.1111/rssa.12125.

Sunny, B.S., Elze, M., Chihana, M., Gondwe, L., Crampin, A.C., Munkhondya, M., Kondowe, S., and Glynn, J.R. (2017). Failing to progress or progressing to fail? Age-for-grade heterogeneity and grade repetition in primary schools in Karonga district, northern Malawi. *International Journal of Educational Development* 52: 68–80. doi:10.1016/j.ijedudev.2016.10.004.

Vail, L. (1989). *The creation of tribalism in southern Africa*. Berkeley: Currey University of California Press. http://ark.cdlib.org/ark:/13030/ft158004rs/.

Yakubu, I. and Salisu, W.J. (2018). Determinants of adolescent pregnancy in sub-Saharan Africa: A systematic review. *Reproductive Health* 15(1): 15. doi:10.1186/s12978-018-0460-4.

Zahra, F. (2020). High hopes, low dropout: Gender differences in aspirations for education and marriage, and educational outcomes in rural malawi. *Comparative Education Review* 64(4): 670–702. doi:10.1086/710778.

Zietz, S., de Hoop, J., and Handa, S. (2018). The role of productive activities in the lives of adolescents: Photovoice evidence from Malawi. *Children and Youth Services Review* 86: 246–255. doi:10.1016/j.childyouth.2018.01.031.

Appendix 1: Assessment of representativeness of complete sequences

A-1.1 Rationale

HDSSs are open cohorts and data are not collected on participants when they are outside the area. This means that any analysis that uses longitudinally linked data from individuals will need to exclude participants that move in and out of the area. This reduces sample size and also potentially introduces bias if those who are excluded differ from those who are not. To assess how representative the relatively complete transition-to-adulthood sequences of those who remain in the area are of the whole population, here we compare two datasets, one which includes only participants with relatively complete sequences (the 'complete sequences' dataset) and one which consists of these complete sequences plus imputed data for those with incomplete data (the 'imputed' dataset), where missing data is imputed from those with complete data. We impute missing data for the marital status variable only. Marital status was chosen for this assessment as initial explorations of the data, and other literature, have shown that marriage is nearly universal among young people in Malawi, and also because it was considered the simplest to attempt predictive multiple imputation. We recognise this is a relatively crude comparison, as it involves assuming that those with incomplete data (i.e., those who leave the area) look similar to those who stay; however, we still consider this an informative analysis in the absence of data on those who leave.

A-1.2 Summary

Three sequence lengths were assessed (9, 7, and 5 years), all of which include the average age at first marriage, as observed in the data (21 years for men and 17 for women). Datasets were created to include participants who were present in the data for at least 1 quarter at the start age, and were born early enough to have been able to complete the full sequence by the end the whole data collection period (2017). The subsets of each dataset with the most complete sequences (those missing up to 4 quarters of follow-up were classed as 'complete') were identified as the 'complete sequences' datasets. These datasets were used to develop predictive models of age at first marriage and age at divorce: the predictions were used to assign ages of marriage and divorce to the remaining dataset when they were missing due to out-migration. The imputed ages were randomly assigned within the 95% confidence interval of the predicted age, and 50 multiply imputed datasets were created. A categorical variable was created using the real and imputed ages (plus estimates of ages of marriage and divorce from data on reasons for migration), grouping individuals into five categories related to whether marriage and/or

divorce were experienced, and the timing of these events. The proportions in each category were compared for the 'complete sequences' and 'imputed' datasets, overall and by time period, separately by sex.

As expected, the shorter the sequence length, the more similar the 'complete sequences' group are to the 'imputed' dataset. However, only including 5 years of data limits the conclusions that can be drawn regarding the dataset. While the longest sequence length (9 years) would provide a more comprehensive analysis of the transitions to adulthood, the 'complete sequences' sample size is small and the differences between the 'complete sequence' and 'imputed' datasets are large. The 7-year sequence allows for a more comprehensive assessment of the transition to adulthood and the 'complete sequence' dataset is relatively representative of the 'imputed' dataset. However, results of this analysis suggest this dataset may underestimate the proportion who do not marry for both men and women (since these proportions were lower in the 'complete sequence' than the 'imputed' dataset), and overestimate the proportion who marry in the mid period for men, and the early period for women (since these proportions were higher in the 'complete sequence' than in the 'imputed').

