Received: 26 October 2020 | Accepted: 4 February 2022

DOI: 10.1111/rssa.12832

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY
Series A Statistics in Society

# Life-course perspective on personality traits and fertility with sequence analysis

**Letizia Mencarini**[1] | **Raffaella Piccarreta**[2] | **Marco Le Moglie**[3]

[1]Department of Social and Political Sciences & Dondena Centre for Research on Social Dynamics and Public Policy, Bocconi University, Milan, Italy

[2]Department of Decision Sciences & Dondena Centre for Research on Social Dynamics and Public Policy & BIDSA (Bocconi Institute for Data Science and Analytics), Bocconi University, Milan, Italy

[3]Department of Economics and Finance, Catholic University of the Sacred Heart, Milan, Italy

**Correspondence**
Raffaella Piccarreta, Department of Decision Sciences & Dondena Centre & BIDSA, Bocconi University, Milan, Italy.
Email: raffaella.piccarreta@unibocconi.it

**Abstract**

We investigate the link between personality traits (PTs) and fertility, accounting for the possible interplay with other key life course events. Using data from German Socio-Economic Panel survey, we build sequence-type representations of fertility, union and job careers between the ages of 20 and 40. We rely on multichannel sequence analysis (MSA) and on the Partitioning around Medoids algorithm to cluster individuals with similar experiences, and relate clusters to PTs via multinomial regression. We also develop a procedure to apply standard and MSA to truncated trajectories. This enables inclusion of individuals whose trajectories were otherwise observed for a limited age span, notably belonging to younger cohorts. We show that PTs relate to these (portions of) life-course trajectories, of which fertility is only one outcome.

**KEYWORDS**

fertility career, job career, multichannel sequence analysis, multinomial logistic regression, personality traits, right-censored (truncated) trajectories, SOEP, union career

## 1 | INTRODUCTION

There is an emerging literature showing that individuals' personality traits (PTs) are linked to fertility behaviour (e.g. Jokela, 2012; Skirbekk & Blekesaune, 2013; Tavares, 2016). There is also a literature linking personality to both union formation and union stability (e.g. Hewitt et al., 2006;

Lundberg, 2012); and a number of studies showing important connections between personality and job career and professional success (e.g. Judge et al., 2006; Seibert et al., 2001; Whyte et al., 2019). Both union formation and vocational choices are often considered antecedents of fertility, because individuals typically make their fertility planning conditional on job and union dynamics. For most people these processes evolve together, and decisions in one domain are connected with events in the others. An inherent weakness of earlier studies on the role of PTs for life course events is that they analysed these processes separately, possibly focusing on specific events. We tackle this issue by studying the relation between PTs and life biographies more generally, as described by the fertility behaviour, job activities and union dynamics.

We contribute to the existing literature in two respects. First, from a substantive point of view, we identify the most typical combinations of trajectories in the three domains and relate them to PTs. To this end, we use data from the German Socio-Economic Panel survey (SOEP), a representative ongoing longitudinal study started in 1984, including all waves up to 2019. We build sequence-type representations of the individuals' partnership, childbearing and work based on the events and activities undertaken between the ages of 20 and 40. We apply multichannel sequence analysis (MSA; Pollock, 2007) to assess the similarity between cases based on the joint evolution of their trajectories in the three domains, and use cluster analysis to identify groups of individuals accordingly. We subsequently relate cluster membership to PTs via multinomial logistic regression, controlling for a set of background socio-demographic covariates. Since 2005, SOEP collects every four years the classical 'Big Five' personality inventory, broadly accepted as a consistent and reliable categorization of peoples' psychological attributes (Goldberg, 1981). We refer to the PTs scores measured at each individual's first occasion (i.e. the occasion closest to the age of 20), including also the age at such occasion among the control variables in regressions.

Our second contribution is methodological. We introduce a novel approach to measure dissimilarity between trajectories of different lengths. This allows including in our analysis also individuals whose trajectories were observed only for a limited age span, either because of censoring or, more importantly, because they were younger than 40 at the last wave. This is particularly convenient for the data at hand because only a relatively low number of people were observed for the entire period of interest (i.e. between the ages of 20 and 40). Our proposal addresses a long-standing issue in standard sequence analysis (SA)—typically based on sequences of the same length—which would take into account only complete trajectories, thus systematically excluding individuals from the most recent cohorts (Piccarreta & Studer, 2019).

The rest of the paper is organised as follows. In the next section, we review the most relevant findings in the literature concerning the relation between PTs and fertility, union formation, and job career. Section 3 describes the data. Section 4 illustrates and contrasts approaches to study the joint evolution of trajectories, even if possibly truncated. Section 5 shows the results of a multinomial regression relating PTs to individual life-course trajectories. Conclusions follow.

## 2 | PERSONALITY TRAITS AND FERTILITY, UNION FORMATION AND JOB CAREERS: A LITERATURE REVIEW

Personality inventories aim at describing stable differences in individual psychological dispositions. There are many alternative taxonomies, but the 'Big Five' self-reported personality inventory (BFI) is broadly accepted as a consistent and reliable categorization of attributes that people find 'important and useful in daily interactions' (Goldberg, 1981), and is widely used in

socio-economic surveys. It aims at measuring the so-called Big-Five dimensions, labelled as: (I) 'Openness' to experience (being imaginative, creative, curious and unconventional); (II) 'Conscientiousness' (being systematic, goal-oriented and self-disciplined); (III) 'Extraversion' (being active, forthcoming and keen to social relationships); (IV) 'Agreeableness' (being friendly, warm and sensitive toward others); and (V) 'Neuroticism' (worrying, being nervous and emotionally unstable). These dimensions are robust to factor analysis extraction and rotation approaches; are stable across different cultures and languages; and are considered reliable to account for substantive co-variations in personality descriptions (Gosling et al., 2003; McCrae & Allik, 2002). Roberts et al. (2007) and Ozer and Benet-Martinez (2006) emphasize the ability of PTs to predict important life outcomes, including health and happiness, the quality of peer and romantic relationships, and occupational choices.

There are a number of studies concerning the relationship between PTs and childbearing. The existing empirical evidence of such relationship mostly relies on individual data, and provides contradictory results. This is both because different measures of personality are employed (which are not fully comparable) and because of the differences in the considered socio-demographic contexts. The relation between the number of children and PTs may take on an evolutionary perspective, aiming at investigating whether personality is associated with key life-history traits (see e.g. Alvergne et al., 2010). However, most of the existing studies relate to low fertility populations, where childbearing is very much a choice, and the birth of children can be controlled through contraception. Eaves et al. (1990) using Australian data find higher completed fertility among women who score high on 'Extraversion' and low on 'Neuroticism' (although with completed fertility higher than the average for the opposite combination of low 'Extraversion' and high 'Neuroticism' scores). Miller (1992) investigates childbearing motivations using the sixteen-item psychological inventory developed by Jackson (1984) to assess four different traits: 'Nurturance', 'Affiliation', 'Autonomy', and 'Achievement'. His results show that PTs predict motivation for childbearing differently by gender. These differences are confirmed by more recent studies based on BFI. Jokela et al. (2009) find that low levels of 'Neuroticism' and high levels of 'Extraversion' are associated with higher fertility in a nine-year follow-up study on a sample of Finnish men and women aged 15–30 years at the base year. Whyte et al. (2019) explore the relation between PTs and the frequency of sexual relations and find a strong positive relationship to 'Extraversion' for both sexes and to 'Conscientiousness' and 'Agreeableness' for men. From the same sample, instead, higher extraverted and less open men tend to have more offspring, whereas the same is true for more agreeable women. 'Agreeableness', together with 'Extraversion', is shown to positively predict parenthood for Dutch women, but not for men (Dijkstra & Barelds, 2009). Moreover, Jokela et al. (2011) conclude that low levels of 'Neuroticism', high levels of 'Extraversion' and high 'Openness' are associated with higher fertility for both genders, whereas high 'Agreeableness' and low 'Conscientiousness' are associated with higher female fertility.

To the best of our knowledge, there are only two studies on the relationship between fertility and PTs based on the German SOEP data set. One, by Lundberg (2009), finds that PTs predict fertility by the age of 30, whereas they apparently do not explain fertility history by the age of 40. This indicates that personality may matter less for fertility than for its timing. In another study, Le Moglie et al. (2015) analyse the effect of subjective well-being on fertility, controlling for PTs. Their results suggest that subjective well-being positively relates to the likelihood of having a child. However, fertility turns out to be related also to PTs, which nonetheless do not unequivocally determine reproductive behaviour. Summarizing, the effect of PTs on controlled fertility seems to be gender and age-specific, and when personality is measured using the BFI,

'Extraversion' tends to increase fertility and 'Neuroticism' tends to be more significant and to depress it.

A related and interesting line of research argues that the relation between fertility and personality might change across cohorts. The argument stems from the Second Demographic Transition (Van de Kaa, 1987), with the idea that the circumstances of childbearing have altered over time. For instance, younger cohorts have a greater freedom to pursue their own fertility intentions, and childbearing is no longer the pillar of social control, as it was in the past. Consequently, one may expect PTs to play a stronger role today than in the past. Particularly for women, personality played a weaker role when society was characterized by the male breadwinner model, with husbands and wives specializing in, respectively, market and household work. Instead, with educational expansion and greater equality in opportunities for men and for women, the role of personality will necessarily become more important. Following this line of argument, Jokela (2012) finds that 'Openness' for both genders and 'Conscientiousness' for women relate to lower fertility among younger cohorts in the US. Similarly, Skirbekk and Blekesaune (2013), using data from the Norwegian Generations and Gender Survey linked with a postal survey that collected information about PTs, conclude that the personality–fertility relation is different for more recently born cohorts, who have experienced adult life in a different historical context. In fact, they show that 'Conscientiousness' is associated with lower fertility for women, whereas for men 'Extraversion' is associated with higher fertility, and 'Openness' and 'Neuroticism' with lower fertility. They also find that personality relates to fertility differently across cohorts, and that 'Neuroticism' is negatively associated with fertility only in the more recently born male cohorts.

Personality can also affect fertility through its relationship with education, which has long been regarded as the main predictor for fertility decline, due to the higher opportunity cost of childrearing and its postponement (for the effect of education on German fertility, see Cygan-Rehm & Maeder, 2013). Yet, educational attainment is also mediated by PTs, which may, in turn, affect fertility. The massive educational expansion that has taken place over the last four decades, again, fuels the idea that the role of PTs on fertility might be different today compared to the 1960s and the 1970s. Tavares (2016) offers an important insight here; namely that reproductive behaviour is interrelated with other careers, such as education. According to her, the influence of PTs both on education and on fertility decisions explains the fertility timing gap between more and less educated women. Using the BHPS, she finds that 'Agreeableness', 'Extraversion' and 'Neuroticism' relate to early childbearing, while 'Openness' and 'Conscientiousness' relate to later childbearing. Even so, individual differences in PTs result in differences in the age at first birth, in particular among better educated women, who postpone childbearing for longer.

