1 Introduction

Writing in their seminal review of the state of subjective well-being research, Diener *et al*. (1999) highlighted the predominance of cross-sectional survey data and the need for future research to embrace other research designs, including longitudinal data collections. This call for a change in research direction has been heeded, with the past decade or so witnessing a marked growth in the amount of published research into subjective well-being that has employed longitudinal data.

Longitudinal (or panel) data, however, have two weaknesses that have the potential to impart serious biases to survey estimates: (i) non-random attrition; and (ii) panel conditioning (Lynn 2009). The former has long been the subject of considerable research interest (see Watson and Wooden 2009). Panel conditioning, on the other hand, has received far less attention.

Panel conditioning is present when the very process of repeatedly administering the same survey questions to the same respondents over a period of time influences responses (Lazarsfeld 1940). This may manifest either in changes in actual behavior or in changes in the way respondents answer questions (Kalton *et al*. 1989).

In the context of reports of subjective well-being, survey participation may stimulate respondents to think more carefully about the subject matter covered by the questions, leading to more thoughtful responses at later survey waves. Repeated survey participation may also be associated with both a growing sense of trust between respondent and interviewer and greater confidence in and identification with the survey more broadly. This, we hypothesize, will have the effect of eliciting more truthful responses over time, and hence reducing the level of social desirability bias (see Paulhus 1991), which we expect to be reflected in a tendency for respondents to overstate their level of well-being. The effect of both of these mechanisms is that data quality improves with frequency of survey participation, but this improved data quality comes at the cost of longitudinal inconsistency. But it should not be assumed that panel conditioning necessarily results in improved data quality. Most obviously, repeated survey participation is often associated with growing disinterest and engagement on the part of the respondent. This is one factor contributing to sample attrition, but among participants that do not drop out it might equally be reflected in responses which are motivated by the desire to minimize interview burden.

It is this relationship between responses to subjective survey questions about well-being and panel conditioning that is at the center of this paper. More specifically, we use longitudinal survey data collected from members of a large national probability sample of households in Australia to examine whether mean scores on both self-reported measures of life satisfaction and mental health, as well as the degree of dispersion in those scores, exhibit any tendency to change over time in a way that might reflect panel conditioning, and to quantify the magnitude of those differences.

The paper makes at least four contributions to the literature. First, the use of Australian data is in its own right a significant contribution, with most previous research on this issue employing data from the same German data set – the German Socio-Economic Panel (GSOEP). Second, this study goes beyond the focus on self-evaluations of life satisfaction to also examine scores on a well established subjective measure of mental health (or emotional well-being). Third, we examine not just changes in the level of subjective well-being but changes in dispersion. Fourth, this study is among the first to attempt to disentangle the effect of panel conditioning from not only aging effects, but also from the effects of non-random attrition.

2 Related Research

The empirical literature on panel conditioning is both relatively small and diverse, reflecting the fact that panel conditioning is content dependent (Lynn 2009). Early studies were mainly concerned with the effects of opinion polling on voter participation, with evidence generally suggesting that voter turnout was enhanced by survey participation, and moreover that this effect was enhanced by repeated interviewing (e.g., Traugott and Katosh 1979). Studies of panel conditioning effects on other behaviors are rare, with most research analyzing changes in reporting rather than changes in actual behaviors. Further, in many cases the size of the effects are found to be small, if not absent entirely, or restricted to quite a narrow range of outcome variables (e.g., Sobol 1959, Klein and Rubovits 1987, Corder and Horvitz 1989, Silberstein and Jacobs 1989, Pennell and Lepkowski 1992, Toepoel *et al*. 2009, Das *et al*. 2011).

Nevertheless, there is research which suggests that panel conditioning effects are present and can matter. It has, for example, long been recognized that recorded unemployment within the rotation sample of the US Current Population Survey falls with length of time in the sample (e.g., Bailar 1975), a finding which is usually assumed to reflect differences in reporting rather than any substantive change in employment situations. Other examples include: (i) Waterton and Lievesley (1989), who reported evidence that attitudes are influenced by previous interviews, and notably that panel respondents are increasingly more likely to report honestly; (ii) Veroff *et al*. (1992), who reported evidence that, in the longer-term, marital satisfaction may be enhanced by participation in a longitudinal survey about marriage; (iii) Wilson and Howell (2005), who reported evidence that the upward trend in the incidence of arthritis within the aged population in the US Health and Retirement Study is spurious, and speculate that panel conditioning may be the cause (e.g., if survey participants are motivated to consult with physicians about health symptoms); (iv) Frick *et al*. (2006), who reported evidence that the quality of reported income data improves with successive interviews; and (v) Torche *et al*. (2012), who found that questioning about substance use is associated with reductions in reported levels of substance use at later survey waves.

The notion of panel conditioning has also recently had some influence on research into subjective well-being. First, Sturgis *et al*. (2009) used data from the British Household Panel Survey (BHPS) to show that the measured reliability of a multi-item life satisfaction scale increased with successive waves of administration. Further, since these questions were not introduced until the 7th wave of survey administration, this finding implies the causal mechanism operates through learning behavior rather than through any reduction in social desirability. Less clear is whether this increase in reliability represents an improvement in data quality. Learning behavior may, for example, lead respondents to respond in ways that facilitate shorter interview times at the expense of more accurate reports.

Second, and very differently, Landau (1993) observed that the distribution of responses to a single-item overall life satisfaction question (scored on a 0 to 10 bi-polar scale) administered as part of the GSOEP changed noticeably over the first four waves of the panel, with far fewer respondents choosing high values on this scale (9 or 10) in wave 4 than in wave 1. Further, no such change could be found over the same period in a repeated cross-section study employing a more or less identical question. These observations are consistent with the hypothesis that panel conditioning is associated with a decline in social desirability bias. As a result, subsequent analyses of life satisfaction using the GSOEP data have included a measure of the number of times interviewed in order to control for this panel conditioning effect (e.g., Frijters *et al*. 2004, Headey *et al*. 2010).

Third, van Landeghem (2012) used data from both the GSOEP and the Swiss Household Panel Survey (SHP) and reported that self-reported life satisfaction scores are, after adjusting for age, higher for first-time respondents to refreshment samples than for more experienced sample members. The weakness in this comparison, however, is that the composition of the sample of experienced respondents is affected by attrition. To get around this problem, the sample of first-time respondents was restricted to persons who subsequently respond to the survey in at least four successive survey waves. Very large differences – of around 0.6 on the 11-point scale – between responses at the first and sixth (or subsequent) interviews were found in the German sample. In contrast, in the Swiss sample the differences were both smaller and more short-lived, possibly reflecting differences in survey administration; unlike the GSOEP, the SHP is administered entirely by telephone.

Finally, Kassenboehmer and Haisken-DeNew (2012) have also shown that life satisfaction scores decline with length of time in the GSOEP panel (with a quadratic specification revealing that the rate of decline gradually lessens over time), though the estimated magnitude of this effect is much smaller than reported by van Landeghem (2012), presumably reflecting the absence of any controls (aside from time) in the latter. More importantly, Kassenboehmer and Haisken-DeNew (2012) demonstrate that the failure to appropriately control for this effect can lead to flawed inferences. Specifically, they show that the often reported U-shaped relationship between life satisfaction and age (e.g., Blanchflower and Oswald 2008) disappears in the GSOEP data once both fixed individual effects and time in panel are controlled for.