A-1.3 Methodology

Three sequence lengths were assessed, and 3 groups of participants identified from the dataset:

a) 9 years (36 quarters)
    1. Female period covers ages 14–23 (included if born between 1990 and 1994 and present in HDSS for at least 1 quarter aged 14)
    2. Male period covers ages 17–26 (included if born between 1987 and 1991 and present in HDSS for at least 1 quarter aged 17)

b) 7 years (28 quarters)
    1. Female period covers ages 15–22 (included if born between 1989 and 1995 and present in HDSS for at least 1 quarter aged 15)
    2. Male period covers ages 18–25 (included if born between 1986 and 1992 and present in HDSS for at least 1 quarter aged 18)

c) 5 years (20 quarters)
    1. Female period covers ages 16–21 (included if born between 1988 and 1996 and present in HDSS for at least 1 quarter aged 16)
    2. Male period covers ages 19–24 (included if born between 1985 and 1993 and present in HDSS for at least 1 quarter aged 19)

From the above 3 datasets, participants with complete sequences (include those with up to 4 missing quarters) were identified (designated the 'complete sequence' sample or 'C'). This sample was used to generate several models predicting age at first marriage in groups of never married participants at all whole ages within the analysis period (separately by sex), using linear regression with the following covariates: calendar year (in categories), school status, relationship to household head, household size (in categories), tarmac road within 1km (binary), household head employment score, population density within 100m (in categories). If actual marriage age was missing for those who did not marry during the period under analysis, it was imputed as the average age in this group among people who did have data on age at marriage. Predictive models were used to assign an imputed age of first marriage to the rest of the dataset based on the information from their last available record (i.e., age, and all variables listed in the model above). This imputed age at first marriage was supplemented with age at migration if the participant was reported to have moved independently and had 'marriage' listed as the reason for the move (see McLean et al. 2023 for details on how 'independent' move was defined).

The 'complete sequence' sample was also used to generate models predicting age at divorce for married participants using linear regression with the following covariates: sex, age at marriage, and year of marriage. If actual divorce age (and year) was missing for those who did not divorce during the period under analysis, it was imputed as the average age in this group among people who did have the data. The predictive model was used to assign imputed age of divorce to the rest of the dataset based on the sex and real or imputed age and year of marriage. This imputed age at divorce was supplemented with age at migration if the participant reported having moved independently and had 'divorce' listed as the reason for the move. Other factors could not be used in the divorce-age model as those given an imputed marriage age in the full dataset would not have this information. It is also recognised that assigning a divorce age to everyone in the dataset is inaccurate; however, the majority of these assigned ages will be after the period under analysis, so should not affect the assessment.

For the imputation of marriage and divorce age, each participant was assigned an age randomly within the confidence interval of the predictive age and 50 imputed datasets were created with slightly different imputed ages for each person. For each imputed dataset a categorical variable using the real or imputed marriage and divorce dates was created with the following categories:

a) No/late marriage: marriage happened in last 25% of analysis period
b) Mid marriage (no divorce): marriage happened in middle 50% of analysis period, no periods of divorce

c) Mid marriage (divorce): marriage happened in middle 50% of analysis period, at least 1 period of divorce
d) Early marriage (no divorce): marriage happened in first 25% of analysis period, no periods of divorce
e) Early marriage (divorce): marriage happened in first 25% of analysis period, at least 1 period of divorce

A comparison was made of the proportions in each category in the 'complete sequence' dataset versus the 'imputed' dataset. These proportions were calculated both for the entire dataset and separately for each calendar year, to assess any trends over time in differences between the 'complete sequence' and the 'imputed' datasets. The 3 sequence lengths were compared, separately by sex, to assess how representative the 'complete sequence' sample was compared to the 'imputed' dataset.