Numerous other studies focus on the relationship between PTs and union formation, finding a significant association, which, as in the case of fertility, differs by gender and by cohort. Diener and Lucas (1999) discover that the PTs related to happiness, such as high 'Extraversion' and low 'Neuroticism', are likely to attract a marriage partner. Lundberg (2012), analyses the effect of PTs both on the formation and on dissolution of partnerships using German SOEP data. She concludes that selection into marriage is associated with distinctly different personality profiles for people born before 1960, whereas for younger cohorts the effects are more similar gender-wise. These results are consistent with the changes in the couple's organization, which has evolved from gender-based specialization, to a more egalitarian partnership. For older cohorts, born between 1945 and 1959, only 'Extraversion' significantly increases the probability of marriage for both genders, while agreeable women and conscientious, antagonistic men,

are more likely to marry. For younger cohorts, born between 1960 and 1970, two PTs, 'Openness' and 'Conscientiousness', make it more likely for men and women to marry by the age of 35. Other studies have specifically investigated the association between PTs and union dissolution (e.g. Kiernan, 1986; Kinnunen et al., 2000; Lowell & Conley, 1987). Personality seems indeed to be linked to long-term relationship quality. Divorce is associated with high 'Neuroticism' and 'Openness' and low Conscientiousness' for both genders, with low 'Agreeableness' for women, and with male 'Extraversion' (see Roberts et al., 2007, for a comprehensive review on this literature). It also seems that, in younger couples, women are more likely than men to monitor relationship quality and to end or avoid unsatisfactory consensual unions (Aberg, 2009). This implies that the characteristics of males may prove more important in terms of union stability. Similarly, and particularly when making decisions around the long-term investment of having children, male personality characteristics may become more relevant than women's personality in family formation and dissolution behaviour (Hewitt et al., 2006; Lillard & Waite, 1993).

The relation between PTs and work career has been investigated in detail too. Judge et al. (2006) find a positive relation between 'Conscientiousness' and work success, both intrinsic (i.e. job satisfaction) and extrinsic (i.e. income and occupational status), and a negative relationship between 'Neuroticism' and extrinsic success. Seibert et al. (2001) survey employees in a diverse set of occupations and organizations and find a positive relation between 'Extraversion' and salary level, promotions, and in-job satisfaction. Instead, there is a negative relation between 'Neuroticism' and work satisfaction, between 'Agreeableness' and career satisfaction, and between 'Openness' and salary level. In particular, several studies point to gender-based differences in the association between PTs and earnings and salary levels. Both Nyhus and Pons (2005) and Mueller and Plug (2006), using respectively Dutch and American data, conclude that emotional stability (i.e. low levels of Neuroticism) positively relates to wages for both men and women, while 'Agreeableness' is associated with lower wages for women. Heineck (2011), instead, finds wage penalties for 'Agreeableness' among all British workers and for 'Neuroticism' only among women. A contrary effect was found of 'Openess'. The returns on personality factors vary both by tenure and by educational group. This suggests that different PTs may enhance productivity in different occupations and that personality effects already start with different tracks and levels of success in education.

An overlooked aspect in the extensive literature on the relationship between personality and union formation and dissolution, working career and fertility, concerns the strong interaction between these life-course events and transitions. Union formation and dissolution, but also education and job career, are all dynamically intertwined, implying that they should be analysed jointly (Elder, 1985). Indeed, the possible association between PTs and fertility may mask the fact that personality affects, at the same time (and even beforehand), participation in education, success on the job market, and the propensity and facility in finding a partner. At the same time, the possible relation between fertility and PTs could derive from the association between PTs and union formation. Likewise, PTs might affect fertility behaviour, which in turn drives individuals' careers (and vice versa). An inherent weakness of the numerous studies on the role of PTs for life course events is that they analysed the relationships between PTs and fertility, work careers, or union formation separately. As far as we are aware, nobody looked into the role of PT and these three processes jointly, as we aim to do here. The close connections between these processes needs to be integrated also in the empirical approach, overcoming the weakness in this regard of the widely used Event History Analysis.

## 3 | THE DATA

Our data arise from the German SOEP, a representative ongoing longitudinal study started in 1984 (Wagner et al., 2007). We include in our study all the 36 waves up to 2019, thus following individuals over a relatively long period of time. For each individual, we build the yearly sequences of fertility, marital, and working status between the ages of 20 and 40.

For fertility, we refer to the birth years of the individuals' children. Since few individuals had more than three children, we only distinguish between 0, 1, 2 and 3 or more children. For the other domains, we rely on the prospective longitudinal data collected at interviews. Therefore, only individuals who started participate in the survey when they were aged 20 or earlier are considered. Since SOEP questionnaire is administered to individuals aged 16 or more, the first year of participation to the survey has limited variation within cohorts.

For the working career, we consider the individuals' labour force status, distinguishing between full-time work, part-time work, in training (or education), unemployment and non-work. For the marital status, we distinguish between being single, married, separated or divorced, and widowed. We used the rich retrospective data available in SOEP concerning the working and the partnership biographies to verify the consistency and coherency of the states declared at the interviews.[1]

We also used the complete information available in SOEP to fill the gaps present in prospect data, in order to avoid discarding cases because of missing answers at the moment of the interview.[2]

Our data include only 984 individuals (523 women and 461 men) whose sequences are *complete* (i.e. continuously observed between the age of 20 and 40) for each domain. For a larger number of cases, only partially complete sequences, truncated at an age lower than 40, are available. Only few individuals (five women and one men) remained widowed before the age of 40. For these cases, we consider only the events experienced before widowhood, because of the very peculiar subsequent marriage sequence. Thus, the sequences of these individuals are truncated and no longer complete.

Note that in SOEP prospective data are available only on marital status and not on cohabitation, which is particularly widespread among younger people. Therefore, we relied on the retrospective information about cohabitation available in SOEP biography histories to build the sequences of *Union*, distinguishing between individuals in a union (cohabiting with their partner, being married or not) and those living as single (possibly after separation or divorce). Unfortunately, this implied focusing on cases whose biographies are available, with a consequent reduction of the number of observations with complete sequences (359 women and 290 men).

As mentioned, our interest lies in relating cluster memberships (derived from sequences' patterns) to PTs via multinomial logistic regression, controlling for a set of variables measured when

[1]For example, if one individual has been married (with the same person) for a number of years but declares to be single or separated at one interview, we modified the status into 'Married'. We followed a similar procedure to recode working activities; for example, if a 'Non-working' status was registered between two identical working states, possibly because of a change of work. This was particularly important to deal with maternal leave. Indeed, some women were interviewed before and after their maternal leave and therefore turned out to be employed without any interruption. Instead, some women were interviewed during their maternal leave. For the sake of consistency, we did not consider the interruption of work due to maternity if it lasted for one year maximum, whereas we coded a longer interruption of the working activity as 'Non-working', even when consequent to maternity.

[2]Specifically, if a (sub-)sequence of identical states is interrupted by a missing value, and retrospective data indicate that the status is maintained for the entire period, the missing status is imputed accordingly. See Appendix 1 for details.

the respondent was about 15 years old. Specifically, we consider the household's income, the macro region of residence (North, South, West or East Germany), the migration background (no migration, direct, or parental migration), and the father's and mother's level of education.

As for the PTs scores, the SOEP collects the Big Five Inventory (BFI) every four years starting from 2005. In each wave, respondents are requested to assess how well 15 descriptive sentences—three for each trait—apply to them, using a 7-points Likert scale. The score on each trait is obtained as the average of the answers to the corresponding set of questions. Appendix 1 reports details about data and variables.

There are contributions in the literature pointing to the stability of personality over the adult lifespan, based on large panel surveys (see among others Caprara & Cervone, 2000; Cobb-Clark & Schurer, 2012; Le Moglie et al., 2015; Specht et al., 2013). In particular, Cobb-Clark and Schurer (2012) argue that average personality changes are small and do not vary substantially across age groups, that intra-individual personality change is generally unrelated to experiencing adverse life events and, therefore, personality can be modelled as a stable input into individuals' decisions. On the contrary, Lucas and Donnellan (2011), using longitudinal German data, showed that differential stability was relatively strong among all age groups but that it increased among young adults, peaked in later life, and then declined among the oldest ones. Similarly, Specht et al. (2011) sustain that—even if personality predicts the occurrence of several objective major life events (selection effects), personality changes throughout the life span in reaction to experiencing these events (socialization effects), with more pronounced changes in young and old ages. Gensowsky et al. (2020) confirm for Danish people changes over the life cycle (even with some unusual paths over the age) and find that the gender gap in terms of PTs widens for 'Openness' but shrinks for 'Neuroticism.' Thus, the stability of PTs at the individual level is still an open and strongly debated issue.

In SOEP, PTs scores are measured at different ages, from one to four times, depending on the duration and continuity of the individuals' participation in the panel, and on the year when they entered the survey. Though there are good reasons to expect a certain level of stability in these traits, we focus on the scores measured at the individual's first occasion (i.e. closest to the age of 20). The age at such occasion varies across cases, and is therefore included among the control variables in the multinomial regressions relating clusters to PTs. In addition, since clusters relate to the activities experienced between 20 and 40 years of age, in the regressions we limit attention to cases whose first PTs measurement was taken before the age of 40. This penalized in particular individuals from older cohorts, who were relatively older in 2005, when PTs scores were measured for the first time. The choice of considering only the PTs at the individual's first occasion is supported by findings in psychology arguing that non-cognitive skills observed at early ages are associated with later differentials in socio-economic conditions and behaviours (Attanasio et al., 2020; Lundberg, 2018).

For the sake of completeness, however, we analysed the relation between the difference in the PTs scores taken at the first and at the last occasions and the background variables, the age at the first PTs measurement, the number of years between the first and the last measurements, and the life events experienced in that period. We focused in particular on the events in the three considered domains (i.e. getting married or starting a cohabitation, having children, moving from employment to unemployment or vice-versa, and moving from training or education to employment). For each of the five PTs scores, both for men and for women, the difference is significantly related only to the score measured at the first occasion. This comforts us in our choice and supports the procedure illustrated above.

