R-Outcomes
MEASURE WHAT MATTERS

Research Paper 17/03

Personal Wellbeing Score (PWS) – a short version of ONS4, testing and validation in social prescribing

11 October 2017

Tim Benson, Joe Sladen, Andrew Liles, Henry Potts

Personal Wellbeing Score (PWS) – a short version of ONS4, testing and validation in social prescribing

Tim Benson[1, 2], Joe Sladen[1, 3], Andrew Liles[1, 4], Henry WW Potts[2]

[1] R-Outcomes Ltd, [2] UCL Institute of Health Informatics, [3] Wessex AHSN, [5] Royal Holloway College, London

For further information please contact:
Tim Benson
R-Outcomes Ltd
tim.benson@r-outcomes.com
Tel: +44 7855 682037

Table of Contents

Abstract

Background

Improving wellbeing is a key policy objective, especially in health and care services. Our aim was to develop and test a short generic measure of wellbeing for use alongside other R-Outcomes outcome measures.

Methods

The Personal Wellbeing Score (PWS) is based on four Office of National Statistics (ONS) personal well-being questions (ONS4) and the four thresholds used in its reporting. The PWS was developed in coproduction with the "Happy, Healthy, at Home" NHS Vanguard project in an iterative way to match the look and feel of other R-Outcomes measures, which are short and generic with four question items and four response options.

PWS was used in the evaluation of five social prescribing projects, with 1,325 respondents having an average age of about 85 years. The data from these studies were pooled for secondary analysis for validation and psychometric testing including examination of distributions, internal consistency and factor analysis. We also analysed the association of PWS with health status (using howRu), health confidence (HCS) and patient experience (howRwe).

Results

Differences from ONS4 include fewer words, four response options rather than eleven, all items positively worded and a summary score. PWS is shorter with a lower reading age than other wellbeing measures widely used in the UK.

The item completion rate was above 98%. Cronbach's alpha was 0.90, which, with factor analysis, indicates that it is appropriate to use a summary score as well as item scores. We found strong correlation between the summary PWS score and health confidence (r=0.60) and health status (r=0.58), moderate correlation with patient experience (r=0.30) and age group (r=0.24), and low negative correlation with the number of medications taken (r=-0.05). There was significant improvement between on referral and after referral, but little difference between men and women.

Conclusions

The Personal Wellbeing Score (PWS) is a new short variant of ONS4, which is suitable for use with older vulnerable people. It is short and quick to use, with good psychometric properties and construct validity.

Background

Wellbeing is probably the most appropriate measure of social progress and goal of public policy.[1] This is especially so for health and social services. Social wellbeing includes a wide range of factors known to support wellbeing, such as caring, freedom, generosity, honesty, health, income and good governance.[2] Personal wellbeing, also called subjective wellbeing, is a narrower concept, which refers to how people experience and evaluate their lives and specific domains and activities in their lives.[3]

Personal wellbeing has several facets: (1) evaluative wellbeing is a reflective assessment on a person's life; (2) eudaimonic wellbeing is a sense of meaning and purpose in life, or good psychological functioning; and (3) experienced wellbeing, or affect, refers to a person's experiences that make life pleasant or unpleasant at that time. Experienced wellbeing is often divided into positive experiences, such as happiness, joy or contentment, and negative experiences, such as anxiety, worry, pain or anger. A key point about personal wellbeing is that only the person involved can provide the information being sought. People's own views are the subject of interest.

In 2009, the influential Stiglitz-Sen-Fitoussi *Commission on the Measurement of Economic Performance and Social Progress* recommended that all national statistical agencies should collect and publish measures of subjective well-being.[4] In 2011, the UK Office of National Statistics (ONS) introduced a set of four subjective wellbeing questions (ONS4) in the Annual Population Survey (APS), based on recommendations from an expert group.[5,6,7] These questions or variants have also been used in other national surveys and are designated National Statistics.[8] ONS uses the term personal wellbeing, because it is thought to be clearer and simpler to understand than subjective wellbeing.

The OECD has also developed a subjective wellbeing measure (OECD core questions), which is similar to ONS4, but with an extra question about depression.[9] For comparison, the ONS4 and the OECD core questions are shown in Box 1.

Two recent reports, commissioned by the Department of Health, describe the conceptual basis[10] and empirical results[11] of six wellbeing measures that have been used in the UK. These include ONS4, the General Health Questionnaire (GHQ-12)[12], Short Warwick-Edinburgh Mental Wellbeing Scale (S-WEMWBS)[13], EuroQol EQ-5D[14], ICECAP-A[15] and Adult Social Care Outcomes Toolkit (ASCOT) [16]. GHQ-12 and S-WEMWBS are mainly used in mental health, EQ-5D is health status measure and ICECAP-A and ASCOT are used mainly in social care.

Box 1

ONS4

Next I would like to ask you four questions about your feelings on aspects of your life. There are no right or wrong answers. For each of these questions I'd like you to give an answer on a scale of nought to 10, where nought is 'not at all' and 10 is 'completely'.

Overall, how satisfied are you with your life nowadays? (where nought is 'not at all satisfied' and 10 is 'completely satisfied').

Overall, to what extent do you feel the things you do in your life are worthwhile? (where nought is 'not at all worthwhile' and 10 is 'completely worthwhile').

Overall, how happy did you feel yesterday? (where nought is 'not at all happy' and 10 is 'completely happy').

On a scale where nought is 'not at all anxious' and 10 is 'completely anxious', overall, how anxious did you feel yesterday?

OECD Core questions

The following question asks how satisfied you feel, on a scale from 0 to 10. Zero means you feel "not at all satisfied" and 10 means you feel "completely satisfied".