A-1.4 Results

The 9-year sequence dataset had 1,597 (810 'complete sequence') men and 2,059 (733 'complete sequence') women, the 7-year dataset had 2,256 (1,325 'complete sequence') men and 2,822 (1,326 'complete sequence') women, and the 5-year dataset had 2,888 (1,967 'complete sequence') men and 3,624 (2,106 'complete sequence') women.

For men, in the mid marriage (divorce), early marriage (no divorce), and early marriage (divorce) categories there was little difference in the proportions in the 'imputed' and 'complete sequence' datasets; however, these were the least common categories. Across all 3 sequence lengths the proportion in the no/late marriage group was lower in the 'complete sequence' dataset than in the total dataset, while the mid marriage (no divorce) group was higher in the 'complete sequence' dataset. The differences were smallest in the dataset with the shortest sequence (Figure A-1).

For women, there were differences in the proportions for all groups with the longest 9-year sequence, while for the 7- and 5-year sequences the proportions were similar for the mid marriage (no divorce), mid marriage (divorce), and early marriage (divorce) categories. The proportion of the no/late marriage group was lower for the 'complete sequence' group, and higher for the early marriage (no divorce) group. These differences were smaller for the shortest sequences; however, for both the differences were smaller than those for men (Figure A-2).

Figure A-1: Proportions in each category in the 'complete sequence' and 'imputed' datasets for men for the 3 sequence lengths

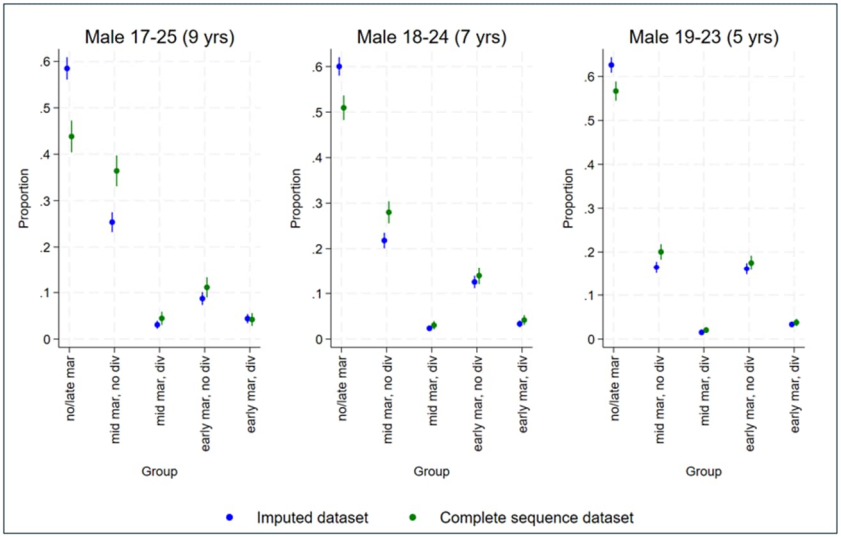
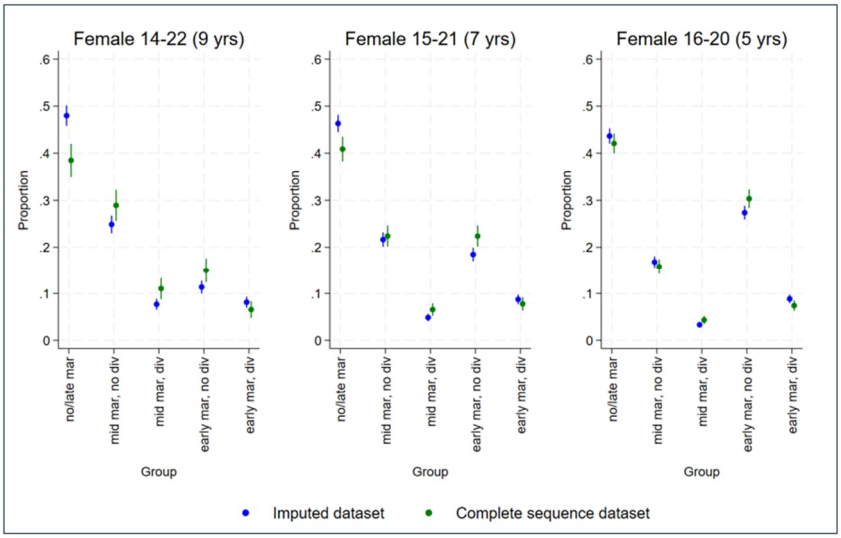
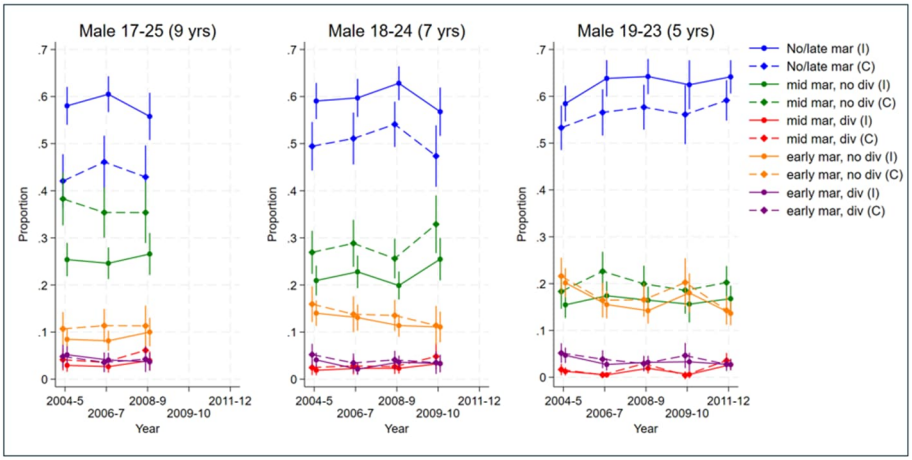