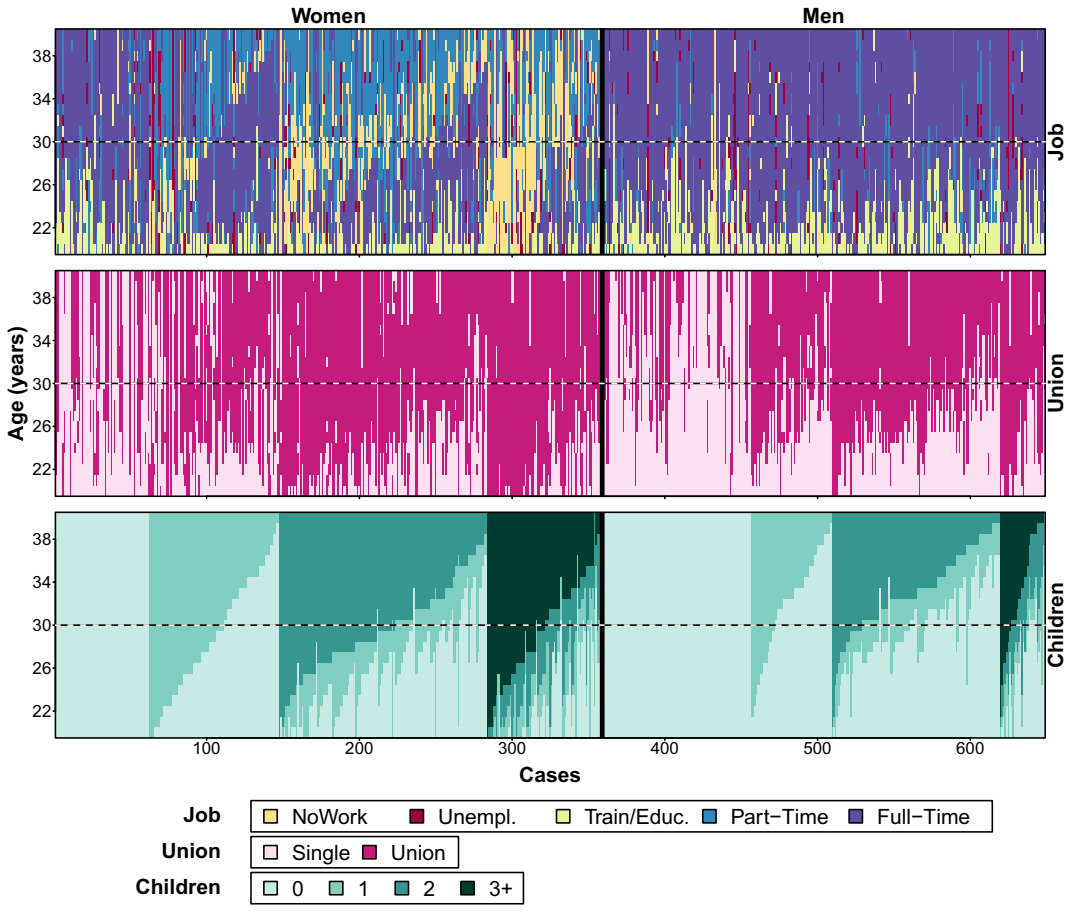


**FIGURE 1** Index plots of the complete sequences—observed between 20 and 40 years of age. For each domain, cases are ordered on the horizontal axis first according to sex, and then based on the number of children and on the parent's age at their birth

## 4 | IDENTIFYING TYPICAL CAREERS USING MSA

To obtain clusters of individuals experiencing similar trajectories in the three domains of interest (*Job*, *Union*, *and Children*) using *standard* MSA (Pollock, 2007), we must limit attention to cases with complete sequences on the three domains. Figure 1 reports the index plots of the sequences by domain, distinguishing between women and men. In each plot, individuals are placed on the horizontal axis.[3] To emphasize the relation between *Union* and *Job* and the *Children* domain, cases are ordered according to their gender, to the number of children and to the age at children's birth. In each domain, each individual is associated to a vertical bar, describing the activities experienced from 20 to 40 years of age, with different colours assigned to different activities. There

[3]In some papers, index plots are built placing individuals on the vertical axis and time on the horizontal one. Here we follow Piccarreta (2017) and flip the axes in order to have a better visualization of sequences in the three domains, which will be particularly useful for the representation of sequences within clusters.

are substantial differences between women and men. Men tend to postpone marriage and parenthood to a later age, and have more stable job trajectories, usually dominated by full-time work. Instead, a relatively high proportion of women experience part-time work, unemployment or NEET (i.e. not in education, training, or employment), with a higher tendency to switch to such activities when they become mothers at an early age or when they have a relatively high number of children. This suggests that cluster analysis should be applied separately for women and men, to avoid gender-based results.

In addition, the three domains appear to be interconnected: individuals with similar trajectories in one domain tend (on average) to experience similar trajectories in others. Particularly for women, job and family formation choices are clearly related. Men's job trajectories seem instead less connected to those in the other domains: most men work, irrespective of whether or not they are in a union and/or have children. This further supports the decision to analyse women and men separately. It also suggests the opportunity to study the *joint* unfolding of trajectories in the three domains, using MSA and cluster analysis to identify their most typical combinations in the sample.

MSA (as in the original proposal of Pollock, 2007) is intrinsically related to *edit* dissimilarities, which quantify the difference between two sequences (in one domain) as the total cost of the operations needed to transform one sequence into the other. Here we refer to the popular Optimal Matching algorithm (OM; Abbott, 1995), that focuses on three basic operations: insertion or deletion of a state, and substitution of one state for another. We postpone the discussion on the choice of the edit costs (Section 4.2), but remark that it is arbitrary, and there are no theoretical results supporting one specific choice over another. MSA combines the (edit) dissimilarities for each domain into one *multichannel* (MC)-dissimilarity, by averaging the costs of the operations needed to align the sequences in each domain, thus preserving the information on each domain (as measured by specific costs).

In principle, MC-dissimilarities can be used to identify clusters of cases with similar trajectories in the three domains. Nonetheless, the relatively low number of complete trajectories available both for women and for men, and the further attrition due to missing values on covariates (Appendix 1) result in a lower possibility of obtaining well distinguished clusters with a decent size. This may also affect power, thereby making it more difficult to establish reliable conclusions drawn based on the multinomial regressions relating clusters to PTs. In addition, this implies excluding from the analysis younger cohorts, who had not yet turned 40 at their last interview. To address this issue, we introduce novel procedures to assess the dissimilarity between trajectories of different lengths, observed from the age of 20 but truncated at an age lower than 40.

## 4.1 | Sequence analysis of trajectories of different lengths

As noted by Piccarreta and Studer (2019) a procedure to deal with truncated or censored trajectories has not been developed yet. Indeed, using a 'missing' state to code the unobserved part of the trajectories would make the length of the sequences an element—possibly the most relevant—of similarity between trajectories, leading to clusters based on the observation time. Therefore, standard SA typically relies on *complete* sequences, observed for the entire period of interest, thus disregarding cases whose trajectories have missing states, possibly because of censoring. When censoring relates to individuals' characteristics, this implies the systematic exclusion of specific strata of the population. Indeed, for our data—as in many applications—the most recent cohorts, that are only partially observed, would be excluded from the analysis.

To address this issue and to avoid the potential weaknesses associated with the exclusion of shorter trajectories, we envision some procedures to evaluate the dissimilarity between sequences of different lengths or, more precisely, right-truncated or censored sequences.

A first possibility consists in clustering the complete sequences, subsequently assigning the truncated sequences to the closest cluster. To assess the dissimilarity between a sequence truncated at $t$ (i.e. observed only between the ages of 20 and $t$) and a cluster of complete sequences, one could consider the dissimilarity between the truncated sequence and a cluster's representative, for example the *medoid* trajectory (which is most similar to all the others in its cluster), truncated at $t$. To avoid an excessive dependence on one specific sequence in the cluster, it is also possible to evaluate the average dissimilarity between the truncated sequence and all the cluster's (complete) sequences truncated at $t$. A possible drawback of this approach is that the measures commonly used to assess clusters' quality (and therefore to evaluate, compare, and choose among different solutions) would necessarily rely on the set of complete sequences and not on all the sequences finally allocated to clusters. In addition, the complete sequences would necessarily drive the clustering process. When shorter sequences (characterizing for example younger cohorts) present features different from those characterizing the (initial track of the) complete ones, it would not be possible to build clusters accounting for such differences.

To fix this issue, we propose to assess the pairwise dissimilarity between two sequences of different lengths (i.e. truncated at different ages) by focusing on the longest period, starting from the age of 20, available for both cases. In standard SA, when a single trajectory is available for each individual, for two sequences $i$ and $h$ truncated at ages $t_i$ and $t_h$ respectively, we consider the dissimilarity between the two sequences both truncated at $t_{ih} = \min(t_i, t_h)$. The size of such dissimilarity might depend (also) on the length of the two sequences, in our case $(t_{ih} - 20 + 1)$, which varies across the pairs of cases, due to the different times at truncation. To normalize the dissimilarity, we therefore divide it by the length of the common period of observation. When more domains are considered, the procedure is identical, but it relies on the MC-dissimilarities calculated on the set (in our case, the triplet) of sequences available for each case, all truncated at $t^*_{ih}$, the minimum time at truncation for the two cases across all the domains. Again, such dissimilarity is normalized and divided by $(t^*_{ih} - 20 + 1)$. Proceeding in this way for all the pairs of cases in data, and arranging the obtained dissimilarities in a matrix, we build what we call the *integrated* MC-dissimilarity matrix, $\mathbf{D}^*$.

Before proceeding, note that typically some elements of $\mathbf{D}^*$ do not satisfy the triangle inequality. This does not prevent its use in cluster analysis. Indeed, many dissimilarity measures are frequently employed that do not satisfy such condition (e.g. the Jaccard coefficient and other dissimilarity measures for categorical data, or the Bray-Curtis coefficient for ecological data; see Everitt et al., 2003), and among them also dissimilarities typically employed in standard SA. Nonetheless, we acknowledge that there are methods to adjust $\mathbf{D}^*$ and to correct for its non-Euclidean portion (see Gower & Legendre, 1986; Legendre & Anderson, 1999). At least for our data, results obtained using the original and the corrected integrated dissimilarities coincide.

As a further, relevant, consideration note that our procedure to obtain the integrated dissimilarities is heuristic—as are in fact all the tools used in SA as well as the most commonly used clustering algorithms. Therefore, the quality and the reliability of the obtained results must be carefully evaluated for each specific situation.

## 4.2 | Clusters of trajectories for the SOEP data

Considering the goal of our analysis, it is important to preliminary assess whether a relation exists between the length of the individuals' sequences and the background variables or the PTs scores. We therefore applied a Poisson regression, which returned a significant coefficient only for the year of birth. Thus, as reasonably expected, truncated sequences typically characterize individuals from recent cohorts, who were younger than 40 at their last interview.

To apply the proposed procedure, it is necessary to set the *minimum* length that truncated sequences must have to be included in the analysis. Note from Figure 1 that for the available *complete* sequences the combinations of trajectories experienced in the three domains between the ages of 20 and 30 are related, at least to a certain extent, to the subsequent unfolding of the trajectories between 30 and 40 years of age. To explore this aspect, observe that for complete sequences it is possible to study the relation between the integrated MC-dissimilarities based on sequences truncated at different ages, say 25, 26, … , 39, and those based on the complete ones. At this aim, we used the Mantel's correlation coefficient, assessing the extent of association between two dissimilarity matrices as the correlation between their elements. Not surprisingly, the correlations increase with the length of the truncated sequences. Moreover, at least 11 years of observation (thus, sequences observed at least until 30 years of age) are necessary to achieve a relatively high correlation coefficient (namely, 0.80 for women and 0.79 for men; see Appendix 4). On the one hand, this does not allow concluding that a relation exists between the *actual* truncated sequences and their (unobserved) future development. On the other hand, this suggests setting to 11 years the minimum age span required for inclusion in our analysis. Similar conclusions arise from the comparison between clusters based on the complete sequences and clusters based on the same sequences truncated at different ages. Specifically, we evaluated the *consensus* between such partitions using the Rand index (Rand, 1971). Our results show that—if the number of clusters is not too small—there is a good match, which again increases with the length of truncated sequences and reaches a relatively high level for sequences truncated at 30 years of age or later (Appendix 4). Importantly, the obtained results hold irrespective of the clustering algorithm and of the chosen dissimilarity criterion (see Appendix 3 and the next section for details on the alternative clustering algorithms and dissimilarity criteria taken into account).

