

DISCUSSION PAPER SERIES

IZA DP No. 10861

**Worries across Time and Age in Germany:
Bringing Together Open- and Close-Ended
Questions**

Julia M. Rohrer
Martin Bruemmer
Juergen Schupp
Gert G. Wagner

JUNE 2017

DISCUSSION PAPER SERIES

IZA DP No. 10861

Worries across Time and Age in Germany: Bringing Together Open- and Close-Ended Questions

Julia M. Rohrer

DIW Berlin, FUB and University of Leipzig

Martin Bruemmer

Independent Researcher, Leipzig

Juergen Schupp

DIW Berlin, FUB and IZA

Gert G. Wagner

DIW Berlin, IZA, MPIB and TUB

JUNE 2017

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Worries across Time and Age in Germany: Bringing Together Open- and Close-Ended Questions*

We investigate how worries in Germany change across time and age, drawing on both closed-ended questions (which typically list a number of worry items) and open-ended questions answered in text format. We find that relevant world events influence worries. For example, worries about peace peaked in 2003, the year of the Iraq War, with a considerable number of respondents also referring to the Iraq war in their text responses. Furthermore, we found that – controlling for these historical effects – worries about various topics such as health and the general economic situation increase with age. With increasing age, respondents also became more likely to answer the open-ended question. This suggests that the age increases in worries we found are not merely a result of an age-biased choice of worry items, but instead also hold for worries self-generated by the respondents.

JEL Classification: C81, C83, I31, Z13

Keywords: life course, worries, satisfaction, German Socio-Economic Panel Study, SOEP

Corresponding author:

Julia Rohrer
DIW Berlin
Mohrenstr. 58
10117 Berlin
Germany

E-mail: jrohrer@diw.de

* We thank Florian Giese (SOEP) for compiling the data files underlying our analyses and the participants of the conference “Subjective well-being over the life course: evidence and policy implications” at the London School of Economics (LSE) in December 2016 for a critical discussion of the first draft of the paper. We are especially grateful to Andrew Clark, Nattavudh Powdthavee and Richard Layard for their helpful comments.

0 Introduction

There are many papers analyzing well-being of people by means of data on happiness and satisfaction with life in general and/or satisfaction in different domains of life. In these papers, income and other resources, as well as events (like birth of a child or getting unemployed), are correlated with happiness and satisfaction. However, there are subjective states beyond mere happiness and satisfaction—such as expectations and e.g. worries—that might be linked to resources and events. In this paper, we are interested in those subjective states. We analyze worries across time and age based on longitudinal data.

We use data from the German Socio-Economic Panel Study (SOEP). Since its inception in 1984, SOEP has collected data on almost a dozen specific concerns that individuals may have about private and public spheres of life. The SOEP data includes items tapping worries about the general economic development, an individual's own financial situation, security of employment, crime, xenophobia, and world peace; answered on an ordinal scale with three response options (no worries, some worries, very worried).

Such “traditional” survey questions result in numbers – a data type that most survey researchers are comfortable with. Even in the case of “qualitative” answers, the categories are specified, such as “yes,” “no,” “strongly disagree,” and “very much the case,” meaning that the answers can be analysed in a quantitative manner, such as percentages. In contrast, open-ended questions, which generate text answers, are rare. In many cases, they are limited to the possibility of filling in an “other” category, which is not specified; and the resulting text answers are rarely analysed. However, as methods of automated text analysis are improving and becoming more popular, this is likely to change.

In this paper, we provide an example of how this methodological expansion of the toolbox for social scientists might look like, illustrating the analytical potential of open-ended texts in social surveys, cohort studies, and panel studies. Our demonstration is embedded in the general topic of this special issue, the dynamics of life-courses and aging, as we investigate people's worries across the life course. The analysis includes answers to

traditional close-ended items about worries as well as text answers to a supplementary item asking for “any other worries.”

First, we will describe our database, the German Socio-Economic Panel Study (SOEP). In the second part, we will describe how worries changed across calendar years by (i) describing how many respondents reported being “very concerned” in different life domains; and by (ii) summarizing the most frequent text answers by year. In the third part, we will employ panel regression models to investigate how worries vary with age, considering (i) single close-ended worry items; (ii) a composite measure reflecting the overall level of worries; and (iii) the odds of reporting additional worries in a text answer. In the fourth part, we explore whether age-associated changes in worries can, in part, account for changes in life satisfaction across the life course. Finally, we will discuss how open-ended questions in combination with a quantitative analysis of the resulting texts can supplement large-scale panel studies in a broader sense.

Our paper is mainly data driven. We will discuss the relevant literature on well-being over the life course in part four where we analyse worries as well as satisfaction with life.

1 German Socio Economic Panel Study (SOEP)

The German Socio Economic Panel Study (SOEP) is a wide-ranging representative longitudinal study of private households in Germany (multi cohort approach) that was started in 1984 (Wagner et al. 2007, Headey et al. 2010). In the last years, more than 10,000 households, covering about 30,000 individuals, have been surveyed by the fieldwork organization Kantar Public, Munich (formerly named TNS Infratest Sozialforschung). The data provide information on all household members, including not only German citizens in East and West Germany, but also foreigners and recent immigrants. The topics included in the survey include, among others, household composition, occupational biographies, employment, earnings, health, political attitudes, subjective indicators on well-being (Schupp

2015), life satisfaction in general and in different domains of life, and, of interest for the present study, worries about different life domains.

For the purpose of this study, we limited analysis to the 2000 to 2016 survey years because the sample size was considerably expanded in 2000. Furthermore, starting from 2000, text answers were saved in electronic form per default, rendering the data more accessible. We selected seven worry items that were included each of these years and were relevant to all respondents, i.e. we excluded the item regarding the security of employment as it only applied to the subsample of employed respondents.

Respondents were queried about their worries concerning (1) the general economic situation; (2) their personal health; (3) the protection of the environment; (4) peace; (5) the development of criminality in Germany; (6) immigration to Germany; and (7) hostility toward foreigners in Germany. Respondents answered on a three-point scale (1 = “very worried,” 2 = “somewhat worried,” 3 = “not worried at all”) and we recoded the variables such that high values indicate more worries (0 = “not worried at all,” 1 = “somewhat worried,” 2 = “very worried”). Including only respondents that answered all of these seven items resulted in a final sample size of 371,649 observations of 62,188 unique respondents with an average of 5.98 records per respondent (Min = 1, Max = 17, SD = 5.27). Mean age across all observations was 48.47 years (Min = 16, Max = 105, SD = 17.55) and 52.63% of all observations were of female respondents.