Figure A-2: Proportions in each category in the 'complete sequence' and 'imputed' datasets for women for the 3 sequence lengths



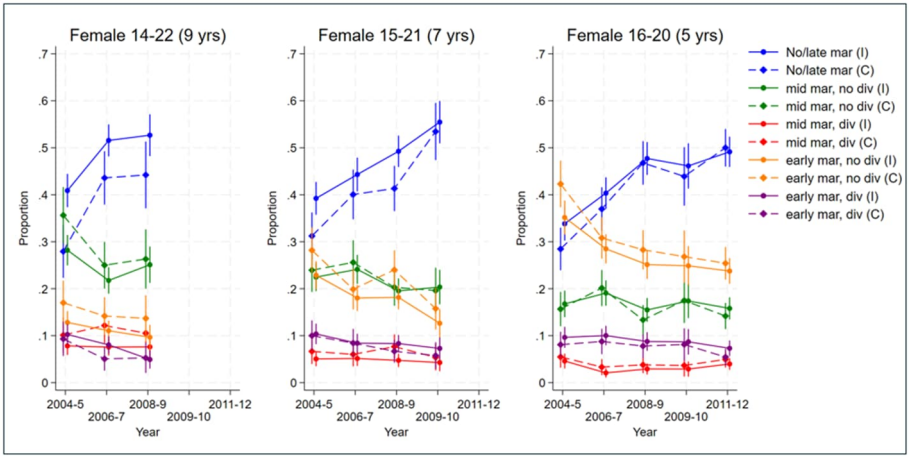
For men, the trends in the proportions in each category over time were similar between the 'complete sequence' and 'imputed' datasets for all 3 sequence lengths (though with the lines lower or higher, as explained above). While the lines are closer together for the shortest sequence length, for the middle sequence length there is overlap of the confidence intervals for most points (Figure A-3).

Figure A-3: Proportions in each category in 2-year bands in the 'imputed' (I) and 'complete sequence' (C) datasets for men for the 3 sequence lengths



For women, the trends are also similar, though there are more differences compared to men; e.g., the increase in the no/late marriage group is steeper in the 'complete sequence' dataset than the 'imputed'. As with the men above, the lines are closer together for the shortest sequence length, but for the middle sequence length, for most points the confidence intervals overlap (Figure A-4).