We remark that the complete sequences truncated at different ages might have features different from those of the *actual* truncated trajectories—possibly observed for younger cohorts. The lack of information on the evolution of such trajectories until the age of 40, does not allow drawing conclusions on the suitability of our choice for partially observed sequences. Nonetheless, our procedure does *not* aim at inferring the characteristics of the complete sequences for individuals only partially observed, but, rather, to group cases based on the similarity of their *common initial track*. A particular attention to this aspect is necessary when evaluating and selecting clusters obtained with alternative approaches.

Based on the above considerations, besides individuals with complete sequences, we included in our analysis also those whose sequences were observed for at least 11 years (between 20 and *at least* 30 years of age). This leads to an increase in the number of individuals included in the sample: 1075 women and 861 men (Appendix 1).

To obtain the integrated MC-dissimilarities using OM, it is necessary to specify the costs of the edit operations for each domain. Note that MC-dissimilarities should suitably account for the dissimilarities in *all* the considered domains. Therefore, to avoid settings of the edit costs inducing prevalence of specific domains over the others, we adopted the same criteria across all the domains. Following a rather standard approach, we defined insertion and deletion (indel) costs

equal to 1 for each domain. As for the substitution costs, representing the effort needed to move from one state to another, we used a data-driven approach, and set them inversely proportional to the frequency of transitions from one state to another. Thus, the cost of substituting state $a$ with state $b$, and *vice-versa* is $s_{ab} = s_{ba} = (2 - f_{a\to b} - f_{b\to a})$, where $f_{a\to b}$ is the sample frequency of transitions from state $a$ to state $b$. In this way, substituting a state with another is less costly when frequent transitions between the two states are observed in the data (see later for alternative settings).

Based on the integrated MC-dissimilarities built for women and men respectively, we extracted a number of clusters ranging between 2 and 10 using the partitioning around medoids algorithm (PAM; Kaufman & Rousseeuw, 1990), that is an extension of the $K$-means partition algorithm to the case when only dissimilarities (rather than measurements on variables) are available. As underlined by Piccarreta (2017), caution is necessary when clusters are based on MC-dissimilarities, which combine information on different domains. Indeed, some domains might prevail over the others and drive the clustering process. It is therefore important to evaluate the quality of partitions with respect both to the integrated MC dissimilarity matrix, $\mathbf{D}^*$, and to the integrated dissimilarity matrices, $\boldsymbol{D}_{\boldsymbol{J}}^*$, $\boldsymbol{D}_{\boldsymbol{P}}^*$ and $\boldsymbol{D}_{\boldsymbol{C}}^*$, obtained for each domain separately. Specifically, we monitor three standard measures of adequacy, namely the R-square, the Point Biserial Correlation, and the average silhouette coefficient (Kaufman & Rousseew, 1990). Results in Figure 2 show that, both for women and for men, the quality of the partitions is lower for the *Job* domain. Indeed, this is the most turbulent domain, with a higher number of states and less stable sequences (Figure 1). Notably, the average silhouette coefficients tend to decrease with the number of clusters, pointing to simpler partitions. Even so, since our intent is to relate clusters to PTs, it is important to extract clusters as much homogeneous as possible across the domains. Therefore, we focused in particular on the R-square (related to the within-compactness of the clusters), which, compared to other criteria, suggests a relatively higher number of clusters. In fact, a close inspection of partitions with a relatively small number of clusters showed that they allowed identifying clear patterns in the family-related domains, but they did not offer a satisfactory representation of the different work trajectories in data. In addition, since we consider sequences of different lengths, even if observed at least until the age of 30, it is important to consider the similarity of clusters with respect to the initial tracks of the careers. Based on these considerations, we selected 8 clusters for women and 9 clusters for men as reasonable compromise solutions.

To evaluate possible alternatives to the selected partitions, we applied the described procedure using a different specification for substitution costs, a different dissimilarity criterion and a different clustering algorithm. Specifically, we followed the data-driven approach introduced by Studer and Ritschard (2016), who propose to (inversely) relate the substitution cost between two states to the frequency with which they are followed by a common state, thus 'sharing similar future' (see Appendix 3). We also considered another edit distance, the Hamming generalized distance, which is based only upon substitution of states (and not on insertions and deletions), using both the illustrated settings of substitution costs. Finally, we extracted clusters using the popular hierarchical Ward's clustering algorithm. Results are quite stable and aligned with those reported above, which are nonetheless preferable—at least based on the criteria we used to evaluate alternative clusters solutions (see Appendix 5 for a discussion on the results).

For the sake of completeness, we also applied the alternative strategies to deal with truncated trajectories described at the beginning of Section 4.1, namely the one based on coding missingness as a specific state and the one based on the assignment of truncated sequences to clusters obtained

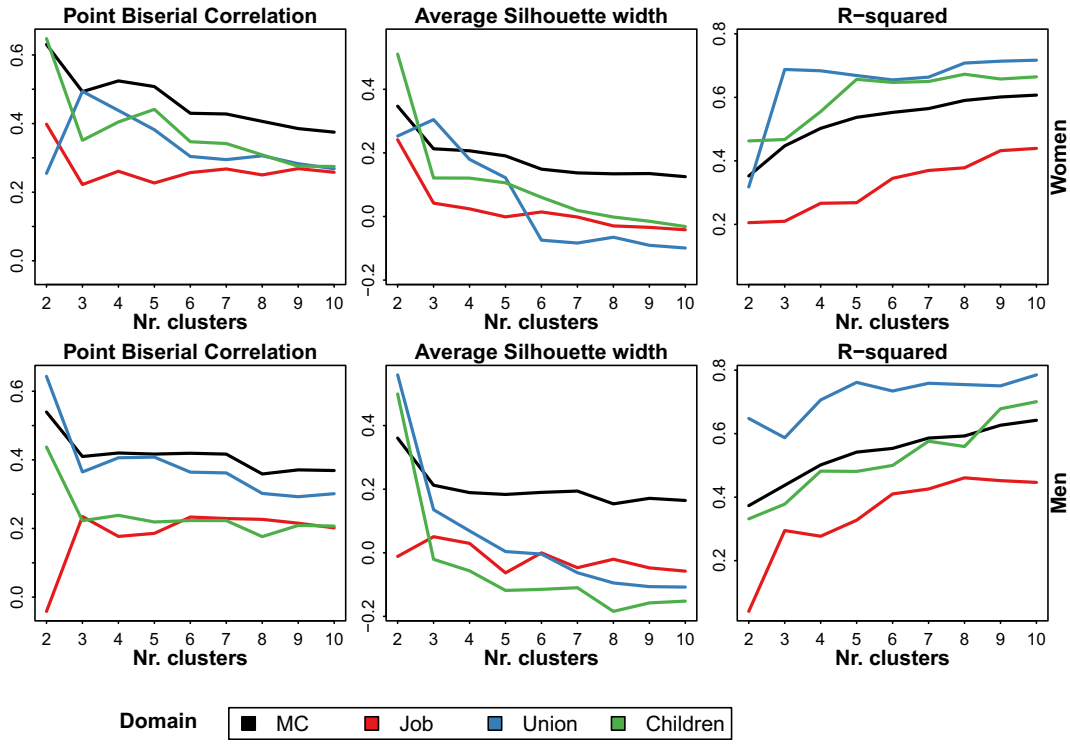


**FIGURE 2** Monitoring Point Biserial Correlation, Average silhouette coefficients and R-squared for a number of clusters ranging from 2 to 10 extracted applying the PAM algorithm to the integrated MC-dissimilarities built for women and men based on individuals' triplets of sequences (possibly truncated) defined over Job, Union and Children domains. For each sub-sample, the quality of the partitions is evaluated both with respect to integrated MC-dissimilarities used to build clusters, and with respect to the domains-specific integrated dissimilarities obtained separately for each domain. This allows assessing the ability of the clusters to address heterogeneity in each specific domain

based on complete sequences only (see Appendix 6). The former approach—as expected—leads to unsatisfactory results and returns clusters based on the length of the sequences. Instead, clusters obtained using the second approach are quite coherent with those found using the approach based on integrated MC-dissimilarities, even if—again in line with expectations—with some differences, due to the fact that the initial clusters are obtained based on complete sequences only (see later).

Figures 3 and 4 report the index plots of the sequences in each domain for each cluster. To enhance the visualization of the sequences, within each cluster cases are arranged along the horizontal axis according to the length of their trajectory. The clusters obtained for women and men are meaningful from a substantive point of view and have well-defined features. Sequences in the same clusters share similar patterns between the ages of 20 and 30, and in particular individuals in the same cluster have about the same number of children around the age of 30. Also, even if the job trajectories are more heterogeneous, clusters show distinctive features with respect to the states experienced in the three domains. For example, women in clusters W1 and W2 are similar with respect to the *Union* and the *Children* domains: most of them tend to postpone family formation, and remain single and without children until 30 years of age or more. Even so, the two clusters present differences with respect to working activities; women in W1 prolonged