1. Overall, how satisfied are you with life as a whole these days? [0-10]

The following question asks how worthwhile you feel the things you do in your life are, on a scale from 0 to 10. Zero means you feel the things you do in your life are "not at all worthwhile", and 10 means "completely worthwhile".

2. Overall, to what extent do you feel the things you do in your life are worthwhile? [0-10]

The following questions ask about how you felt yesterday on a scale from 0 to 10. Zero means you did not experience the feeling "at all" yesterday while 10 means you experienced the feeling "all of the time" yesterday. I will now read out a list of ways you might have felt yesterday.

3. How about happy? [0-10]

4. How about worried? [0-10]

5. How about depressed? [0-10]

Methods

Development

Working with the North-East Hampshire and Farnham NHS Vanguard project (known as *Happy, Healthy at Home*)[17], we identified a requirement for a short easy to use measure of personal wellbeing.

R-Outcomes Ltd has developed, validated and used a coherent family of short generic patient-reported outcome measures during the past decade. These cover health status (*howRu*),[18] patient experience (*howRwe*)[19] and health confidence (HCS)[20]. We sought a short measure of personal wellbeing for use in combination with other R-Outcomes measures in surveys, which could also have broad use.

We reviewed the literature on and decided to explore the feasibility of adapting the ONS4 questions to the R-Outcomes format, initially because it appeared relatively easy to adapt.

The feasibility study and first drafts were done quite quickly and the wording was refined over about eighteen months, initially in collaboration with staff at North-East Hampshire and Farnham, who asked patients to complete early versions of the questionnaire. This was to ensure that the measure was as simple to understand and as easy to use as possible.

All R-Outcomes measures share a strong family resemblance. They are generic (applicable to people with any conditions), short, with a low reading age. Each measure has four response options, which are labelled, colour-coded and use smiley pictographs, with the best option on the left and the least desirable on the right. The measures usually have four question items, although exceptions are allowed. By making measures short and easy to use, the aim has been to increase their usability across different contexts and to improve data completion rates, so that these constructs can be collected and reported routinely for a wide variety of purposes.

The design criteria included clarity, brevity, suitability for frequent use, multi-modality (suitability for use with multiple data collection modalities including smart-phones), responsiveness, good psychometric properties and easily understood scoring.

Length and Readability

The length and readability of PWS was compared with six widely used measures of health and well-being in the UK (GHQ-12, S-WEMWBS, ONS-4, ICECAP-A, ASCOT and EQ-5D).[10] The text analysed is that shown therein. The version of ONS4 used is less verbose than that shown in Box 1. The OECD Core Questions were also measured.

The number of items and words were counted, excluding superfluous statements such as copyright ownership. Readability was measured using the Flesch-Kincaid Readability Grade (FKG) provided in Microsoft Word. It has been recommended that patients should not be asked to complete questionnaires with a reading age of more than ten,[21] which corresponds roughly to readability grade FKG=5.

Testing and Validation

To examine the psychometric properties and construct validity of PWS, we used secondary analysis of data collected as part of the evaluation of five social prescribing services in Hampshire, the Isle of Wight and Surrey during 2016 and early 2017.

Social prescribing (or community referral) is a relatively new approach in health care, aiming to create referral pathways that enable the GP or a health care practitioner to refer patients with social or practical needs to a local provider of nonclinical services.[22] These services are often offered by voluntary or the community organisations and cover a wide range of interventions including educational sessions, exercise training, dietary advice, creative activities, self-help groups, emotional or social support, and stress management.[23] The social prescribing interventions were broadly similar but with local differences in terms of case mix, support skills and on-call availability. Some services discharged people but others did not.

The evaluation of each project used a mixed methods approach including economic, qualitative and survey methods, recognising the difficulties of evaluating this type of innovation.[24] Similar data was collected for each service. The choice of measures and method of data collection was discussed and agreed with each local service in advance. Each service had its own survey, with minor differences between them.

Each survey included PWS, the howRu health status measure, howRwe patient experience measure, the health confidence score (HCS) Astonhall46 and two additional items on service integration ('services talk to each other' and 'I don't have to repeat my story'). The surveys also included gender (male or female), age in deciles, and number of medications being taken (none, 1 or 2, 3 to 5, 6 to 9,

10 or more). All items were optional, although people were encouraged to complete all items. All responses were anonymous.

For each service, responses were collected from two cohorts: (1) people referred to each service over a period of about 4 weeks (on referral), and (2) from current or previous patients (after referral). A member of the social prescribing team visited each participant in their own home soon after referral, where the on referral survey was collected. Some after referral surveys were collected over the telephone, but the mode of data collection was not recorded. The interval between referral and after referral ratings was not recorded, but was usually between 4 and 10 weeks. The responses were recorded on a paper copy of the survey and either they, or another member of the service, transcribed these later onto the R-Outcomes server. There was no linkage between responses on referral and after referral. For this analysis, the data has been pooled from the five separate services.

Sample size and missing data: We measured the number of responses, the number and proportion of missing items on referral and after referral. A small number of responses which did not include type of response (on referral or after referral) were excluded from the analysis.

Distribution: The response distributions on referral and after referral were examined for each item and summary statistics calculated including the overall summary score, means, standard deviations and proportion of responses in floor (lowest) and ceiling (highest) states. We also examined the distributions by gender, age group and, as a proxy for co-morbidities, the number of medications taken.

Internal consistency: We expected the degree of interrelatedness among the items, assessed by correlations between the items, would be in the range 0.4 to 0.6, with the strongest correlation between the pairs of items on positive and negative experience, then life evaluation and worthwhileness (convergent validity).

