1235
2025

SOEPpapers
on Multidisciplinary Panel Data Research

Order out of chaos: A specification curve analysis of age and wellbeing

Kausik Chaudhuri and Alan Piper

DIW BERLIN

SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.
The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.
Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at **http://www.diw.de/soeppapers**

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)
DIW Berlin
Anton-Wilhelm-Amo-Strasse 58
10117 Berlin, Germany
Contact: soeppapers@diw.de

Order out of chaos: A specification curve analysis of age and wellbeing

Kausik Chaudhuri

Economics Department, University of Leeds

Alan Piper

Economics Department, University of Leeds

November 2025

Abstract: The empirical literature on the relationship between age and well-being is characterised by an unusually persistent series of disagreements over data, method, and interpretation. Previous attempts to advance the discussion have involved different scholars' specific prescriptions, which were often in near total contradiction to other scholars' attempts to do the same. Instead, we use specification curve analysis to provide a structured and transparent resolution to these disputes. This also helps to illuminate the sensitivity of findings to key analytical decisions. With twenty-five years of panel data from the UK and Germany, we show that most of the specifications are consistent with a turning point for wellbeing in midlife, with a decline to that point and increase thereafter. The consistency of the finding renders some of the previous debate moot. Furthermore, this robust result is supportive of much theoretical discussion from different disciplines and areas of enquiry.

JEL codes: I30

Key words: Age, Ageing, lifespan development, wellbeing

Email addresses: K.Chaudhuri@leeds.ac.uk, A.T.Piper@leeds.ac.uk (corresponding author)

Acknowledgements: We thank Nick Adnett, Clemens Hetschko, Peter Howley, Margie Lachman, Anthony Lepinteur, Svenja Miltner, Geoff Pugh, and Ying Zhou for valuable comments, as well the audience of an ISQOLS webinar, participants of the 6th BeWell workshop, and seminars at the University of Leeds and the University of Surrey. We declare no conflict of interest.

Order out of chaos: A specification curve analysis of age and wellbeing

> "[Our] literature review… discloses a puzzling pattern of mixed empirical evidence. This is especially disturbing since the question on how happiness is affected by ageing is a fundamental question on the "conditio humana". The social and behavioral sciences seem unable to answer such a fundamental question despite much research effort!" Kratz and Brüderl (2021, p.2)

> "Because of different intellectual traditions, occasionally the debate in related literatures has been at cross-purposes and perhaps needlessly confrontational." Giuntella et al. (2022 p.66)

1. Introduction and background

One of the most extensively studied—and contested—relationships in the subjective well-being literature is the link between age and well-being. First explored in psychology, this topic has since attracted sustained interest from economists and other social scientists. The most common empirical finding is a midlife low or nadir reflecting a U-shaped trajectory in which self-reported well-being declines through early and middle adulthood before rising again in later years. However, this pattern is neither universal nor uncontroversial. As Tobias and Bond (2025) note, the age–well-being relationship is arguably the most debated topic in the well-being literature; not only because of conflicting empirical results, but also divergent methodological choices, and underlying theoretical ambiguities. In recent decades, the availability of large-scale survey data has enabled more systematic empirical investigations, but consensus has remained elusive. In this paper, we aim to bring clarity to this debate using specification curve analysis (SCA), a method designed to transparently map the robustness of empirical findings across the full range of defensible analytic choices.

Specification curve analysis is designed to resolve empirical debates. As Simonsohn et al. (2020) argue, SCA is especially useful when empirical findings are sensitive to specification choices, as is often the case in contentious literatures. Beyond introducing the methodology and its conceptual underpinnings, Simonsohn et al. (2020) demonstrate its practical utility in clarifying a number of disputed findings. We briefly recount one such example here, given its similarity to the age–well-being methodological debate. The case in question concerns a study by Jung et al. (2014), published in Proceedings of the National Academy of Sciences, which reported that hurricanes with female names resulted in significantly higher death tolls than those with male names—an effect attributed to implicit gender biases in risk perception. This claim sparked considerable academic controversy, including a series of critical commentaries, author responses, and further rebuttals. Voracek et al. (2019) describe this exchange as a "daisy chain" of arguments, many of which hinged on diverging methodological assumptions or preferences and statistical choices, often without resolution. By applying specification curve analysis, Simonsohn et al. (2020) were able to clarify which results were robust and which were dependent on contested analytic decisions—effectively settling the debate. This example illustrates the value of SCA not only as a diagnostic tool, but also as a means of resolving empirical disputes that might otherwise persist indefinitely.[1] Other studies undertaking the same exercise for various

[1] Their figure, the specification curve, illustrated that the statistically significant results reported by Jung et al. (2014) represented only a narrow slice—approximately 2%—of the possible analytical specifications. This empirical approach provided strong evidence that the original findings lacked robustness. Emphasising the potential of specification curve analysis to conclude drawn-out debates, Simonsohn et al. (2020, p. 1211) write: "PNAS could have published nearly 1,700 letters showing individual specifications that make the effect disappear… It also could have published 37 responses with individual specifications showing the robustness of

psychological and, increasingly, economic debates are Akey et al. (2021), Frey et al. (2021), Gao et al. (2021), Han et al. (2021), Rauvola and Rudolph (2023) and Kleven (2024). To our knowledge, this is the first use of specification curve analysis in the broad wellbeing literature.[2]

Returning to the debate about age and wellbeing, one central point of contention is the appropriate control set for analyses. A foundational contribution, Blanchflower and Oswald (2008), has been described by Glenn (2009) as "potentially one of the most important studies of age and well-being that has ever been conducted." In their work, Blanchflower and Oswald control for demographic and economic variables but deliberately exclude physical health. Their results identify a midlife low in well-being—commonly referred to as the "U-shape"—in many, though not all, countries studied. Glenn (2009), however, criticised their specification, describing it as "inappropriate and questionable." He specifically challenged the inclusion of controls for marital status, education, and income, asserting that only birth cohort should be controlled for. Based on his preferred specification, Glenn found no evidence of a U-shape in American data (Glenn, 2009, p. 481). Blanchflower and Oswald (2009) responded by presenting U-shaped patterns both with and without the contested controls, reinforcing the empirical regularity of their original claim. Crucially, they acknowledged that the selection of control variables involves "a matter of methodological judgment," and allowed that "there is scope on this for reasonable people to disagree reasonably" (p. 488). This acknowledgement underscores the importance of methods—such as specification curve analysis—that can systematically explore the impact of different control sets. Since then, the debate over control variables has not abated. Bartram (2023), for example, argued for specifications with only wave controls, repeating the position of Glenn (2009). Kratz and Brüderl (2021) challenged this position, questioning both the methodological rationale, the lack of consideration of the sample, and the implications for inference.[3]

A second main argument in the debate concerns whether the observed midlife dip in well-being is a cross-sectional artefact or a longitudinal reality. Infurna et al. (2020) contend that the U-shape arises from comparing individuals of different ages at a single point in time and that longitudinal data do not support the same pattern. In contrast, Cheng et al. (2017) find evidence consistent with a U-shape using longitudinal data and methods; declining life satisfaction of the young was found longitudinally by Piper (2015). More recently, Oparina et al. (2025) and Bauer and Kaiser (2025) also present longitudinal evidence of a midlife low. For their contention, Infurna et al. rely in part on Galambos et

the findings. It would have been better to publish a single specification curve in the original paper." They exaggerate the point but highlight well the value of specification curves to resolve debates and bring order to what might seem like chaos.

[2] Specification curve analysis is not the only option, but it is arguably the best one given the nature and content of the debate. Other methods designed to handle model uncertainty and researcher choices include (i) Extreme Bounds Analysis, (ii) Bayesian Model Averaging, and (iii) Meta-Regression Analysis. All have systematic ways to identify and measure the effects of variant model specifications instead of relying on a single arbitrary model. SCA is preferred to each for the several reasons including the following: it plots the entire distribution of results (every model, every estimate, and its significance level) rather than producing a summary measure like a weighted average or the lower and upper bound; it enables an easy identification of the choices that are particularly important for a given found relationship and in doing so provides a clear interpretable indication of model uncertainty; and can – in comparison to (i) in particular – test robustness to more than different control sets, though control sets are an important source of debate in the extant literature under consideration.

[3] Interestingly, some of these studies consider their methodological choices the correct ones or the best ones and these choices are in strong opposition to the correct or best prescriptions of other scholars. As just one example, the prescriptions of Kratz and Brüderl (2021) involves specific choices regarding controls and methodology, and rules out for example the work of both Bartram (2023) and Blanchflower and Oswald (2008) and others, for quite different reasons. In fact, no other study that has investigated the age and wellbeing relationship follows their prescriptions for how it should be researched and is thus ruled out. The same could easily be said of other prior research aiming to end the debate. Much more preferable is specification curve analysis which permits a variety of methodological choices and allows easy assessment of how these choices affect the relationship found.

al. (2015), one of three studies, to support their argument, but Blanchflower et al. (2023) highlight substantial problems with the dataset they use —including attrition, irregular measurement intervals, and changes in survey mode—which cast doubt on its reliability for adjudicating on this question. In short, this is another argument not easily resolved by consulting the literature. Adding to the complexity, Kassenboehmer and DeNew (2012) report that the U-shape disappears once individual fixed effects and survey experience are controlled for in longitudinal data. Yet again, these findings were met with counter-arguments, with Blanchflower and Piper (2022), critiquing their methodology and showing a different result. The point for the present study is that, once again, the lack of consensus across these studies reinforces the value of methods that can consider the full range of plausible specifications and compare their associated outcomes.

A related, and equally unresolved, issue concerns whether the U-shape reflects true ageing (a lifecycle effect), differences across cohorts, or some interaction between the two. This concern leads into the well-known age-period-cohort (APC) identification problem, stemming from the exact linear dependency among the three variables: i.e., age = period − cohort. Empirical strategies to disentangle these effects have been proposed (see Clark 2002; Glenn 2009; De Ree and Alessie 2011; Piper 2015; Cheng et al. 2017), but consensus – again – remains elusive. Each study tends to rest on assumptions or constraints that others find problematic. As a result, APC debates remain open-ended, further muddying conclusions about the true shape of the relationship between age and subjective well-being.

In sum, the literature on age and well-being is characterised by an unusually persistent series of disagreements over data, method, and interpretation. The accumulation of contradictory claims has arguably hindered theoretical progress and policy application. We bring order to this apparent chaos with our specification curve analysis, which most often finds wellbeing declining to a low in midlife before increasing again; a result that is an ageing effect as well as being found by a comparison of the wellbeing of individuals of different ages. In the remainder of this paper, we show that specification curve analysis can provide a structured and transparent resolution to these disputes by illuminating the sensitivity of findings to key analytical decisions. Specifically, in Section 2, after we introduce the data, we discuss the specification curve solutions to resolving the debate, along with other sources of contention, in the empirical literature. Section 3 presents the results. Section 4 is a discussion and Section 5 offers concluding remarks.