Following the block of close-ended worry items, since the very beginning of the study in 1984, the SOEP questionnaire also contains an open-ended question asking for “any other worries.” Depending on the survey mode, either respondents or interviewers write down the text answers that typically only contain a small number of key words. These text data were cleaned according to the procedure described in Rohrer et al. (2017) and in the appendix of this paper.

In each survey wave, at the very end of the questionnaire, respondents reported their life satisfaction on a single item measure ranging from 0 (completely dissatisfied) to 10 (completely satisfied).

2 Worries across the years: Close-ended items and text answers

Figure 1 shows the percentage of respondents who reported being “very worried” about a specific domain (Panels A-G), as well as the percentage of respondents who answered the open-ended question (Panel H), across calendar years, thus offering a descriptive summary of time trends in the items.

Tables 1 and 2 show the 10 most frequently used words in the reported “other worries” for each of the survey years from 2000 to 2016 based on the unweighted data. To illustrate the responses, Figure 2 offers a visualization of the 25 most frequently reported worries for the most recently available survey wave, 2016, scaled according to their frequency.

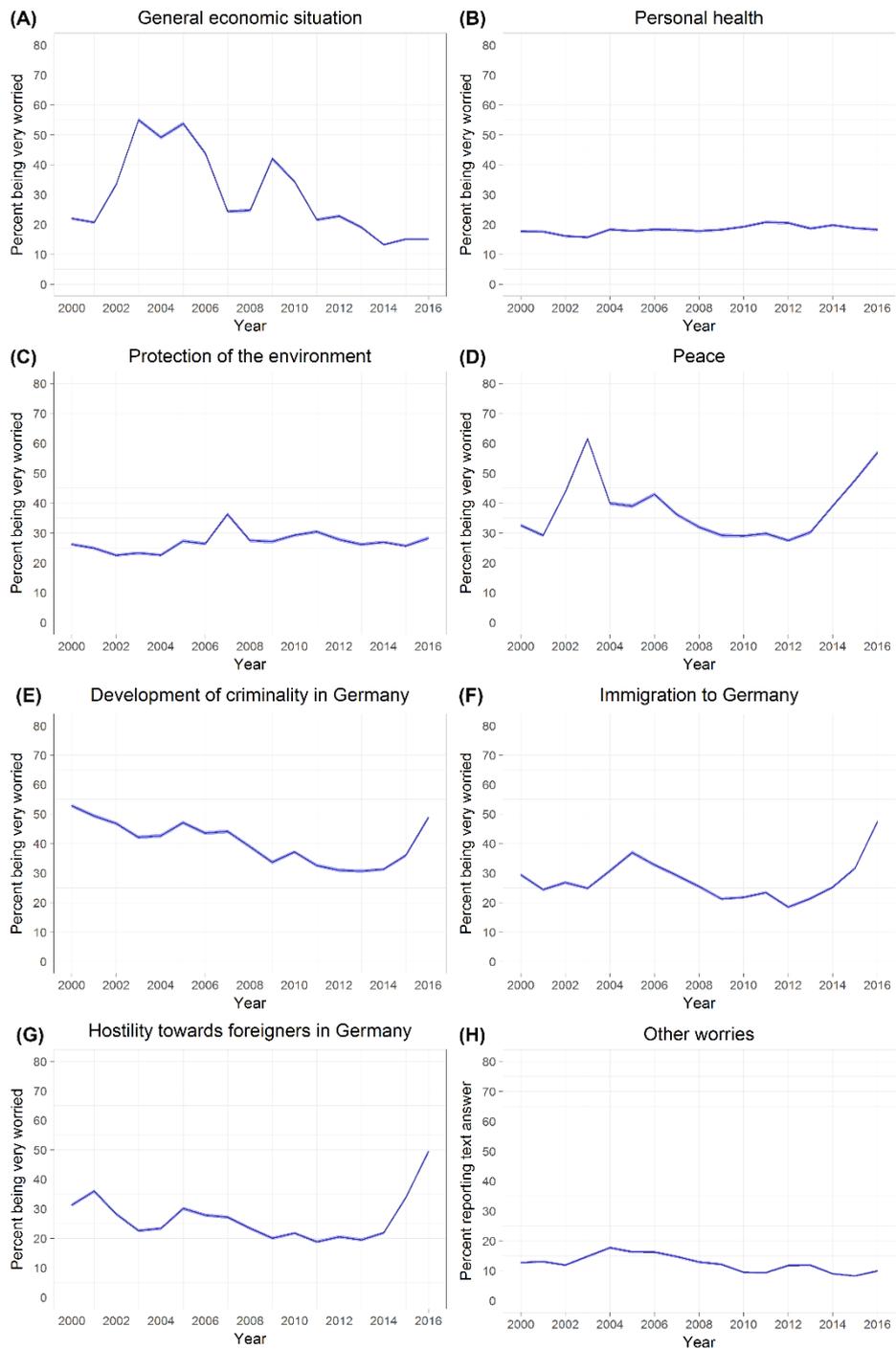


Figure 1: Percentage of SOEP respondents reporting being very worried about a specific domain (Panel A-G) and percentage of respondents additionally reporting “other worries” in text format (Panel H) between 2000 and 2016.

Table 1: The ten most commonly used words in the reply to the open-ended question asking for “other worries” for the years 2000 to 2007 and the proportion of text answers in which they were used. Highlighted words are discussed in the main text.