To our knowledge, such results have, with one exception, not been replicated using other panel survey data sets that collect data on subjective well-being. The exception is Frijters and Beatton (2012), who estimate regression models of life satisfaction using data from the GSOEP, BHPS, and Household, Income and Labour Dynamics in Australia (HILDA) Survey. While the focus of their research is on the relationship between age and life satisfaction, the results presented in their supplementary tables (available only on the journal website) reveal that there is a significant negative time-in-panel effect in all three data sets. The magnitude of this effect, however, is most pronounced in the GSOEP data.

3 Hypotheses

A priori we do not have strong hypotheses. However, if social desirability bias effects dominate, then we would expect initial survey reports of subjective well-being to be overstated and for these reports, other things held constant, to converge on the true level of well-being (and so decline) with repeated survey participation. This seems to be consistent with the findings reported in previous studies, and especially those using the GSOEP data.

Very differently, and not previously examined, the effects of panel conditioning might also manifest in changes in the distribution of reported scores. Again it is difficult to predict *ex ante* the direction of those effects. If social desirability bias dominates, then it might be expected that distributions would widen as respondents become more comfortable with interviewers and the survey more generally, and so feel less compelled to choose scores closer to some social norm. Alternatively, repeated survey participation will be associated with learning which, in turn, might lead respondents to both avoid extreme values in future survey waves and to condition responses on responses at previous survey waves. Repeated survey participation would thus be associated with a reduction in the within-person dispersion of reported well-being scores.

4 The HILDA Survey Data

Described in more detail in Watson and Wooden (2012), the HILDA Survey is a household panel survey with a focus on work, income and family, but with coverage that extends to many other topic areas, including health and subjective well-being. The data used in this analysis are (mostly) drawn from Release 10 of the HILDA Survey in-confidence unit record file, which covers the first ten years (or waves) of data collection (or roughly the period 2001 to 2010).[1]

4.1 Sample and Survey Administration

The survey commenced in 2001 with a national probability sample of Australian households. Personal interviews were completed at 7,682 of the 11,693 households identified as in scope for wave 1, which provided an initial sample of 13,969 individual respondents. While non-response was considerable, the characteristics of the sample appear to match the broader population quite well. The main weaknesses of the initial unweighted sample are a slight over-representation of females, and an under-representation of both immigrants from a non-English-speaking background and residents of Australia's largest city, Sydney (Wooden *et al.* 2002).

The members of these participating households form the basis of the panel pursued in the subsequent waves of interviews, which are conducted approximately one year apart (with the

[1] Data from Release 11, which adds results from wave 11 (conducted between August 2011 and February 2012), are used to construct the measure of panel attrition that is used in the regression analyses. They are also used to identify whether there has been any population-wide change in subjective well-being levels.

fieldwork concentrated into the period between September and December). Interviews are conducted with all adults (defined as persons aged 15 years or older on the 30th June preceding the interview date) who are members of the original sample, as well as any other adults who, in later waves, are residing with an original sample member.

A large population refreshment sample was introduced in wave 11 (2011) which added a further 2153 responding households (see Watson and Wooden 2013). But given we do not yet have longitudinal data from this sample, they are (with one small exception) not used in the analyses reported here.

To study the impact of panel conditioning it is important to be able to separate out the effect of one further year of survey participation from any aging effect. This is not possible in a panel study employing a fixed cohort design and where non-response at any survey wave represents a permanent exit from the study. Like other household panels, however, the HILDA Survey employs an indefinite life design. It thus augments its sample each wave with household members who reach interview age (15 years in the HILDA Survey) and with other persons who join the household, and attempts to re-contact persons who have been non-respondents at previous survey waves (Watson and Wooden forthcoming). As a result, with these data it is possible to separate time-in-panel effects from aging effects.

The sample used here begins with the unbalanced panel of responding persons, a dataset comprising a total of 130,211 observations from 21,280 people.

The principal mode of data collection is face-to-face personal interviews. Telephone interviews are conducted both as a last resort and to reach sample members that move to locations not covered by the network of face-to-face interviewers. The proportion of interviews conducted by telephone in wave 1 was negligible. By wave 8, however, this proportion had reached 10%, before falling back to 8.4% by wave 10.

Annual re-interview rates (the proportion of respondents from one wave who are successfully interviewed the next) are high, rising from 87% in wave 2 to over 94% by wave 5. Over the next 5 years (waves 6 to 10) the annual re-interview rate was relatively stable and averaged 95.5%. These relatively high rates are the result of numerous factors, but perhaps most significant are the amount of effort devoted to tracking and locating sample members, the use of cash incentives, and a multi-phase fieldwork period designed to maximize opportunities to gain cooperation.[2] The maximum number of call attempts made to any household in a single survey wave, for example, has ranged from 19 (in wave 2) up to 50 (in

[2] For further details about the HILDA Survey and its administration, see Summerfield *et al.* (2012).

wave 9). The mean number of call attempts per households, of course, is far less, averaging between 4.8 and 6.1.

All interview respondents are also given a separate self-completion questionnaire (SCQ). This instrument is handed directly to respondents by interviewers, the majority of which (70% on average) are also collected by those interviewers. If not completed by the time of collection, respondents are asked to return them by mail. These instruments also have to be distributed by mail to telephone interviewees. As a result, there is additional non-response associated with this component of the survey; on average, about 90% of all interviewees complete and return the SCQ.

4.2 Measuring Subjective Well-being

The HILDA Survey regularly collects data on a number of subjective well-being measures. Most notably, as part of the personal interview component of the survey, respondents are asked to rate, on a 0 to 10 scale, their satisfaction with eight aspects of life – the home in which they live, their employment opportunities, their financial situation, how safe they feel, the extent to which they feel part of their local community, their health, the neighborhood in which they live, and the amount of free time they have. This is then followed by a question about overall life satisfaction. The question reads: "All things considered, how satisfied are you with your life? Again, pick a number between 0 and 10 to indicate how satisfied you are". A visual aid is used in the administration of these questions, which involves a pictorial representation of the scale with the extreme points labelled "totally dissatisfied" and "totally satisfied". The approach adopted is similar to that used in other large surveys, including both household panels (notably the GSOEP) and cross-sectional surveys (e.g., the World Values Survey).[3]

In addition, as part of the SCQ, the Short Form (SF36) Health Survey is administered. Described in more detail by Ware *et al*. (2000, 2007), the SF36 is a survey of generic health concepts that has been extensively tested and used around the world. It comprises 36 items that can then be used to construct multi-item scales measuring eight different health concepts. One of those sub-scales is the Mental Health Inventory (MHI-5), a measure of psychological well-being that has proven to be an effective screening instrument for persons with mental health problems in large populations (e.g., Rumpf *et al*. 2001; Hoeymans *et al*. 2004). It comprises five items that assess frequency (on a 6-point scale) of symptoms of anxiety and

[3] The World Values Survey employs a 10 point scale, rather than an 11-point scale. It also only measures satisfaction with one other life domain – the household's financial situation.

mood disturbance over the 4-week period preceding interview. The response options range from “all of the time” to “none of the time”, with all response options fully labelled. Like all SF36 sub-scales, raw scores on each item are summed and then standardized so that the scale values range from 0 to 100. Relatively low scores are indicative of a poor mental health state.