2. Data and methodology

This section has two parts. Firstly, we introduce the data, and provide a description of the datasets we use, which are nationally representative panel surveys: the British Household Panel Survey (BHPS); Understanding Society (UKHLS); and the German Socioeconomic Panel (SOEP). Secondly, we develop the discussion on the empirical debate regarding age and wellbeing and explain how we use this data to account for these key aspects within the specification curve analysis framework, in pursuit of robust findings for the relationship.

2.1 The data

The data we use come from surveys of individuals in two large European countries, the UK and Germany. The UK data are nationally representative and come from a harmonised British Household Panel survey and follow up Understanding Society survey; and the German data is the nationally representative German Socioeconomic Panel. They are constructed with multi-stage random sampling, and contain much information about individuals and households, collected annually. They are well suited for investigations of age and wellbeing. Frijters and Beatton (2012), for example, describe these datasets (the BHPS and the SOEP) as rich and appropriate for the purposes of investigating the relationship between age and wellbeing. See Benzeval and Crossley (2023) and Goebel et al. (2019) for more information on the UK and German data respectively. We use all waves since 1996, the first wave of the UK data with wellbeing information, up until 2021.[4] Descriptive statistics are presented in Table 1. Alongside dummy variables are life satisfaction as a percentage of maximum possible (POMP) scores, age, and real income.

[4] The particular UK dataset used is University of Essex, Institute for Social and Economic Research (2024) Understanding Society: Waves 1-14, 2009-2023 and Harmonised BHPS: Waves 1-18, 1991-2009, 19th Edition, UK Data Service. SN: 6614, DOI: http://doi.org/10.5255/UKDA-SN-6614-20; the particular German dataset used is Socio-Economic Panel (SOEP), data for years 1984-2021, SOEP-Core v38, EU Edition, 2023, doi:10.5684/soep.core.v38eu

Table 1: Descriptive Statistics, person-year observations (1996-2021) for the UK and Germany

		UK			Germany		
	Nobs	Mean	SD		Nobs	Mean	SD
Life Satisfaction (POMP score)	457,893	68.771	23.595	Life Satisfaction (POMP score)	481,558	71.951	17.353
Age	457,893	44.582	14.490	Age	481,558	43.975	14.095
Employment Status				Employment Status			
Self-employed	457,893	0.086	0.281	Self-employed	481,558	0.067	0.250
Employed	457,893	0.581	0.493	Employed	481,558	0.524	0.499
Unemployed	457,893	0.045	0.208	Unemployed	481,558	0.074	0.262
Retired	457,893	0.132	0.338	Retired	481,558	0.123	0.328
Family care	457,893	0.067	0.250	Not Working	481,558	0.089	0.284
Sick	457,893	0.042	0.200	In Education	481,558	0.044	0.206
Others (as base)	457,893	0.047	0.210	Others (as base)	481,558	0.079	0.270
Marital Status				Marital Status			
Single	457,893	0.208	0.406	Single	481,558	0.257	0.437
Married	457,893	0.685	0.465	Married	481,558	0.606	0.489
Separated/Divorced	457,893	0.084	0.278	Separated/Divorced	481,558	0.111	0.314
Others (as base)	457,893	0.023	0.151	Others (as base)	481,558	0.026	0.159
Household Income (Real, month)	457,893	4137.14	2908.41	Household Income (Real, month)	481,558	4217.35	4401.91
Household Size				Household Size			
One member	457,893	0.120	0.325	One member	481,558	0.120	0.325
Two members	457,893	0.327	0.469	Two members	481,558	0.320	0.467
Three-four members	457,893	0.424	0.494	Three-four members	481,558	0.425	0.494
Others (as base)	457,893	0.129	0.334	Others (as base)	481,558	0.135	0.342
Educational Status				Educational Status			
No education	457,893	0.101	0.301	Secondary	481,558	0.189	0.391
University education	457,893	0.255	0.436	Secondary Complete	481,558	0.403	0.490
Other higher education	457,893	0.117	0.322	Intermediate	481,558	0.123	0.329
A-level	457,893	0.227	0.419	University	481,558	0.233	0.423
GCSE	457,893	0.214	0.410	Less than University	481,558	0.053	0.224
Others (as base)	457,893	0.086	0.280				

Table 1 shows that these nationally representative samples from the UK and Germany are rather similar with the biggest difference being related to marital status: German respondents are more likely to be single and less likely to be married than those in the UK.[5] Unemployment is more prevalent in the German sample, with a corresponding higher rate of employment and self-employment in the UK. The average age is very similar, and those in the UK report slightly more life satisfaction. For our analysis we calculate Percentage of maximum possible (POMP) scores (generated as: [life satisfaction -min(life satisfaction)/max(life satisfaction)- min(life satisfaction)]*100) so that the specification curves for each country are directly comparable.

2.2 Methodology

Specification Curve Analysis (SCA) works by first identifying all defensible analytical decisions—those debated in the literature—that could be made when testing a hypothesis. It then runs combinations of these decisions to generate a comprehensive set of results, plotting them to show how effect sizes and statistical significance vary across specifications. This process reveals whether findings are robust across reasonable choices or rather hinge on specific, potentially biased decisions, thereby improving transparency and credibility in empirical research. Bryan et al. (2019) similarly state that specification curve analysis is "designed with the express purpose of maximizing transparency and minimizing the influence of arbitrary analytical choices…" (p.25536). Given the variety of opinions regarding the range of possible analytical choices central to the debates about the age and well-being relationship, specification curves are clearly useful. Rather than adjudicating on which scholar's opinion is correct about control variables, for example, specification curves can help the researcher move towards playing the role of a 'disinterested scientist' and show the results for all judgements at the same time, allowing all possibilities. This is the 'show everything' transparency of specification curve analysis; this transparency also makes possible an indication of the importance of various methodological choices.

Our discussion of the various issues within the general debate about age and wellbeing began with differing opinions about the choice of control variables. These perspectives were quite varied, reflecting differing methodological judgements (and presumably disciplinary backgrounds). Rather than weighing in on this debate and favouring one or another approach, specification curves consider all perspectives and show how they influence the outcome. Our 'show everything' solution with respect to control variables is to vary the contents of the control set employed in each specification. In practice this means that, as well as a baseline set with no control variables, we allow the specifications to include various combinations (including all and none) of the following: income, employment, education, marital status, and household size.

With respect to the claim that the midlife low (or U-shape) finding is only a cross-sectional one (and not a longitudinal, or ageing one), we allow our specifications to vary – together with the control set variation – by whether they consider individual fixed effects or not. This can provide evidence for or against this claim: if it is correct, we will see a fundamentally different outcome for the age-wellbeing relationship specifications when fixed effects are included than when not. This helps us address another aforementioned debate too: individual fixed effects controls for birth year, so when they are considered, we are dealing with a lifecycle effect having taken into account influences common to a cohort. In such specifications, we can learn how wellbeing changes as an individual ages (though see the discussion below about the age-period-cohort issue). In contrast, cross-section estimates compare the wellbeing of different people at different ages at a particular point or in time, with no individuals

[5] Official statistics support this. See Office of National Statistics (2023) and Statisches Bundesamt Destatis (2025).

ageing within the dataset. Too often discussions of the literature make simplistic comparisons between studies of age and wellbeing that are fundamentally different in these (and other) ways and conclude that prior work has little to say, part of the chaos of our title. For our analysis, the obtained coefficients come from regressions that either take no fixed effects into account, just individual fixed effects into account, or estimates that consider both individual and regional fixed effects; this enables an easy comparison and can inform about this particular aspect of the overall debate. Furthermore, we vary our estimations by different waves also capturing different cohorts. In doing so, we use eight different time periods, as well as the full range of data.

There are some other issues not discussed in the introduction. We turn to these now. Previous research has sometimes found that individuals give elevated life satisfaction scores in their first appearance in panel data (e.g. Ehrhardt et al., 2000). This is potentially important for age and wellbeing investigations; in particular, because many individuals have their first appearance when they are old enough to enter the sample. This might mean individuals reporting more life satisfaction at the start of adulthood contributing, at least partially, to an approximate U-shape finding. [6] Some studies do not consider this possibility; here we consider it in every specification. To do so we include a dummy variable indicating whether or not it is an individual's first appearance in the panel. The inclusion of this dummy variable does not really alter the results so, unlike with the other matters of methodological judgement, we do not allow this to vary, and instead always control for first appearance in the panel. Doing so would double the main specifications from 750 to 1500, without any additional explanatory benefit.[7]

Importantly, we never consider the relationship between age and wellbeing to have a quadratic functional form, the source of much – if rather outdated – consternation; thus, all of our subsequent results do not suffer from the popular critique that the uncovered relationship is a statistical artefact, formed because of the use of age and age-square. Rather than using a quadratic functional form, we treat age as linear and split the sample up into groups. For the main curves, we show results for 18-50 and 51-70; this split was chosen because fixed effects estimations, in several specifications and for both countries, treating age as a quadratic show a turning point at around 50. However, we also present results for other age ranges too including those over 70, specifically: 18-30, 18-35, 18-45, 31-50, 51-60, 51-65, 71 and above. With our main results, i.e. 18-50 and 51-70, if a midlife low, or an approximate U-shape, is found we would expect the obtained coefficients for age to be negative for the first age group, and positive for the second. This is similar to Cheng et al. (2017), who make use of the linear properties of a quadratic and find longitudinal evidence of a midlife low while avoiding the still common accusation of quadratic bias. As briefly discussed at the end of the results section, we also allow the method to vary and estimate a fixed effects ordered logit model. The results are consistent with our main results obtained from treating age as linear.

In general, we have the following equation, which is estimated 750 times with slight changes to the precise specification to consider the variety of credible methodological choices:

[6] Our data do indeed show higher average wellbeing for the first appearance. A simple paired t-test for UK (18-70) reveals that average life satisfaction for the first appearance (POMP score: 70.68) is significantly higher than for subsequent appearances (POMP score: 69.08); a simple paired t-test for Germany (18-70) also reveals that average life satisfaction for the first appearance (POMP score: 74.23) is significantly higher than for their subsequent appearances (POMP score: 71.65).

[7] The 750 specifications come from all the possible combinations of the following: 25 sets of controls, 3 FE (none, individual, individual and regional), and 9 sub-samples and (1) full sample i.e. 25*3*10.

$$lfsat_{it} = \alpha_i + \beta age_{it} + \gamma' X_{it} + e_{it} \quad (1)$$

In sum, our set of control variables X_{it} varies as outlined above. Some of the specifications include individual fixed effects α_i and some do not. We also control for regional fixed effects (with and without individual fixed effects). The time period investigated is allowed to vary too. We also either cluster the standard errors at the individual level or both the individual and region level. In total, this gives us the 750 specifications, all of which reflect plausibly credible methodological choices. Our main interest is in what β is obtained by every specification, and whether there is a consistency in the results; this β is the estimate of the linear relationship between life satisfaction and age, and assessed in this way, again and again taking into account the variety of debated methodological choices, is the transparent 'show all' approach of specification curves. This removes the arbitrariness and irreconcilable views on methodology which, as indicated above, are numerous. So far unaddressed, we discuss the age-period-cohort problem in an upcoming subsection.