Figure A-4: Proportions in each category in 2-year bands in the 'imputed' (I) and 'complete sequence' (C) datasets for women for the 3 sequence lengths



A-1.5 Conclusions

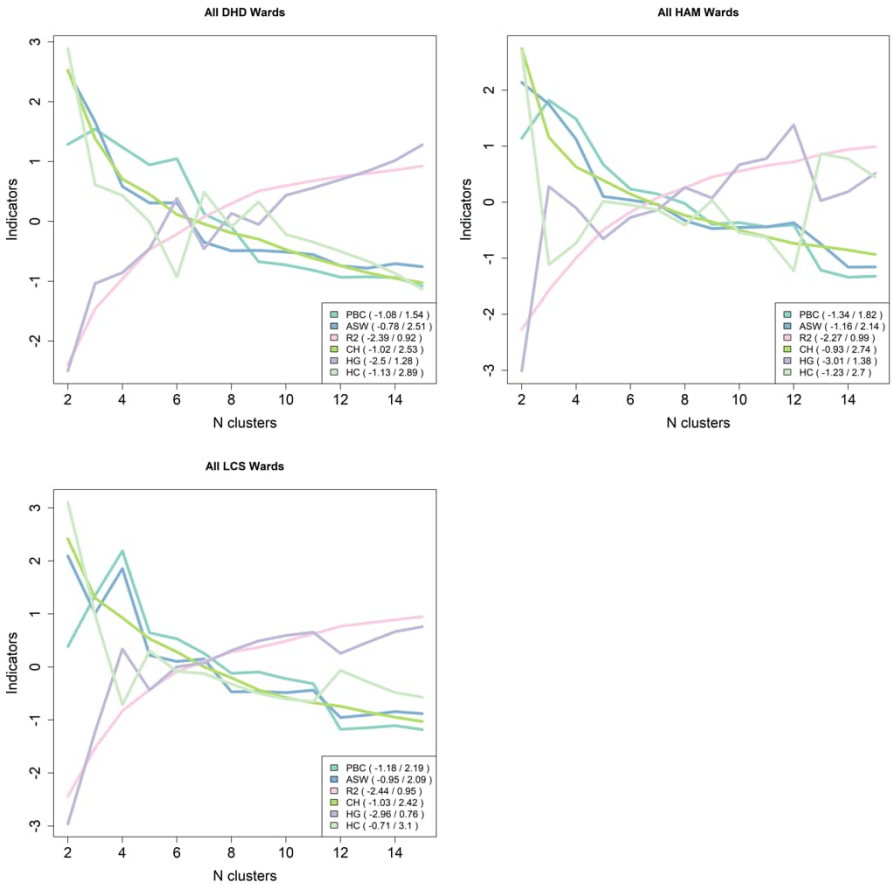
As expected, the shorter the sequence length, the more similar the 'complete sequence' group are to the 'imputed' dataset. However, only including 5 years of data limits the conclusions that can be made about the dataset. While the longest sequence length (9 years) provides a more comprehensive analysis of the transitions to adulthood, the 'complete sequence' sample size is small and the differences between the 'complete sequence' and 'imputed' datasets are large. Using the 7-year sequence length provides a balance between having the most representative sample and being able to study the transition to adulthood in a more useful time period. Given the pattern of differences between the 'complete sequence' and 'imputed' dataset, it needs to be borne in mind that the 'complete sequence' dataset is likely to underestimate the proportion who do not marry for both men and women, and overestimate the proportion who marry in the mid period for men, and the early period for women.

We recognise that the predictive models used to impute the age at marriage and divorce for those who leave the area are relatively crude, and also the imputation assumes that those who leave remain similar to those who remain. If those who leave are less likely to marry, then the underestimation of the no/late marriage group would increase. However, our previous analysis of the data on young people's migration found no evidence that there was a trend of young people moving to cities for work or education,

which would likely delay marriage (McLean et al. 2023), so the assumption that those that leave are similar to those who stay may be reasonable.

