



THE UNIVERSITY OF
MELBOURNE

HILDA PROJECT DISCUSSION PAPER SERIES
No. 2/09, February 2009

**Use of the Kessler Psychological Distress Scale in the
HILDA Survey**

Mark Wooden

**The HILDA Project was initiated, and is funded, by the Australian
Government Department of Family and Community Services**

Acknowledgements

The author thanks Markus Hahn and Simon Freidin for assistance with preparation of data used in this paper.

Introduction

Like other major household panel surveys, the content of the HILDA Survey is dominated by economic and demographic data. Nevertheless, the HILDA Survey also includes a range of subjective measures of health. In particular, the Short Form 36 (or SF-36) health questionnaire has been administered in every wave as part of the self-completion questionnaire (SCQ). Described in more detail in Ware et al. (2000), the SF-36 comprises 36 items which can then be combined to produce eight sub-scales measuring different health concepts. The properties of the SF-36 data collected during wave 1 of the HILDA Survey have been analysed by Butterworth and Crosier (2004). They concluded that the “eight scales were ... psychometrically sound, with good internal consistency, discriminant validity and high reliability”.

Inclusion of the SF-36 in the HILDA Survey was justified in part by its widespread use internationally, and in part by the inclusion of the instrument, as a self-administered supplement, in the 1995 National Health Survey (NHS) conducted by the Australian Bureau of Statistics (ABS). In subsequent rounds of the NHS, however, the SF-36 was not included, possibly because of the burden, both on respondents and on interviewers, associated with administering a separate questionnaire comprising 36 different items. In its place, the 2001 NHS included the 10-item measure of non-specific psychological distress created by Ron Kessler and Dan Mroczek from the Harvard Medical School for use in the redesigned United States National Health Interview Survey (see Kessler et al. 2002, 2003). The measure was retained in the 2004-05 round of the NHS. It has also been included in the 1997 and 2007 National Surveys of Mental Health and Wellbeing (NSMHWB), also conducted by the ABS, but this time alongside much more extensive measures of mental disorders and symptoms collected using the World Health Organizations’s Composite International Diagnostic Interview (CIDI).

The NHS, of course, is a large survey dedicated to the measurement of individual health and health behaviours. As a result, its need for subjective measures of physical health is not great (unlike the HILDA Survey which does not have the capacity to collect extensive data on health every year). However, the assessment of the ABS was that the measurement of mental health within the NHS was difficult – measures based on medical diagnosis were certainly not possible. The ABS thus sought a short subjective measure of symptoms of psychological distress, with the 10-item version of the measure developed by Kessler and colleagues (or K10) deemed the most appropriate.

Many stakeholders have thus argued that the value of the HILDA Survey data, especially to public health researchers and policy-makers, would be greatly enhanced by the inclusion of the K10. On the other hand, the items that comprise the K10 appear to overlap closely with the items that comprise the 5-item mental health sub-scale within the SF-36, suggesting that there may be some redundancy from its inclusion alongside the SF-36. Despite this, the K10 was added to the SCQ in wave 7, with the expectation that it would be repeated at regular intervals, perhaps every two years. It is scheduled to be included in the wave 9 instrument.

This short paper briefly describes the K10 and the measures derived from it and included in the HILDA Survey data release 7.0, makes some brief comparisons with the mental health sub-scale of the SF36, and reports how population estimates from the HILDA Survey compare with those from ABS sources (namely, the NHS and the NSMHWB).

The Construction, Administration and Scoring of the K10

Item Development

As summarised in ABS (2003), the K10 was originally designed for inclusion in the US National Health Interview Survey (NHIS) and “consists of 10 questions about non-specific psychological distress and seeks to measure the level of current and depressive symptoms a person may have experienced in the four weeks prior to interview”. Its key strength, however, is that it was designed to have optimal sensitivity at the upper end of the population distribution with respect to psychological distress, and thus has been shown to be extremely effective at screening for serious mental disorders (Kessler et al. 2003, Furukawara et al. 2003). It has subsequently been included in the US National Comorbidity Survey Replication as well as in all the national surveys that comprise the World Health Organization’s World Mental Health Initiative, represented in Australia by the National Survey of Mental Health and Wellbeing,

The process by which the 10 items were selected is described at length in Kessler et al. (2002), but in brief, the K10 items were developed from an initial pool of 612 questions that was gradually pared back to a list of 45 items, which were then included in a series of pilot surveys undertaken by mail and by telephone. Results from these surveys were then analysed using item response theory, with the scales validated in a subsequent two-stage survey, with the second stage involving in-person interviews.