2000		2001		2002		2003		2004		2005		2006		2007	
%	Word	%	Word	%	Word	%	Word								
14.0	Children	12.2	Children	13.2	Children	12.4	Children	13.3	Children	13.6	Children	11.7	Children	16.0	Children
10.2	Future	9.4	Future	8.4	Politics	10.7	Politics	13.0	Politics	12.1	Politics	10.7	Future	13.0	Future
9.2	Politics	8.6	BSE	8.2	Future	9.7	Future	10.5	Future	11.2	Future	7.7	Politics	9.4	politics
4.9	Unemployment	6.5	Politics	6.0	Unemployment	6.6	Unemployment	6.4	Unemployment	8.4	Unemployment	7.3	Unemployment	5.0	Health
4.4	Germany	5.7	Health	4.2	Germany	4.7	Germany	5.4	Germany	5.7	Germany	5.6	Germany	4.3	Germany
4.3	Youth	4.0	Unemployment	3.9	Youth	4.4	War	4.3	Development	3.8	Health	4.8	Health	4.0	Youth
4.1	Health	3.8	Germany	3.7	Health	3.7	Youth	4.1	Health reform	3.7	Job	3.9	Pension	3.7	Unemployment
3.6	Development	3.4	Job	3.2	Worries	3.6	Health	3.9	Health	3.5	Social	3.9	Development	3.6	Worries
3.6	Worries	3.1	Development	3.1	Job	3.2	Job	3.6	Youth	3.4	Development	3.9	Youth	3.2	Development
3.6	Job	3.0	Youth	3.0	People	3.2	Irak	3.1	Job	3.3	Worries	3.2	Job	3.2	Family

Table 2: The ten most commonly used words in the reply to the open-ended question asking for “other worries” for the years 2008 to 2016 and the proportion of text answers in which they were used. Highlighted words are discussed in the main text.

2008		2009		2010		2011		2012		2013		2014		2015		2016	
%	Word	%	Word	%	Word	%	Word	%	Word	%	Word	%	Word	%	Word	%	Word
13.7	Children	19.6	Children	16.1	Children	11.5	Children	17.2	Children	15.6	Children	19.6	Children	17.3	Children	14.8	Children
10.3	Future	15.2	Future	11.8	Future	9.2	Politics	12.0	Future	11.0	Health	12.2	Future	11.0	Health	9.1	Future
9.1	Politics	8.2	Politics	10.6	Politics	8.1	Future	10.0	Politics	9.6	Future	10.6	Health	10.4	Future	7.8	Politics
5.4	Germany	6.7	Health	7.6	Health	6.6	Health	8.8	Health	8.7	Politics	7.9	Politics	6.4	Family	7.1	Health
5.0	Health	5.2	Family	4.4	Germany	4.9	Germany	5.4	Family	5.9	Family	6.3	Family	5.2	Politics	5.6	Germany
3.9	Family	4.0	Development	4.3	Family	3.7	Family	4.3	Worries	4.6	Worries	4.9	Development	5.0	Worries	5.3	Refugees
3.5	Development	3.9	People	3.7	Youth	3.6	Development	3.8	Germany	3.5	Germany	4.8	Worries	4.5	Germany	4.6	Family
3.2	Pension	3.7	Youth	3.5	People	2.8	People	3.8	People	3.0	Development	3.7	Germany	3.9	Social	4.4	Worries
3.2	Worries	3.6	Germany	3.4	Worries	2.3	Society	3.3	Development	2.9	People	2.7	Job	3.2	Development	3.0	People
3.1	Youth	3.3	Worries	3.3	Social	2.3	Worries	3.3	Job	2.7	Social	2.6	People	2.8	People	2.7	Development



Figure 2: 25 most frequently used words in 2016 describing “other worries,” with font size reflecting word frequency. Notice that “refugee politics” corresponds to a single word in German (“Flüchtlingspolitik”), which we decided not to decompose to preserve more detailed information

The numbers show that the worries of SOEP respondents changed over the course of time. From an analytical point of view, this implies that analyses of age effects should apply means to control for the apparent calendar effects.

Worries about the general economic situation peaked from 2003 to 2006, and then again in 2009 and 2010 before they reached a comparably low and stable level over the last years (Figure 1, Panel A). This trend mirrors the peak of unemployment in Germany in the years after 2000 and the worldwide financial crisis in 2008 and 2009. Across the years, the mean level of worries about the general economic situation was highly correlated with the unemployment rate as reported by the German Federal Employment Agency, $r = .80$, $p < .001$.

In contrast, worries about peace were especially pronounced in 2003 (Figure 1, Panel 4), the year of the Iraq War. Furthermore, we can observe the general and more recent trend that respondents reported increased worries about peace, the development of criminality in Germany, immigration to Germany, and hostility toward foreigners in Germany starting from 2014/2015; a pattern that likely captures the respondent's concerns in the context of the so-called European migrant crisis, including both worries about the immigrants as well as reactions to immigration. This becomes quite clear when we look at the top 25 words in 2016, as depicted in Figure 2: respondents worried about refugees, the state of Europe and the European Union, as well as about the new German right-wing party, AfD ("Alternative for Germany", Kroh and Fetz 2016), which has become increasingly popular since it was founded in April 2013.

In contrast to these domains of concern, which appear to be affected by current events, respondents reported approximately constant levels of worries about their personal health across the different survey waves. This indicates that worries about health, first and foremost, are a private issue that are decoupled from larger political events. Furthermore, this might underline the quality of the survey data, since unfortunate placement of the worry items could have introduced systematic but invalid differences between the years. For

example, more detailed assessments of health in some, but not all of the survey years, *could* have led to temporal spikes in health-related worries (as the health items preceded the worry items); however, data show that the worry items were not affected by such changes in questionnaire content.

Interestingly, the response rate to the open-ended question about "other worries" showed comparably little change over time, with a peak in 2004 at about 18% followed by a slow decline with a bump in 2012 and 2013 that levels off slightly above 10%. The peak for reporting of "other worries" was reached when the unemployment rate peaked in Germany, probably because it directly affected the lives of a large number of respondents; in contrast, more recent worries about immigration and refugees co-occur with only minor increases of "other worries."

This might again lead to the suspicion that these "other worries" are, first and foremost, private matters. Indeed, as seen in Table 1 and Table 2, worries about children and their future ("future [of] children" was the most frequent bigram, i.e. sequence of two words) dominated the "other worries" across all years, while other rather private matters, such as the health of the family¹ and its various members, were also frequently mentioned. This could be taken as a sign that the close-ended worry items do not capture at least two areas of life that are a matter of concern to a substantial number of respondents: The future of one's children and the health of family members.

However, politics and the development of Germany, as well as the future of younger generation in general (as opposed to the future of one's children) and concerns about job loss and unemployment are also frequently mentioned. Thus, the "other worries" capture a

¹ Notice that worries about *one's own* health were included as a close-ended item. A large number of mentions of health in the open-ended question could be attributed to health of family members, such as "my husband's health," "my mother's health," etc., as seen in the results of the topic detection in Rohrer et al. (2017).

wealth of concerns about various distinct life domains. Noticeably, a number of worries considering contemporary events show up in the list of the most frequent words: In 2001, respondents reported worries about BSE (commonly known as mad cow disease), of which the first confirmed case in Germany was reported in November 2000; worries about the Iraq war in 2003; worries about a major health reform in 2004; and lastly, worries about refugees in 2016. Thus, responses to the open-ended items mirror current events.