In this study we examine two outcome variables: overall life satisfaction (LS) and the MHI-5. Conceptually the two measures are distinctly different – LS provides a cognitive appraisal of overall well-being while the MHI-5 is intended to capture moods and emotions, or affective reactions to life circumstances (emotional well-being). Further the MHI-5 has a specific reference period (the preceding 4-weeks) whereas there is no specified time frame for the overall life satisfaction measure. Despite these differences, however, the two measures are quite strongly correlated, with the wave specific Pearson correlations ranging from .44 to .49.

5 Analyses and Results

5.1 Descriptive Statistics

We begin, in Table 1, by presenting mean scores on the overall life satisfaction measure (and standard deviations) cross-tabulated by the number of times responding in the panel. As hypothesized, mean scores decline over time, consistent with the notion that social desirability may cause responses in the first wave to be upwardly biased. The magnitude of the decline, however, while statistically significant, is quite small, with mean life satisfaction falling by only 0.15 of a point (or less than 2%) over the observation period. Given the unbalanced panel is affected by attrition, we also report scores after restricting the sample to the balanced panel (persons that responded at all ten survey waves). As can be seen, the observed decline in mean life satisfaction for this sub-group is almost identical to that observed in the unbalanced panel, suggesting that selective attrition is not behind the observed decline.

While the change in the mean scores is quite small, this is not true of the change in the dispersion of scores. The standard deviation of scores at later interviews (sixth through tenth) is noticeably less (up to 13.8% less) than at the initial interview. Such changes in distributions cannot be explained by recourse to social desirability arguments, but are consistent with explanations grounded in learning behavior. Repeated survey participation may thus lead respondents to avoid selection of extreme scores. In these data, for example, 19.3% of respondents in their first wave of participation select the maximum score on the scale (10); by the 10th wave of participation this fraction is just 10.2%.

We next report, in Table 2, like statistics for the MHI-5 measure. As should be apparent, there is no evidence here that means scores on the MHI-5 decline with repeated interviewing. Indeed, if anything, scores increase with successive interviewing. Some of this may reflect non-response and attrition bias, with persons with poor mental health being more likely to discontinue participation and less likely to complete and return the SCQ. Nevertheless, even within the balanced panel, the trend is upwards, though the magnitude of that increase is both very small and statistically insignificant. There is also some weak evidence that the distribution of scores narrows with repeated interviewing.

Overall, these descriptive findings are consistent with the notion that self-reported well-being measures may be subject to modest panel conditioning effects. Further, panel conditioning may manifest less in trends in mean scores, and more in changes in the distributions of those scores.

5.2 Modelling the Level of Subjective Well-being

We next test whether these observed differences in mean scores by time-in-panel are robust to the inclusion of other influences on subjective well-being.

Our model for estimating relationships between variables when using panel data takes the form:

$$SWB_{it} = \mu_i + \beta' X_{it} + \varepsilon_{it} \qquad (1)$$

where SWB_{it} is a measure of the subjective well-being of individual i at time t, X is a set of time-varying conditioning variables, and μ_i is an unobserved individual-specific effect which is assumed to be constant over time. This is the fixed effects (FE) model.[4]

As noted earlier, we use two measures of SWB: overall life satisfaction (LS) and the MHI-5. LS is an ordinal variable, suggesting the need for a non-linear estimator such as ordinal logit. While non-linear estimators that incorporate fixed effects are now available in some software packages (notably LIMDEP), it is still relatively difficult to interpret estimates from these models. We thus follow many others (including Kassenboehmer and Haisken-DeNew 2012) and treat LS as if it is continuous. This can be justified given other evidence showing that, in the presence of fixed effects, the assumption of cardinality or ordinality makes little

[4] An alternative approach is to use a random effects (RE) model, which replaces the individual-specific constant with an individual-specific disturbance term. The major drawback to this approach is that it requires the individual disturbances to be uncorrelated with the explanatory variables, X_{it}. If these conditions are not met then the RE model will not control for the effects of the unobserved differences between individuals. Since the assumption that the unobserved differences be uncorrelated with all observed variables seems quite unrealistic, we prefer the results from FE estimation. Furthermore, in the data analyzed here this is strongly supported by large values on the Hausman test statistic.

difference to the results (Ferrer-i-Carbonell and Frijters 2004). MHI-5 is also not linear, being a standardized scale where values must lie in the range 0 to 100. Nevertheless, the absence of many cases at the limits suggests that it is reasonable to also apply linear regression methods.

The selection of conditioning variables is based closely on the analysis of life satisfaction using HILDA Survey data reported in Wooden *et al*. (2009), but there are a number of important differences. First, we include age rather than time dummies. Further, and slightly different to much of the literature, we find that the best fit is usually found when age is specified as a cubic function (compared with the more usual linear or quadratic specifications).

Second, we also include additional variables to control for: (i) possible wealth effects (home ownership and home equity variables), given the evidence that life satisfaction is responsive to both income and wealth (Headey *et al*. 2008); (ii) location differences; and (iii) interview mode, given some interviews are conducted on the telephone.

Third, to simplify the presentation of results we reduce the set of labor force status variables used in Wooden *et al*. (2009) to just five. These represent the three conventional states, not in the labor force, unemployed and employed, with the latter broken into a further three states based on the difference between usual weekly hours worked and stated working hours preferences – prefer fewer hours, prefer more hours, and prefer to work about the same hours (with the latter serving as the reference category and hence omitted). This more parsimonious list of labor force status variables was also used in the analysis of GSOEP data by Headey *et al*. (2010).

Finally, given the possibility that estimates are biased due to endogenous sample attrition, we include an additional regressor identifying whether the sample member is a non-respondent at the next survey wave. As Verbeek and Nijman (1992) have shown, inclusion of a variable that is a function of observed response behavior provides a simple test of, and control for, selectivity bias without the need for estimating a model that explicitly specifies the response mechanism. It also seems especially important to control for attrition in this study given it is, by definition, correlated (inversely) with the duration of time an individual spends in the panel.

The full list of conditioning variables, together with summary statistics, is provided in Table 3.

5.3 Results

5.3.1 Life Satisfaction – Basic Model

Results from the estimation of the basic model of life satisfaction are presented in Table 4. For comparative purposes, we also present results from a simple pooled data model which does not allow for individual effects (but does assume errors within i are correlated over t). Further, we also present results from the fixed effects estimation after splitting the sample by gender.

Focusing first on the results for life satisfaction using the entire sample, the estimated coefficients are mostly sensible and in line with expectations, and mostly much smaller in the presence of fixed effects. Nevertheless, even with fixed effects most parameter estimates are statistically significant, with the only exceptions being the number of persons in the household, the dummy variables indicating whether an individual lives in an outer regional or remote part of Australia (as defined in the Australian Standard Geographical Structure [see ABS 2001], which in turn are based on a location's score on the Accessibility / Remoteness Index of Australia), and the dummy variable indicating interview mode.