3. Results

Figure 1 and 2 shows the specification curve analysis for females and males, aged 18-50, for the UK and Germany, respectively. Following the transformation of life satisfaction described above, the vertical scale reflects the point estimates for age. Each point estimate comes from a different specification, with the specific specification indicated by the dots in the lower panel of the figure, ordered by coefficient size. For specifications shown in the left-hand side panel, i.e. for the UK, 750 out of 750 models have negative point estimates significant at the 10% level (hence the curve being fully blue), 700 are significant at the 1% level, with the coefficients ranging from -0.871 to -0.096. As an example of how to read it, the first of the 750 specifications on Figure 1 (UK), and therefore the one with the largest negative coefficient, has a control set which includes income, marital status and household size, controls for fixed effects (both individual and region), and covers waves 17-21. The other blue dots represent all of the other possible specifications mentioned above: a control set consisting of none, some or all of the controls, which take into account income (Inc, in the figures), employment (Employ), education (Educ), marital status (Marital), household size (HH Size); no fixed effects, individual fixed effects only, or individual and region fixed effects, and a variety of waves including all. Note well that all of these include a dummy variable for first appearance in the sample, however all results are very similar if this is not included. The specification curve for Germany, the right-hand side panel, and is like that for the UK with obtained coefficients for linear age ranging from -1.026 to 0.257. Here, some of the coefficients are not significantly different from zero (the red points), and some are positive. Of the 750 specifications, 686 are negative with 570 of these being statistically significant at the 10% level; 500 are statistically significant at the 1% level and 537 at the 5% level. The bottom panel of this curve indicates that the specifications that only include waves 13-17 appear to be responsible for the positive coefficient. This unusual result could be seen as a reminder to not generalise too firmly from cross-sectional work, or work from one particular short era, for overall relationships. The overall pattern, for both countries, indicates that, on average, wellbeing declines as people age, at least until 50 years old.

Figures 1 and 2: specification curve analysis for the UK, 18-50, Harmonised BHPS and Understanding Society data 1996/7 – 2021/2; and Figure 2: Germany, 18-50, German Socioeconomic Panel (1996-2021)

Fig 1: Life Satisfaction & Age (UK: 18-50)

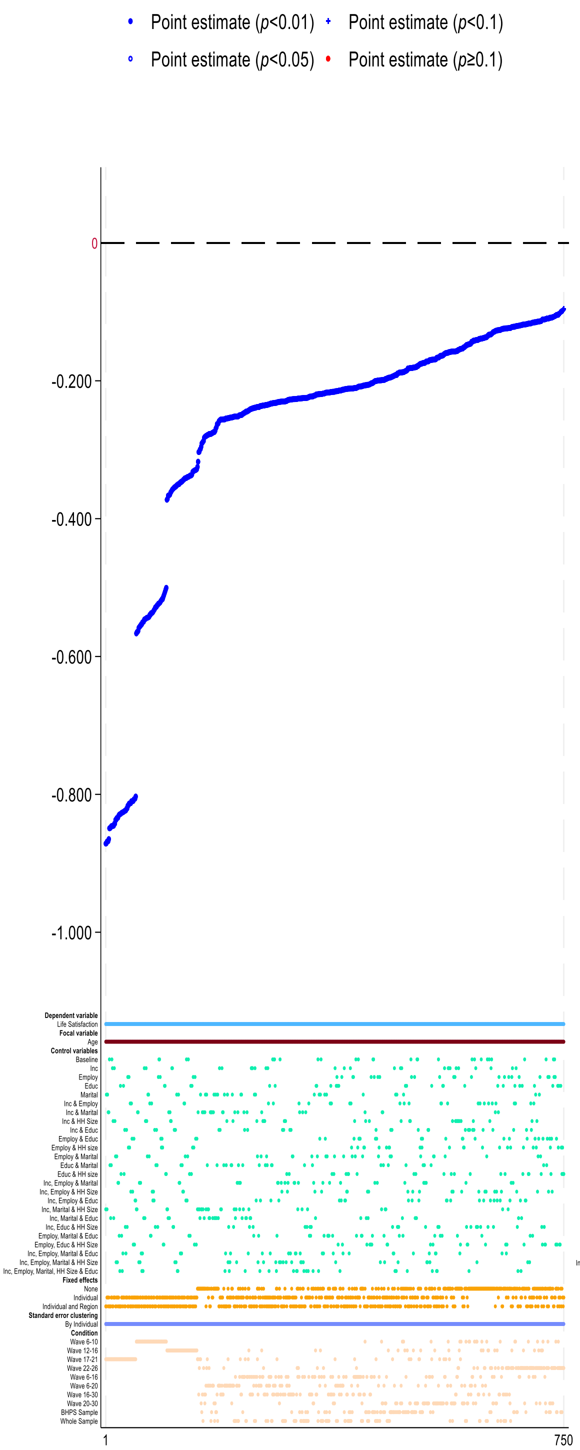


Fig 2: Life Satisfaction & Age (Germany: 18-50)

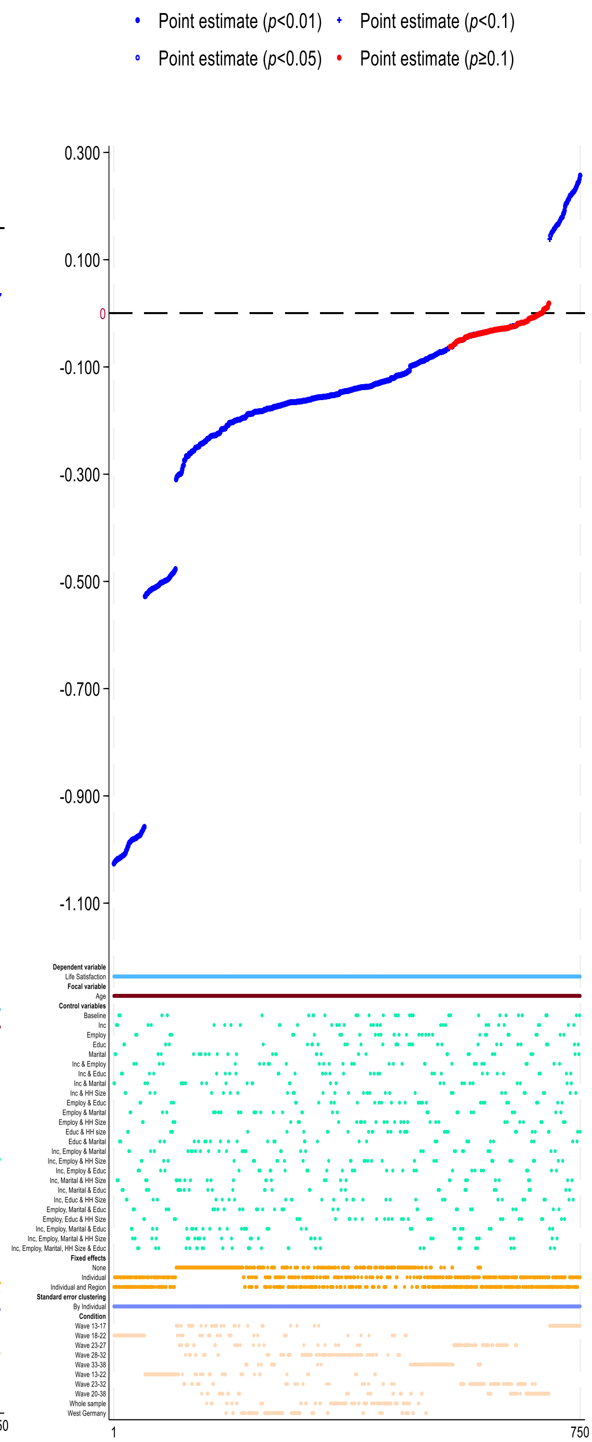


Figure 3 shows the specification curves for ages 51-70. In general, the majority finding for the UK is positive indicating that wellbeing increases, on average, as one ages from 51 to 70. This represents 585 of the 750 specifications. The full range of point estimates for the obtained coefficients goes from -0.763 to 0.824. The obtained negative coefficients appear to reflect some particular waves, 6-10, and 17-21. The insignificant coefficients also reflect specifications covering a small range of waves. For

Germany, the specification curve of Figure 4 indicates a majority finding for a positive coefficient on linear age, with 650 of the full 750 being positive. A large majority of these 650 are statistically significant at conventional levels. Some of the specifications result in negative and significant findings that relate to wave restrictions. Indeed, the wave restrictions are more responsible for the found variety in the coefficient size than the control set used, where no patterns seem to emerge. A pattern is found regarding the use of fixed effects: broadly, higher coefficients are found for the specifications that do not control for them. This indicates that the overall effect (based on between and within variation) is larger than the pure ageing effect (based on within variation only).

Figures 3 and 4: specification curve analysis for the UK, 51-70, Harmonised BHPS and Understanding Society data 1996/7 – 2021/2, and Figure 4: Germany, 51-70, German Socioeconomic Panel (1996-2021)

Fig 3: Life Satisfaction & Age (UK: 51-70)

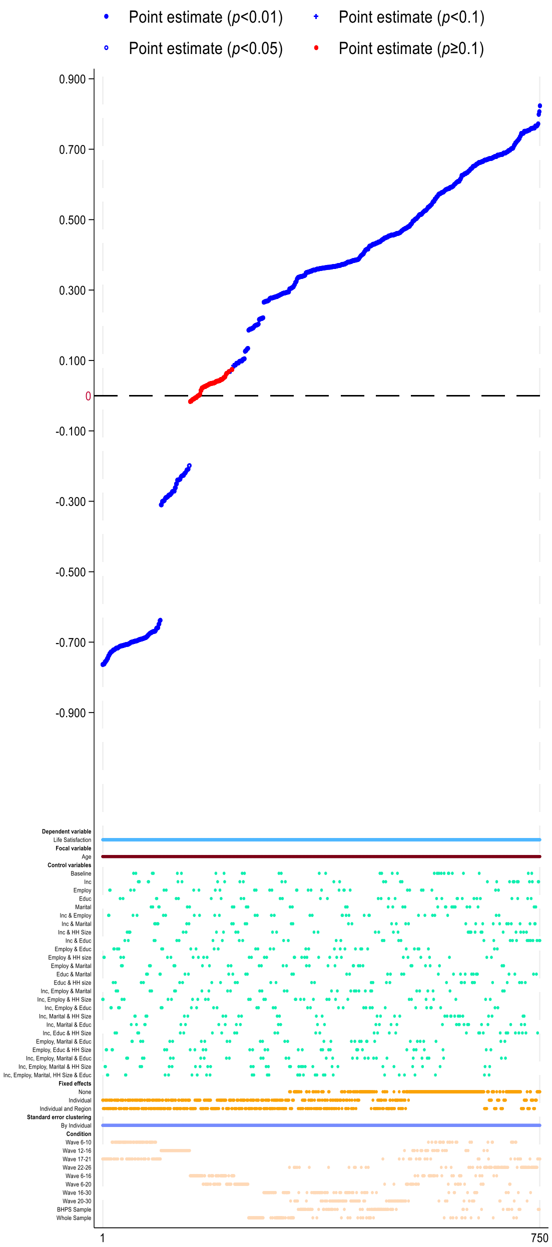


Fig 4: Life Satisfaction & Age (Germany: 51-70)