3 Worries across the life course

To investigate how worries changes with the respondent age, we ran multilevel models with respondents nested within the observations using the R package lme4, using a restricted maximum likelihood estimator. All models include age, age^2 , and age^3 in order to allow for the estimation of smooth age trajectories. Furthermore, in order to capture historical events that reflect in effects of the calendar year, we control for survey year by using dummy variables for all years except the reference year of 2000 in all models. The model includes individual intercepts for each respondent, but slopes were fixed so as to not vary between respondents.

Notice that we did not include any time-varying covariates such as income, household size, or number of children. The effects of chronological age – the passing of time – are necessarily mediated by other processes, such as role changes (e.g. job market entry and retirement, family formation) or age-associated differences in health and cognition. Controlling for these mediating variables would obscure the actual age effect (overadjustment bias, cf. e.g. Schisterman, Cole, & Platt, 2009) and, in the most extreme case, make any age effects disappear: An age effect net of everything that changes with age would be zero, as the sheer passing of time in itself cannot have a meaningful effect. To offer an illustration: The effect of age on mortality (death by natural causes) would disappear if we “controlled for” all conceivable health parameters down to the cellular level, and we

might predict an infinite lifespan based on such a model. Thus, we report all analyses without the inclusion of such time-varying covariates, i.e. the unconditional effects (cf. Baetschmann, 2014).

First, we ran generalized linear mixed-effects models with a logit-link function for the single worry items, predicting whether a respondent reported being “very worried” (as opposed to “not worried” or “somewhat worried”). In almost all cases, the coefficients of the dummy-variables for survey year indicated significant differences between the calendar years, which is not surprising given the large sample size (see Appendix for the numerical results of all multilevel models). Furthermore, they closely mirrored the purely descriptive results in Figure 1, which is why we do not discuss the effects of the calendar year here, instead focusing on the model-implied age trends.

Figure 3 shows the predicted odds ratio of respondents being “very worried” about a specific domain (i.e. odds of reporting being “very worried” divided by odds of being less than very worried; Panel A-G) and the odds of reporting “other worries” (Panel H) in the reference year, 2000. For example, the model implies that the odds of reporting being very worried about criminality (Panel E) for respondents over the age of 70 exceeded two, which means that they were twice as likely to report being very worried than to not report being very worried. Notice that all age trends were statistically significant at $p < .001$, which is not a surprise given the large sample size.

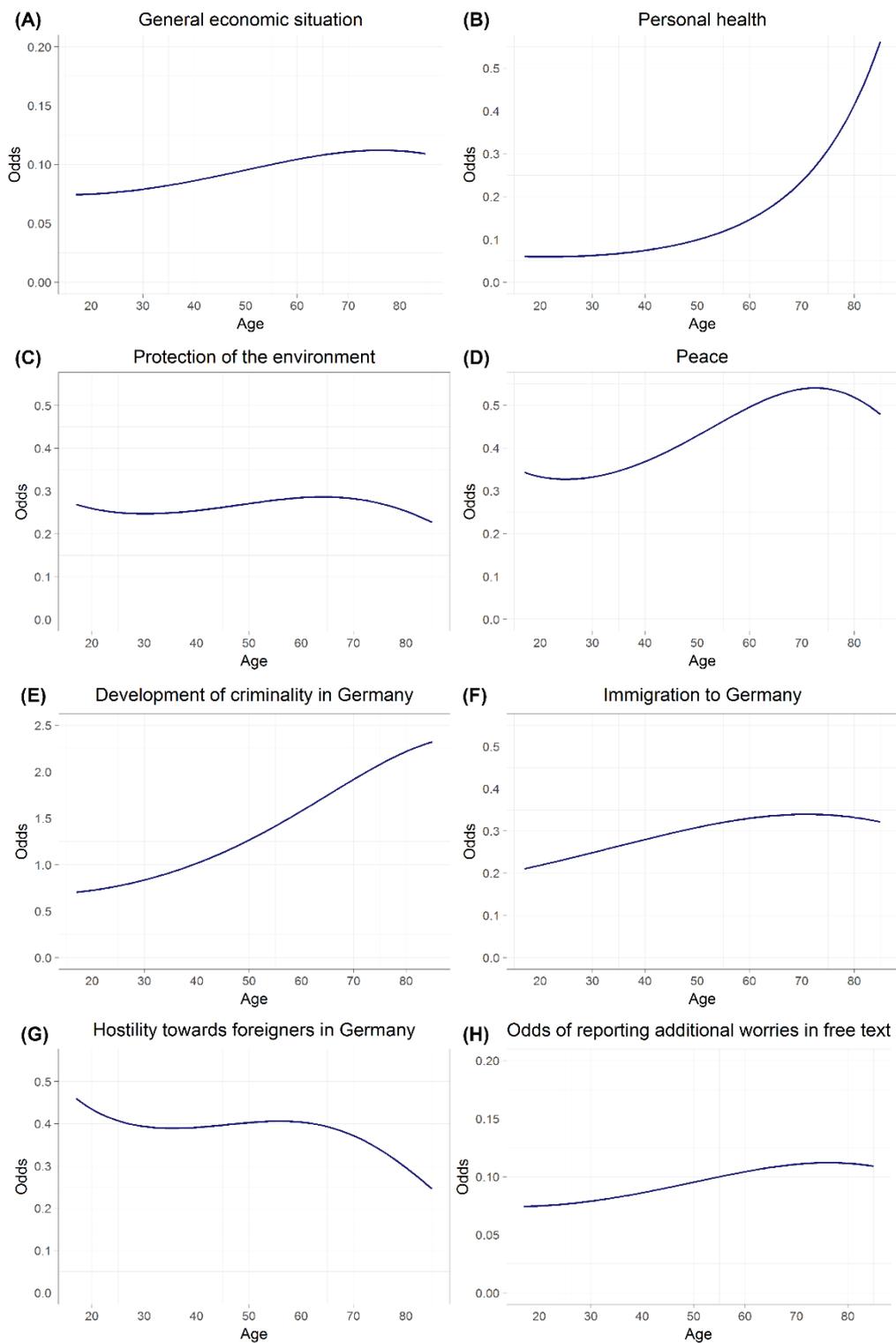


Figure 3: Age trajectories of odds of being very worried about a specific domain, i.e. odds of reporting being “very worried” divided by odds of being less than very worried (Panel A-G) or reporting “other worries” (Panel H) in the reference year 2000, as implied by multilevel

models. Notice that the range of the y-scales deviates for panel A, E and H to magnify statistically significant age trends

Second, we generated a worry score by averaging all seven worry items for each observation. Across all observations, this measure had a satisfying reliability with Cronbach's $\alpha = .72$. The grand mean of this score is 1.10, indicating that the overall mean answer across all respondents, all waves and items is close to "somewhat worried" ($SD = .42$). This worry score was used as the dependent variable in a linear multilevel model. Figure 4 shows the model implied worry scores by age in the reference year 2000.