In summary, life satisfaction scores vary significantly with changes in: partnership status (marriage and cohabitation is associated with higher life satisfaction); the number of children; disability and long-term health conditions, and more importantly the severity of that condition; employment status (with unemployment being the most adverse state, but with employed persons reporting preferences for both fewer hours and more hours also being significantly less satisfied than other employed persons); income; and both home ownership and home equity. We also find significant associations with age. More importantly, when we plot the estimated relationship with age, we find that there is a marked U-shaped relationship between age and life satisfaction throughout most of life in the pooled data model, which disappears with fixed effects estimation. Instead, life satisfaction falls with age both early and late in life, but is very flat during much mid-life (that is, between roughly 40 and 70 years of age). Such results are consistent with those reported by both Frijters and Beaton (2012) and Kassenboehmer and Haisken-DeNew (2012).

We also find that reported life satisfaction scores can be influenced by the interview situation, with the presence of other adults in the household at the time of interview associated with slightly higher scores. Finally, there is evidence of sample selectivity, with respondents that cease survey participation at the next wave reporting significantly lower scores in the current wave.

Turning briefly to the gender-specific results, the differences in parameter estimates between men and women are mostly quite small. A notable exception to this is the influence of labor market status, with the results indicating that men that move out of the labor force (from an employment state where they were working hours consistent with preferences) suffer much larger declines in satisfaction than women that make the same move. This result suggests that, compared with women, men that exit the workforce are more likely to do so involuntarily.

All of the significant relationships embodied in these fixed effects models, however, can still be argued to be quite small. Severe disability, for example, only reduces life satisfaction scores by a little more than half a point, and entry into unemployment by about half that. Further, the results presented here say nothing about adjustment. It could be still that even these relatively small variations are not sustained over time.

Finally, it should be noted that our set of explanatory variables explains relatively little of the variance in life satisfaction. It is the time-invariant (and mostly unobserved) characteristics that explain the bulk of the variation across individual life satisfaction (reflected in the rho parameter) and hence account for the large estimated adjusted R-squared value.[5] In contrast, within-person variation in our explanatory variable set accounts for less than 3% of the total variation in life satisfaction. This might reflect omitted variables. Alternatively, it could just signal that there is a considerable amount of transitory noise in self-reported life satisfaction scores.

5.3.2 MHI-5 – Basic Model

We then repeated this regression estimation exercise with MHI-5 as the outcome variable. The results are reported in Table 5. The overall explanatory power of these models is similar to those estimated for LS, with the individual fixed effects again accounting for most of the 'explained' variance. Perhaps surprisingly, the estimated coefficients (in terms of relative size and significance) on most variables are also very similar to those estimated for LS. Notable differences are the lack of significance of children and home ownership in the MHI-5 equations, and the significant negative effect of the number of adults living in the household. The estimated association with age is also slightly different. MHI-5 scores exhibit more of a wave effect, with scores tending to rise in mid-life (and thus consistent with a U-shaped pattern pattern) before declining sharply in old age.

5.3.3 Time-in-Panel Effects on Levels

We next augmented our basic fixed effects specification with the variable, T, the number of times the respondent has been interviewed up to, and including, the current interview. We

[5] Derived from the estimation of a linear regression model designed to handle a large number of groups (using the 'areg' command in STATA). The coefficient estimates from this model are identical to those derived from fixed effects estimation.

experimented with a number of different specifications, the results from three of which are reported in Table 6. Focusing first on life satisfaction (reported on in the top half of Table 6), we begin in specification 1 by simply adding the variable T in it linear form. The estimated coefficient is small and statistically insignificant, which stands in marked contrast to the significant negative effect found by Frijters and Beatton (2012). Following, Kassenboehmer and Haisken-DeNew (2012), we then specified T as a quadratic. The two terms are now statistically significant, but only for men. Note further that the implied turning point in this specification occurs at 8 years, which is close to the observed maximum. In other words, the effect of panel conditioning is to cause men to lower their reported life satisfaction over time, but with this effect concentrated mostly in the first few waves of survey participation, which is exactly what would we expect. Less obvious is why no such relationship is found among female respondents.

Finally, we tried the most flexible functional form – one that allows the effect of T to vary with every value that T can take. For men, and entirely consistent with the results from the quadratic specification, we find an immediate drop in life satisfaction scores (of about 0.06 of a point) at the second interview, followed by a more gradual deterioration over the course of the panel. The magnitude of the effect, however, is small and imprecise (though a test of joint significance of the nine dummy terms does lead us to reject the null hypothesis that all coefficients are equal to zero). By the tenth interview, life satisfaction scores are, on average, just 0.1 of a point lower than at the first interview, and this difference is not statistically significant. Interestingly, this relationship is both much larger (double the magnitude) and much more precise, if the variable controlling for attrition is omitted. In other words, it appears that some of what might appear to be a panel conditioning effect is due to selectivity bias. Among women there is also an immediate drop in life satisfaction at the second interview, but this is not sustained. Indeed, life satisfaction scores rise with further participation, and the estimated coefficients are jointly significant (p <.001). The magnitudes of these positive coefficients, however, are still very small (the largest is .139).

These findings provide little support for the hypothesis that reported SWB will decline with repeated survey participation. We are thus drawn to the conclusion that any panel conditioning effects on the level of life satisfaction scores in the HILDA Survey are small, and for most purposes ignorable.

The identical exercise performed with respect to MHI-5 (and reported on in the bottom half of Table 6) suggests a different conclusion, at least for women. If we focus first on all persons, it can be seen that none of the parameters of interest are statistically significant. It turns out, however, that pooling males and females is hiding significant associations. Among men the pattern of coefficients in each of the three specifications is very similar to that reported for life satisfaction. Thus after controlling for all individual-specific time-invariant characteristics, as well as numerous time-varying influences, scores on the MHI-5 scale tend to decline with time in the panel, though this effect is estimated with a relatively low degree of precision; the coefficients in specification 3 are jointly insignificant. The findings for women, however, are very different. For women, scores on the MHI-5 rise with length of time in the panel, as reflected in the large and positive coefficient on T in both specifications 1 and 2, and in the rising magnitude of the dummy terms in specification 3 (which are jointly significant; $p = .018$). Indeed, by the 9th wave of participation the time-in-panel effect adds close to 5 points to the reported score (and is highly significant). Further, this effect is, relative to other estimated effects, very large. It is larger than both the effect of women entering unemployment (1.7 points) or of acquiring a moderate (that is, work-limiting) disability (3.9 points).

Again, findings with respect to MHI-5 provide little support for the hypothesis that reported SWB will decline with repeated survey participation. Indeed, for women we find strong evidence of a large effect in the opposite direction.

5.3.4 Time-in-Panel Effects on Dispersion

We next return to the issue of dispersion, and test whether the differences observed in the raw descriptive data are still apparent once we control for other influences, and especially time-invariant traits and characteristics. To this do this we once again estimate fixed effects regressions, but where the dependent variable is the absolute deviation in subjective well-being from the within-person mean. In the absence of any strong theoretical model explaining such deviations, we opt to use the same specification as used in estimating models of the level of well-being.

In the interest of brevity we only report the results on the coefficients of the variable of key interest – length of time in the panel. These are provided in Table 7. Focusing first on life satisfaction, it should be immediately apparent that there are significant time-in-panel effects. These, however, are not so obvious when using a simple linear specification for time; the relationship is clearly non-linear. Among males we find a strongly significant relationship ($p < .001$), with the average size of the deviation from the within-person mean declining over time until about the 6th wave of participation, before beginning to grow again. The magnitude of the effect, however, is arguably still quite small, reaching about .17 of a point at its maximum. Among women, however, the size of the effect is much larger (and again highly significant), reaching almost 0.4 of a point, with little evidence of a turning point in the relationship being reached; the size of the decline in deviations gradually increases, but at a declining rate.