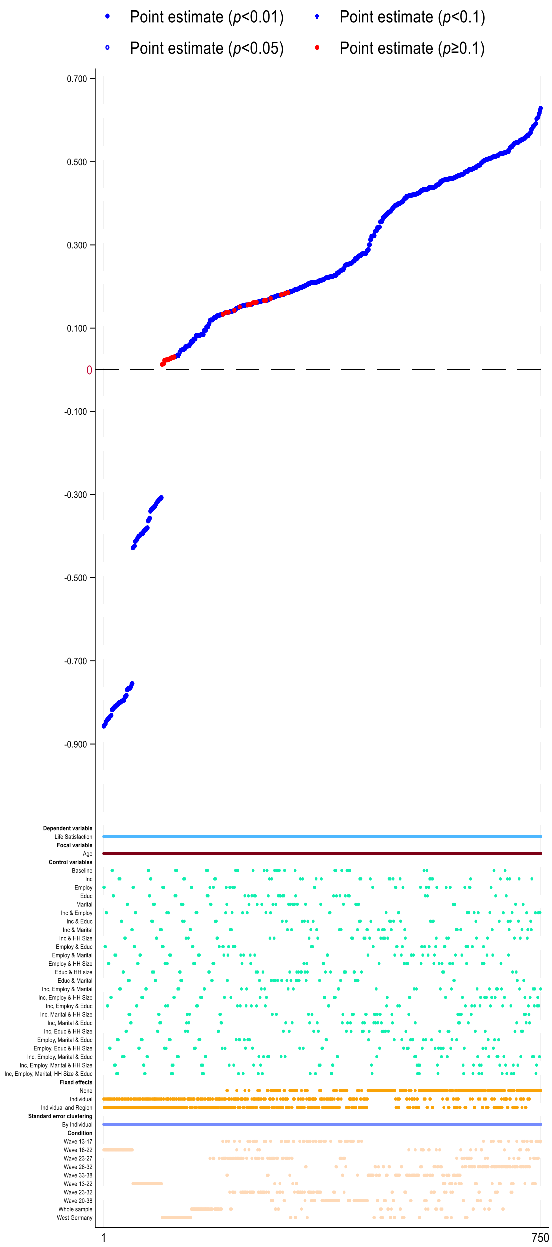


Table A1 in the Appendix shows the outcomes for different age ranges and genders. The pattern is supportive of the findings from the figures above: life satisfaction, in most specifications, declines in the first half of adulthood before turning upwards. Of note is the decline – again broadly – in life satisfaction of those over 70, thus supporting the analysis of Wunder et al. (2013) and others. There is limited difference by gender, though more coefficients obtained for females are statistically insignificant at the youngest age range (18-30), though the majority finding is negative and statistically significant. One reason for this might be the happiness bump found in some datasets at around 30, discussed below and as clearly seen in the de Ree and Alessie (2011) correction graphs also below (Figures 5 and 6).

(iii) The Age-Period-Cohort identification problem and the de Ree and Alessie correction.

Both cross-sectional and longitudinal studies finding a U-shaped and other patterns between life satisfaction and age can suffer from a unique identification problem for age. This is due to the linear dependencies between age, period, and cohort effects, with age being the difference between survey (calendar) year and year of birth. See in particular, de Ree and Alessie (2011), but also Ferrer-i-Carbonell and Frijters (2004), McKenzie (2006), Glenn (2009), Wunder et al. (2013), and Beja (2018). Although the use of panel data enables an analysis to control for differences across cohorts by taking into account individual fixed effects (i.e. the fixed effects controls for the individual's year of birth (cohort), the issue of precise identification of age and time effects remains. We follow de Ree and Alessie (2011) and estimate the following equation to solve the identification problem[8]:

$$lfsat_{it} = \alpha_i + \beta age_{it} + \gamma' X_{it} + \sum_{k=t_{min}+2}^{t_{max}} \widetilde{\phi_k} \widetilde{D_k^T}(t) + \sum_{\partial=age_{min}+2}^{age_{max}} \widetilde{\theta_\partial} \widetilde{D_\partial^A}(age_{it}) + e_{it} \qquad (2)$$

where we define

$$\widetilde{D_k^T}(t) = D_k^T(t) + (k - t_{min} - 1) D_{t_{min}}^T(t) - (k - t_{min}) D_{t_{min}+1}^T(t); k = t_{min} + 2, \ldots., t_{max}$$

$$\widetilde{D_\partial^A}(age) = D_\partial^A(age_{it}) + (\partial - age_{min} - 1) D_{a_{min}}^A(age_{it}) - (\partial - age_{min}) D_{age_{min}+1}^A(age_{it});$$

$$\partial = age_{min} + 2, \ldots., age_{max}$$

For our main estimations, we have $t_{min} = 1996, t_{max} = 2021, a_{min} = 18, and\ a_{max} = 70$.[9]

The discussion above regarding the control sets and individual fixed effects also applies here, and our interest remains in the coefficients each individual specification obtains for β, the marginal effect of age specified as a linear variable. In this way, we can control for cohort effects and thus compare the outcomes for those specifications where such a control is in place and those where it is not; this aids us in discussing the robustness of the age and wellbeing relationship. The de Ree and Alessie (2011)

[8] These normalizations imply the coefficients on the dummy variables sum to zero over the available range and are orthogonal to a linear trend.

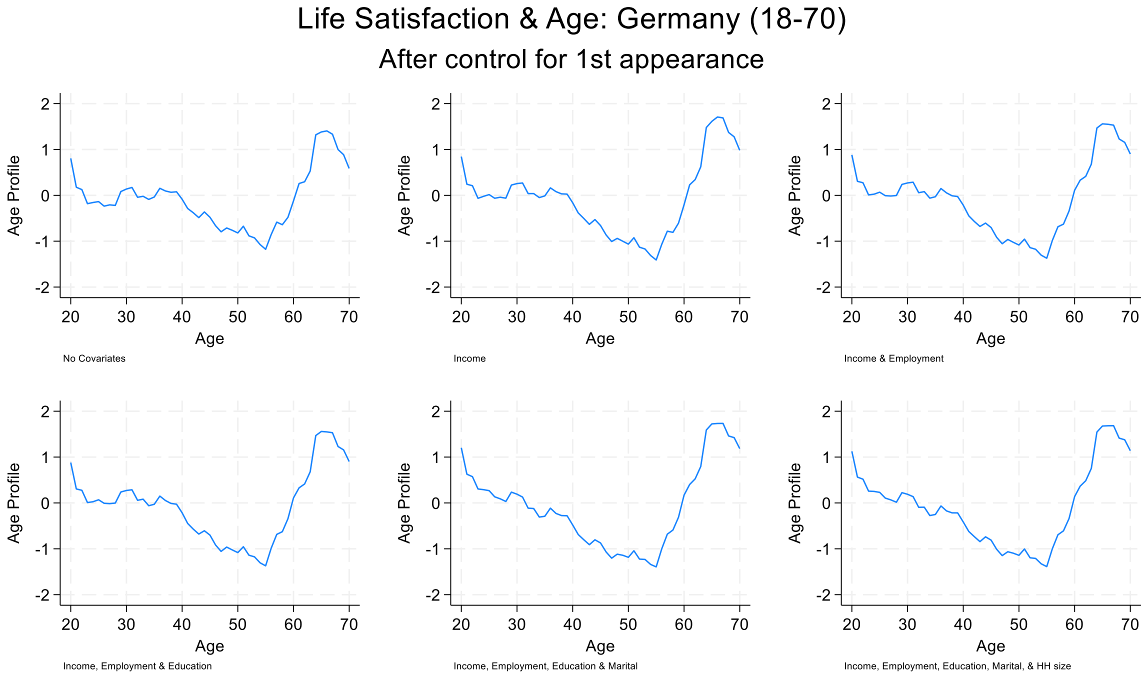
[9] The intuitive explanation is as follows: interest is in knowing the effect of aging on life satisfaction, however this is shaped by three interconnected factors which are an age effect, a time effect and a cohort effect. An assumption is needed to separate these three effects, and equation (2) includes individual fixed effects (αi), which control for all time-invariant individual characteristics, including birth cohort. Following de Ree and Alessie (2011), we impose the standard identifying assumption that period effects deviate from a linear trend. This allows us to estimate how life satisfaction evolves as individuals age, net of transient historical shocks and all their own time-invariant traits.

methodology also enables graphs to be produced showing age and wellbeing demeaned of time, removing cohort influence, thus giving an indication of a pure ageing effect.

Table A2 in the Appendix presents the outcomes from specifications that make use of the de Ree and Alessie correction. The results are remarkably like those already discussed, though with more consistency, and thus very supportive of a longitudinal finding of a midlife low in wellbeing. This use of the de Ree and Alessie correction method also enables us to show the de-trended age profiles, with the vertical axis showing the implied coefficient associated with age. Figures 5 and 6 show this for six different control sets for the UK and Germany respectively. Of note is that the nadir in life satisfaction is reached earlier in the UK than in Germany, as also shown by Cheng et al. (2017), and that there is a minor bump around the age of 30, which is perhaps surprising, although also reported by Cheng et al. (2017). This latter result is less commonly found than the midlife nadir and, as yet, unexplained, though Figure 3 below does seem to indicate that taking into account marital status flattens the bump at around 30 years old. In general, these de-trended age profiles indicate evidence consistent with a midlife low, with different broad trajectories by country. The de-trended profiles by gender do not indicate much difference between females and males (not shown).