Appendix 2: Additional figures and tables

Figure A-5: Standardised cluster statistics using Wards algorithm



Note: Cluster statistics where maximum value is preferable: PBC (Point Biserial Correlation): Measure of the capacity of the clustering to reproduce the distances; ASW (Average Silhouette Width): Coherence of assignments. High coherence indicates high between-group distances and strong within-group homogeneity; R2 (Pseudo R2): Share of the discrepancy explained by the clustering solution (only to compare partitions with identical number of groups); CH (Calinski-Harabasz index): Pseudo F computed from the distances; HG (Hubert's Gamma): Measure of the capacity of the clustering to reproduce the distances (order of magnitude); Cluster statistics where minimum value is preferable: HC (Hubert's C): Gap between the partition obtained and the best partition theoretically possible with this number of groups and these distances.

Table A-1: Cluster statistics for 3 dissimilarity matrix methods and 2 clustering algorithms

	PBC	ASW	R2	CH	HG	HC
Dynamic Hamming Distance (DHD): Wards algorithm						
3	0.65	0.38	0.38	816.76	0.80	0.09
4	0.63	0.32	0.43	677.70	0.81	0.09
5	0.61	0.30	0.49	625.99	0.82	0.08
6	0.61	0.30	0.51	556.94	0.85	0.07
7	0.54	0.27	0.54	524.52	0.82	0.09
8	0.52	0.26	0.57	494.09	0.84	0.08
9	0.48	0.26	0.59	472.84	0.84	0.09
10	0.47	0.26	0.60	436.84	0.85	0.08
Hamming Distance (HAM): Wards algorithm						
3	0.71	0.40	0.36	740.08	0.86	0.07
4	0.68	0.35	0.42	629.85	0.85	0.07
5	0.61	0.28	0.47	579.78	0.83	0.08
6	0.57	0.28	0.50	528.60	0.84	0.08
7	0.56	0.27	0.53	487.87	0.85	0.08
8	0.55	0.25	0.54	448.82	0.86	0.08
9	0.52	0.24	0.56	425.38	0.85	0.08
10	0.52	0.25	0.57	394.01	0.87	0.07
Longest Common Sequence (LCS): Wards algorithm						
3	0.60	0.33	0.37	767.50	0.75	0.12
4	0.66	0.37	0.44	695.34	0.86	0.07
5	0.56	0.29	0.48	619.06	0.80	0.10
6	0.55	0.29	0.52	571.15	0.83	0.08
7	0.53	0.29	0.54	516.86	0.84	0.08
8	0.51	0.26	0.56	477.34	0.85	0.08
9	0.51	0.26	0.57	433.15	0.87	0.07
10	0.50	0.26	0.58	404.84	0.87	0.07
Dynamic Hamming Distance (DHD): Partition around medoids						
5	0.55	0.30	0.50	660.77	0.79	0.10
6	0.54	0.31	0.54	624.48	0.82	0.09
7	0.52	0.28	0.56	561.13	0.83	0.08
Hamming Distance (HAM): Partition around medoids						
5	0.55	0.29	0.48	621.87	0.79	0.10
6	0.54	0.30	0.53	586.06	0.82	0.09
7	0.52	0.27	0.54	525.61	0.83	0.09
Longest Common Sequence (LCS): Partition around medoids						
5	0.55	0.31	0.50	650.48	0.80	0.10
6	0.54	0.31	0.54	617.34	0.84	0.08
7	0.52	0.28	0.56	552.89	0.84	0.08

Table A-2: Average Silhouette Width for groups identified through the 3 dissimilarity matrix methods and 2 clustering algorithms