The final list of items is shown in Table 1.

Table 1: The K10, as Administered in the HILDA Survey Wave 7

		The following questions are about your feelings in the past 4 weeks. In the <u>last four weeks</u> , about how often did you feel ... (Cross one box on each line.)				
		<i>All of the time</i>	<i>Most of the time</i>	<i>Some of the time</i>	<i>A little of the time</i>	<i>None of the time</i>
a	tired out for no good reasons?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b	nervous?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c	so nervous that nothing could calm you down?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d	hopeless?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e	restless or fidgety?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f	so restless that you could not sit still?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g	depressed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h	that everything was an effort?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i	so sad that nothing could cheer you up?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
j	worthless?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Administration in the HILDA Survey

The K10 was originally designed to be administered by personal interview, which is the approach adopted by the ABS in both of its large health surveys. For the HILDA Survey this was not an option, with no room for any new content in the personal interview content. Thus for the HILDA Survey the only available option was inclusion in the self-administered component (SCQ), and so we employed the self-administered version as recommended on the National Comorbidity Survey website at Harvard University.¹ In our view, however use of the self-administered version should be seen as an advantage. Questions on psychological health are almost certainly subject to social desirability bias, with many respondents likely to be reluctant to admit to being affected by distress symptoms (such as depression and anxiety). We would expect such reluctance to be reduced when the questions are completed privately and in complete anonymity, as is more likely the case with a self-administered questionnaire, than when administered by an interviewer.

Self-administration, however, did necessitate one significant departure from the original design. In the original design, two of the items (c and f in Table 1) were not intended to be asked if the answer to the preceding question is “None of the time”. This, however, is too complicated for implementation in a self-administered instrument and hence no such skipping instruction is included.

The other main difference from the original or recommended version is the time frame. In the original version, and used today in the US, the referent time frame is the “last 30 days”. We, however, follow the ABS and employ a reference period of the “last four weeks”.

Scoring

While the analysis of Kessler et al. (2002) suggests that the different items within the K10 should be weighted differently to reflect the fact that the discriminating ability of each item within that part of the distribution where it has most information value differs, most researchers have ignored this and treat the data as if all items are equally sensitive. The reason for this is pragmatic – it means that an overall score can be derived by simply summing scores on the 10 items.²

Following Andrew and Slade (2001) and the ABS (2003) we thus derived a unique K10 score by scoring responses on each of the ten items using a simple linear scale running from 5 to 1, and summing across all items. The overall score thus ranges from 10 to 50.³

As noted above, an important difference between the self-administered and interviewer-administered version is that in the latter, two of the items (c and f in Table 1) are not asked if the answer to the preceding question is “None of the time”. To ensure both internal consistency in the data and to maintain comparability with other data sources administering the K10, scores to items c and f have all been set to the value 1 if the answer to the preceding question was “None of the time”.

Like all questions in the SCQ, there is some item non-response. For each item in the K10 this averages around 1.5%, which compares more than favourably with other SCQ items. Nevertheless, when constructing a score across all 10 items we lose 3.3% of cases due to non-response on at least one item. To minimise the effect of missing data we have thus applied a

¹ Go to: http://www.hcp.med.harvard.edu/ncs/k6_scales.php.

² The NCS website refers to more complicated calibration rules being made available at some point in the future, but at the time of writing these had still to be posted.

³ This differs from the scoring used in the US, where items are scored on a 0 to 4 scale and so range from 0 to 40.

rule that computes a K10 score for any observation that provides valid responses to at least five of the eight items – a, b, d, e, g, h, i and j. In these cases, the mean value of the valid responses is taken and then multiplied by 10. This rule (together with the rule concerning the treatment of responses to items c and f) had the effect of reducing the incidence of missing data on the K10 score from 3.3% to 1.4%.