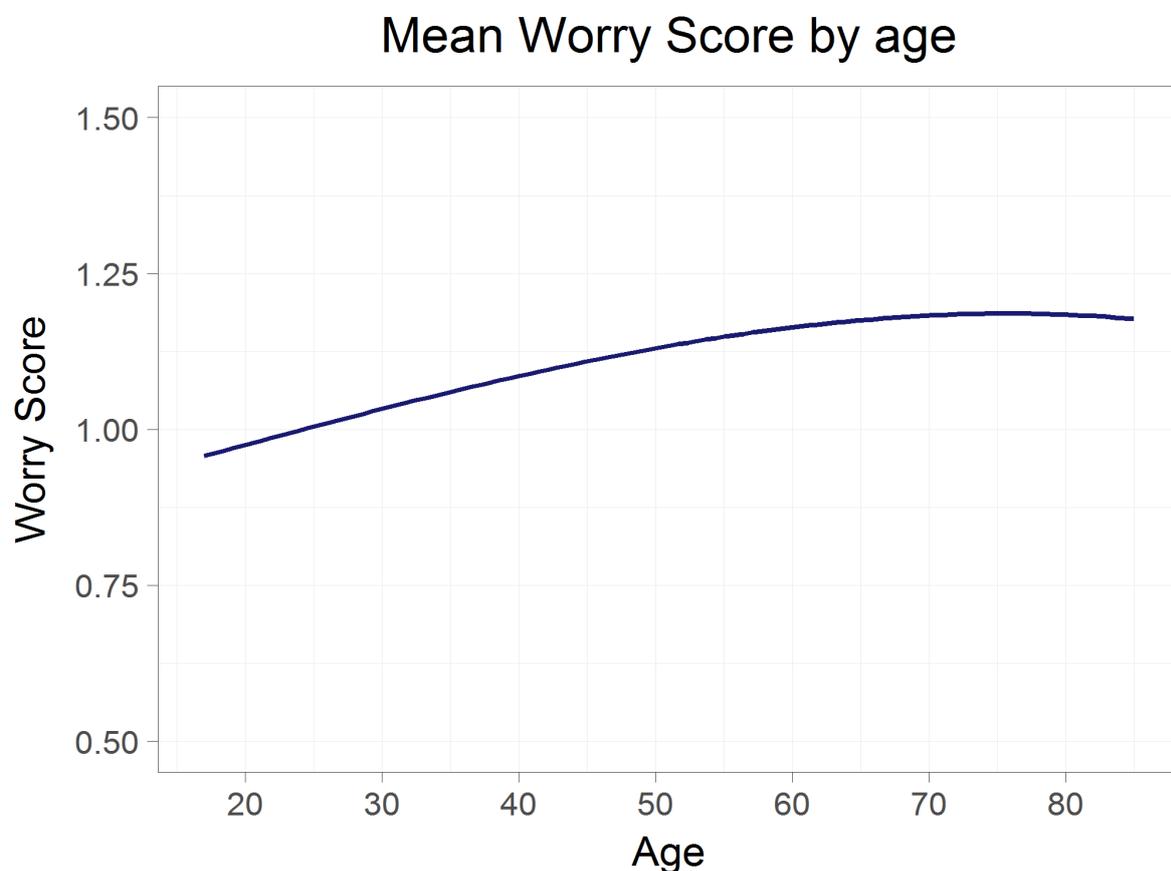


Figure 4: Model implied mean worry scores by age in the reference year 2000, scale ranges from 0 ("not worried at all" in all domains) to 2 ("very worried" in all domains)

Mean worry scores increase across the age span, with an average change of 0.09 SD per ten years of age. This trend is sustained by five of the seven items: Worries about the general economic situation, personal health, peace, the development of criminality, and immigration to Germany all significantly increased with age. Additionally, older respondents are increasingly likely to use the open-ended question to express their worries. However, this general trend is somewhat qualified by the item concerning worries about the protection of the environment that follows a more complex age pattern with only slight age changes, and the item concerning worries about hostility against foreigners, which shows a clear negative age trend.

The age profile of the mean worry score (Figure 4) is in some sense a counterpart to the well-known age profile of well-being: the strong increase of worries up to age 50 is mirrored by a decrease of satisfaction with life. Then, the almost flat line beyond age 60 is in line with the development of satisfaction with life in older age as we show in the following section.

4 The relationship between worries and life satisfaction

The analyses of well-being by means of “happiness research” is a booming research field, especially in economics (Frey 2008), but also in psychology, sociology, political science, and public policy (Diener et al. 2009). As its name implies, “happiness research” often focuses on the positive side of humans’ subjective experience, such as affective well-being or life satisfaction as a more cognitive measure. Age trends on such positive constructs are investigated in depth, but the conclusions are not as clear as one might wish.

For example, Blanchflower and Oswald (2008) found, for the US and European countries, consistent evidence across different birth cohorts of a robust U-shape of happiness in age (although in most surveys the very old people are not well represented). In line with such a pattern, Cheng, Powdthavee and Oswald (2017) find evidence for a midlife

nadir in life satisfaction in their investigation of trends *within* respondents using longitudinal data from four different longitudinal studies. Stone et al. (2010) suggest that global well-being and positive hedonic well-being generally has U-shaped age profiles showing increased well-being after the age of 50; in contrast, negative hedonic wellbeing (including worries) was elevated through middle age and then declined. In line with the peak in *negative* hedonic states, Oswald and Tohamy (2017) provide evidence for increased suicide rates of women in midlife. Schwandt (2016) suggested that the overall U-shaped pattern arises because of unmet expectations in middle age: Young adults might overestimate their future life satisfaction, leading to regret in midlife.