HAM		LCS		DHD	
Wards algorithm					
Early marriage (primary)	0.4	Early marriage (primary)	0.31	Early marriage (primary)	0.46
Early marriage (secondary)	0.11	Early marriage (secondary)	0.1	Early marriage (secondary)	0.27
Mid marriage (primary)	0.12	Mid marriage (primary)	0.24	Mid marriage (primary)	0.12
Mid marriage (secondary)	0.19	Mid marriage (secondary)	0.14	Mid marriage (secondary)	0.14
No/late marriage (primary)	0.61	No/late marriage (primary)	0.36	No/late marriage (primary)	0.64
No/late marriage (secondary)	0.3	No/late marriage (left secondary)	0.29	No/late marriage (secondary)	0.36
No/late marriage (late move to secondary)	0.31	No/late marriage (in secondary)	0.4		
Partition Around Medoids					
Early marriage (primary)	0.34	Early marriage (primary)	0.34	Early marriage (primary)	0.35
Mid marriage (primary)	0.24	Mid marriage (primary)	0.25	Mid marriage (primary)	0.24
Mid marriage (secondary)	0.2	Mid marriage (secondary)	0.22	Mid marriage (secondary)	0.2
No/late marriage (primary)	0.46	No/late marriage (primary)	0.48	No/late marriage (primary)	0.49
No/late marriage (left secondary)	0.34	No/late marriage (left secondary)	0.28	No/late marriage (left secondary)	0.34
No/late marriage (in secondary)	0.24	No/late marriage (in secondary)	0.34	No/late marriage (in secondary)	0.27

Table A-3: Full model results from multinomial multivariate logistic regression model with sequence group as outcome and including all listed variables as co-variates (b = binary, c = continuous), run separately for men and women

	Men						Women					
	RRR	95% CI		SD	Z	p-value	RRR	95% CI		SD	Z	p-value
Earlier marriage (primary)												
Parents' HH (b)	1.2	0.7	1.9	0.3	0.7	0.468	1.0	0.7	1.5	0.2	0.1	0.952
Fem HH head (b)	0.8	0.5	1.4	0.2	–0.7	0.487	1.2	0.7	2.0	0.3	0.7	0.466
HH size (c)	0.9	0.7	1.0	0.1	–2.1	0.033	0.9	0.8	1.1	0.1	–0.9	0.374
Father sec edu (b)	0.7	0.5	1.2	0.2	–1.2	0.222	1.5	1.0	2.2	0.3	1.8	0.069
Mother sec edu (b)	1.8	1.1	3.1	0.5	2.2	0.031	1.1	0.7	1.7	0.2	0.4	0.725
Dwelling (c)	1.1	1.0	1.3	0.1	1.4	0.170	0.9	0.7	1.0	0.1	–1.9	0.060
Water (b)	0.9	0.5	1.3	0.2	–0.7	0.494	0.8	0.5	1.2	0.2	–1.2	0.213
1km of road (b)	1.5	0.9	2.4	0.4	1.5	0.130	1.4	0.9	2.2	0.3	1.6	0.108
Pop density (c)	1.1	0.9	1.2	0.1	0.8	0.450	1.0	0.9	1.2	0.1	0.6	0.528
Mid marriage (primary) [baseline]												
Mid marriage (secondary)												
Parents' HH (b)	1.7	1.0	2.7	0.4	2.0	0.044	1.7	1.0	2.6	0.4	2.1	0.032
Fem HH head (b)	1.1	0.7	1.9	0.3	0.5	0.642	1.2	0.7	2.1	0.3	0.8	0.442
HH size (c)	0.9	0.8	1.0	0.1	–1.6	0.100	1.0	0.9	1.2	0.1	0.7	0.508
Father sec edu (b)	2.3	1.4	3.7	0.6	3.3	0.001	2.5	1.5	4.0	0.6	3.6	<0.001
Mother sec edu (b)	1.3	0.8	2.3	0.4	1.0	0.300	1.2	0.8	1.9	0.3	0.8	0.433
Dwelling (c)	1.4	1.2	1.6	0.1	3.6	<0.001	1.0	0.9	1.2	0.1	0.5	0.603
Water (b)	1.5	1.0	2.5	0.4	1.8	0.070	1.7	1.1	2.8	0.4	2.2	0.028
1km of road (b)	1.3	0.8	2.1	0.3	1.1	0.264	2.2	1.4	3.4	0.5	3.3	0.001
Pop density (c)	0.9	0.8	1.0	0.1	–1.6	0.101	1.0	0.8	1.1	0.1	–0.5	0.645

Table A-3: (Continued)