Finally, in a relatively small proportion of cases multiple answers were given to the same item. In these cases we applied the same data cleaning rule that is applied throughout the SCQ (and which in turn was based on the rules applied in the SF-36). That is, if two responses were checked and these responses were adjacent, one response was randomly selected. In all other situations the response was treated as missing.

Distress Categories

Following the ABS (2003), we also provide in the data set a derived categorical variable that uses the K10 scores to divide the population into four groups based on their level of psychological distress. These categories are:

Level of psychological distress	K10 score
Low	10-15
Moderate	16-21
High	22-29
Very high	30-50

Comparisons with the MH5

As noted above, the HILDA Survey regularly includes another self-reported measure of mental health – the mental health sub-scale (MH5) within the SF-36. The items that make up the MH5, like the K10, relate to the frequency of feelings experienced in the past 4 weeks, and are as follows:

- (i) Have you been a nervous person?
- (ii) Have you felt so down in the dumps nothing could cheer you up?
- (iii) Have you felt calm and peaceful?
- (iv) Have you felt down?
- (v) Have you been a happy person?

With the exception of item v, these items are not dissimilar to the sorts of items included in the K10. Indeed at least three of the items in the MH5 (i, ii and iv) have close analogs within the K10 (b, i, and g). We thus would expect, and find, a very strong correlation (Pearson $r = -.80$) between the K10 score and the score on the MH5 (bearing in mind that poor mental health is indicated by a high score on the K10 but a low score on the MH5). In part, this large correlation is driven by the many respondents reporting few symptoms of psychological distress. Once we focus on that part of the distribution for which the K10 was really designed – those experiencing serious mental illness – the magnitude of the correlation declines. Thus for the 10.8% of the sample with a K10 score in excess of 23 the correlation is only $-.49$, and for the 5.2% of the sample with a K10 score in excess of 28 the correlation is just $-.39$. While still sizeable, these much smaller correlations suggest that the K10 and MH5 are tapping into slightly different concepts, at least among those reporting the most frequent symptoms of psychological distress.

The two measures also have quite different distributions. As reported in Table 2, the distribution of the K10 scores is more skewed and exhibits more kurtosis (the distribution is more peaked) than the MH5.

Table 2: Descriptive Statistics, K10 and MH5 (HILDA Survey Wave 7)

<i>Measure</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>Skewness</i>	<i>Kurtosis</i>
K10	10	50	15.66 (0.06)	1.82 (0.02)	3.91 (0.05)
MH5	0	100	74.46 (0.16)	-1.02 (0.02)	1.01 (0.05)

Note: Standard errors in brackets.

Comparisons with ABS Surveys

Data users may also be interested in knowing the extent to which scores on the K10 from the HILDA Survey compare with those obtained in other large national population surveys in Australia. In Table 3, therefore, we report the distribution of scores from wave 7 of the HILDA Survey (conducted mostly from August 2007 to December 2007) and from the 1997 and 2007 NSMHWB and the 2001 and 2004-05 NHS. For comparison purposes we restrict the population to persons aged 18 years or older. The only exception to this is the 2007 NSMHWB, where the reference population was persons aged 16 to 85 years. Analysis of the HILDA data suggests that this will group will be associated with a slightly higher mean K10 score than the population of persons aged 18 years or older. The difference, however, is relatively small.

As should be immediately obvious, the HILDA Survey suggests a greater risk of depression and anxiety disorder in the Australian population than is suggested by the ABS. According to the HILDA data, over 15% per cent of the Australian population is at considerable risk (high or very high) of psychological distress, with 4.4% at very high risk. The comparable figures from the most recent NHS, conducted in 2004-05, are 13% and 3.8%, respectively. Much further behind is the 2007 NSMHWB, where the comparable proportions are just 9.5% and 2.6%.

In trying to explain differences between the NHS and NSMWB estimates, the ABS (2003) speculated that one reason may have been changes in the incidence of reported symptoms over time. The results in Table 2 suggest this explanation can account for, at most, a very small fraction of the difference. While it is true that both the NHS and the NSMHWB indicate some increase over time in K10 scores, this modest increase cannot explain the marked differences between these two sources. The ABS also speculated that differences in question placement within the different instruments might be to explain. While plausible, we think the most likely explanation is one the ABS did not canvass – the NSMHWB is largely a voluntary survey and thus associated with much lower response rates than the NHS (around 60% in the NSMWHB compared with rates in the order of 90% for the NHS). It seems to us highly likely that non-response will be correlated with mental illness and psychological distress.