Very similar patterns are also found for the MHI-5 measure. Among both men and women, repeated interviewing is associated with a gradual regression towards the within-person mean, which does not stabilize until about the 8^{th} wave of participation.

These findings are consistent with the hypothesis that responses are affected by learning behavior. They are entirely inconsistent with the hypothesis that reports are affected by social desirability bias.

6. Conclusions and Discussion

This paper was motivated by the possibility that self-reported answers to questions about subjective well-being in panel surveys may change over time for reasons unrelated to any change in a person's underlying feelings of well-being. Rather, changes might reflect either learning behavior on the part of respondents or a reduction in social desirability bias.

Our analysis of HILDA Survey data suggests that any panel conditioning effect on the level of self-reported overall life satisfaction is very small, especially once non-random attrition is controlled for. This conclusion is obviously good news for users of HILDA Survey data given it implies that panel conditioning effects can be largely ignored when analyzing life satisfaction data. It, however, is in marked contrast to that reported by Kassenboehmer and Haisken-DeNew (2012) for an almost identically worded life satisfaction measure using the GSOEP data, suggesting that previous findings may be survey or context specific.

Very differently, self-reported scores on the MHI-5, which is more a measure of emotional well-being, demonstrate much larger variations with the number of times interviewed. These variations, however, are only statistically significant for women and, moreover, indicate that self-reported scores improve with time, counter to expectations. That is, women on average tend to report fewer symptoms of anxiety and mood disturbance as survey participation lengthens. Why this might be, and why this finding is specific to women, is not obvious, but it does suggest the possibility that the higher incidence of mental health problems among women found in most studies employing the SF-36 (e.g., Australian Bureau of Statistics 1997; Strand *et al.* 2003; Hoeymans *et al*. 2004; Ware *et al*. 2007) may be a statistical artifact. That is, women may have a tendency to exaggerate symptoms indicative of mental health problems when responding to cross-section surveys (or to longitudinal surveys on the first occasion).

One potential problem with our analysis is that, while we are able to hold constant aging effects, we are not able to simultaneously hold constant any change in overall well-being due to society wide change. Most notably, the data covers a period that includes the 2008 global economic crisis and its aftermath, which might be expected to have increased economic uncertainty and thus feed into reduced levels of SWB. On the other hand, and in contrast to most other Western nations, Australia largely escaped the worst effects of the economic crisis – economic growth slowed but did not go into decline and median house prices remained fairly steady. One simple way to identify whether changes in population-wide levels of SWB may have affected the data used here is to compare differences in the population means reported by the original sample in wave 1 (i.e., 2001) with those reported by the population refreshment sample interviewed in wave 11 (i.e., 2011). As it turns out, the mean levels of both overall life satisfaction and the MHI-5 were slightly higher in the refreshment sample. After applying weights that both adjust estimates for differential selection probabilities and ensure the distribution of selected sample characteristics match known population totals, mean life satisfaction rises from 7.94 to 8.00 while the mean MHI-5 score rises from 73.5 to 74.0. Both of these differences are small, and so suggest our results are unlikely to be affected by some unobserved population-wide change in SWB.[6]

We have also shown that repeated participation in a panel survey is associated with a clear and gradual reduction in the dispersion of scores on both well-being measures, and for both men and women. Such findings are consistent with learning behavior, which might lead respondents to avoid extreme values. Alternatively, it might reflect declining respondent engagement in the study, which, in turn, might be reflected in some respondents choosing what they think as the social norm when answering, rather than reflecting deeply about their own life experiences. This will tend to cause stability in SWB scores over time to be overstated. Note further that this finding of declining dispersion is inconsistent with arguments that panel survey participation is associated with a gradual reduction in social desirability bias, given reductions in such bias should lead respondents to move away from social norms. Perhaps, most significantly, our results suggest that the tendency for panel data observations to exhibit regression to the mean may in part be due to panel conditioning processes.

Overall, our findings suggest that panel conditioning may or may not be a problem when using longitudinal data to examine variations in the level of subjective well-being, but that

---

[6] Despite being small, the difference in weighted means on life satisfaction is statically significant (p = .044). The difference in the weighted mean MHI-5 scores, however, is insignificant (p = .202).

wherever possible, researchers should attempt to identify the extent of the problem and control for it. Further, researchers need to consider not just how panel conditioning affects levels of well-being but how it might impact on the distribution of reported outcomes. The evidence reported here shows very clearly that panel conditioning exacerbates the tendency in panel data for repeated observations to regress towards the mean, which might have significant implications for research focused on understanding both stability in individual well-being scores over time, and how the position of individuals in the well-being distribution changes.

Finally it needs to be acknowledged that the data analyzed here were not created with the specific purpose of identifying and quantifying panel conditioning effects. An experimental design might, for example, involve randomly assigning the initial sample into sub-groups differentiated by the number of times to be interviewed. Interviews would be attempted at all waves with members of one group and only intermittently, or rarely, with members of other groups. This would arguably provide stronger conditions under which to identify panel conditioning effects. It is, however, rarely possible to implement such designs on large population samples. More feasible is to draw comparisons between different population samples drawn at different times, as van Landeghem (2012) does. This is now also possible in the HILDA Survey following the introduction of the large refreshment sample in wave 11. Such comparisons, of course, require comparable populations, which is rendered difficult by changes in population composition (e.g., immigration) and, more importantly, by the attrition that will affect the earlier sample. Future research in this space will thus inevitably have to pay much more attention to modeling the process that determines survey participation.

Acknowledgments The research reported on in this paper was supported by an Australian Research Council Discovery Grant (#DP1095497). The paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Survey Project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The authors also thank Robert Cummins and Nicole Watson for helpful advice and comments on an earlier version of this paper. The findings and views reported in this paper, however, are those of the authors and should not be attributed to any of the aforementioned.

References

Australian Bureau of Statistics (1997). *1995 National Health Survey: SF-36 Population Norms, Australia* (ABS cat. No. 4399.0). Canberra: Australian Bureau of Statistics.

Bailar, B. A. (1975). The effects of the rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, 70(349), 23-30.

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

Corder, L.S., & Horvitz, D. G. (1989). Panel effects in the National Medical Care Utilization and Expenditure Survey. In D. Kasprzyk, G. Duncan, G Kalton & M. P. Singh (eds), *Panel Surveys* (pp. 304-318). New York: John Wiley & Sons.

Das, M., Toepoel, V. & van Soest, A. (2011). Nonparametric tests of panel conditioning and attrition bias in panel surveys. *Sociological Methods & Research*, 40(1), 32-56.

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125(2), 276-302.

Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *Economic Journal*, 114(497), 521-530.

Frick, J. R., Goebel, J., Schechtman, E., Wagner, G. G., & Yitzhaki, S. (2006). Using analysis of gini (ANOGI) for detecting whether two subsamples represent the same universe: the German Socio-Economic Panel Study (SOEP) experience. *Sociological Methods & Research*, 34(4), 427-468.

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.