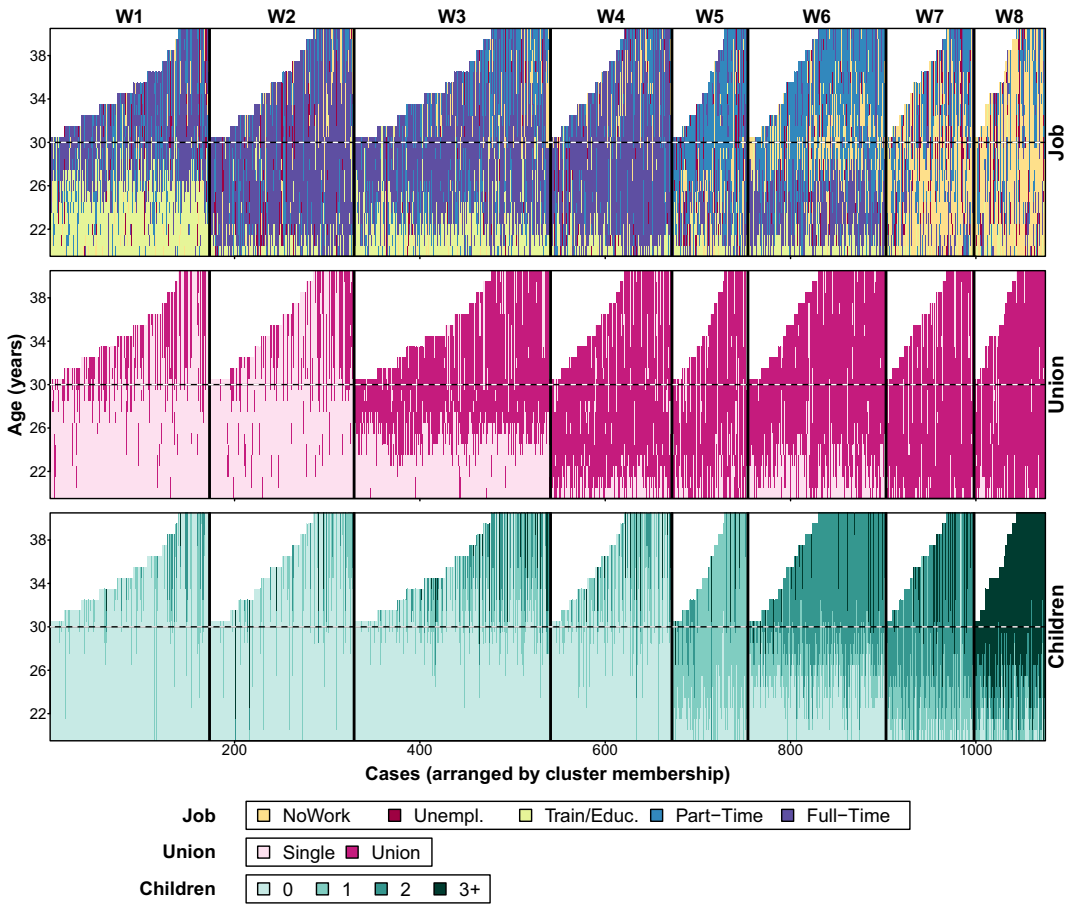


**FIGURE 3** Index plots of the sequences for women in the three considered domains within the-clusters extracted using the PAM algorithm. Clusters: (W1) Long education & full-time job, with late or no family; (W2) Short education & full-time job, with late or no family; (W3) Medium education & full/part-time job, with late family; (W4) Low education & full-time job, with early union/late motherhood; (W5) Part-time job, with early union/one child; (W6) Part-time job, with early union &two children; (W7) No-work, with very early formation of medium-sized family; (W8) No-work, with very early formation of large family

education (or started their career with a relatively long period of training), whereas (most of the) women in W2 entered the labour market with a full-time job relatively early. Note in addition from Figure 1 that only a low proportion of women with complete trajectories prolonged education or training. In fact, when clusters are obtained based on complete sequences only (with subsequent assignment of truncated trajectories to clusters), there is no cluster dedicated to women with this type of sequences (see Appendix 6). Such a cluster is instead identified (W1, Figure 3) when truncated trajectories are directly included in the analysis. Thus, our procedure based on the integrated dissimilarities allows incomplete sequences to effectively contribute to the formation of clusters.

Caution is necessary when analysing clusters including sequences truncated at different ages. Indeed, the similarity between cases in the same cluster as well as their dissimilarity to cases in other clusters is limited to the period of observation available for each case. For example, younger women in cluster W4 could have their first child few years after the age of 30 and their second

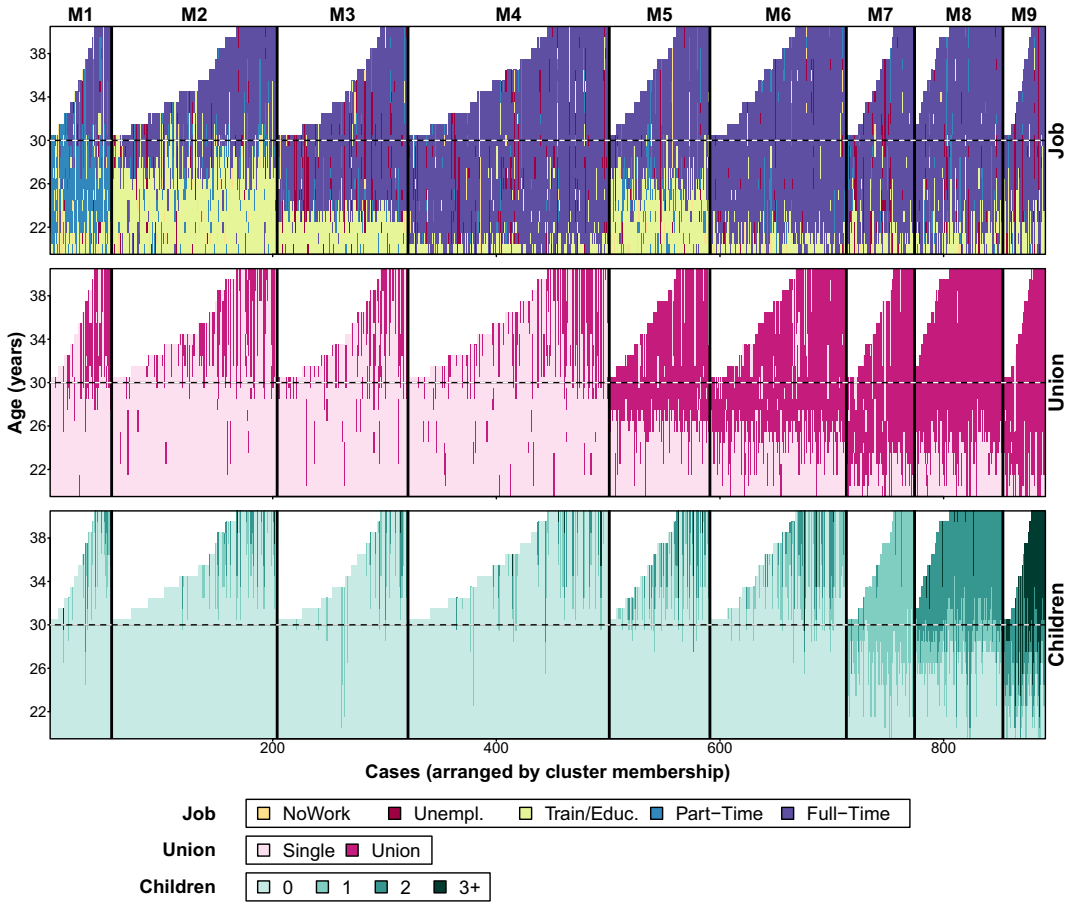


**FIGURE 4** Index plots of the sequences for men in the three considered domains within the-clusters extracted using the PAM algorithm. Clusters: (M1) Early part-time & late full-time job, with late or no family; (M2) Long education & full-time job, with late or no family; (M3) Medium educated with job discontinuity and late or no family; (M4) Low education & full-time job, with late or no family; (M5) Long education & full-time job, with early union and later fatherhood; (M6) Low/intermediate education & full-time job, with early union and later fatherhood; (M7) Low/intermediate education & full-time job, with early formation of small family; (M8) Low/intermediate education & full-time job, with early formation of medium-sized family; (M9) Low/intermediate education & full-time job, with early union and very early formation of large family

child at a short distance, and could switch from full-time to part-time job, thus showing the same patterns observed in cluster W6 though delayed by a few years. In addition, the patterns observed for longer trajectories cannot be considered as the probable evolution of the truncated ones, unless one assumes that truncation is unrelated to socio-demographic characteristics and/or that the possible differences in such characteristics have no impact on the trajectories' unfolding. For our data, truncation is related to birth cohort. Even if the patterns of evolution within clusters are quite homogeneous as the length of the sequence increases, the future features of early-truncated trajectories might differ from those characterizing the complete trajectories in the same cluster. For example, most of the women in cluster W4 start a union and enter the labour market with a full-time job relatively early, and have their first child around the age of 30. The trajectories

observed for a relatively long period show a tendency of these women to maintain their working status. Nonetheless, younger women might decide to switch to a part-time job at a certain age.

Turning to the substantive interpretation of the clusters, for the women we can summarize their most relevant features as follows.

*Cluster W1: Long education & full-time job, with late or no family*. This cluster is unique mainly with respect to job trajectories. It is characterized by an initial long track spent in education or training, on average at least until 26 years old, and by a subsequent entry in the labour market with a full-time job. Most of the women in this cluster remain childless and single at least until the age of 30 (and a very high proportion also later). It is not possible to relate causally the delayed entry into stable work with delayed—or possibly never experienced—family formation. Indeed, this cluster might include women who faced difficulties accessing the labour market, as well as highly-educated women who decided to wait for a job meeting their expectations.

*Cluster W2: Short education & full-time job, with late or no family*. This cluster is similar to W1, except for the job trajectories, characterized by a shorter education track before full-time working. The trajectories of partnership and parenthood show a net prevalence of women who remain single and childless until the age of 30, few of whom enter a union (marriage or cohabitation) and/or have children even thereafter.

*Cluster W3: Medium education & full/part-time job, with late family*. Compared to the first two clusters, most of the women in this cluster enter a union relatively earlier (but not before the age of 25 or 26) and become mothers around the age of 30 (very few did so earlier). As for the job trajectories, the cluster presents a high level of variability, with differences in the length of the education (or training) tracks (on average ending around the age of 22–23), followed by full-time job and by frequent transitions from full-time to part-time job, especially after the age of 30, probably after the birth of the first child. In any case, the switch to part-time work after motherhood seems to be linked to the tendency to remain in the labour market after the age of 30.

*Cluster W4: Low education & full-time job, with early union/late motherhood*. A high proportion of women in this cluster start working full-time and enter a union relatively early (around or before the age of 22); nonetheless very few have children before the age of 30, and only a few have a second child. Most of them are permanently full-time worker, and only a few switch to part-time work or exit the labour market after children's birth.

*Cluster W5: Part-time job, with early union/one child*. Similarly to those in W4, women in this cluster enter a union relatively early (around or before the age of 22), but they have their first child at the same age, without further increasing their family's size. The initial tracks of the job careers—before the age of 30—are rather turbulent, but most of the women have a part-time job after the age of 30. Note that for this and the subsequent clusters, we do not refer to the level of education because of substantial alignment.

*Cluster W6: No-work, with very early formation of medium-sized family*. This cluster includes women who start a union relatively early (but on average later than those in W4 and W5), have their first child at an intermediate age compared to the other clusters (around the age of 26), and have two children at the age of 30. They combine family and part-time job (especially after the age of 30 and the birth of their children). Indeed, some of them enter the labour market with a full-time job, before having children. This might explain the slight postponement of family formation as well as the later 'renounce' to a full-time career as their family grew.

*Cluster W7: No-work, with very early formation of medium-sized family*. This cluster reflects a more traditional gender role set, where women do not participate strongly in labour market (only few have a part-time job at late age, after 35). This is probably because they start a union and have their first child quite early (around or before the age of 22), and have a second child before turning 30 (mostly around the age of 26), in few cases giving birth to a third child around the age of 30.

*Cluster W8: No-work, with very early formation of large family*. This cluster is similar to W7 as for the early age at family formation. Notably, nonetheless, the age at the second child is anticipated, and almost all the women have three or more children before turning 30. The tendency to stay out of the labour market is very strong in this case: most of these women never started a job (or worked only for very short spells).