Cronbach's alpha: We expected Cronbach's α to be between 0.7 and 0.9, which would support the use of an aggregate summary score.[25]

Construct validity is the degree to which the scores of an instrument are consistent with hypotheses, such as internal relationships, relationships to scores of other instruments or differences between relevant groups, based on the assumption that the instrument validly measures the construct to be measured.[26] This was assessed by the measure being sensitive to clinical interventions, such as the social prescribing service. We hypothesised that:

- Personal wellbeing would be lower on referral to social prescribing than after referral.
- There would be little difference in personal wellbeing between men and women.
- Personal wellbeing would be positively associated health status, health confidence and, less strongly, with patient experience.
- For people using these services, personal wellbeing would be higher in older people, because older people tend to report higher wellbeing than those of working age.[27]
- For people using these services, personal wellbeing would decrease with the number of medications taken, because wellbeing is positively correlated with health.

Regression analysis was performed to help understand the relationship between PWS summary scores and health status (howRu), health confidence (HCS), age and number of medications taken.

Responsiveness is the ability of an instrument to detect change over time in the construct to be measured. This was assessed by comparing the results of the on referral and after referral cohorts.

Factor analysis was applied to the whole data set (using an oblique rotation, Promax, as we expected constructs to be correlated) for the individual questions in howRu, PWS, HCS, howRwe and the two additional experience questions asked.

Ethics Statement

We performed secondary analysis of existing anonymous data that had been collected as part of the routine service evaluation of social prescribing services. This did not require ethics approval because data collection was undertaken to define or judge the current services without randomisation and were anonymous. No data was collected by the services until patients had consented and there was no risk of substantial damage or distress to individual participants.[28]

Results – The PWS Measure

The final version of the Personal Wellbeing Score (PWS) is shown in Figure 1.

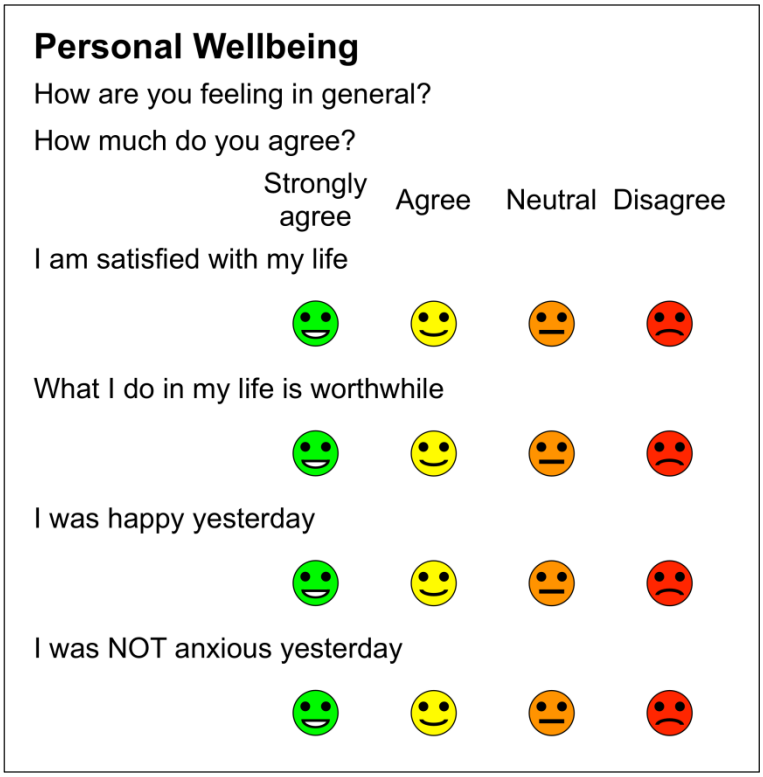

Figure 1 Personal Wellbeing Score

The principal differences between ONS4 and PWS and the reasons for these, are described below in terms of options, items and scoring.

Items

Brevity is a feature of all R-Outcomes measures and the length of ONS4 items was reduced as follows.

The life evaluation question (*Overall, how satisfied are you with your life nowadays? where nought is 'not at all satisfied' and 10 is 'completely satisfied'*) became *I am satisfied with my life* with a scale from *Disagree* to *Strongly agree*.

The worthwhile (eudaimonia) question (*Overall, to what extent do you feel the things you do in your life are worthwhile? where nought is 'not at all worthwhile' and 10 is 'completely worthwhile'*) became *What I do in my life is worthwhile* with a scale from *Disagree* to *Strongly agree*.

The positive experience question (*Overall, how happy did you feel yesterday? (where nought is 'not at all happy' and 10 is 'completely happy'*) became *I was happy yesterday* with a scale from *Disagree* to *Strongly agree*.

The negative experience question (*On a scale where nought is 'not at all anxious' and 10 is 'completely anxious', overall, how anxious did you feel yesterday?*) became *I was NOT anxious yesterday* with a scale from *disagree* to *strongly agree*. The scale direction was reversed from negative to positive by adding *NOT*. The potential problems of a double negative (not anxious) are offset by consistency between questions.

Options

The main scale was modified, from an 11-point scale anchored at *not at all* = 0 and *completely* = 10, to an agree/disagree format with four options: *Strongly agree*, *Agree*, *Neutral* and *Disagree* to match other R-Outcomes measures. There is no *strongly disagree* option, because scales should approximate the actual distribution of the characteristic in the population and this option would be rarely used.[29] For example, a study using a 5-point variant of ONS4, found that only 2.1% of responses chose the option closest to *strongly disagree*.[30]

The four PWS response options *Strongly agree*, *Agree, Neutral* and *Disagree* correspond roughly with the way that ONS4 results are usually published, using four thresholds for each question.[31] For life satisfaction, worthwhile and happiness scores, ONS4 responses 9-10 are grouped as *Very high*, 7-8 as *High*, 5-6 as *Medium* and 0-4 as *Low*. For anxiety scores ONS4 responses 6-10 are grouped as *High*, 4-5 as *Medium,* 2-3 as *Low* and 0-1 as *Very low.*

The response options in PWS are ordered left to right from best (*Strongly agree*) to worst (*Disagree*), but in ONS4 the response options are ordered left to right from nought (*not at all*) to 10 (*completely*), which is from worst to best for three items and best to worst for negative experience.