However, the narrative of the U-shaped trajectory is not uncontested. Kassenboehmer and Haisken-DeNew (2012) suggest that accounting for fixed effects and respondent experience in the panel led to a flat trajectory. Newer studies attempt to separate age, period, and cohort effects on the level of life satisfaction and to consider the different durations longitudinal study participation (Baetschmann, 2014). Baetschmann suggests that the total U-shape effect is small in magnitude, thus also compatible with findings of flat trajectories. Based on SOEP data, he finds that, for Germany, life satisfaction is on average mildly decreasing up to age 55 followed by a hump shape with a maximum at age 70, followed by a steep decline toward the end of respondents' lives (Gerstorff et al 2008).

So far, we have investigated worries that are arguably located at the negative end of the well-being continuum.² As we find negative age trends in worries, it might be interesting to consider whether the age trends in worries can, in part, explain age trends in well-being. Thus, we ran additional exploratory analyses, this time modelling life satisfaction as a function of age and calendar year, following the statistical approach that does not “control”

² For example, across all observations, we found a negative correlation between the worry score and life satisfaction, $r = -.18, p < .001$.

for mediators as outlined before. Considering the more complex age trajectory of well-being suggested by Baetschmann (2014), the model additionally contained age to the power of four. We then re-ran the model including all seven worry items (dummy-coded) into the analyses. All worry items had significant main effects on life satisfaction with negative effects for worries about the general economic situation, health, and immigration to Germany, but positive effects for worries about the environment, peace, criminality, and foreigners. These positive correlations between life satisfaction and worries are likely explained by confounding variables such as socio-economic status. For example, respondents with higher education and higher levels of satisfaction with life are more worried about the environment than less educated respondents. However, these somewhat unexpected main effects are not the focus of the present study. Nonetheless, it might be noteworthy that the inclusion of worries in our analysis of satisfaction might partially indirectly control for time-varying socio-economic variables like education and income.

Figure 5 shows the model-implied age trajectories for both versions of the model. The blue line shows that in our standard model (without including worries), life satisfaction steadily decreases until age 50. Instead of a hump with a peak at age 70 (cf. Baetschmann, 2014), in our model, an almost flat line follows prior to the well-known terminal decline in very old age. In the model including the worry items, the age trend appears to be different (and as expected compared to the literature): The decline until age 50 is followed by a hump shape that peaks at 70 prior to the terminal decline, which now appears less pronounced. Note that the red line, i.e. the age trajectory controlling for the seven worry items, indicates a higher level of life satisfaction only because it is estimated for “no worries” on all seven worry items.

Importantly, even after statistical control for worries, age trends on life satisfaction overall persisted ($p < .001$). This indicates that negative and positive indicators of well-being are not interchangeable and cannot be reduced to a simple one-dimensional construct: While the age trends in both measures complement each other, they provide non-redundant

information. Multiple measures with different focuses should be taken into account to gain a more complete picture of SWB (cf. Diener, Scollon & Lucas, 2003).



Figure 5: Model implied life satisfaction scores in the reference year 2000 with and without statistical control for the seven worry items, scale ranges from 0 (“completely dissatisfied”) to 10 (completely satisfied). Values in model with worries at 0 (“no worries”) on all seven items.

5 General discussion

To summarize our results, we investigate how worries in Germany change across time and age, drawing on both close- and open-ended questions. We find that world events influence worries. For example, worries about peace peaked in 2003, the year of the Iraq War, which was further sustained by the fact that Iraq and war were among the top words that

respondents used in reply to the question whether they had “any other worries,” while corresponding patterns are observable for other items. Importantly, worries about both immigration and hostility against foreigners have increased substantially since 2014 with refugees and the rising right-wing political party, AfD, the objects of “other worries” as reported in text format in 2016.

Furthermore, results indicated that worries about different domains increased with age, which held true for worries about the general economic situation, one’s health, peace, criminality, and immigration; but not for worries regarding hostility toward foreigners in Germany. It is possible that this age-related increase in worries is a result of a kind of biased selection of items that are more relevant to older adults more than younger adults. However, if this was true, one would probably expect that younger adults would increasingly use the open-ended question to voice additional worries that plague them, whereas we observed in our data that the chances of answering the open-ended question also increases with age. Lastly, we find that worries are not able to account for the age trends in life satisfaction, underlining that these two measures are not redundant.

Regarding the potential of analysing text answers to open-ended questions, we demonstrate that open-ended questions can capture relevant domains that were not covered by the items explicitly included in a questionnaire; in this case worries about the future of children and the health of one’s own family. Likewise, open-ended questions can help researchers identify topics that they did not consider in their item selection but that are important to respondents. However, in many cases, current world events will bring up topics that slow-paced surveys cannot capture by design: For example, it would have not been possible to predict the importance of the 2001 BSE epidemic when designing the questionnaire multiple months in advance. Furthermore, topics might be rather specific, such as worries about the emergence of one particular party, such that the inclusion of specific items for each of these topics would result in an unwieldy questionnaire; yet a more general item (such as worries about the political development) would be too vague to capture the

underlying phenomenon. Thus, open-ended question in combination with automated text analysis should become indispensable for survey studies trying to capture recent developments with a certain accuracy.

References

- Baayen, H.R. (2001), *Word Frequency Distributions*. Dordrecht: Kluwer Academic Publishers.
- Baetschmann, G. (2014). Heterogeneity in the relationship between happiness and age: Evidence from the german socio-economic panel. *German Economic Review*, 15(3): 393–410.
- Blanchflower DG, Oswald AJ (2008) Is well-being U-shaped over the life cycle? *Social Science & Medicine*, 66: 1733–1749.
- Blei DM, Ng AY, Jordan MI. (2003), Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3: 993–1022.
- Cheng, T.C., Powdthavee, N. and Oswald, A.J., 2017. Longitudinal evidence for a midlife nadir in human well-being: results from four data sets. *Economic Journal*, 127(599): 126-142.
- Diener, E., Lucas, R.E., Schimmack, U. & Helliwell, J. (2009). *Well-being for public policy*. New York: Oxford University Press.