It is worthwhile reflecting on the substantive implications of these patterns. Looking at clusters from W1 to W4, it seems clear that full-time work for women is hardly compatible with childbearing, or at least with early motherhood (i.e. before age 30). Interestingly, long education does not appear to be the main driver to postpone or forego family formation. Indeed, cluster W1 and W2 show similar family formation patterns even if they differ as for the length of the education (or training) period. Only in clusters W3 and W4 we see larger proportions of women having children, combined nonetheless with an increase in part-time work. On the other extreme of the work-family balance are clusters W6, W7 and in particular W8, grouping women who have children earlier and more of them. The most traditional groups (W7 and W8), where women have two or more children and do not work, are quite limited in size. Interestingly, clusters W1 and W2 present the highest proportions of truncated sequences, and clusters W1 to W4 indeed include the highest proportion of women from the youngest cohorts (Appendix 2). This is in line with the fact that younger cohorts show a tendency to postpone marriage and childbirth, and indeed the high-fertility clusters include lower proportions of young women. Interestingly W6—characterized by an intermediate age at first childbirth compared to the others—presents the smallest proportion of truncated sequences. In fact, younger individuals will generally show late family formation or—even if in fewer cases—by an opposite anticipation of family formation. Thus, an intermediate age at childbirth is more difficult to observe for younger cohorts, who in a sense have 'less time' to have their status changed.

As for the clusters extracted for men (Figure 4), note that the first four clusters include men who remain single and without children for almost the entire period of observation, whereas the last ones identify men who had children before the age of 30, even if with different ages at their first child. Focusing on the most salient features of the clusters for men, we can label them as follows.

*Cluster M1: Early part-time & late full-time job, with late or no family*. This cluster is unique mainly with respect to job trajectories: most of the men in it start to work very early (around the age of 22) but part-time. After the age of 30 some of them show a troubled working career, but the majority has a full-time job. A clear tendency can be noted to remain childless and single at least until the age of 30 (and, for a very high proportion of cases, also later).

*Cluster M2: Long education & full-time job, with late or no family*. This cluster (similarly to W2 for women) is characterized by an initial long track spent in education or training, on average at least until the age of 27, and by a subsequent entry in the labour market with a full-time job. Most of the men in this cluster remain childless and single at least until the age of 30 (and a very high proportion also later). As noted for cluster W2, this cluster might include highly-educated men who wait for a job meeting their human capital formation.

*Cluster M3: Medium education & job discontinuity, with late or no family*. Most of the men in this cluster continued education (or training) until 22 to 23 years of age, and started then a

full-time job, with some episodes of unemployment. Also in this cluster, most of the men remain childless and single at least until the age of 30 (and a very high proportion also later).

*Cluster M4: Low education & full-time job, with late or no family*. This cluster is very similar to the previous ones with respect to family formation, despite an early entry into the job market, mainly with full-time jobs stable over time.

*Cluster M5: Long education & full-time job, with early union and later fatherhood*. Men in this cluster enter the labour market with a full-time job around 26 to 27 years of age, after a long period in education or training. They start a union relatively early (at the completion of the education/training period or before), experiencing in few cases a later union dissolution. Only some of them become fathers, but always after the age of 30.

*Cluster M6: Low/intermediate education & full-time job, with early union and later fatherhood*. This cluster is very similar to M5, but includes men with shorter education and earlier entry into the job market.

*Cluster M7: Low/intermediate education & with full-time job, with early formation of small family*. In this cluster men have an intermediate level of education or training. Compared to individuals in M6, they start a union earlier (only few with a later union dissolution) and have their first (and unique) child very early, definitively below the age of 30.

*Cluster M8: Low/intermediate education & full-time job, with early formation of medium-sized family*. This cluster includes men similar to those in M7 in terms of education, work, early union and parenthood, but most of them have two children. None in this group experience union dissolution.

*Cluster M9: Low/intermediate education & full-time job, with early union and very early formation of large family*. This group differs from the previous one only by the higher number of children (three or more).

Note that men in the first four clusters tend to remain single and childless, even if they differ with respect to the length of education or training, and to the age and the work conditions (part-time in the first cluster, full-time in the others) at their entry into the labour market. It is also of interest to note that these four clusters make up about half of the sample. Thus, a substantial part of men have very few children despite working full time. Turning to clusters from M5 to M9 more traditional patterns can be noted, though still with important differences. Men in clusters M5 and M6—who differ as for the time spent in education or training—enter their first union before the age of 30, and compared to those in the first clusters show a somewhat higher rate of (late) fatherhood. Clusters M7 to M9 are those characterized by higher fertility, and substantially differ by the number of children. Notably, the number of cases in these clusters is lower than that of cases in clusters M1 to M5, where few have children.

As already observed for women, the proportion of younger men is higher in low or late fertility clusters (Appendix 2) and the reverse holds for the last three clusters. Also in this case, the cluster characterized by an intermediate age at first childbirth compared to the others, M8, presents the smallest proportion of truncated sequences, for the same reasons illustrated for the cluster W6 in the women sample.

## 5 | MULTINOMIAL ANALYSIS: EFFECT OF PTS ON INDIVIDUALS' TRAJECTORIES

We analysed the relation between PTs and the clusters via multinomial logistic regressions, controlling for the set of socio-demographic variables introduced in Section 3 and described

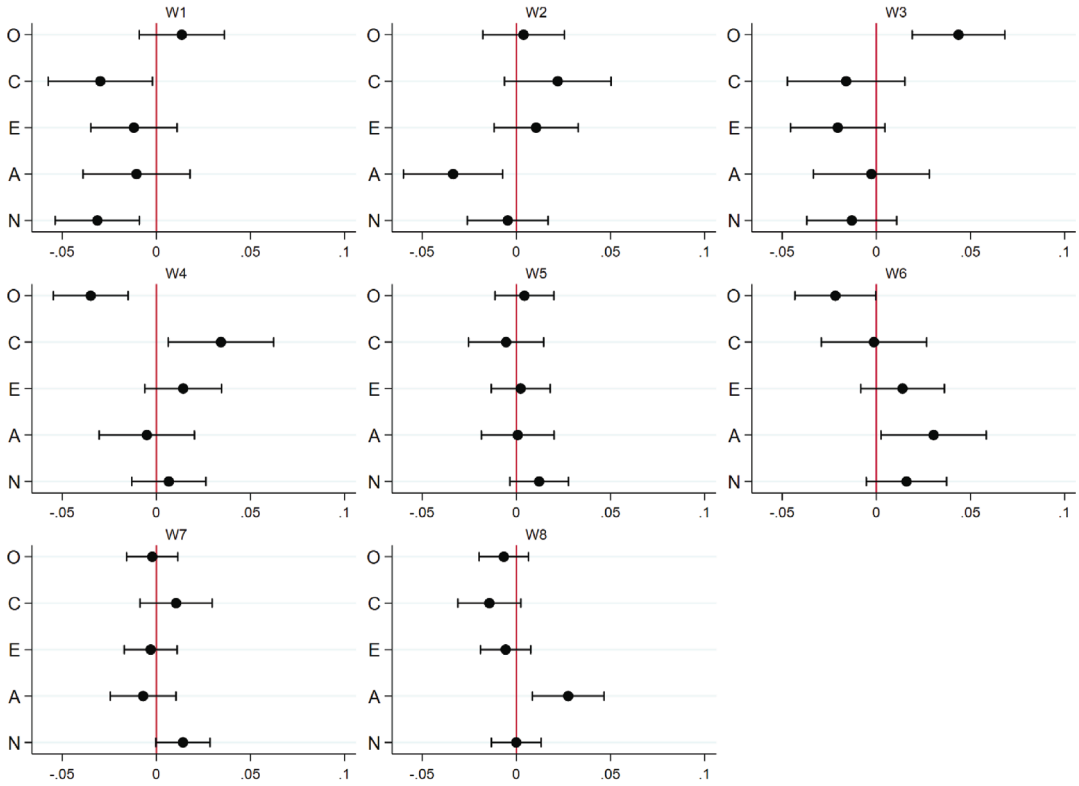


**FIGURE 5** Women: Marginal effect of PTs on the probability to belong to a specific cluster together with the 95% confidence intervals. Personality traits: (O): 'Openness'; (C): 'Conscientiousness'; (E): 'Extraversion'; (A): 'Agreeableness'; (N): 'Neuroticism'. Clusters: (W1) Long education & full-time job, with late or no family; (W2) Short education & full-time job, with late or no family; (W3) Medium education & full/part-time job, with late family; (W4) Low education & full-time job, with early union/late motherhood; (W5) Part-time job, with early union/one child; (W6) Part-time job, with early union & two children; (W7) No-work, with very early formation of medium-sized family; (W8) No-work, with very early formation of large family

in Appendix 1: birth cohort; migration background, father's and mother's level of education; household income; and geographical macro-region of residence. As discussed (Section 3), we also include among controls the individuals' age at the first measurement of the PTs scores. Figures 5 and 6 report, respectively for women and for men, the average marginal effects of one unit-increase in the scale measuring each PT trait on the probability to belong to each cluster, which are easier to interpret compared to the regression coefficients of the non-linear model.[4]

At a first glance, results in Figures 5 and 6 suggest that personality relates both with women's and with men's trajectories. For women, the probability of being in seven out of eight clusters is significantly associated to at least one PT, while for men this holds only for six out of nine clusters. Skirbekk and Blekesaune (2013) instead observed a different relevance of PTs gender-wise, with stronger effects for men. That said, it must be kept in mind that our study has peculiar features under some relevant respects. First, we analyse the relation between PTs and the *combinations*

[4]For the sake of brevity, we avoid presenting the regression results of the multinomial regression. Nonetheless, they are available upon request, together with the numerical evaluations of the marginal effects used to create Figures 5 and 6.