PWS response options are colour-coded (*Strongly agree* is green, *Agree* is yellow, *Neutral* is orange and *Disagree* is red) and annotated with smiley pictographs. However, colour and pictographs are not essential components of the measure.

Scoring

A high score is always better than a low score. Each PWS item is scored as follows: *Disagree* = 0, *Neutral* = 1, *Agree* = 2 and *Strongly Agree* = 3.

In addition to each item score, the PWS calculates a summary score from the 4 items, although ONS4 does not. The PWS summary score is the sum of the four component scores giving a 13-point scale from 0 (4 x *Disagree*) to 12 (4 x *Strongly agree*).

For cohorts, the mean item scores and the summary score are transformed linearly to a 0-100 scale. Here a score of 100 is obtained if all respondents choose the highest possible score (ceiling) and 0 if all respondents choose the lowest possible score (floor).

Results – Testing

Length and Readability

Table 1 shows the number of items, word count, Flesch-Kincaid Grade and approximate reading age for PWS and other measures. PWS has the lowest word count and reading age.

Table 1. Number of items, word count, Flesch Kincaid Grade and reading age for related measures.

Measure	Number of Items	Word count	Flesch Kincaid Grade	Reading Age
PWS	4	42	3.7	9 years
GHQ-12	12	324	6.3	11 years
S-WEMWBS	7	89	3.8	9 years
ONS4 (concise version)	4	62	6.5	12 years
ICECAP-A	5	264	5.1	10 years
ASCOT	8	415	5.3	10 years
EQ-5D (including VAS)	6	263	5.9	11 years
OECD Core Questions	5	177	6.4	11 years

Distribution

We analysed data for all participants (n=1324), on referral (n=647) and after referral (n=677). The number of missing values for PWS items was low, between 0.8% (*I am satisfied with my* life) and 1.3% (*What I do in my life is worthwhile*). For the PWS summary score, which requires all four items, there were 1.9% missing values.

Table 2 shows the overall frequency distribution for each item and response option, showing a broad spread in this population. The highest proportion is for *Agree* that *I am satisfied with my life* (48.2%). The lowest proportion, is *Disagree* that *What I do in my life is worthwhile* (7.4%).

Table 2 Frequency counts (%) for each PWS item.

	Strongly agree	Agree	Neutral	Disagree
I am satisfied with my life	315 (24.0%)	633 (48.2%)	239 (18.2%)	126 (9.6%)
What I do in my life is worthwhile	311 (23.8%)	574 (43.9%)	325 (24.9%)	97 (7.4%)
I was happy yesterday	312 (23.8%)	623 (47.5%)	234 (17.8%)	142 (10.8%)
I was NOT anxious yesterday	303 (23.1%)	525 (40.1%)	276 (21.1%)	205 (15.7%)

Table 3 shows the mean and standard deviation of each item on 0-100 scale for all participants, on referral and after referral. Mean scores range from *I was NOT anxious yesterday* on referral (52.5) to *I am satisfied with my life* after referral (66.3). The standard deviations show a wide distribution for all items, ranging from *What I do in my life is worthwhile* after referral (27.1) to *I was NOT anxious yesterday* on referral (33.1). This illustrates the heterogeneous case mix of people using these services.

Table 3 Mean scores (Standard Deviation) on 0-100 scale for each PWS item and summary score, for all participants, on referral and after referral

Variable	**All participants mean (SD)**	**On referral mean (SD)**	**After referral mean (SD)**	**Mean difference**
I am satisfied with my life	62.0 (29.6)	57.9 (31.1)	66.3 (27.5)	8.34
What I do in my life is worthwhile	61.4 (29.0)	57.2 (30.5)	65.2 (27.1)	8.01
I was happy yesterday	61.4 (30.3)	56.7 (32.1)	66.0 (27.9)	9.30
I was NOT anxious yesterday	56.9 (33.1)	52.5 (34.6)	61.1 (31.1)	8.67
PWS Summary Score	60.4 (26.0)	56.0 (26.9)	64.6 (24.3)	8.66

Table 4 shows the distribution of PSW summary scores for all participants, on referral and after referral. In this population 2.8% were at the floor (lowest option on all four items) and 15.2% were at the ceiling (highest option on all four items). This does not suggest a problematic floor or ceiling effect, whereby a measure is unable to detect whether participants become worse or better, because they are already in an extreme class.

Table 4 Distribution of PWS summary scores, for all participants, on referral and after referral.

PWS score	All responses		On referral		After referral	
	n	%	n	%	n	%
0	37	2.8%	27	4.3%	10	1.5%
1	23	1.8%	15	2.4%	8	1.2%
2	44	3.4%	31	4.9%	13	2.0%
3	53	4.1%	34	5.4%	19	2.9%
4	109	8.4%	57	9.0%	52	7.8%
5	84	6.5%	47	7.4%	37	5.6%
6	136	10.5%	72	11.4%	64	9.6%
7	123	9.5%	61	10.1%	62	9.3%
8	311	23.9%	137	21.6%	174	26.1%
9	87	6.7%	36	5.7%	51	7.7%
10	66	5.1%	26	4.1%	40	6.0%
11	28	2.2%	9	1.4%	19	2.9%
12	198	15.2%	81	12.8%	117	17.6%
Total	1,299		633		666	

Figure 2 shows the distribution of PWS summary scores for social prescribing patients, on referral and after referral, grouped in four classes, very low (0-3), low (4-6), medium (7-9) and high (10-12).