- Diener, Ed, Scollon, C.N. and Lucas, R.E. (2009). *The evolving concept of subjective well-being: The multifaceted nature of happiness*. Springer Netherlands, 2009. 67-100.
- Enquete Commission (2013), Final report by the Enquete Commission: "Wachstum, Wohlstand, Lebensqualität – Wege zu nachhaltigem Wirtschaften und gesellschaftlichem Fortschritt in der Sozialen Marktwirtschaft". *Bundestag printed matter 17/13300*, Berlin: <http://dip21.bundestag.de/dip21/btd/17/133/1713300.pdf>
- Frey, B. S. (2008): *Happiness: A Revolution in Economics*. Cambridge, MA: MIT Press.
- Garcia J, Evans J, Reshaw M. (2004), "Is There Anything Else You Would Like to Tell Us" – Methodological Issues in the Use of Free-Text Comments from Postal Surveys. *Quality & Quantity*, 38(2): 113–25. doi: 10.1023/B:QUQU.0000019394.78970.df.
- Gerstorf, D., Estabrook, R., Ram, N. Schupp, J, Wagner, G.G., Lindenberger, U. (2008), Life Satisfaction Shows Terminal Decline in Old Age: Longitudinal Evidence from the German Socioeconomic Panel Study. *Developmental Psychology*, 44(4): 1148-1159.
- Giesselmann, M., Hilmer, R., Siegel, N.A., Wagner, G.G. (2013), Measuring Well-Being: W3 Indicators to Complement GDP. *DIW Economic Bulletin*, No. 5: 10-19.
- Headey, B., Muffels, R. & Wagner, G.G. (2010). Long-running German panel survey shows that personal and economic choices, not just genes, matter for happiness. *Proceedings of the National Academy of Sciences of the United States of America* (PNAS,) 107(42): 17922-17926.
- Hornsey, M. J., Harris, E.A., Bain, P.G. Fielding, K. S. (2016), Meta-analyses of the determinants and outcomes of belief in climate change, in: *Nature Climate Change*, doi:10.1038/nclimate2943.

- Hsieh, H. and SE Shannon (2005), Three approaches to qualitative content analysis. *Qualitative Health Research*. 15(9):1277–88. doi: 10.1177/1049732305276687.
- Huschka, D. Wagner, G.G. (2012), Statistical Problems and Solutions in Onomastic Research – Exemplified by a Comparison of Given Name Distributions in Germany throughout the 20th Century. *Onoma – Journal of the International Council of Onomastic Science*, Vol. 7: 329-365.
- Kassenboehmer, S. C., & Haisken-DeNew, J. P. (2012). Heresy or enlightenment? The well-being age U-shape effect is flat. *Economics Letters*, 117(1): 235-238.
- Kroh, M., Fetz, K. (2016), Das Profil der AfD-AnhängerInnen hat sich seit Gründung der Partei deutlich verändert. *Wochenbericht des DIW Berlin*, No. 34: 711-719.
- Malterud K. (2001) Qualitative research: standards, challenges, and guidelines. *Lancet*, 358(9280): 483–488. doi: 10.1016/S0140-6736(01)05627-6.
- Mayring, P. (2000), Qualitative Content Analysis. *Forum: Qualitative Social Research*, 1(2). Available from: <http://www.qualitative-research.net/index.php/fqs/article/view/1089>.
- Mehl MR, Gill AJ. (2010), Automatic Text Analysis. In: Gosling S, Johnson JA (eds.) *Advanced methods for conducting online behavioral research*. 1st ed. Washington, D.C.: American Psychological Association.
- Minozzi, W., Neblo, M.A., Esterling, K.M., Lazer, D.M.J. (2015), Field experiment evidence of substantive, attributional, and behavioral persuasion by members of Congress in online town halls. *PNAS*, Vol. 112, No. 13: 3937-3942.
- O’Cathain A, Thomas KJ. (2004), “Any other comments?” Open questions on questionnaires - a bane or a bonus to research? *BMC Medical Research Methodology*, 4:25. doi: 10.1186/1471-2288-4-25.

- Oswald, A. J.; Ahmed, T. (2017) : Female Suicide and the Concept of the Midlife Crisis, IZA Discussion Paper No. 10759, Bonn.
- Priem, M., Schupp, J. (2014): Everyone Happy: Living Standards in Germany 25 Years after Reunification, *DIW Economic Bulletin*, No. 11, 65-71.
- Priem, M., Schupp, J., Wagner, G.G. (2015): 25 Jahre nach der Wende: Gibt es Unterschiede bei Sorgen und Zufriedenheiten zwischen Ost- und Westdeutschland? *ifo Schnelldienst*, 68. Jg., Nr. 22.
- Rich JL, Chojenta C, Loxton D. (2013), Quality, rigour and usefulness of free-text comments collected by a large population based longitudinal study - ALSWH. *PLoS one*, 8(7):e68832. doi: 10.1371/journal.pone.0068832.
- Rohrer, J. M., Bruemmer, M. Schmukle, S.C., Goebel, J. Wagner, G.G. (2017). What Else Are You Worried About? – Integrating Textual Responses into Quantitative Social Science Research. <http://doi.org/10.17605/OSF.IO/KSG7D>.
- Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*, 20(4), 488.
- Schupp, J. (2015): Forty Years of Social Reporting and Quality of Life Research in Germany. In: Trommsdorff, G., Assmann, W.R.(Eds.), *Forschung fördern. Am Beispiel von Lebensqualität im Kulturkontext*. Konstanz und München: UVK: 107-126.
- Schwandt, H., 2016. Unmet Aspirations as an Explanation for the Age U-shape in Wellbeing. *Journal of Economic Behavior & Organization*, 122: 75-87.

Stone, A. A., Schwartz, J.E., Broderick, J.E., Deaton, A. (2010) A snapshot of age distribution of psychological well-being in the United States. *PNAS - Proceedings of the National Academy of Sciences of the United States of America* (early edition).

<http://doi.org/10.17605/OSF.IO/KSG7D>.

Wagner, G. G., Frick, J.F., Schupp, J. (2007), The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancement. *Schmollers Jahrbuch*, Vol. 127, No. 1: 139-169.

Wagner, G. G., Bruemmer, M, Glemser, A., Rohrer, J., Schupp, J. (2017), Dimensions of Quality of Life in Germany: Measured by Plain Text Responses in a Representative Survey (SOEP), SOEPpaper No. 893, Berlin.

http://www.diw.de/documents/publikationen/73/diw_01.c.552497.de/diw_sp0893.pdf.