**Table 3: Correspondence between K10 Scores – HILDA W7 and ABS Surveys
(persons aged 18 years or over, except where stated otherwise)**

<i>Survey / Level of psychological distress</i>	<i>Males</i>	<i>Females</i>	<i>Persons</i>
1997 NSMHWB			
Low	76.2	71.8	73.8
Moderate	16.6	19.2	18.1
High	5.2	6.6	6.0
Very high	2.0	2.4	2.2
2007 NSMHWB (aged 16-85)			
Low	75.0	67.3	71.1
Moderate	17.8	21.2	19.5
High	5.2	8.5	6.9
Very high	2.0	3.1	2.6
2001 NHS			
Low	68.9	59.9	64.4
Moderate	21.2	24.7	23.0
High	7.1	10.9	9.0
Very high	2.7	4.4	3.6
2004-05 NHS			
Low	66.7	59.2	62.9
Moderate	22.5	25.7	24.1
High	7.5	10.8	9.2
Very high	3.3	4.3	3.8
2007 HILDA			
Low	66.0	61.9	63.9
Moderate	19.6	21.7	20.7
High	10.9	11.0	11.0
Very high	3.5	5.3	4.4

Notes: For purposes of comparability, both persons under the age of 18 years and persons residing in non-private dwellings have been excluded from the HILDA Survey estimates. All estimates are population weighted.

NHS denotes National Health Survey.

NSMHWB denotes National Survey of Mental Health and Wellbeing.

Sources: ABS, *Information Paper: Use of the Kessler Psychological Distress Scale in ABS Health Surveys, Australia, 2001* (ABS cat. no. 44817.0.55.001). ABS: Canberra.

ABS, *2007 National Survey of Mental Health and Wellbeing: Summary of Results* (ABS cat. no. 4326.0), Table 14. ABS: Canberra.

ABS, *2004-05 National Health Survey: Summary of Results* (ABS cat. no. 4364.0), Table 9. ABS: Canberra.

This, however, cannot help explain why, of the three sources, the HILDA Survey provides the largest K10 estimates; after all, the HILDA Survey is also a voluntary survey. Instead we suspect the primary explanation lies in the mode of data collection. As already discussed, responses to questions about symptoms of psychological distress are likely to be subject to social desirability bias. It has long been recognised that many people see depression and other mental health problems as a sign of personal weakness and so will be reluctant to admit to such problems. Such tendencies can be expected to be more pronounced when a person is required to admit to those problems in the presence of some third party, especially if that party is not a trained medical or psychology specialist. This, of course, is a key characteristic of both of the ABS Surveys – all questions are administered in person by an interviewer, and while these people may be extremely well trained in how to deliver the interview questions, they nevertheless are not specialists who can (or should) assist respondents to deal with any emotional or psychological problems they may have. In contrast, as we have emphasised repeatedly, in the HILDA Survey the instrument which contains the K10 questions is a self-administered questionnaire. It is handed to the respondent who is then able to complete it on their own and in their own time, seal it in an envelope that is provided and then either return it to the interviewer (typically at a subsequent visit) or put in the post.

Relatedly, panel conditioning effects might also mean that respondents to the HILDA Survey are more likely to tell the truth and admit to suffering from symptoms of psychological ill health. Unlike respondents to the ABS population surveys, HILDA Survey sample members are part of a panel, the majority of whom have been participating for a number of years (up to seven) at the time the Kessler 10 data were collected. As a result they may have become more comfortable with, and less suspicious of, the survey and its administrators.

In short, while the HILDA Survey will be subject to the same sorts of non-response bias that affect the NSMHWB, responses will be less subject to contamination from external influences and from respondent fears about the purposes of the survey. We thus both expect higher K10 scores in the HILDA Survey and expect these higher scores to be subject to less bias.