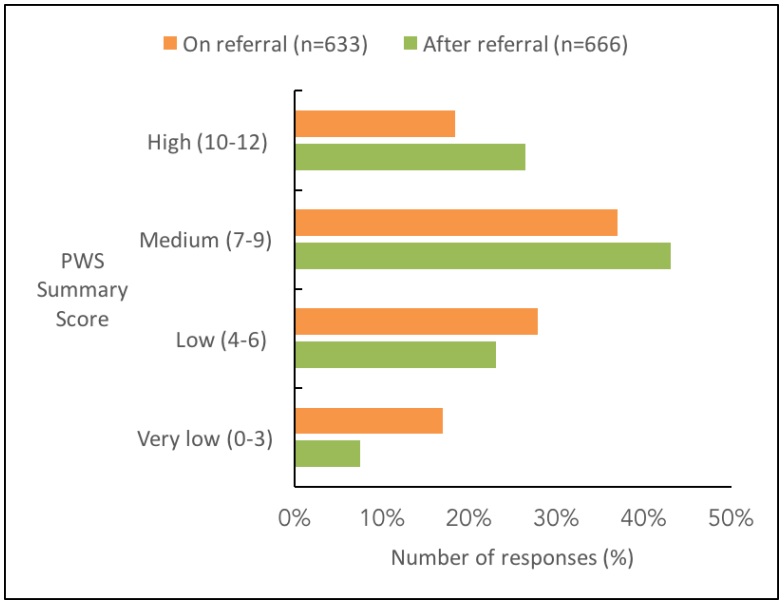

Figure 2 Distribution of grouped PWS summary scores for social prescribing patients on referral and after referral

Table 5 shows measures of central tendency (mean, median, standard deviation and inter-quartile range) for the PWS summary score on both raw (0-12) and (0-100) scales, for all responses, on referral and after referral. The mean PWS summary score is 7.25 on the 0-12 scale, with standard deviation 3.12, range 0 - 12, interquartile range 5 to 9, median 8, mode 8, negative skew -0.266 and

kurtosis -0.435, indicating a relatively flat distribution, as shown in Figure 2. On the 0-100 scale, mean PWS is 60.4 (95% CI 59.0 to 61.8).

Table 5 Mean, median, standard deviation and interquartile range of PWS summary score for all responses, on referral and after referral.

PWS score	**All responses**		**On referral**		**After referral**	
	0-12 scale	**0-100 scale**	**0-12 scale**	**0-100 scale**	**0-12 scale**	**0-100 scale**
Mean	7.25	60.4	6.72	56.0	7.76	64.6
Median	8	67	7	58	8	67
Standard deviation	3.12	26.0	3.23	26.9	2.92	24.3
Interquartile range (IQR)	5 to 9	42 to 75	4 to 8	33 to 67	6 to 10	50 to 83

Internal Consistency

Table 6 shows inter-item correlations. All items are highly correlated, with the highest correlations between evaluation items (*I am satisfied with my life* and *What I do in my life is worthwhile)* r=0.77, and affect items, (*I was happy yesterday* and *I was NOT anxious yesterday)* (r=0.73). Cronbach's alpha is 0.903.

Table 6 PWS inter-item correlation matrix

Variable	**I am satisfied with my life**	**What I do in my life is worthwhile**	**I was happy yesterday**	**I was NOT anxious yesterday**
I am satisfied with my life	1.0			
What I do in my life is worthwhile	0.77	1.0		
I was happy yesterday	0.61	0.65	1.0	
I was NOT anxious yesterday	0.51	0.52	0.73	1.0
PWS Summary Score	0.84	0.85	0.88	0.83

Construct Validity

Table 7 shows the correlations between PWS items and summary score and health status (howRu), health confidence (HCS) and experience (howRwe) summary scores. PWS items and summary score are strongly correlated with the howRu and health confidence scores and moderately with patient experience (howRwe).

Table 7 Pearson correlations between PWS items and summary score and summary scores for health status, health confidence and experience. All correlations are significant (p<0.001)

Item	Health status (howRu) r	Health confidence (HCS) r	Experience (howRwe) r
I am satisfied with my life	0.49	0.49	0.28
What I do in my life is worthwhile	0.49	0.55	0.27
I was happy yesterday	0.51	0.52	0.26
I was NOT anxious yesterday	0.49	0.47	0.23
PWS Summary Score	0.58	0.60	0.30

People with low health status report low personal well-being and those with high health status tend to report high personal well-being (Figure 3).

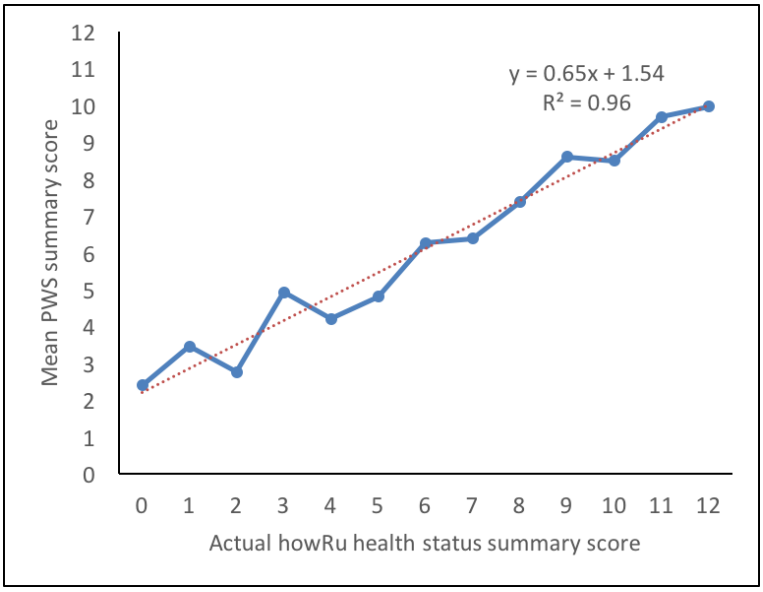

Figure 3 Shows the relationship between actual health status (howRu) summary scores and mean PWS summary scores.