Appendix

Pre-processing of text data*

Open-ended questions are commonly used as complements in surveys that otherwise rely heavily on closed-ended questions. Yet, to this day, there are no established routines for the analysis of the large numbers of textual answers typically generated by such questions. For the purpose of the present study, text answers were pre-processed following the process sketched below, which includes many steps that are considered standard in the field of Natural Language Processing (NLP):

- Tokenization: split text answers into individual words and remove punctuation
- Set all letters to lowercase: align the heterogenous spellings
- Removal of so-called stop words: remove the most common words (e.g. articles, pronouns, conjunctions) because they have no semantic content in this application
- Spelling correction: will remove the most frequent and obvious spelling errors to further reduce the heterogeneity of the text answers
- Replace abbreviations using a custom dictionary of the most frequently used abbreviations in our text answers
- Stemming: Reduce words to their stem to unify different words that share semantics
Words were manually translated into English just before the visual representations were created so that the translation process had no impact on the results of any analysis.

The aim of the pre-processing of text data was to reduce words with similar meanings to identical strings in order to reduce the inherent heterogeneity of the language. In the case of texts written in German, this involves taking into account the peculiarities of this language, in particular the high degree of inflection. Note that our analyses do not incorporate any syntactic information as the answers were very brief and lacked syntactic features; in fact, most respondents only provided one or multiple keywords.

* This appendix is a much shorter version of the respective chapter in Rohrer et al. (2017).

Numerical results of multilevel models

Results of general linear mixed models predicting worries and life satisfaction from age (including compound test of all three age coefficients, all $p < .001$) and calendar year.

	Economic situation		Health		Environment		Peace		Criminality		Immigration		Hostility		Text answer		Worry score		Life satisfaction	
	<i>b</i>	<i>p</i>	<i>b</i>	<i>p</i>	<i>B</i>	<i>P</i>	<i>b</i>	<i>p</i>	<i>b</i>	<i>p</i>	<i>B</i>	<i>p</i>	<i>b</i>	<i>p</i>	<i>b</i>	<i>p</i>	<i>b</i>	<i>p</i>	<i>b</i>	<i>p</i>
<i>Age – standardized</i>																				
Age	0.063	<.001	0.569	<.001	0.114	<.001	0.278	<.001	0.394	<.001	0.156	<.001	0.050	.001	0.176	<.001	0.072	<.001	-0.018	.058
Age^2	-0.197	<.001	0.160	<.001	-0.008	0.269	0.003	.679	0.007	.358	-0.044	<.001	-0.025	<.001	-0.009	.321	-0.017	<.001	0.017	<.001
Age^3	0.048	<.001	-0.010	.168	-0.041	<.001	-0.050	<.001	-0.023	<.001	-0.009	.150	-0.054	<.001	-0.020	<.001	-0.003	<.001	-0.069	<.001
Overall	X(3) = 1,500.70		X(3) = 4,613.10		X(3) = 97.54		X(3) = 719.85		X(3) = 2,225.20		X(3) = 389.35		X(3) = 230.26		X(3) = 340.19		X(3) = 2898.00		X(3) = 1407.02	
<i>Year – Reference 2000</i>																				
2001	-0.126	<.001	-0.047	.139	-0.110	<.001	-0.220	<.001	-0.202	<.001	-0.395	<.001	0.274	<.001	0.012	.707	-0.022	<.001	0.034	.005
2002	0.749	<.001	-0.116	<.001	-0.306	<.001	0.636	<.001	-0.296	<.001	-0.159	<.001	-0.217	<.001	-0.148	<.001	0.028	<.001	-0.117	<.001
2003	1.944	<.001	-0.173	<.001	-0.244	<.001	1.616	<.001	-0.598	<.001	-0.326	<.001	-0.605	<.001	0.168	<.001	0.074	<.001	-0.181	<.001
2004	1.631	<.001	0.143	<.001	-0.293	<.001	0.434	<.001	-0.564	<.001	0.132	<.001	-0.551	<.001	0.461	<.001	0.041	<.001	-0.353	<.001
2005	1.879	<.001	0.088	.007	0.069	.009	0.373	<.001	-0.300	<.001	0.547	<.001	-0.090	<.001	0.319	<.001	0.092	<.001	-0.219	<.001
2006	1.328	<.001	0.135	<.001	0.001	.974	0.576	<.001	-0.543	<.001	0.267	<.001	-0.231	<.001	0.266	<.001	0.062	<.001	-0.280	<.001
2007	0.143	<.001	0.113	.001	0.666	<.001	0.196	<.001	-0.512	<.001	0.000	.987	-0.280	<.001	0.098	.003	0.012	<.001	-0.241	<.001
2008	0.171	<.001	0.080	.017	0.080	.003	-0.054	.030	-0.845	<.001	-0.301	<.001	-0.546	<.001	-0.117	.001	-0.042	<.001	-0.212	<.001
2009	1.281	<.001	0.150	<.001	0.030	.266	-0.249	<.001	-1.199	<.001	-0.639	<.001	-0.797	<.001	-0.239	<.001	-0.042	<.001	-0.264	<.001
2010	0.869	<.001	0.396	<.001	0.208	<.001	-0.209	<.001	-0.889	<.001	-0.533	<.001	-0.668	<.001	-0.416	<.001	-0.037	<.001	-0.052	<.001
2011	-0.044	.107	0.400	<.001	0.265	<.001	-0.218	<.001	-1.276	<.001	-0.432	<.001	-0.888	<.001	-0.576	<.001	-0.079	<.001	-0.222	<.001
2012	0.057	.039	0.363	<.001	0.045	.106	-0.397	<.001	-1.391	<.001	-0.859	<.001	-0.770	<.001	-0.322	<.001	-0.105	<.001	-0.164	<.001
2013	-0.248	<.001	0.172	<.001	-0.090	.002	-0.228	<.001	-1.422	<.001	-0.597	<.001	-0.865	<.001	-0.341	<.001	-0.116	<.001	-0.106	<.001
2014	-0.710	<.001	0.380	<.001	0.018	.482	0.325	<.001	-1.244	<.001	-0.204	<.001	-0.646	<.001	-0.474	<.001	-0.082	<.001	-0.085	<.001
2015	-0.527	<.001	0.269	<.001	-0.076	.005	0.783	<.001	-0.953	<.001	0.247	<.001	0.128	<.001	-0.599	<.001	-0.005	.113	0.024	.075
2016	-0.555	<.001	0.189	<.001	0.102	<.001	1.278	<.001	-0.227	<.001	1.289	<.001	1.019	<.001	-0.600	<.001	0.109	<.001		