Further information on how the reported K10 scores differ from those reported in the 2004-05 NHS is provided in Table 4, which disaggregates the data by age group. This table reveals that while it is true that for all age groups the likelihood of reporting *high* levels of distress is greater in the HILDA Survey, the pattern is not so even when we focus on the proportion of people reporting *very high* levels. Specifically, it is only among the younger members of the population (those aged under 35 years) that the differences between the HILDA Survey estimates and the ABS estimates are pronounced.⁴ Indeed, for most of the older age groups the rates of very high psychological distress are less in the HILDA Survey (though not significantly so). We cannot be sure why the HILDA Survey and ABS data would differ in this way, but speculate that it is among the youngest members of the population where the tendency to under-report signs of mental illness is likely to be greatest. Thus if the HILDA Survey does better at reducing social desirability biases it could be expected that it would have a greater impact on the scores reported by young people.

⁴ The differences in the incidence of reporting very high levels of psychological among persons aged 18 to 24 and 25 to 34 are statistically significant (based on simple t-tests which assume no sample variance around the ABS estimates).

Table 4: Correspondence between K10 Scores by Age Group: HILDA W7 and 2004-05 NHS (persons aged 18 years or over)

Level of psychological distress	Age group							Total
	18-24	25-34	35-44	45-54	55-64	65-74	75+	
2004-05 NHS								
Low	55.0	59.8	61.0	63.7	69.8	70.5	66.5	62.9
Moderate	29.5	28.3	25.0	22.6	18.0	18.3	22.7	24.1
High	12.1	9.0	9.8	8.9	7.8	8.1	7.4	9.2
Very high	3.4	2.9	4.3	4.8	4.4	3.1	3.4	3.8
2007 HILDA								
Low	52.7	59.2	63.8	63.7	69.5	70.7	70.1	63.7
Moderate	26.5	24.0	21.1	20.2	16.7	17.9	18.5	20.9
High	14.4	11.6	11.2	11.0	9.4	9.1	8.9	11.0
Very high	6.4	5.3	3.8	5.2	4.3	2.3	2.4*	4.5

* Based on a sample size that may be too small to generate a reliable population estimate.

Concluding Remarks

The Kessler 10 inventory is almost certainly working as well as could be expected. Indeed, the mode of administration used in the HILDA Survey means that it is likely that it is generating more reliable estimates than that produced by interviewer-administered surveys. What is less clear is whether, given the presence of the SF36, the administration of the K10 is enhancing the ability of researchers to identify variations across HILDA Survey respondents in mental health.

References

- Andrews, G. and Slade, A. (2001), 'Interpreting Scores on the Kessler Psychological Distress Scale (K10)', *Australian and New Zealand Journal of Public Health* 25, 494-497.
- Australian Bureau of Statistics [ABS] (2003), *Information Paper: Use of the Kessler Psychological Distress Scale in ABS Health Surveys, Australia, 2001* (ABS cat. no. 44817.0.55.001). ABS: Canberra.
- Butterworth, P. and Crosier, T. (2004), 'The Validity of the SF-36 in an Australian National Household Survey: Demonstrating the Applicability of the Household Income and Labour Dynamics in Australia (HILDA) Survey to Examination of Health Inequalities', *BMC Public Health* 4 [available at: www.biomedcentral.com/1471-2458/4/44].
- Furukawa, T.A., Kessler, R.C., Slade, T. and Andrews, G. (2003), 'The Performance of the K6 and K10 Screening Scales for Psychological Distress in the Australian National Survey of Mental Health and Well-Being', *Psychological Medicine* 33, 357-362.
- Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S.-L.T., Walters, E.E. and Zaslavsky, A.M. (2002), 'Short Screening Scales to Monitor Population Prevalences and Trends in Non-specific Psychological Distress', *Psychological Medicine* 32, 959-976.

Kessler, R.C., Barker, P.R., Colpe, L.J., Epstein, J.F., Gfroerer, J.C., Hiripi, E., Howes, M.J, Normand, S-L.T., Manderscheid, R.W., Walters, E.E. and Zaslavsky, A.M. (2003), 'Screening for Serious Mental Illness in the General Population', *Archives of General Psychiatry* 60, 184-189

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