Figure 5: Age and wellbeing, time demeaned graphs, for the UK

Figure 6: Age and wellbeing, time demeaned graphs, for Germany



Given the above results, we assess the robustness of our findings to the assumption of a linear model with normally distributed additive errors, recognising that the outcome variables are measured on discrete ordinal scales. Thus to test whether our conclusions are sensitive to functional form assumptions, we re-estimate the specification in Equation (1) using a fixed-effects ordered logit model, which is appropriate for ordinal outcomes. We implement the blow-up and cluster (BUC) estimator introduced by Baetschmann et al. (2015), available via the feologit command in Stata (Baetschmann et al., 2020). Unlike the standard random-effects ordered logit or probit models—which impose restrictive assumptions about the independence of individual-specific effects and covariates, and the normality of the error distribution—the BUC estimator provides consistent estimates of β from Equation (1). While the BUC estimator does not recover threshold parameters or individual fixed effects and therefore precludes direct estimation of predicted probabilities or marginal effects, the sign and significance of β (and its associated odds ratio) can be meaningfully interpreted. Thus, an odds-ratio less than 1 implies life satisfaction and age are negatively associated whereas odds-ratio greater than 1 reveals the converse. Our findings from the fixed-effects ordered logit estimator presented in Figures A1-A4 (in the Appendix) reinforce the conclusions drawn from the linear specification.

4. Concluding Discussion

This study uses specification curve analysis to evaluate the robustness of the relationship between age and subjective well-being, while addressing several key methodological disputes in the existing literature. Using longitudinal data from the United Kingdom and Germany, covering the period from 1996/7 to 2021, we find consistent evidence of a midlife dip in well-being. Across a wide range of model specifications, well-being declines into midlife before rising in later life—a pattern that holds across alternative control sets, functional forms, and cohort adjustments. These results directly challenge the claim that the midlife low is a statistical artefact driven by cross-sectional bias, inappropriate controls, cohort effects, or resulting from the imposition of a quadratic specification.

While some variation emerges in the depth or timing of the dip across specifications, a decline to midlife and a subsequent uptick proves to be rather robust. Additional analyses further corroborate the pattern. The findings contribute transparent, comprehensive evidence to a long-standing and often fractious debate. This empirical evidence, supportive of midlife struggles, is also supportive of much thinking and investigation prior to longitudinal datasets and corresponding analytic techniques.[10]

Some of our specifications did not support this overall pattern. Our analysis suggests that, for example, a study using the SOEP waves between 1996 and 2000 (i.e. waves 13-17), along with some other methodological choices, might uncover increasing wellbeing until midlife in Germany. Such a study might have been another entry in the debate or, to use the analogy of Simonsohn et al. (2020) (see footnote 1 above), another 'letter to the editor'. However, in the context of the full specification curve analysis we see that it is an exception, rather than a general finding. This difference from the fundamental pattern may (or may not) be interesting. Is it something to do with the sample, perhaps one of its occasional refreshments; or with Germany during this period; or more reflective of a type 1 error? In the curve, it is clearly an outlier and does not reflect the overall pattern. In any case, it is less interesting than the overall, much more common, finding. These findings are average effects, and future work could investigate when, for some individuals, the common trajectory is not followed. One recent example is Piper (2025) who shows that males exhibiting traits associated with antagonistic grandiose narcissism do not seem to recover from their midlife lows.

Many prior empirical studies examining the relationship between age and wellbeing both acknowledge and add to the debate about the overall pattern, often with their own highly specific and prescriptive 'best-practice' or 'correct' way to analyse it; prescriptions that are often in direct contrast with the specific prescriptions of other scholars. Though they come with claims to advance the debate, they cannot and instead push the debate down increasingly narrow one-way streets, with their prescriptions rendering almost all other attempts to analyse the relationship invalid. Instead of the different individual one-way streets of previous research, specification curve analysis presents the whole map. Indeed, and in contrast, this prior research is often forcefully (and merely) arguing for one particular point on the specification curve, in the face of other equally forceful arguments for other particular points. This adds to the appearance of more debate and confusion within the social sciences regarding age and wellbeing than there actually is. Slightly beyond the scope of our article, but

[10] The notion of a midlife low is at least a century old. As one example, in his autobiography (and elsewhere), Carl Jung wrote at length about his midlife struggles which occurred in his late 30s between 1914 and 1917 (Jung 1965). O'Connor (1981) discusses what he calls Freud's midlife crisis, approximately twenty years earlier. Just before the second world war, an introduction to psychotherapy discusses problems in midlife in such a way as to leave the reader with the impression that they were commonplace (Brown 1938). A recent history of midlife crises shows how midlife travails were used again and again in marketing campaigns in the 1940s and 50s (Jackson 2021). And as is increasingly well known, the coinage of the term midlife crisis is frequently attributed to Elliot Jacques for a paper presented at the British Psychoanalytical Society in 1957, work eventually published as Death and the Midlife Crisis years later (Jacques 1965). Since then, many books have been written on the subject (Fried 1967; Sheehy 1977; Hollis 1993; Polden 2002; Jameson 2022 – a far from exhaustive list). This literature, however, while interesting and thought provoking, is not able to provide systematic evidence on the phenomena of a midlife low but is suggestive of such a phenomenon. Thus, rather than, for example, the anecdote of Jung's midlife struggles, the 'random sample' of over three hundred creative artists that Jacques concerned himself with, and Sheehy's best-selling account of multiple qualitative interviews with individuals, modern-day quantitative research has a better chance at a systematic establishment of the fundamental pattern of the relationship between age and wellbeing. This is the aim of our specification curve analysis.

analogously, sometimes rather limited literature reviews of this topic, often of the same studies, amplify this appearance of much confusion in several ways. These ways include the following: miscategorising the previous literature (which is equivalent misreading a coefficient's sign on the specification curve); being very partial (relying on very few dots of the curve to make a case for a particular pattern); and comparing 'apples and oranges' (studies that control for almost nothing and are then directly compared to a literature that controls for co-variates; and cross-section results used to challenge panel evidence – in both cases different things are being analysed) and then wondering why they are sometimes different. Furthermore, studies are mentioned which should not be generalised from are and are additionally sometimes given an outsized importance; often studies which have their own clear limitations. Future research should be more careful in its use of previous literature and not repeat the standard lines. With such a context, it is unsurprising that some scholars can assert (falsely in our opinion) that social science has almost nothing to say about age and wellbeing (as in the opening epigraph). In contrast, specification curve analysis, by encompassing a wide range of methodological choices and datasets, offers a more comprehensive, robust conclusion, finding that midlife is a difficult phase of the lifecycle.

Finally, the future of research regarding age and wellbeing should also pay attention to cohort-based changes, and the wellbeing of different age groups and generations. The focus on the whole of life pattern can obscure important trends. Relatedly, the notion of the U-shape has been recently challenged by one of its chief advocates, David Blanchflower. Blanchflower et al. (2024a) find evidence that since about 2011 ill-being measures like anxiety and depression have caught up with – and for some measures even sometimes exceed – those in midlife. Given the history of this research, and the consistency of the midlife low, this is a surprising result, and alarming. Blanchflower et al. (2024b) discuss cohort reasons for this finding including social media and the COVID-19 pandemic as well as the measures taken to combat it. Banks and Xu (2020), similarly, presented early evidence that in the first few months of the pandemic, the mental health of the young was particularly negatively affected. Importantly, and missed in the discussion, none of the reasons put forward for this cohort-based finding (which also include social media use) address any of the issues thought to be behind the travails of midlife.

The finding of current distress of the young does not indicate that the idea of a midlife crisis or struggle is old news, as is sometimes claimed. Ceteris paribus, it should make it worse. Those in midlife will often be the parents of the currently distressed young. Perhaps the U-shape language is unhelpful; after all, no age range has a monopoly on misery. Instead, what is important is who is suffering: those in midlife as they often have done; and, over the last decade, the young. This may mean that the suffering of the young is more amenable to policy measures by being less 'natural' than the low wellbeing of those in midlife. Despite midlife suffering, it remains to be seen which of the following of several possibilities for the low wellbeing of the young is the case: (1) a temporary blip for this cohort; (2) a long-lasting cohort effect that will accompany them along at least some of the lifecycle (and maybe all); or (3) a fundamental lifecycle effect, indicating a permanent change in the age-wellbeing relationship. Other possibilities exist too. Alternatively, Lenzen et al. (2025) provide empirical evidence that the wellbeing crisis among the young mainly reflects reporting biases, and that "once we adjust for reporting bias, the relationship between age and poor mental health returns to the well-known hump-shape curve with worse mental health in mid-life" (p.1). These issues can be fully investigated in the future with additional waves of data covering the next few years: clearly another important avenue for future research. When these data exist, a specification curve analysis would be useful in helping us understand what is going on.

References

Akey, P., Heimer, R. Z., & Lewellen, S. (2021). Politicizing consumer credit. *Journal of Financial Economics*, 139(2), 627-655.

Baetschmann, G., Ballantyne, A., Staub, K. E., & Winkelmann, R. (2020). feologit: A new command for fitting fixed-effects ordered logit models. *The Stata Journal*, 20(2), 309–384.

Baetschmann, G., Staub, K. E., & Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(3), 685–703.

Bartram, D. (2023). IS HAPPINESS U-SHAPED IN AGE EVERYWHERE? A METHODOLOGICAL RECONSIDERATION FOR EUROPE. *National Institute Economic Review*, 263, 61–75.

Bauer, J. M., & Kaiser, M. (2025). Can major life events explain the U-shape in subjective well-being over the lifespan? *SSRN*.

Beja, E. L. Jr. (2018). The U-shaped relationship between happiness and age: Evidence using world values survey data. *Quality & Quantity*, 52(4), 1817-1829.

Benzeval, M., & Crossley, T. F. (2023). A symposium on Understanding Society, the UK Household Longitudinal Study: Introduction. *Fiscal Studies*, 44(3), 317–340.

Blanchflower, D. G., Graham, C., & Piper, A. (2023). HAPPINESS AND AGE—RESOLVING THE DEBATE. *National Institute Economic Review*, 263, 76–93.

Blanchflower, D.G. & Oswald, A.J. (2008). Is well-being U-shaped over the life cycle? *Social Science and Medicine*, 66 (8), 1733-1749.

Blanchflower, D.G. & Oswald, A.J. (2009). The U-shape without controls: a response to Glenn. *Social Science and Medicine*, 69, 486–488.

Blanchflower, D. G., & Piper, A. (2022). There is a mid-life low in well-being in Germany. *Economics Letters*, 214, 110430.

Brown, W. (1938). Psychological Methods of Healing: An Introduction to Psychotherapy. *University of London Press Limited*.

Bryan, C. J., Yeager, D. S., & O'Brien, J. M. (2019). Replicator degrees of freedom allow publication of misleading failures to replicate. *Proceedings of the National Academy of Sciences*, 116(51), 25535–25545.

Cheng, T. C., Powdthavee, N., & Oswald, A. J. (2017). Longitudinal evidence for a midlife nadir in human well-being: Results from four data sets. *The Economic Journal*, 127(599), 126–142.

Clark, A. (2002). Born to be mild: cohort effects in subjective well-being. *DELTA mimeo*.

de Ree, J., & Alessie, R. (2011). Life satisfaction and age: Dealing with underidentification in age-period-cohort models. *Social Science & Medicine*, 73(1), 177–182.