Low health confidence scores (HCS) are strongly associated with low personal wellbeing scores, and high health confidence scores are associated with high personal wellbeing scores (Figure 4).

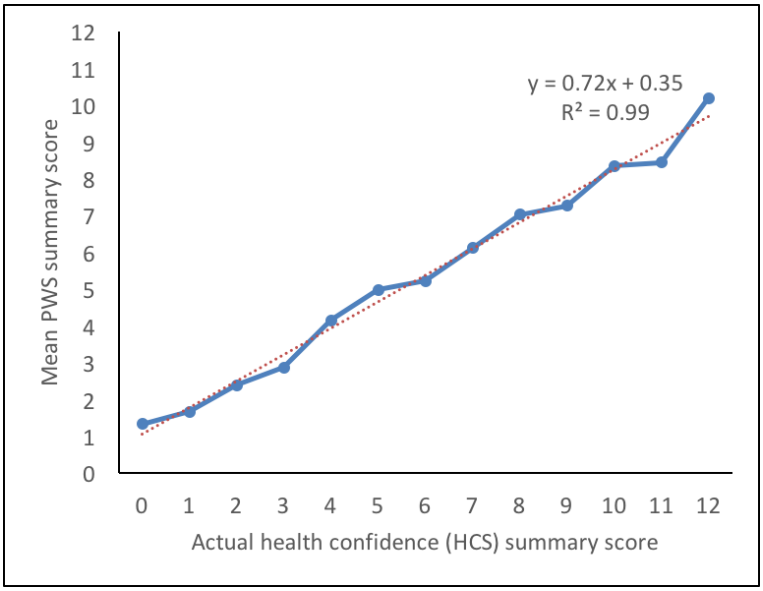

Figure 4 Relationship between actual health confidence (HCS) summary score and mean PWS score

Regression analysis on 1224 records with complete data for all variables (health status (howRu) summary score, health confidence (HCS) summary score, age decade (Age) and rank order of number of medications taken (Meds)), generated the following regression equation:

PWS = 0.446*howRu + 0.496*HCS + 0.376*Age + 0.112*Meds - 3.715

Coefficients for howRu, HCS and Age are significant ($p<0.00001$); coefficient for Meds is not significant ($p=0.128$). $R^2 = 0.484$.

We found no significant differences between males and females for any item apart from for *I am satisfied with my* life ($p=0.043$) (2-tailed t-test).

Table 8 shows the distribution, mean PWS scores and standard deviations on 0-100 scale, broken down by all participants, on referral and after referral, and by gender, age group and number of medications.

The proportion of female participants (60.7%) is greater than the number of males, but there is no statistical difference in their mean PWS score ($p=0.196$).

Age is skewed to older age groups with 64.3% of participants over 80 years old and 15.8% over 90. Older participants tend to report higher wellbeing than younger. In this population of people referred to social prescribing, the mean PWS scores for the minority of participants under the age of 70 (n=146) was 43.6, which is low; for the large majority over 70 (n=1118), PWS 62.5.

Table 8 Counts, mean PWS and standard deviations (SD) for all participants with complete PWS scores, on referral (A) and after referral (B) by sex, age group and number of medications taken

Variable	**All participants**		**On referral (A)**		**After referral (B)**		**PWS Difference**
	Count	Mean PWS	Count	Mean PWS	Count	Mean PWS	(B-A)
Overall	1,299	60.4	633	56.0	666	64.6	8.6
Sex							
Female	781	59.5	367	55.0	414	63.5	8.5
Male	506	61.4	260	56.9	246	66.2	9.3
Age Group							
20 to 29	11	38.6	7	34.5	4	45.8	11.3
30 to 39	15	32.8	8	26.0	7	40.5	14.5
40 to 49	7	56.0	6	50.0	1	91.7	41.7
50 to 59	35	41.9	27	37.3	8	57.3	19.9
60 to 69	78	46.0	37	39.6	41	51.8	12.2
70 to 79	253	57.3	126	51.3	127	63.3	12.0
80 to 89	638	63.6	297	61.0	341	66.0	5.0
90 to 99	224	64.9	97	61.2	127	67.8	6.6
100 +	3	66.7	1	16.7	2	91.7	75.0
Medications							
None	25	60.7	12	64.6	13	57.1	-7.53
1 or 2	118	55.9	71	54.2	47	58.3	4.11
3 to 5	530	63.2	276	59.2	254	67.5	8.21
6 to 9	410	60.6	161	53.4	249	65.3	11.88
10 or more	182	54.8	87	51.6	95	57.7	6.09

Table 9 shows the results of the factor analysis. A scree plot implies four or six factors, while Kaiser's criterion implies four. This supports the design of the 4 scales as measuring distinct constructs, although in this population it sub-divides howRu into 'disability' and 'distress', similar to Rosser's seminal classification of disability and distress.[32] A factor analysis of the 4 PWS questions shows no indication of internal structure. Results were broadly the same when repeated on just the on referral or post-referral data.

Table 9 Factor analysis results, using oblique rotation, Promax. Only weights over 0.3 are shown.