Frijters, P., Haisken-DeNew, J. P., & Shields, M. A. (2004). Investigating the patterns and determinants of life satisfaction in Germany following reunification. *Journal of Human Resources* 39(3), 649-674.

Headey, B., Muffels, R., & Wagner, 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*, 107(42), 17922-17926

Headey, B., Muffels, R., & Wooden, M. (2008). Money doesn't buy happiness: or does it? a reassessment based on the combined effects of wealth, income and consumption. *Social Indicators Research*, 87(1), 65-82.

Hoeymans, N., Garssen, A. A., Westert, G. P., & Verhaak, P. F. (2004). Measuring mental health of the Dutch population: A comparison of the GHQ-12 and the MHI-5. *Health and Quality of Life Outcomes*, 2(23). [Available from: www.hqlo.com]

Kalton, G., Kasprzyk, D., & McMillen, D. B. (1989). In D. Kasprzyk, G. Duncan, G Kalton & M. P. Singh (eds), *Panel Surveys* (pp. 249-270). New York: John Wiley & Sons.

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.

Klein, D. N., and Rubovits, D. R. (1987). The reliability of subjects' reports on stressful life events inventories: a longitudinal study. *Journal of Behavioral Medicine*, 10(5), 501-512.

Landua, D. (1993). Changes in reports of satisfaction in panel surveys: an analysis of some unintentional effects of the longitudinal design. *Kolner Zeitschrift fur Soziologie und Sozialpsychologie*, 45(3), 553-571.

Lynn, P. (2009). Methods for longitudinal surveys. In P. Lynn (ed.), *Methodology of Longitudinal Surveys* (pp. 1-19). Chichester, UK: Wiley.

Lazarsfeld, P. (1940). Panel studies. *Public Opinion Quarterly*, 4(1), 122-128.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-40). San Diego, CA: Academic Press.

Pennell, S., & Lepkowski, J. (1992). Panel conditioning effects in the Survey of Income and Program Participation. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 566-571

Rumpf, H.-J., Meyer, C., Hapke, U., & John, U. (2001). Screening for mental health: Validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard. *Psychiatry Research*, 105(3), 243-253.

Silberstein, A. R., & Jacobs, C. A. (1989). Symptoms of repeated interview effects in the Consumer Expenditure Interview Survey. In D. Kasprzyk, G. Duncan, G Kalton & M. P. Singh (eds), *Panel Surveys* (pp. 289-303). New York: John Wiley & Sons.

Sobol, M. G. (1959). Panel mortality and panel bias. *Journal of the American Statistical Association*, 54(285), 52-68.

Strand, B. H., Dalgard, O. D., Tambs, K., & Rognerud, M. (2003). Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic Journal of Psychiatry*, 57(2), 113-118.

Sturgis, P., Allum, N., & Brunton-Smith. I. (2009). Attitudes over time: the psychology of panel conditioning. In P. Lynn (ed.), *Methodology of Longitudinal Surveys* (pp. 113-126). Chichester, UK: Wiley.

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys*. Aldine: Chicago.

Summerfield, M., Freidin, S., Hahn, M., Ittak, P., Li, N., Macalalad, N., Watson, N., Wilkins, R., & Wooden, M. (2012). *HILDA User Manual – Release 11*. Melbourne: Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Toepoel, V., Das, M., & van Soest, A. (2009). Relating question type to panel conditioning; comparing trained and fresh respondents. *Survey Methods Research*, 3(2), 73-80.

Torche, F., Warren, J. R., Halpern-Manners, A. & Valenzuela, E. (2012). Panel conditioning in a study of adolescents' substance use: evidence from an experiment. *Social Forces*, 90(3), 891-

Traugott, M. W., & Katosh, J. P. (1979). Response validity in surveys of voter behavior. *Public Opinion Quarterly*, 43(3), 359-377.

Van Landeghem, B. (2012). Panel conditioning and self-reported satisfaction: Evidence from International panel data and repeated cross-sections. SOEPpapers (on Multidisciplinary Panel Data Research) no. 484. Berlin: DIW.

Verbeek, M., & Nijman T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, 33(3), 681-703.

Veroff, J., Hatchett, S., & Douvan E. (1992). Consequences of participating in a longitudinal study of marriage. *Public Opinion Quarterly*, 56(3), 315-327.

Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (2000). *SF-36 Health Survey: Manual and Interpretation Guide*. Lincoln, RI: QualityMetric Inc.

Ware, J. E., Kosinski, M., Bjorner, J. B., Turner-Bowker, D. M., Gandek, B., & Maruish, M. E. (2000). *Users's Manual for the SF-36v2 Health Survey* (2nd ed.). Lincoln, RI: QualityMetric Inc.

Waterton, J., & Lievesley, J. (1989). Evidence of conditioning effects in the British Social Attitudes. In D. Kasprzyk, G. Duncan, G Kalton & M. P. Singh (eds), *Panel Surveys* (pp. 319-339). New York: John Wiley & Sons.

Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (ed.), *Methodology of Longitudinal Surveys* (pp. 157-182). Chichester, UK: Wiley.

Watson, N. & Wooden, M. (2013). Adding a top-up sample to the HILDA Survey. *The Australian Economic Review*, 46(4), forthcoming.

Watson, N., & Wooden, M. (forthcoming). Re-engaging with survey non-respondents: Evidence from three household panels. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Watson, N., & Wooden, M. (2012). The HILDA Survey: a case study in the design and development of a successful household panel study. *Longitudinal and Life Course Studies*, 3(3), 369-381.

Wilson, S. E., & Howell, B. L. (2005). Do panel surveys make people sick? US arthritis trends in the Health and Retirement Study. *Social Science & Medicine*, 60(11), 2623-2627.

Wooden, M., Freidin, S., & Watson, N. (2002). The Household, Income and Labour Dynamics in Australia (HILDA) Survey: wave 1. *The Australian Economic Review*, 35(3), 339-348.

Wooden, M., Warren, D., & Drago, R. (2009). Working time mismatch and subjective well-being. *British Journal of Industrial Relations*, 47(1), 147-179.

Panel Conditioning and Subjective Well-being

Mark Wooden and Ning Li

Melbourne Institute of Applied Economic and Social Research, University of Melbourne, Victoria 3010, Australia

+61 3 8344 2089

+ 61 3 8344 2111

m.wooden@unimelb.edu.au

Table 1 Mean general life satisfaction scores (and standard deviations) by frequency of survey participation

Years in panel	Unbalanced panel			Balanced panel		
	Mean	SD	N	Mean	SD	N
1	8.00	1.64	21280	7.98	1.61	7460
2	7.93	1.55	17537	7.90	1.53	7460
3	7.97	1.51	15470	7.99	1.49	7460
4	7.93	1.50	14022	7.96	1.49	7460

5	7.88	1.47	12871	7.91	1.44	7460
6	7.86	1.46	11801	7.89	1.45	7460
7	7.84	1.46	10856	7.88	1.46	7460
8	7.84	1.42	9956	7.89	1.39	7460
9	7.86	1.45	8958	7.89	1.43	7460
10	7.85	1.43	7460	7.85	1.43	7460

Table 2 Mean MHI-5 scores (and standard deviations) by frequency of survey participation