Ehrhardt, J. J., Saris, W. E., & Veenhoven, R. (2000). Stability of life-satisfaction over time: Analysis of change in ranks in a national population. *Journal of Happiness Studies*, 1, 177–205.

Ferrer-i Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114(497), 641-659.

Frijters, P., & Beatton, T. (2012). The mystery of the U-shaped relationship between happiness and age. *Journal of Economic Behavior & Organization*, 82(2–3), 525–542.

Frey, R., Richter, D., Schupp, J., Hertwig, R., & Mata, R. (2021). Identifying robust correlates of risk preference: A systematic approach using specification curve analysis. *Journal of Personality and Social Psychology*, 120(2), 538–557.

Galambos, N.L., Fang, S., Krahn, H.J., Johnson, M.D. and Lachman, M.E. (2015). Up, not down: The age curve in happiness from early adulthood to midlife in two longitudinal studies. *Developmental Psychology*, 51, 1664–1671.

Gao, M., Leung, H., & Qiu, B. (2021). Organization capital and executive performance incentives. *Journal of Banking & Finance*, 123, 106017.

Giuntella, O., McManus, S., Mujcic, R., Oswald, A. J., Powdthavee, N., & Tohamy, A. (2022). The Midlife Crisis. *Economica*, 90(357), 65-110.

Glenn, N. (2009). Is the apparent U-shape of well-being over the life course a result of inappropriate use of control variables? A commentary on Blanchflower and Oswald. *Social Science and Medicine*, 69(4), 481–485.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel Study (SOEP). *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics*, 239(2), 345–360.

Han, Q., Zheng, B., Agostini, M., Bélanger, J. J., Gützkow, B., Kreienkamp, J., Reitsema, A. M., van Breen, J. A., Leander, N. P., & PsyCorona Collaboration. (2021). Associations of risk perception of COVID-19 with emotion and mental health during the pandemic. *Journal of Affective Disorders*, 284, 247–255.

Hollis, J. (1993). The Middle Passage: From misery to meaning in midlife. *Inner City Books*.

Infurna, F. J., Gerstorf, D., & Lachman, M. E. (2020). Midlife in the 2020s: Opportunities and challenges. *The American Psychologist*, 75(4), 470–485.

Jackson, M. (2021). Broken dreams: An intimate history of the midlife crisis. *Reaktion Books*.

Jamieson, A. (2022). Midlife: Humanity's Secret Weapon. London: *Notting Hill Editions*.

Jaques, E. (1965). Death and the midlife crisis. *International Journal of Psychoanalysis*, 46(4), 502-514.

Jung, C. G. (1965). Memories, dreams, reflections. New York, NY: *Vintage Books*.

Jung, K., Shavitt, S., Viswanathan, M., & Hilbe, J. M. (2014). Female hurricanes are deadlier than male hurricanes. *Proceedings of the National Academy of Sciences* of the USA, 111, 8782-8787.

Kassenboehmer, S.C., & Haisken-DeNew, J. (2012). Heresy or enlightenment? The well-being age U-shape effect is flat. *Economics Letters*, 117, 235-238.

Kleven, H. (2024). The EITC and the extensive margin: A reappraisal. *Journal of Public Economics, 236*, 105135.

Kratz, F., & Brüderl, J. (2021). The Age Trajectory of Happiness, working paper.

Lenzen, S., Connelly, L., & Whittaker, W. (2025). The good old days? Mental health across birth cohorts. *SSRN*.

McKenzie, D. J. (2006). Disentangling age, cohort and time effects in the additive model. *Oxford Bulletin of Economics and Statistics*, 68(4), 473-495.

Neulinger, A., & Radó, M. (2018). The impact of household life-cycle stages on subjective well-being: Considering the effect of household expenditures in Hungary. *International Journal of Consumer Studies*, 42(1), 16–26.

O'Connor, P. A. (1981). *Understanding the mid-life crisis.* Melbourne: Sun Books.

Office for National Statistics. (2023). Marriage and civil partnership status, England and Wales: Census 2021. *ONS*.

Oparina, E., Kaiser, C., Gentile, N., Tkatchenko, A., Clark, A. E., De Neve, J.-E., & D'Ambrosio, C. (2025). Machine learning in the prediction of human wellbeing. *Scientific Reports*, 15, Article 1632.

Oshio, T., & Shimizutani, S. (2024). Well-being paradox: Comparing the age-happiness relationship across Japan, China, and the US. *The Japanese Economic Review*, 75(4), 547-562.

Polden, J. (2002). Regeneration – Journey through the Mid-life Crisis. *London: Continuum*.

Piper, A. T. (2015). Sliding down the U-shape? A dynamic panel investigation of the age–well-being relationship, focusing on young adults. *Social Science & Medicine*, 143, 54–61.

Piper, A. T. (2025). Narcissism and wellbeing in midlife and beyond. *SSRN*

Rauvola, R. S., & Rudolph, C. W. (2023). Worker aging, control, and well-being: A specification curve analysis. *Acta Psychologica*, 233, 103833.

Sheehy, G. (1977). Passages: Predictable crises of adult life. *Bantam Books*.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11), 1208-1214.

Socio-Economic Panel (SOEP). (2023), data for years 1984-2021, SOEP-Core v38, EU Edition, doi:10.5684/soep.core.v38eu

Statistisches Bundesamt (Destatis). (2025). Marriages, divorces and life partnerships.

Tobias, J. L., & Bond, T. N. (2025). Semiparametric Bayesian estimation in an ordinal probit model with application to life satisfaction across countries, age, and gender. Journal of Econometrics.

University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Waves 1-14, 2009-2023 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. *19th Edition.* UK Data Service. SN: 6614, DOI: http://doi.org/10.5255/UKDA-SN-6614-20

Voracek, M., Kossmeier, M., & Tran, U. S. (2019). Which data to meta-analyze, and how? A specification-curve and multiverse-analysis approach to meta-analysis. *Zeitschrift für Psychologie*, 227(1), 64–82.

Wunder, C., Wiencierz, A., Schwarze, J., & Küchenhoff, H. (2013). Well-being over the life span: Semiparametric evidence from British and German longitudinal data. *The Review of Economics and Statistics*, 95, 154–167.

APPENDIX

Table A1

	Without de Ree and Alessie (2011) correction					
	UK			Germany		
	All	Male	Female	All	Male	Female
18-30	(-0.864 - 0.047), 745 is negative and 5 positive of total 750, out of 745 negative coefficients 545 (591) (637) significant at 1% (5%) and (10%), none of the positive coefficients are significant	(-0.854 - 0.049), 740 is negative and 10 positive of total 750, out of 740 negative coefficients 527 (608) (619) significant at 1% (5%) and (10%), none of the positive coefficients are significant	(-0.985 - 0.216), 699 is negative and 51 positive of total 750, out of 699 negative coefficients 288 (329) (423) significant at 1% (5%) and (10%), 0 (5) (9) of the positive coefficients are significant at 1% (5%) (10%)	(-0.858 - 0.220), 670 is negative and 80 positive of total 750, out of 670 negative coefficients 452 (478) (524) significant at 1% (5%) and (10%), 2 out of 80 positive coefficients are significant at 10%	(-1.042 - 0.282), 700 is negative and 50 positive of total 750, out of 700 negative coefficients 505 (585) (616) significant at 1% (5%) and (10%), none of the positive coefficients are significant	(-0.708 - 0.225), 605 is negative and 105 positive of total 750, out of 605 negative coefficients 316 (341) (357) significant at 1% (5%) and (10%), 24 (35) (44) of the 105 positive coefficients are significant at 1% (5%) (10%)
18-35	(-0.002 - -0.933), 750 is negative of total 750, out of 750 negative coefficients 617 (705) (723) significant at 1% (5%) and (10%)	(-0.026 - -0.848), 750 is negative of total 750, out of 750 negative coefficients 546 (610) (634) significant at 1% (5%) and (10%)	(-1.090 - 0.054), 734 is negative and 16 positive of total 750, out of 734 negative coefficients 465 (630) (686) significant at 1% (5%) and (10%), none of the positive coefficients are significant	(-0.922 - 0.320), 642 is negative and 108 is positive of total 750, out of 642 negative coefficients 432 (443) (454) significant at 1% (5%) and (10%), out of 108 positive coefficients 20 (50) (68) significant at 1% (5%) and (10%)	(-1.071 - 0.366), 678 is negative and 72 is positive of total 750, out of 678 negative coefficients 469 (513) (531) significant at 1% (5%) and (10%), out of 72 positive coefficients 0 (24) (24) significant at 1% (5%) and (10%)	(-0.789 - 0.295), 596 is negative and 154 is positive of total 750, out of 596 negative coefficients 375 (401) (405) significant at 1% (5%) and (10%), out of 154 positive coefficients 4 (29) (72) significant at 1% (5%) and (10%)

18-45	(-0.087 - -0.870), 750 is negative of total 750, and all are significant at 1%	(-0.053 - -0.786), 750 is negative of total 750, out of 750 negative coefficients 691 (700) (716) significant at 1% (5%) and (10%)	(-0.092 - -1.009), 750 is negative of total 750, all are significant at 1%	(-0.992 - 0.309), 700 is negative and 50 is positive of total 750, out of 700 negative coefficients 533 (581) (595) significant at 1% (5%) and (10%), out of 50 positive coefficients 46 (50) (50) significant at 1% (5%) and (10%)	(-1.074 - 0.314), 696 is negative and 54 is positive of total 750, out of 696 negative coefficients 485 (508) (534) significant at 1% (5%) and (10%), out of 54 positive coefficients 14 (26) (44) significant at 1% (5%) and (10%)	(-0.942 - 0.318), 646 is negative and 104 is positive of total 750, out of 646 negative coefficients 506 (561) (570) significant at 1% (5%) and (10%), out of 104 positive coefficients 14 (45) (50) significant at 1% (5%) and (10%)
18-50	(-0.096 - -0.871), 750 is negative of total 750, out of 750 negative coefficients 700 (730) (750) are significant at 1% (5%) and (10%)	(-0.073 - -0.687), 750 is negative of total 750, out of 750 negative coefficients 700 significant at 1%	(-0.084 - -1.049), 750 is negative of total 750, out of 750 negative coefficients 700 (700) (724) are significant at 1% (5%) (10%)	(-1.026 - 0.257), 686 is negative and 64 is positive of total 750, out of 686 negative coefficients 500 (537) (570) significant at 1% (5%) and (10%), out of 64 positive coefficients 28 (49) (50) significant at 1% (5%) and (10%)	(-1.084 - 0.227), 682 is negative and 68 is positive of total 750, out of 682 negative coefficients 449 (490) (518) significant at 1% (5%) and (10%), out of 54 positive coefficients 0 (11) (23) significant at 1% (5%) and (10%)	(-0.976 - 0.294), 630 is negative and 120 is positive of total 750, out of 630 negative coefficients 500 significant at 1%, out of 120 positive coefficients 13 (46) (50) significant at 1% (5%) and (10%)
31-50	(-0.884 - 0.005), 747 is negative of total 750 and 3 are positive, out of 747 negative coefficients 672 (682) (684) are significant at 1% (5%) and (10%), none of the positive	(-0.003 - -0.651), 750 is negative of total 750, out of 750 negative coefficients 585 (669) (695) are significant at 1% (5%) and (10%)	(-1.098 - 0.053), 732 is negative of total 750 and 18 are positive, out of 750 negative coefficients 607 (665) (671) are significant at 1% (5%) and (10%), none of the positive	(-1.117 - 0.236), 636 is negative and 114 is positive of total 750, out of 636 negative coefficients 500 (512) (530) significant at 1% (5%) and (10%), out of 114 positive	(-1.150 - 0.229), 605 is negative and 145 is positive of total 750, out of 605 negative coefficients 444 (454) (486) significant at 1% (5%) and (10%), out of 145 positive coefficients 0 (0) (3)	(-1.086 - 0.267), 598 is negative and 152 is positive of total 750, out of 598 negative coefficients 500 (500) (508) significant at 1% (5%) and (10%), out of 152 positive coefficients 0 (1) (28)