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
howRu 1				0.51		
howRu 2				0.51		
howRu 3					0.77	
howRu 4					0.82	
PWS 1			0.69			
PWS 2			0.67			
PWS 3			0.87			
PWS 4			0.82			
HCS 1		0.61				
HCS 2		0.46				
HCS 3		0.71				
HCS 4		0.71				
howRwe 1	0.85					
howRwe 2	0.91					
howRwe 3	0.88					
howRwe 4	0.91					
Services talk to each other	0.36					0.42
No need to repeat story	0.61					0.39

Discussion

This study describes the development and testing of the Personal Wellbeing Score (PWS). This measure has been adapted from the Office of National Statistics ONS4. PWS is shorter (42 words) with a lower reading age (9 years) than other measures of personal wellbeing. PWS includes an overall unweighted summary score as well as separate scores for each item. Mean scores for a population are reported on a scale from 0 to 100.

In the PWS all items are worded positively, but the anxiety item in ONS is worded negatively, while the other three items are worded positively. Factor analysis and analysis of internal correlations show that the PWS anxiety item is not a separate factor and behaves in a similar way to the other items. However, factor analysis on ONS4 data sets has shown that positively and negatively worded items relate to different factors.[11]

The validation study used secondary analysis of data on 1,324 patients, collected as part of the evaluation of five social prescribing projects in Hampshire, Surrey and the Isle of Wight. 1,299 people (98.1%) completed all four PWS items. The results also show good psychometric properties and strong construct validity.

As expected, there was a significant difference between scores on referral (PWS = 56.0) and several weeks after referral (PWS = 64.6). The distribution of responses covers the whole range, with no indication of problematic floor or ceiling effects.

In high-income English-speaking countries, subjective wellbeing generally follows a U-shaped pattern, lower in middle age and higher as people get older.[27,33] We found this pattern in our population. The mean PWS of people under 70 years old was considerably lower (43.6) than those over 70 (62.5). The age effect may be enhanced because the people referred to these social prescribing projects usually have a case mix with multiple problems, which are less common in younger people.

Internal correlations between PWS items were high, in the range r=0.51 to 0.77. Cronbach's alpha was 0.903, at the top end of the expected range. This suggests that it is appropriate to use a single summary score for this instrument, in addition to individual item scores.

The results were consistent with the hypotheses suggested to test construct validity. The mean scores for people who have been receiving social prescribing and proactive care services were significantly higher after the intervention than on referral.

Regression analysis shows that PWS scores are strongly related to health confidence (r=0.60) and health status (r=0.58), moderately with age group (r=0.24), but not with number of medications taken. Studies using ONS4 have shown that personal wellbeing has a strong relationship with self-reported health, employment status and living alone, and a moderate association with age.[34] Our data strongly supports this for health status, but we did not record employment status nor whether people lived alone. A strong association between successful goal pursuit, which is likely to be closely associated with health confidence, and subjective wellbeing has been reported.[35] There was little difference between the mean scores for men and women.

The correlations between PWS dimensions and howRu (health status) dimensions were moderate or strong. As expected, the strongest correlations were between howRu *Feeling low and worried* and PWS positive and negative affect, *I was happy yesterday* (r=0.55) and *I was NOT anxious yesterday* (r=0.56). The lowest correlations were between howRu *Pain or discomfort* and PWS *What I do in my life is worthwhile* (r=0.29) and between howRu *Require help from others* and PWS *I was NOT anxious yesterday* (r=0.30).

Factor analysis shows that the PWS items are related and distinct form other questions asked in the survey.

Practical Issues and Limitations

The PWS questions were asked as part of a longer survey covering health status, health confidence and patient experience as well as personal wellbeing. Some staff expressed concern before asking people about their personal wellbeing, but in practice almost all people were happy to answer these questions. The main difficulties occurred with people with advanced dementia and at their end of life, where discretion was used.

There is some evidence that telephone surveys may elicit slightly higher ratings for wellbeing than face-to-face interviews.[36] In our data, the on-referral ratings were collected face to face, but a proportion of after referral ratings were done by telephone. We did not collect mode of administration, so the mode effect may have impacted the results to an unknown degree.

The population being tested comprised mainly older people with multiple conditions. The data has been combined from five smaller projects each with slightly different case mix and ways of working. The population in this study is older and in poorer health than that of national studies of wellbeing such as the ONS Annual Population Survey (APS).[31] More than 68% of people who completed our surveys were more than 80 years old, and many were in poor health. This demonstrates the practicality of using the PWS with old and vulnerable people, although it is not designed specifically for older people.

Useful comparisons can be made with the APS results, which use ONS4. These show that poor self-reported health has a large impact on all ONS4 items and we have found the same. Old age has a moderate impact on life satisfaction and a smaller impact on the other items. Gender has only a small or very small impact. Our findings using PWS are consistent with these findings. Other variables in the APS which have a large contribution to personal wellbeing are economic activity and marital status, but our study did not record this information.

Conclusions

The Personal Wellbeing Score (PWS) is a new short variant of ONS4, which has been shown to be easy and suitable for use with vulnerable older people and is responsive to social prescribing interventions. It is shorter and has a lower reading age than other widely used instruments. In evaluation studies, it had few missing values, good psychometric results with good correlation with concurrent measures of health status and health confidence, and construct validity.

Declaration of conflicting interests

TB and AL are directors of R-Outcomes Ltd, which owns the copyright of the Health Confidence Score and provides quality improvement and evaluation services using it. Please contact R-Outcomes Ltd if you wish to use the PWS.

HP has done consultancy work for Crystallise Ltd, unrelated to the work reported herein. The authors declare that they have no other conflicting interests.

Authors' contributions

TB designed the questionnaire, and wrote the first draft of the paper. TB, HWWP and JS performed the analyses. JS and AL were actively involved in the data collection. All authors contributed to the final text, read and approved the final manuscript.