Years in panel	Unbalanced panel			Balanced panel		
	Mean	SD	N	Mean	SD	N
1	73.57	17.35	19017	75.23	16.80	5368
2	73.94	17.16	15814	75.52	16.63	5368
3	73.96	17.13	13923	75.55	16.59	5368
4	74.12	16.99	12565	75.70	16.31	5368
5	74.25	16.84	11419	75.63	16.23	5368
6	74.51	16.90	10569	75.85	16.48	5368
7	74.47	17.15	9685	75.88	16.39	5368
8	74.81	16.88	8764	75.83	16.39	5368
9	75.12	16.87	8060	76.22	16.31	5368
10	74.91	16.67	6950	75.64	16.31	5368

Table 3 Conditioning variables – definitions and summary statistics

Variable	Definition	Mean	SD (b/w; within)	N
Age	Age (in years) at 30th June in year prior to interview	43.80	18.36 (18.90; 2.57)	130211
Partnered	Equals 1 if married or in de facto relationship	.619	.486 (.462; .205)	130168
# children	Number of own children aged less than 15 years living with respondent	1.65	1.57 (1.57; .32)	130211
# adults	Number of persons aged 15 years or more living in the household	2.29	1.01 (.96; .55)	130211
Mild disability	Equals 1 if has long-term health condition that does not limit work	.080	.271 (.179; .226)	130211
Moderate disability	Equals 1 if has long-term health condition that limits work, but not totally	.167	.373 (.295; .242)	130211
Severe disability	Equals 1 if has long-term health condition that prevents any work being undertaken	.017	.128 (.086; .103)	130211
NLF	Equals 1 if not in the labor force (i.e., not employed and not actively seeking work)	.327	.469 (.413; .242)	130211
Unemployed	Equals 1 if employed and actively seeking work	.036	.186 (.164; .149)	130211
Prefers fewer hours	Equals 1 if employed but prefers to work fewer hours	.175	.380 (.276; .276)	130211
Prefers more hours	Equals 1 if employed but prefers to work more hours	.095	.293 (.227; .234)	130211
Ln income	Log of real equivalized disposable household income ($000) (with missing values imputed and non-positive incomes set to $1)	9.824	1.006 (.739; .741)	130211
Non-positive income	Equals 1 if household income is non-positive	.006	.078 (.047; .067)	130211
Homeowner	Equals 1 if lives in a household where a member owns, or is paying the mortgage on, the place of residence	.707	.455 (.426; .233)	130156
Home equity	Real value of household's net equity in home ($m) (with missing values on home value imputed)	.139	.244 (.203; .134)	130156
Inner regional	Equals 1 if lives in inner regional Australia (as defined in the Australian Standard Geographical Classification [ASGC])	.245	.430 (.407; .150)	130211
Outer regional	Equals 1 if lives in outer regional Australia (as defined in the ASGC)	.117	.321 (.309; .110)	130211
Remote	Equals 1 if lives in remote or very remote location in Australia (as defined in the ASGC)	.021	.143 (.134; .061)	130211
Telephone	Equals 1 if interviewed by telephone	.063	.243 (.185; .191)	130211
Others present	Equals 1 if other adults were present during the interview	.372	.483 (.359; .369)	130211
NR at t+1	Equals 1 if did not respond at next survey wave	.062	.241 (.285; .184)	130211

Variable	Definition	Mean	SD (b/w; within)	N
T	Number of years responding to the survey	4.64	2.82 (1.82; 2.49)	130211

Table 4 Regression results, overall life satisfaction – basic model

	Pooled: All persons		Fixed effects: All persons		Fixed effects: Males		Fixed effects: Females	
	Coeff.	S.E.[a]	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Age	-.178	.008**	-.190	.008**	-.223	.012**	-.163	.011**
Age squared (/100)	.328	.017**	.367	.018**	.433	.026**	.314	.025**
Age cubed (/10000)	-.167	.011**	-.230	.012**	-.270	.018**	-.200	.016**
Partnered	.395	.019**	.404	.015**	.421	.022**	.388	.021**
# children	.010	.007	-.030	.010**	-.028	.013*	-.039	.015**
# adults	.006	.008	-.010	.006	-.006	.008	-.014	.008
Mild disability	-.207	.019**	-.063	.013**	-.060	.019**	-.068	.019**
Moderate disability	-.719	.022**	-.288	.013**	-.298	.019**	-.279	.018**
Severe disability	-1.346	.066**	-.552	.030**	-.544	.042**	-.555	.043**
NLF	-.063	.019**	-.084	.013**	-.158	.021**	-.040	.017*
Unemployed	-.425	.034**	-.268	.021**	-.305	.029**	-.237	.030**
Prefer fewer hours	-.327	.015**	-.162	.011**	-.164	.015**	-.159	.016**
Prefer more hours	-.230	.019**	-.089	.013**	-.088	.019**	-.091	.018**
Ln income	.120	.012**	.051	.008**	.044	.011**	.057	.011**
Non-positive income	1.074	.124**	.402	.086**	.322	.123**	.478	.119**
Homeowner	.141	.018**	.078	.013**	.090	.019**	.067	.019**
Home equity	.179	.025**	.073	.023**	.064	.034	.080	.030**
Inner regional	.161	.019**	.078	.021**	.053	.030	.100	.030**
Outer regional	.250	.025**	.022	.029	.026	.041	.018	.042
Remote	.334	.050**	.059	.050	.047	.071	.067	.071
Telephone	-.049	.022*	.008	.016	-.008	.022	.024	.022
Others present	.079	.011**	.041	.008**	.038	.011**	.045	.011**
NR at t+1	-.091	.019**	-.081	.016**	-.071	.022**	-.089	.023**
Constant term	9.092	.153**	10.328	.137**	10.797	.195**	9.941	.194**
R-sq. within			.024		.027		.022	
R-sq. between			.028		.049		.015	
Rho			.604		.612		.596	
R-squared	.109		.595		.612		.580	
Adj R-squared			.515		.533		.500	
F statistic	184.61**		116.2**		62.31**		56.13**	
F-test that all $\mu_i = 0$			6.13**		6.36**		5.90**	
N (observations)	130022		130022		61536		68486	
N (individuals)	21263		21263		10316		10947	

a Standard errors adjusted for within-person clustering of observations.
* *p* < .05; ** *p* < .01.