	coefficients are significant		coefficients are significant	coefficients 3 (25) (52) significant at 1% (5%) and (10%)	significant at 1% (5%) and (10%)	significant at 1% (5%) and (10%)
51-60	(-0.932 - 0.765), 150 is negative of total 750 and 600 are positive, out of 150 negative coefficients 84 (100) (100) are significant at 1% (5%) and (10%) are significant, out of 525 positive coefficients 500 (507) (535) are significant at 1% (5%) and (10%)	(-1.157 - 0.739), 201 is negative of total 750 and 549 are positive, out of 201 negative coefficients 52 (101) (101) are significant at 1% (5%) and (10%) are significant, out of 549 positive coefficients 336 (429) (463) are significant at 1% (5%) and (10%)	(-0.836 - 0.949), 109 is negative of total 750 and 641 are positive, out of 109 negative coefficients 45 (100) (100) are significant at 1% (5%) and (10%) are significant, out of 641 positive coefficients 502 (543) (576) are significant at 1% (5%) and (10%)	(-0.901 - 0.700), 155 is negative and 595 is positive of total 750, out of 155 negative coefficients 100 significant at 1%, out of 595 positive coefficients 377 (420) (468) significant at 1% (5%) and (10%)	(-0.938 - 0.738), 121 is negative and 629 is positive of total 750, out of 121 negative coefficients 72 (85) (101) significant at 1% (5%) and (10%), out of 629 positive coefficients 292 (353) (384) significant at 1% (5%) and (10%)	(-0.885 - 0.714), 158 is negative and 592 is positive of total 750, out of 158 negative coefficients 100 (100) (101) significant at 1% (5%) and (10%), out of 592 positive coefficients 324 (398) (448) significant at 1% (5%) and (10%)
51-65	(-0.770 - 0.833), 150 is negative of total 750 and 600 are positive, out of 150 negative coefficients 124 (145) (150) are significant at 1% (5%) and (10%) are significant, out of 600 positive coefficients 519 (534) (550) are significant at 1% (5%) and (10%)	(-0.881 - 0.741), 150 is negative of total 750 and 600 are positive, out of 150 negative coefficients 50 (76) (93) are significant at 1% (5%) and (10%) are significant, out of 600 positive coefficients 490 (512) (546) are significant at 1% (5%) and (10%)	(-0.836 - 0.949), 163 is negative of total 750 and 587 are positive, out of 163 negative coefficients 100 (100) (118) are significant at 1% (5%) and (10%) are significant, out of 587 positive coefficients 507 (522) (524) are significant at 1% (5%) and (10%)	(-0.805 - 0.748), 100 is negative and 650 is positive of total 750, out of 100 negative coefficients 100 significant at 1%, out of 650 positive coefficients 534 (607) (625) significant at 1% (5%) and (10%)	(-0.745 - 0.749), 100 is negative and 650 is positive of total 750, out of 100 negative coefficients 85 (100) (100) significant at 1% (5%) and (10%), out of 650 positive coefficients 487 (514) (555) significant at 1% (5%) and (10%)	(-0.852 - 0.760), 100 is negative and 650 is positive of total 750, out of 158 negative coefficients 100 significant at 1%, out of 650 positive coefficients 534 (556) (570) significant at 1% (5%) and (10%)
51-70	(-0.763 - 0.824), 165 is negative of total	(-0.811 - 0.780), 150 is negative of total 750 and	(-0.830 - 0.867), 195 is negative of total	(-0.857 - 0.629), 100 is negative and 650 is	(-0.867 - 0.728), 100 is negative and 650 is	(-0.837 - 0.595), 100 is negative and 650 is

	750 and 585 are positive, out of 165 negative coefficients 134 (150) (150) are significant at 1% (5%) and (10%) are significant, out of 585 positive coefficients 515 (524) (533) are significant at 1% (5%) and (10%)	600 are positive, out of 150 negative coefficients 90 (126) (145) are significant at 1% (5%) and (10%) are significant, out of 600 positive coefficients 513 (546) (558) are significant at 1% (5%) and (10%)	750 and 555 are positive, out of 195 negative coefficients 100 (131) (147) are significant at 1% (5%) and (10%) are significant, out of 555 positive coefficients 500 (506) (506) are significant at 1% (5%) and (10%)	positive of total 750, out of 100 negative coefficients 99 significant at 1%, out of 650 positive coefficients 543 (582) (609) significant at 1% (5%) and (10%)	positive of total 750, out of 100 negative coefficients 100 significant at 1%, out of 650 positive coefficients 509 (524) (530) significant at 1% (5%) and (10%)	positive of total 750, out of 100 negative coefficients 100 significant at 1%, out of 650 positive coefficients 527 (551) (554) significant at 1% (5%) and (10%)
71 plus	(-1.783 - 0.450), 637 is negative of total 750 and 113 are positive, out of 637 negative coefficients 435 (449) (474) are significant at 1% (5%) and (10%) are significant, out of 113 positive coefficients 50 are significant at 1%	(-1.579 - 0.323), 648 is negative of total 750 and 102 are positive, out of 648 negative coefficients 384 (405) (411) are significant at 1% (5%) and (10%) are significant, out of 102 positive coefficients 26 are significant at 10%	(-2.161 - 0.538), 557 is negative of total 750 and 193 are positive, out of 557 negative coefficients 391 (416) (445) are significant at 1% (5%) and (10%) are significant, out of 193 positive coefficients 50 (58) (77) are significant at 1% (5%) and (10%)	(-1.489 - 0.357), 700 is negative and 50 is positive of total 750, out of 700 negative coefficients 601 (635) (659) significant at 1% (5%) (10%), out of 50 positive coefficients all significant at 1%	(-1.243 - 0.160), 590 is negative and 160 is positive of total 750, out of 590 negative coefficients 400 (409) (431) significant at 1% (5%) and (10%), out of 160 positive coefficients none is significant	(-1.659 - 0.599), 700 is negative and 50 is positive of total 750, out of 700 negative coefficients 633 (639) (647) significant at 1% (5%) (10%), out of 50 positive coefficients all significant at 1%

Note: For each one of the above specifications, we control for the first appearance dummy for the individuals in the wave. The results reported with whole-sample with no/ individual/individual-region fixed effects. Significance has been reported with clustered standard error at the individual level.