Funding acknowledgement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgement

We are grateful for the valuable suggestions of Alexis Foster on an earlier draft of this paper.

References

1. Helliwell J, Layard R, Sachs J. *World Happiness Report 2017*, New York: Sustainable Development Solutions Network 2017.
2. Thin N. *Social Happiness: Theory into Policy and Practice*. Bristol: The Policy Press 2012.
3. National Research Council. *Subjective well-being: measuring happiness, suffering and other dimensions of experience.* The National Academies Press 2014
4. Stiglitz J, Sen A, Fitoussi J-P. *Report by the commission on the measurement of economic performance and social progress*. 2009 www.stiglitz-sen-fitoussi.fr
5. Dolan P, Metcalfe R. Measuring subjective wellbeing: recommendations on measures for use by national governments. *Journal of social policy* 2012; 41 (2): 409-427.
6. Tinkler L. The Office for National Statistics Experience of Collecting and Measuring Subjective Well-Being. *Statistics in Transition new series* 2015; 16 (3): 373-396.
7. Allin P, Hand D. New statistics for old? – measuring the wellbeing of the UK. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2017; 180(1): 3-43.
8. ONS. *Personal Well-being in the UK QMI – Quality and Methodology Information*. Newport: Office for National Statistics 8 July 2016.
9. OECD. *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing 2013. http://dx.doi.org/10.1787/9789264191655-en
10. Peasgood T, Brazier JE, Mukuria C, Rowen D. A conceptual comparison of well-being measures used in the UK. Policy Research Unit in Economic Evaluation of Health and Social Care Interventions (EEPRU) (26) 2014.
11. Mukuria C, Rowen D, Peasgood T, Brazier JE. An empirical comparison of well-being measures used in the UK. Policy Research Unit in Economic Evaluation of Health and Social Care Interventions (EEPRU) (27) 2016.
12. Goldberg D, Williams P. *A user's guide to the GHQ*. Windsor, NFER Nelson. 1988

13. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh mental well-being scale (WEMWBS): a Rasch analysis using data from the Scottish health education population survey. *Health and quality of life outcomes*. 2009; 7(1): 15.
14. EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health policy* 1990; 16(3): 199.
15. Al-Janabi H, Flynn T, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research*. 2012; *21*(1): 167-176.
16. Malley JN, Towers AM, Netten AP, Brazier JE, Forder JE, Flynn T. An assessment of the construct validity of the ASCOT measure of social care-related quality of life with older people. Health and Quality of life Outcomes. 2012; 10(1): 21.
17. Happy healthy at home project web site http://www.happyhealthyathome.org (accessed 4 April 2017).
18. Benson T, Whatling J, Arikan S, Sizmur S, McDonald D, Ingram D. Evaluation of a new short generic measure of HRQoL: howRu. *Informatics in Primary Care 2010*, 18:89-101.
19. Benson T, Potts HWW. A short generic patient experience questionnaire: howRwe development and validation. *BMC Health Services Research* 2014; 14:499.
20. Benson T, Bowman C, Sladen J, Liles A, Potts HWW. Development and validation of the Health Confidence Score (HCS). Submitted for publication 2017.
21. Paz S, Jiu H, Fongwa M, Morales L, Hays R: Readability estimates for commonly used health-related quality of life surveys. *Qual Life Res* 2009; 18: 889-900.
22. Brandling J, House W. Social prescribing in general practice: adding meaning to medicine. *Br J Gen Pract* 2009; 59 (563): 454-456.
23. Pilkington K, Loef M, Polley M. Searching for real-world effectiveness of health care innovations: scoping study of social prescribing for diabetes. *J Med Internet Res* 2017; 19 (2): e20.
24. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, Greaves F, Harper L, Hawe P, Moore L, Petticrew M. The need for a complex systems model of evidence for public health. Lancet. 2017 Jun 13. doi: 10.1016/S0140-6736(17)31267-9. [Epub ahead of print]
25. Streiner D, Norman G. *Health measurement scales: A practical guide to their development and use*. 4th edition. Oxford: Oxford University Press; 2008.
26. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, Bouter L, de Vet H. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology* 2010; 63 (7): 737-745.
27. Steel M. *Measuring national well-being: at what age is personal well-being the highest?* Newport: Office for National Statistics, 2 February 2016.
28. NHS Health Research Authority. *Defining Research: Research Ethics Service guidance to help you decide if your project requires review by a Research Ethics Committee.* UK Health Departments' Research Ethics Service 2016.
29. Dillman DA, Smyth JD, Christian LM. Internet, phone, mail, and mixed-mode surveys: the tailored design method. John Wiley & Sons; 2014
30. Palmer V, Evans E. *Opinions and Lifestyle Survey: methodological investigation into response scales in personal well-being*. ONS 2015.
31. ONS. *Personal Well-being in the UK QMI*. Office of National Statistics 8 July 2016.
32. Rosser RM, Watts VC. The measurement of hospital output. *International Journal of Epidemiology* 1972; 1:361-368
33. Steptoe A, Deaton A, Stone AA. Subjective wellbeing, health, and ageing. *The Lancet* 2015; 385(9968): 640-648.
34. Oguz S, Merad S, Snape D. *Measuring national well-being – What matters most to personal well-being.* Newport: Office for National Statistics 30 May 2013.
35. Klug H, Maier G. Linking Goal Progress and Subjective Well-Being: A Meta-analysis. *Journal of Happiness Studies* 2015; 16 (1): 37-65.
36. Dolan P, Kavetsos G. Happy talk: Mode of administration effects on subjective well-being. *Journal of Happiness Studies* 2016; 17 (3): 1273-1291.