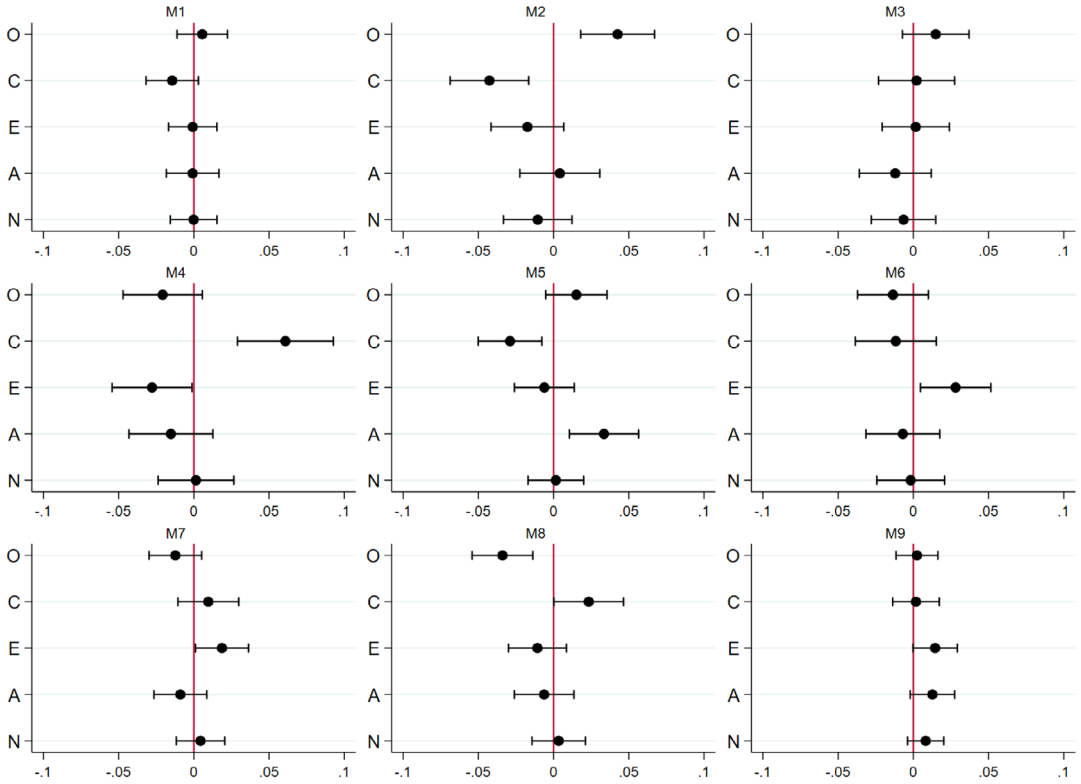


**FIGURE 6** Men: Marginal effect of PTs on the probability of belonging to a specific cluster together with the 95% confidence intervals. Personality traits: (O): 'Openness'; (C): 'Conscientiousness'; (E): 'Extraversion'; (A): 'Agreeableness'; (N): 'Neuroticism'. Clusters: (M1) Early part-time & late full-time job, with late or no family; (M2) Long education & full-time job, with late or no family; (M3) Medium educated with job discontinuity and late or no family; (M4) Low education & full-time job, with late or no family; (M5) Long education & full-time job, with early union and later fatherhood; (M6) Low/intermediate education & full-time job, with early union and later fatherhood; (M7) Low/intermediate education & full-time job, with early formation of small family; (M8) Low/intermediate education & full-time job, with early formation of medium-sized family; (M9) Low/intermediate education & full-time job, with early union and very early formation of large family

of events experienced in three domains over a portion of the life course. In addition, our sample includes more recent cohorts compared to those used in previous studies. As we hypothesized, PTs might be more relevant for younger cohorts, for whom the socio-demographic context is different. Specifically, Skirbekk and Blekesaune (2013) analysed individuals living in Norway, characterized by relatively high fertility and by strong gender equality in terms of employment. We instead rely on data collected in Germany, where fertility is generally lower and gender equality is weaker.

Focusing on the significant associations observed for women, only cluster W5 (*Part-time job*, *with early union/one child*) does not show any discernible association with any PT. The probability of belonging to W1 (*Long education & full-time job*, *with late or no family*) is negatively associated to 'Conscientiousness', whereas membership to W2 (*Short education & full-time job*, *with late or no family*) is negatively associated to 'Agreeableness'. Inclusion in W3

(*Medium education & full/part-time job*, *with late family*) instead relates positively to 'Openness'. The likelihood of experiencing a trajectory in W4 (*Low education & full-time job*, *with early union/late motherhood*) is positively associated to 'Conscientiousness', which might explain the close attachment and commitment to work for these women, and negatively to 'Openness', which may play a role in their relatively low level of education. 'Openness' negatively relates with being assigned to W6 (*Part-time job*, *with early union & two children*), even if with a borderline level of significance, whereas 'Agreeableness' is positively associated with this cluster. Cluster W7 (*No-work*, *with very early formation of medium-sized family*) is positively, though only weakly, associated with 'Neuroticism', while the last cluster, W8 (*No-work*, *with very early formation of large family*), relates strongly and positively to 'Agreeableness'. Thus, 'Agreeableness' appears to be a relevant trait for women ending up with the very traditional family form.

Among men, three clusters seem to be particularly and significantly associated to PTs. In particular, men in M2 (*Long education & full-time job*, *with late or no family*) score relatively high on 'Openness' but low on 'Conscientiousness'. Membership to M4 (*Low education & full-time job*, *with late or no family*) is strongly and positively associated to 'Conscientiousness', but negatively (though to a weaker extent) associated to 'Extraversion'. The probability of being placed in M5 (*Long education & full-time job*, *with early union and later fatherhood*) relates negatively to 'Conscientiousness' and positively to 'Agreeableness'. Cluster M6 (*Low/intermediate education & with full-time job*, *with early union and later fatherhood*) is positively associated only to 'Extraversion'. One very interesting aspect of these findings is that clusters from M1 to M4 are all characterized by strong attachment to work, and in particular, M2 *(Long education & full-time job*, *with late or no family)* to M4 *(Low education & full-time job*, *with late or no family)* consist of men with stable full-time jobs. Still, the PTs relate differently with these clusters, possibly because of their distinctive features. Cluster M2 *(Long education & full-time job*, *with late or no family)*, for instance, relates to 'Openness' and 'Conscientiousness', whereas no significant relation with PTs is observed for clusters M3 *(Medium educated with job discontinuity and late or no family)* and M1 *(Early part-time & late full-time job*, *with late or no family* where admittedly men engaged in part-time work before entering full-time work). Another interesting point concerns the different relation with PTs observed for clusters M2 *(Long education & full-time job*, *with late or no family)* and M4 *(Low education & full-time job*, *with late or no family)*, whose key difference lies in the longer length of the education spell in M2. The average marginal effect of 'Consciousness' is negative for M2 and positive for M4, whereas 'Openness' is positively related to M2 and negatively (though not significant at the 95% level) to M4. 'Consciousness' also matters for M5 *(Long education & full-time job*, *with early union and later fatherhood)*, again with a negative effect, and indeed, we see that the length of education for men in this cluster is higher. The distinctive feature of M5 lies in its positive association with 'Agreeableness', and indeed, men in this cluster start a union at a younger age, and a number of them have children, even if at a later age.

Four clusters out nine are not associated significantly (or just with quite a low significance) with any PTs, namely M1 (*Early part-time & late full-time job*, *with late or no family*), M3 (*Medium educated men with job discontinuity and late or no family*), M7 (*Low/intermediate education & full-time job*, *with early formation of small family*) and *M9* (*Low/intermediate education & full-time job*, *with early union and very early formation of large family*). The latter group is remarkable, because it includes men with a traditional family, with early union and high fertility. Interestingly, for women we observed a positive association between 'Agreeableness' and the probability of being placed in clusters of women with traditional families. Thus, as for

traditional families with many children, the 'Agreeableness' of women apparently drives the higher fertility.

Our results are partly in line with previous literature (see e.g. Lundberg, 2012, for Germany). Even so, it is clear that by considering the relevant domains jointly we are also able to identify different relations between PTs and clusters which are similar in one domain but differ in others.

As an overarching conclusion, note that among the five traits, 'Conscientiousness' significantly and similarly matters for both genders, reducing the probability to delay (or not experience) the entry in the labour market and increasing the probability of building a solid work career for men (M2—*Long education & full-time job*, *with late or no family*, M4—*Low education & full-time job*, *with late or no family*) and also for women (W4—*Low education & full-time job*, *with early union/late motherhood*). In addition, low levels of 'Conscientiousness' seem linked also with the tendency to avoid of delay family formation (W1—*Long education & full-time job*, *with late or no family*, and M2—*Long education & full-time job*, *with late or no family* or high fertility (W8—*No-work*, *with very early formation of large family*), whereas high 'Conscientiousness' is associated with late motherhood (W4—*Low education & full-time job*, *with early union/late motherhood*). Interestingly, previous literature (Skirbekk & Blekesaune, 2013) points to a systematic, across cohorts, negative relation between 'Neuroticism', family formation and fertility. Partially in line with these findings, in our case 'Neuroticism' seems to play a role only for women, being negatively associated to W1 (*Long education & full-time job*, *with late or no family*) and positively (even if less significantly) to W7 (*No-work*, *with very early formation of medium-sized family*).

One interesting direction of further analysis refers to differences across cohorts. Both the idea of the Second Demographic Transition and that of individualization suggest that younger cohorts have a stronger sense of self-realization, and are freer from social norms. This implies that life events are much more down to individuals' personal choice. As a consequence, the relation between PTs and the unfolding of life courses should be stronger for the younger cohort than for the older ones, who were presumably more exposed to normative pressures in their family and work life. Analysing and comparing the trajectories of younger and older cohorts is complicated when attention is limited to sequences of the same length. Our approach, which caters also for truncated sequences, makes a comparison across cohorts feasible. Specifically, we classify the interviewees into three large cohorts: individuals born in 1964–69, 1970–79 and 1980–88. We then applied multinomial regressions including in the models also the interactions between PTs and cohorts. Results are reported in Appendix 7 (Figures A7.1 and A7.2). In line with previous studies (e.g. Skirbekk & Blekesaune, 2013), we find that the relation between PTs and clusters is indeed stronger for younger cohorts. In particular, for women born in the 1980s we observe a significant positive association between 'Openness' and the probability of belonging to cluster W1 (*Long education & full-time job*, *with late or no family*), and a negative effect of both 'Conscientiousness' and 'Neuroticism' on it. Also, 'Agreeableness' positively relates with cluster W8 (*No-work*, *with very early formation of large family*). Similarly, for younger men born in the 1980s 'Openness' and 'Conscientiousness' are respectively positively and negatively related to the likelihood of being part of cluster M2 (*Long education & full-time job*, *with late or no family*). Instead, membership to M4 (*Low education & full-time job*, *with late or no family*) is negatively associated with 'Openness' and 'Conscientiousness', the latter relating negatively also with M5 (*Long education & full-time job*, *with early union and later fatherhood*). For the older cohorts, instead, the relations between PTs and clusters are weaker.

## 6 | CONCLUSION

In this study, we discuss the association between PTs, here measured by the 'Big Five' inventory, and individuals' paths through eventual changes in three important domains, namely job, union and fertility, characterizing individuals' life between the ages of 20 and 40. The idea is that personality does correlate with these paths. Previous literature has focused on specific events such as the length of education, working onset and spells, union formation and dissolution and not least, childbearing. Here instead we consider the relation between PTs and the probability of experiencing specific life-course trajectories between the ages of 20 and 40. This is an important contribution, since it is not obvious that personality matters directly for individual patterns through different life-course domains. Rather, personality appears to be associated on certain types of individuals' life-course trajectories—of which education, jobs, union dynamics, and, fertility are components. In other words, personality may matter for the frequency of events across all domains, or, personality may matter in terms of postponing some event while accelerating others. Our results allow analysing the relationship between Pts and life-course from a different perspective and in a multifaceted way. For instance, our results suggest that PTs are potentially more important for women than for men. For women, the 'Conscientiousness' trait correlates with relevant aspects of their life course. Women scoring high on this trait are much less likely to experience trajectories characterized by postponement (or avoidance) of family formation and by work instability. At the same time, for women, higher 'Conscientiousness' is also associated to a lower probability to become 'dedicated to family', that is non-working, early married and with large families. Differently from Jokela et al. (2011), who finds a general negative relationship between 'Conscientiousness' and the offspring number, our results point towards a relation which is instead mediated by women's marital and labour status. Indeed, the low 'Conscientiousness' score associated with higher fertility for women, is also in part explained by weaker positions of these women in terms of family formation and work. As such, it appears that personality is related to those clusters which represent deviations from the full-time working norm. In sum, our work confirms, extends and in part refines, some of the insights highlighted by previous studies linking specific PTs to fertility. Our results suggest that those same PTs are intrinsically linked with union formation and individuals' work.