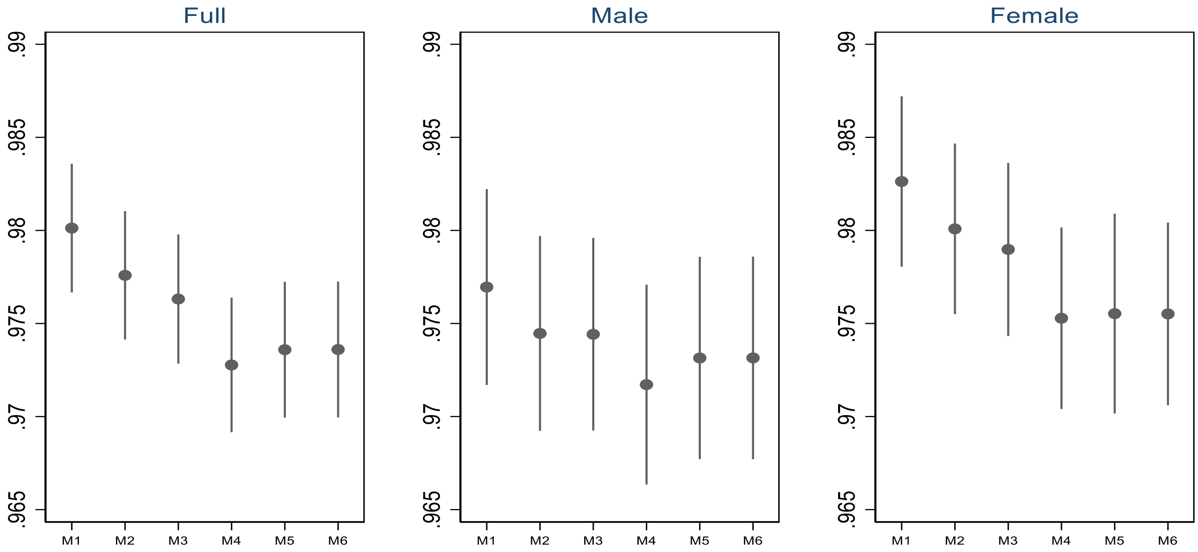
Table A2

	With de Ree and Alessie (2011) correction					
	UK			Germany		
	All	Male	Female	All	Male	Female
18-30	(-0.385 - -0.230), out of 50, all are negative and significant at 1%	(-0.524 - -0.329), out of 50, all are negative and significant at 1%	(-0.295 - -0.148), out of 50, all are negative and significant at 1%	(-0.216 - -0.311), out of 50, all are negative and significant at 1%	(-0.336 - -0.422), out of 50, all are negative and significant at 1%	(-0.123 - -0.205), out of 50, all are negative and significant at 1%
18-35	(-0.305 - -0.179), out of 50, all are negative and significant at 1%	(-0.389 - -0.244), out of 50, all are negative and significant at 1%	(-0.255 - -0.130), out of 50, all are negative and significant at 1%	(-0.134 - -0.203), out of 50, all are negative and significant at 1%	(-0.188 - -0.245), out of 50, all are negative and significant at 1%	(-0.097 - -0.167), out of 50, all are negative and significant at 1%
18-45	(-0.302 - -0.217), out of 50, all are negative and significant at 1%	(-0.318 - -0.213), out of 50, all are negative and significant at 1%	(-0.299 - -0.223), out of 50, all are negative and significant at 1%	(-0.095 - -0.156), out of 50, all are negative and significant at 1%	(-0.107 - -0.170), out of 50, all are negative and significant at 1%	(-0.083 - -0.138), out of 50, all are negative and significant at 1%
18-50	(-0.281 - -0.218), out of 50, all are negative and significant at 1%	(-0.288 - -0.203), out of 50, all are negative and significant at 1%	(-0.285 - -0.225), out of 50, all are negative and significant at 1%	(-0.087 - -0.142), out of 50, all are negative and significant at 1%	(-0.094 - -0.150), out of 50, all are negative and significant at 1%	(-0.077 - -0.127), out of 50, all are negative and significant at 1%
31-50	(-0.302 - -0.246), out of 50, all are negative and significant at 1%	(-0.251 - -0.188), out of 50, all are negative and significant at 1%	(-0.344 - -0.285), out of 50, all are negative and significant at 1%	(-0.105 - -0.149), out of 50, all are negative and significant at 1%	(-0.094 - -0.127), out of 50, all are negative and significant at 1%	(-0.111 - -0.152), out of 50, all are negative and significant at 1%
51-60	(0.028 - 0.087), out of 50, all are positive and 6 significant at 10%	(-0.073 - -0.029), out of 50, all are negative and none is significant	(0.116 - 0.189), out of 50, all are positive and 6 (38) (46) significant at 1% (5%) (10%)	(-0.005 - 0.048), out of 50, 6 (44) are negative (positive) and none is significant	(0.005 - 0.050), out of 50, all are positive and none is significant	(-0.027 - 0.043), out of 50, 22 (28) is negative (positive) and none is significant
51-65	(0.083 - 0.206), out of 50, all are positive and 43 (50) (50) significant at 1% (5%) (10%)	(0.072 - 0.169), out of 50, all are positive and 24 (36) (48) significant at 1% (5%) (10%)	(0.088 - 0.237), out of 50, all are positive and 32 (43) (50) significant at 1% (5%) (10%)	(0.062 - 0.124), out of 50, all are positive and significant at 1%	(0.081 - 0.130), out of 50, all are positive and significant at 1%	(0.046 - 0.119), out of 50, all are positive and(23) (36) (42) significant at 1% (5%) (10%)
51-70	(0.083 - 0.216), out of 50, all are positive and significant at 1%	(0.087 - 0.218), out of 50, all are positive and 43 (50) (50) significant at 1% (5%) (10%)	(0.077 - 0.215), out of 50, all are positive and 36 (50) (50) significant at 1% (5%) (10%)	(0.053 - 0.132), out of 50, all are positive and significant at 1%	(0.063 - 0.140), out of 50, all are positive and 48 (50) significant at 1% (5%)	(0.046 - 0.128), out of 50, all are positive and 36 (44) (50) significant at 1% (5%) (10%)

71 plus	(-0.491 - -0.473), out of 50, all negative and none is significant	(-0.498 - -0.470), out of 50, all negative and significant at 1%	(-0.553 - -0.525), out of 50, all negative and none is significant	(-1.051 - -1.121), out of 50, all are negative and none is significant	(-0.846 - -0.970), out of 50, all are negative and none is significant at 1%	(-0.762 - -0.797), out of 50, all are negative and 25 significant at 1%

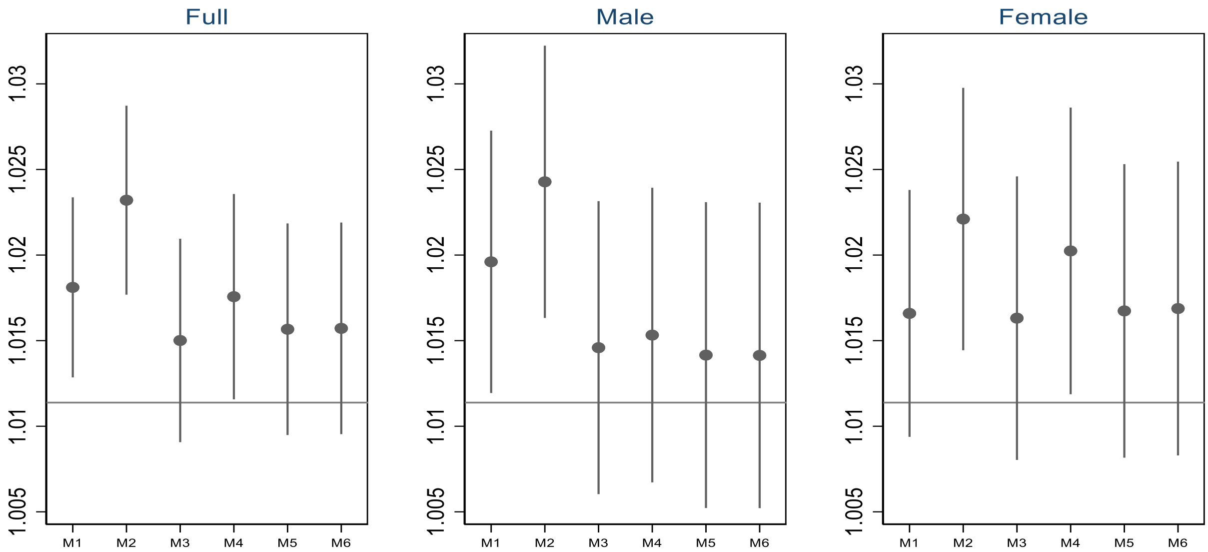
Note: For each one of the above specifications, we control for the first appearance dummy for the individuals in the wave. The results reported with whole-sample with individual/individual-region fixed effects. Significance has been reported with clustered standard error at the individual level

Fig A1: Odds Ratio associated with Age (Germany, 18-50)



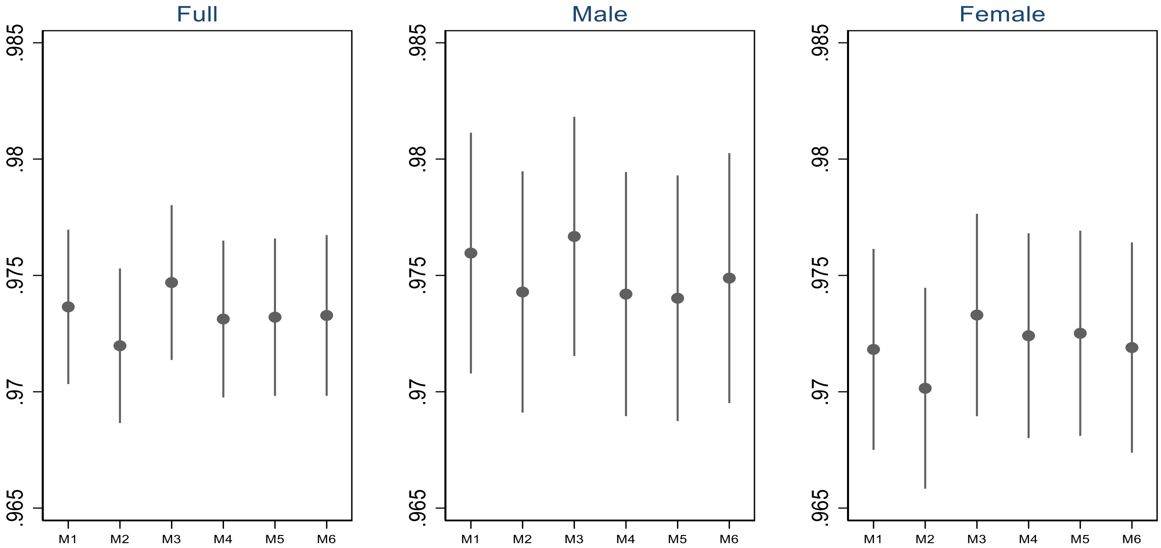
M1: Base
M2: Base+Inc
M3: Base+Inc+Employ
M4: Base+Inc+Employ+Marital
M5: Base+Inc+Employ+Marital+HH Size
M6: Base+Inc+Employ+Marital+HH Size+Educ

Fig A2: Odds Ratio associated with Age (Germany, 51-70)



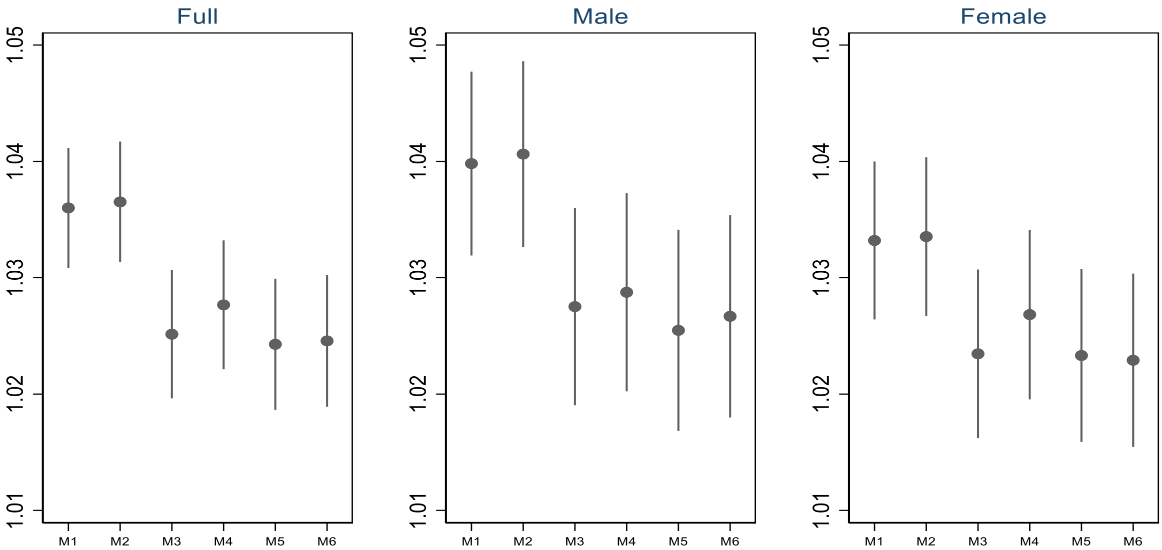
M1: Base
M2: Base+Inc
M3: Base+Inc+Employ
M4: Base+Inc+Employ+Marital
M5: Base+Inc+Employ+Marital+HH Size
M6: Base+Inc+Employ+Marital+HH Size+Educ

Fig A3: Odds Ratio associated with Age (UK, 18-50)



M1: Base
M2: Base+Inc
M3: Base+Inc+Employ
M4: Base+Inc+Employ+Marital
M5: Base+Inc+Employ+Marital+HH Size
M6: Base+Inc+Employ+Marital+HH Size+Educ

Fig A4: Odds Ratio associated with Age (UK, 51-70)



M1: Base
M2: Base+Inc
M3: Base+Inc+Employ
M4: Base+Inc+Employ+Marital
M5: Base+Inc+Employ+Marital+HH Size
M6: Base+Inc+Employ+Marital+HH Size+Educ