	Men						Women					
	RRR	95% CI		SD	Z	p-value	RRR	95% CI		SD	Z	p-value
Later/no marriage (primary)												
Parents' HH (b)	1.2	0.7	1.9	0.3	0.7	0.500	1.6	0.8	3.3	0.6	1.2	0.228
Fem HH head (b)	1.0	0.6	1.7	0.3	0.0	0.963	1.4	0.6	3.1	0.6	0.8	0.434
HH size (c)	0.9	0.8	1.0	0.1	–2.0	0.051	1.0	0.8	1.3	0.1	0.1	0.901
Father sec edu (b)	0.9	0.6	1.5	0.2	–0.3	0.772	1.0	0.5	1.9	0.3	0.0	0.971
Mother sec edu (b)	1.0	0.6	1.7	0.3	0.0	0.976	0.7	0.4	1.6	0.3	–0.8	0.448
Dwelling (c)	1.0	0.8	1.2	0.1	0.0	0.981	0.9	0.7	1.2	0.1	–0.7	0.454
Water (b)	1.3	0.8	2.0	0.3	1.2	0.232	1.2	0.6	2.5	0.4	0.6	0.575
1km of road (b)	1.9	1.2	3.0	0.5	2.7	0.007	1.7	0.9	3.5	0.6	1.5	0.121
Pop density (c)	1.0	0.8	1.1	0.1	–0.7	0.493	1.0	0.8	1.3	0.1	0.2	0.824
Later/no marriage (left secondary)												
Parents' HH (b)	1.4	0.9	2.1	0.3	1.5	0.147	2.0	1.2	3.3	0.5	2.7	0.006
Fem HH head (b)	1.0	0.6	1.6	0.2	0.0	0.967	1.1	0.6	2.0	0.3	0.4	0.703
HH size (c)	1.0	0.9	1.1	0.1	–0.3	0.737	0.9	0.8	1.1	0.1	–0.8	0.402
Father sec edu (b)	1.8	1.2	2.7	0.4	2.6	0.009	3.4	1.9	6.0	1.0	4.1	<0.001
Mother sec edu (b)	1.3	0.8	2.2	0.3	1.2	0.244	1.1	0.7	1.8	0.3	0.5	0.596
Dwelling (c)	1.4	1.2	1.6	0.1	4.1	<0.001	1.4	1.2	1.6	0.1	3.8	<0.001
Water (b)	2.3	1.5	3.5	0.5	3.7	<0.001	1.4	0.8	2.3	0.4	1.2	0.248
1km of road (b)	2.5	1.6	3.8	0.5	4.1	<0.001	2.7	1.6	4.5	0.7	3.9	<0.001
Pop density (c)	0.9	0.8	1.0	0.1	–1.7	0.086	0.9	0.8	1.1	0.1	–1.0	0.327
Later/no marriage (in secondary)												
Parents' HH (b)	1.1	0.7	1.7	0.3	0.3	0.782	1.9	1.3	2.8	0.4	3.1	0.002
Fem HH head (b)	1.2	0.7	2.1	0.3	0.8	0.412	1.6	1.0	2.5	0.4	2.0	0.049
HH size (c)	1.0	0.9	1.2	0.1	0.7	0.507	1.0	0.9	1.2	0.1	0.4	0.685
Father sec edu (b)	0.9	0.6	1.4	0.2	–0.6	0.566	1.6	1.1	2.3	0.3	2.2	0.027
Mother sec edu (b)	1.1	0.6	1.9	0.3	0.3	0.784	0.7	0.5	1.1	0.2	–1.5	0.146
Dwelling (c)	1.2	1.0	1.4	0.1	2.0	0.041	1.2	1.0	1.3	0.1	2.4	0.017
Water (b)	1.5	1.0	2.4	0.3	1.9	0.054	2.0	1.3	3.0	0.4	3.2	0.001
1km of road (b)	1.6	1.0	2.6	0.4	2.0	0.044	1.7	1.1	2.5	0.3	2.5	0.011
Pop density (c)	0.8	0.7	1.0	0.1	–2.6	0.010	0.9	0.8	1.1	0.1	–1.0	0.325