The important differences between the results produced by our study and those presented in the literature mostly rely on the general uniqueness of our work, both substantively and methodologically. On the one hand, our analysis allows considering different kinds of transitions within individuals' life-course trajectories simultaneously, instead of analysing them separately as the literature on these topics has done so far. On the other hand, we introduced an important technical innovation. As it is common in this type of studies, observed sequences might have different lengths, for a number of reasons (e.g. respondent's age or exit from the study). The method applied here addresses this relevant issue, thereby increasing the sample size by including also individuals with shorter sequences, notably belonging to younger cohorts. All in all, our study considers the complexity and interconnection of different domain careers for life-course trajectories, showing how similar PTs relate to the trajectories in the single domains and to their joint unfolding.

### ACKNOWLEDGEMENTS

The authors are grateful to Matthias Studer for interesting discussion and contribution on approaches for the treatment of truncated sequences, and in particular on the idea about the assignment of truncated trajectories to clusters built based on complete ones. To build and cluster sequences we used the R packages TraMineR (Gabadinho et al., 2011) and WeightedClusters

(Studer, 2013). The procedures to obtain plots and to build integrated dissimilarity matrices were developed in R by one of the authors. Multinomial regression was applied using Stata.

## ORCID

*Letizia Mencarini* https://orcid.org/0000-0001-5070-6069
*Raffaella Piccarreta* https://orcid.org/0000-0002-8876-3656
*Marco Le Moglie* https://orcid.org/0000-0003-3123-9818

## REFERENCES

Abbott, A. (1995) Sequence analysis: new methods for old ideas. *Annual Review of Sociology*, 21(1), 93–113.
Aberg, Y. (2009) The contagiousness of divorce. In: Hedstroem, P. & Bearman, P. (Eds.) *Analytical psychology: women more likely to initiate divorce*. New York: Oxford University Press.
Alvergne, A., Jokela, M. & Lummaa, V. (2010) Personality and reproductive success in a high-fertility human population. *Proceedings of the National Academy of Sciences*, 107(26), 11745.
Attanasio, O., Blundell, R., Conti, G. & Mason, G. (2020) Inequality in socio-emotional skills: a cross-cohort comparison. *Journal of Public Economics*, 96(4), 898–912.
Caprara, G.V. & Cervone, D. (2000) *Personality: determinants, dynamics, and potentials*. New York, NY: Cambridge University Press.
Cobb-Clark, D.A. & Schurer, S. (2012) The stability of big-five personality traits. *Economics Letters*, 115(1), 11–15.
Cygan-Rehm, K. & Maeder, M. (2013) The effect of education on fertility: evidence from a compulsory schooling reform. *Labour Economics*, 25, 35–48.
Diener, E. & Lucas, R.E. (1999) Personality and subjective well-being. In: Kahneman, D., Diener, E. & Schwarz, N. (Eds.) *Well-being: the foundations of hedonic psychology*. New York: Russell Sage Foundation, pp. 213–229.
Dijkstra, P. & Barelds, D.P. (2009) Women's well-being. The role of individual differences. *Scandinavian Journal of Psychology*, 50(4), 309–315.
Eaves, L.J., Martin, N.G., Heath, A.C., Hewitt, J.K. & Neale, M.C. (1990) Personality and reproductive fitness. *Behavior Genetics*, 20(5), 563–568.
Elder, G.H. (1985) Perspectives on the life course. In: Elder, Jr G.H. (Ed.) *Life course dynamics, trajectories, and transitions, 1968–1980*, Ithaca, NY: Cornell University Press, pp. 23–49.
Everitt, B.S., Landau, S., Leese, M. & Stahl, D. (2003) *Cluster analysis*, 5th Edition, Hoboken, NJ: John Wiley & Sons.
Gabadinho, A., Ritschard, G., Studer, M. & Müller, N.S. (2011) *Mining sequence data in R with the TraMineR package: a user's guide*. Geneva: University of Geneva.
Gensowski, M., Gørtz, M. & Schurer, S. (2020) Inequality in personality over the life cycle. IZA Discussion Papers 13378, Institute of Labor Economics (IZA).
Goldberg, L.R. (1981) Language and individual differences: the search for universals in personality lexicons. In: Wheeler, L. (Ed.), *Review of personality and social psychology*. Vol. 2. Beverly Hills, CA, Sage, pp 141–165.
Gosling, S.D., Rentfrow, P.J. & Swann, W.B. Jr (2003) A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6), 504–528.
Gower, J. & Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, 3(1), 5–48.
Heineck, G. (2011) Does it Pay to Be Nice? Personality and Earnings in the United Kingdom. *ILR Review*, 64(5), 1020–1038. https://doi.org/10.1177/001979391106400509
Hewitt, B., Western, M. & Baxter, J. (2006) Who decides? The social characteristics of who initiates marital separation. *Journal of Marriage and the Family*, 68, 1165–1177.
Jackson, D.N. (1984) *Personality research form manual*. Port Huron, MI: Research Psychologist Press.
Jokela, M. (2012) Birth-cohort effects in the association between personality and fertility. *Psychological Science*, 23(8), 835–841.
Jokela, M., Alvergne, A., Pollet, T.V. & Lummaa, V. (2011) Reproductive behavior and personality traits of the five factor model. *European Journal of Personality*, 25(6), 487–500.
Jokela, M., Kivimäki, M., Elovainio, M. & Keltikangas-Järvinen, L. (2009) Personality and having children: a two-way relationship. *Journal of Personality and Social Psychology*, 96(1), 218–230.

Judge, T.A., Higgins, C.A., Thoresen, C.J. & Barrick, M.R. (2006) The big five personality traits, general mental ability, and career success across the life span. *Personnel Psycology*, 52(3), 6212–6652.

Kaufman, L. & Rousseeuw, P. (1990) *Finding groups in data: an introduction to cluster analysis*. Hoboken, NJ: John Wiley & Sons.

Kiernan, K. (1986) Teenage marriage and marital breakdown: a longitudinal study. *Population Studies*, 40(1), 35–54.

Kinnunen, U., Rytkonen, O., Miettinen, N. & Pulkkinen, L. (2000) Personality and marriage: personality characteristics as predictors of marital quality and stability. *Psykologia*, 35(4), 332–345.

Le Moglie, M., Mencarini, L. & Rapallini, C. (2015) Is it just a matter of personality? On the role of subjective well-being in childbearing behavior. *Journal of Economic Behavior & Organization*, 117, 453–475.

Legendre, P. & Anderson, M.J. (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, 69, 1–24.

Lillard, L. & Waite, J. (1993) A joint model of marital childbearing and marital disruption. *Demography*, 30, 653–681.

Lowell, K.E. & Conley, J.J. (1987) Personality and compatibility: a prospective analysis of marital stability and marital satisfaction. *Journal of Personality and Social Psychology*, 52(1), 27–40.

Lucas, R.E. & Donnellan, M.B. (2011) Personality development across the life span: longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology*, 101(4), 847–861.

Lundberg, S. (2009) Psychological traits and family outcomes. IZA, mimeo.

Lundberg, S. (2012) Personality and marital surplus. *IZA Journal of Labor Economics*, 1(1), 1–21.

Lundberg, S. (2018) Non-cognitive skills as human capital. In: Hulten, C.R. & Ramey, V.A. (Eds.) *Education, skills, and technical change: implications for future US GDP growth*. NBER/CRIW, Chicago, IL: University of Chicago Press, pp. 219–243.

McCrae, R.R. & Allik, L. (Eds.) (2002) *The five-factor model of personality across cultures*. Berlin: Springer.

Miller, W.B. (1992) Personality traits and developmental experiences as antecedents of childbearing motivation. *Demography*, 29(2), 265–285.

Mueller, G. & Plug, E. (2006) Estimating the effect of personality on male and female earnings. *Industrial and Labor Relations Review*, 60(1), 3–22.

Nyhus, E. & Pons, E. (2005) The effects of personality on earnings. *Journal of Economic Psychology*, 26, 363–384.

Ozer, D.J. & Benet-Martinez, V. (2006) Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, 57, 401–421.

Piccarreta, R. (2017) Joint sequence analysis: association and clustering. *Sociological Methods & Research*, 46(2), 252–287.

Piccarreta, R. & Studer, M. (2019) Holistic analysis of the life course: methodological challenges and new perspectives. *Advances in Life Course Research*, 41, 100251.

Pollock, G. (2007) Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis. *Journal of the Royal Statistical Society. Series A*, 170(1), 167–183.

Rand, W.M. (1971) Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336), 846–850.

Roberts, B.W., Kuncel, N.R., Shiner, R., Caspi, A. & Goldberg, L. (2007) The power of personality: the comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), 313–345.

Seibert, S.E., Kraimer, M.L. & Liden, R.C. (2001) A social capital theory of career success. *Academy of Management Journal*, 44, 219–237.

Skirbekk, V. & Blekesaune, M. (2013) Personality traits increasingly important for male fertility: evidence from Norway. *European Journal of Personality*, 28(6), 521–529. Article first published online: 5 August 2013: https://doi.org/10.1002/per.1936

Specht, J., Egloff, B. & Schmukle, S.C. (2011) Stability and change of personality across the life course: the impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology*, 101(4), 862–882.

Specht, J., Egloff, B. & Schmukle, S.C. (2013) Everything under control? The effects of age, gender, and education on trajectories of perceived control in a nationally representative German sample. *Developmental Psychology*, 49(2), 353–364.

Studer, M. (2013) WeightedCluster library manual: a practical guide to creating typologies of trajectories in the social sciences with R. LIVES Working Papers, NCCR LIVES 24.
Studer, M. & Ritschard, G. (2016) What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Association A*, 179, 481–511.
Tavares, L.P. (2016) Who delays childbearing? The associations between time to first birth, personality traits and education, *European Journal of Population*, 32(4), 575–597.
Van de Kaa, D.J. (1987) Europe's second demographic transition. *Population Bulletin*, 42, 1–59.
Wagner, G.G., Frick, J.R. & Schupp, J. (2007) *The German socio-economic panel study (SOEP): scope, evolution and enhancements, SOEP papers on multidisciplinary panel data research*. No. 1, Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).
Whyte, S., Brooks, R.C., Chan, H.F. & Torgler, B. (2019) Do certain personality traits provide a mating market competitive advantage? Sex, offspring & the big 5. *Personality and Individual Differences*, 139, 158–169.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Mencarini, L., Piccarreta, R. & Le Moglie, M. (2022) Life-course perspective on personality traits and fertility with sequence analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1–26. Available from: https://doi.org/10.1111/rssa.12832