Table 5 Regression results, mental health (MHI-5) – basic model

	Pooled: All persons		Fixed effects: All persons		Fixed effects: Males		Fixed effects: Females	
	Coeff.	S.E.[a]	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Age	-1.080	.098**	-.878	.094**	-1.314	.134**	-.508	.133**
Age squared (/100)	2.022	.213**	2.152	.203**	2.890	.292**	1.542	.284**
Age cubed (/10000)	-1.138	.142**	-1.476	.136**	-1.873	.198**	-1.154	.189**
Partnered	2.452	.229**	2.262	.170**	2.424	.245**	2.126	.236**
# children	-.072	.086	-.139	.112	-.068	.146	-.288	.172
# adults	-.362	.093**	-.236	.065**	-.121	.090	-.345	.092**
Mild disability	-4.479	.235**	-1.357	.148**	-1.013	.203**	-1.702	.215**
Moderate disability	-11.219	.275**	-3.815	.147**	-3.707	.212**	-3.914	.204**
Severe disability	-18.898	.751**	-7.388	.355**	-7.700	.486**	-7.037	.514**
NLF	-4.070	.235**	-1.248	.148**	-1.690	.232**	-.995	.194**
Unemployed	-5.769	.373**	-2.053	.235**	-2.401	.324**	-1.769	.339**
Prefer fewer hours	-2.421	.181**	-1.163	.121**	-1.187	.161**	-1.141	.180**
Prefer more hours	-2.825	.225**	-.541	.147**	-.490	.208*	-.606	.207**
Ln income	1.839	.134**	.406	.089**	.468	.125**	.357	.125**
Non-positive income	15.872	1.494**	.193	.973**	3.240	1.377*	2.852	1.373*
Homeowner	1.506	.219**	.193	.150	.137	.209	.245	.213
Home equity	1.832	.418**	.680	.257**	.754	.388*	.631	.345
Inner regional	1.311	.227**	.467	.237*	-.097	.334	.961	.336**
Outer regional	1.397	.309**	.142	.334	-.136	.463	.403	.480
Remote	2.682	.629**	.133	.583	-.734	.825	.871	.821
Telephone	-.080	.306	-.095	.200	.159	.279	-.339	.284
Others present	.261	.137	.415	.089**	.524	.123**	.328	.127*
NR at t+1	-1.608	.237**	-.598	.193**	-.488	.270	-.671	.276*
Constant term	70.634	1.847**	79.793	1.559**	87.664	2.207**	72.997	2.205**
R-sq. within			.015		.018		.014	
R-sq. between			.115		.133		.095	
Rho			.628		.635		.619	
R-squared	.120		.658		.668		.648	
Adj. R-squared			.587		.596		0.576	
F statistic	167.81**		64.58**		35.29**		32.29**	
F-test that all $\mu_i = 0$			7.52**		7.57**		7.40**	
N (observations)	116716		116716		54618		62098	
N (individuals)	20215		20215		9717		10498	

a Standard errors adjusted for within-person clustering of observations.
* $p < .05$; ** $p < .01$.

Table 6 The impact of time in panel (T) on level of subjective well-being: estimates from FE models (standard errors in parentheses)

	Variable	Men	Women	Persons
Life satisfaction (LS)				
Specification 1	T	-.005 (.014)	.018 (.015)	.006 (.010)
Specification 2	T	-.037 (.017)*	.017 (.018)	-.009 (.012)
	T-squared	.002 (.0006)**	.000 (.001)	.001 (.0005)*
	F-test	6.52, p = .002	0.69, p = .503	3.06, p =.047
Specification 3	T=2	-.063 (.023)**	-.054 (.023)*	-.059 (.016)**
	T=3	-.034 (.035)	.041 (.036)	.004 (.025)
	T=4	-.071 (.048)	.057 (.051)	-.006 (.035)
	T=5	-.110 (.062)	.049 (.066)	-.029 (.045)
	T=6	-.118 (.076)	.060 (.080)	-.028 (.055)
	T=7	-.119 (.089)	.064 (.095)	-.028 (.066)
	T=8	-.123 (.104)	.101 (.106)	-.011 (.076)
	T=9	-.102 (.118)	.139 (.120)	.018 (.086)
	T=10	-.115 (.131)	.128 (.134)	.005 (.096)
	F-test	2.98, p = .002	4.19, p = .000	5.83, p =.000
Mental health (MHI-5)				
Specification 1	T	-.266 (.176)	.608 (.189)**	.175 (.129)
Specification 2	T	-.377 (.201)	.574 (.215)**	.101 (.148)
	T-squared	.008 (.007)	.002 (.007)	.005 (.005)
	F-test	1.79, p = .168	5.21, p = .005	1.44, p =.236
Specification 3	T=2	-.236 (.256)	.322 (.270)	.035 (.186)
	T=3	-.748 (.413)	1.023 (.440)*	.145 (.303)
	T=4	-.965 (.581)	1.594 (.623)*	.318 (.428)
	T=5	-1.341 (.751)	2.244 (.807)**	.463 (.553)
	T=6	-1.504 (.922)	2.830 (.993)**	.672 (.679)
	T=7	-1.982 (1.095)	3.378 (1.180)**	.720 (.807)
	T=8	-2.094 (1.266)	3.980 (1.365)**	.956 (933)
	T=9	-2.114 (1.436)	4.995 (1.548)**	1.466 (1.059)
	T=10	-2.752 (1.604)	4.932 (1.730)**	1.109 (1.182)
	F-test	0.93, p = .500	2.22, p = .018	1.54, p =.127

* $p < .05$; ** $p < .01$.

Table 7 The impact of time in panel (T) on dispersion in subjective-wellbeing: estimates from FE models (standard errors in parentheses)

	Variable	Men	Women	Persons
Life satisfaction (LS)				
Specification 1	T	.006 (.008)	-.018 (.009)*	-.006 (.229)
Specification 2	T	-.067 (.009)**	-.100 (.010)**	-.084 (.007)**
	T squared	.005 (.000)**	.006 (.000)**	.005 (.000)**
	F-test	105.7, *p* = .000	129.3, *p* = .000	233.3, *p* =.000
Specification 3	T=2	-.080 (.012)**	-.113 (.013)**	-.097 (.009)**
	T=3	-.120 (.019)**	-.194 (.021)**	-.157 (.014)**
	T=4	-.141 (.027)**	-.239 (.029)**	-.190 (.020)**
	T=5	-.164 (.035)**	-.284 (.037)**	-.224 (.025)**
	T=6	-.167 (.042)**	-.309 (.046)**	-.237 (.031)**
	T=7	-.155 (.050)**	-.329 (.054)**	-.242 (.037)**
	T=8	-.156 (.058)**	-.363 (.062)**	-.259 (.043)**
	T=9	-.148 (.066)*	-.390 (.071)**	-.269 (.048)**
	T=10	-.121 (.074)	-.363 (.079)**	-.240 (.054)**
	F-test	25.68, *p* = .000	2.65, *p* = .000	57.65, *p* =.000
Mental health (MHI-5)				
Specification 1	T	-.104 (.098)	-.080 (.105)	-.093 (.072)
Specification 2	T	-.679 (.112)**	-.757 (.119)**	-.721 (.082)**
	T squared	.041 (.004)**	.047 (.004)**	.044 (.003)**
	F-test	55.87, *p* = .000	73.22, *p* = .000	128.8, *p* = .000
Specification 3	T=2	-.826 (.142)**	-.940 (.149)**	-.885 (.103)**
	T=3	-1.241 (.230)**	-1.529 (.244)**	-1.394 (.168)**
	T=4	-1.630 (.323)**	-1.753 (.345)**	-1.695 (.237)**
	T=5	-1.934 (.418)**	-2.102 (.447)**	-2.024 (.306)**
	T=6	-2.024 (.513)**	-2.271 (.550)**	-2.157 (.377)**
	T=7	-2.129 (.609)**	-2.453 (.653)**	-2.305 (.447)**
	T=8	-2.471 (.704)**	-2.581 (.756)**	-2.533 (.517)**
	T=9	-2.312 (.798)**	-2.574 (.858)**	-2.455 (.587)**
	T=10	-2.391 (.892)**	-2.586 (.958)**	-2.499 (.656)**
	F-test	14.25, *p* = .000	18.61, *p* = .000	32.36, *p* =.000

* $p < .05$; ** $p < .01$.