
# Public health interventions to promote mental well-being in people aged 65 and over: systematic review of effectiveness and cost-effectiveness

**Gill Windle[1], Dyfrig Hughes[2], Pat Linck[2], Ian Russell[4,] Rhodri Morgan[1], Bob Woods[3], Vanessa Burholt[1], Rhiannon Tudor Edwards[2], Carla Reeves[1], Seow Tien Yeo[2],**

1 **Centre for Social Policy Research & Development, Institute of Medical & Social Care Research, University of Wales Bangor.**
2 **Centre for Economics & Policy in Health, Institute of Medical & Social Care Research, University of Wales Bangor.**
3 **Professor of the Clinical Psychology of the Elderly, Institute of Medical & Social Care Research, University of Wales Bangor.**
4 **Professor of Public Health, Institute of Medical & Social Care Research, University of Wales Bangor.**

## List of contents

## List of Tables (T) & Figures (F)

## Summary

A systematic search of 21 data bases and 11 websites sought evidence, published between January 1993 and February 2007, of the effectiveness or cost-effectiveness of interventions to promote mental well-being in later life. The search was restricted to the English language. In principle all study designs were considered for inclusion. In total 15,388 citation titles and abstracts were screened for relevance. By this process 250 articles were identified for further appraisal for inclusion in either review – 220 for effectiveness and 30 for cost-effectiveness. Application of inclusion criteria selected 98 papers for the review – 97 for effectiveness and two for cost-effectiveness. The 97 effectiveness papers described four meta-analyses, 14 trials of good quality (one of which generated two papers), 69 quantitative studies of poor quality (one of which generated two papers) and eight qualitative papers (including six of good quality).

Thus most included studies were of poor quality. Many used small samples that may not represent the population of interest, and certainly lack statistical power. Many recruited participants through advertisements, probably recruiting more motivated individuals, and again making findings less representative. The frequent use of self selection means that women predominate. Few included studies focused on frail older people or people over 80. Few interventions were targeted at alleviating poverty, and none at older people from ethnic or sexual minorities. Few studies answered sub-questions in full, including who delivered the intervention and where.

We divided the 97 studies into 15 categories – six concerned with different types of exercise, three with different types of health promotion, one each with psychological interventions, computer use, gardening, support groups and volunteering, and one residual category. Three of these categories generated useful evidence statements from meta-analyses, and another three from single rigorous trials. Unfortunately two categories, each with three rigorous trials, did not generate a useful evidence statement because the evidence from these trials was conflicting. Another six categories generated no rigorous evidence. While the four good qualitative studies in the final category – mixed health promotion – are helpful, they cannot estimate strength of evidence. Hence the review has generated six robust positive evidence

statements – nos. 1 to 4 relating to exercise, no. 7 relating to health promotion and no. 10 relating to psychological interventions (pp. 9 to 11). Of the two identified cost-effectiveness papers, one added to the evidence on exercise and the other to the evidence on health promotion. To address the lack of economic papers, this review also shows how economic modelling can extend studies concerned solely with effectiveness so as to throw light on cost-effectiveness.

In summary there is a shortage of robust evidence for the effectiveness and cost-effectiveness of interventions to improve the mental well-being of older people. Better research is needed to estimate the value of most interventions. Research into cost-effectiveness is especially sparse, with little economic research even into programmes with evidence of effectiveness. Nevertheless this review has generated six useful positive evidence statements

## Evidence statements

| | |
|---|---|
| 1 | **Mixed exercise**<br>Two meta-analyses (Arent et al., 2000, MA**+**; Netz et al., 2005, MA**+**), together comprising 68 controlled trials from many developed countries, since augmented by four other rigorous trials in the Netherlands (2), Norway and the US, together provide strong evidence that mixed exercise programmes generally have small-to-moderate effects on mental well-being. As the reported exercise programmes cover a range of types, settings and countries, firm conclusions about the duration of programmes and the frequency of sessions are difficult. It is clear, however, that exercise of moderate intensity (not well defined in the meta-analyses) has beneficial effects on physical symptoms and psychological well-being.<br><br>The programmes evaluated were generally community-based, well organised and run by trained instructors. The findings apply to similar populations (relatively healthy and independent, and motivated to take exercise) in similar community settings in the UK. The sole qualitative study (Hardcastle & Taylor, 2001; Q**+**) highlights the importance of appropriate facilities and good supervision. |
| 2 | **Strength & resistance exercise**<br>Meta-analysis of four US trials that included a total of 1733 independent frail older people aged 65+ living in the community. Four of the SF-36 scales were used to evaluate similar resistance exercise interventions. A significant small-to-moderate improvement in emotional health was found (Schechtman & Ory, 2001; MA+). The findings are likely to be broadly applicable to frail older people in a range of settings in the UK.<br><br>Of six smaller controlled studies evaluating the benefit of resistance exercise for older people in general, five reported significant positive effects, mostly on the POMS measure (a self-reported measure of general mood over the past week). As all six were of poor quality, this finding should not be considered robust. |
| | |

| | |
|---|---|
| 3 | **Aerobic exercise** A medium-sized RCT in the US showed that both interventions – supervised aerobic brisk walking and 'toning & stretching' – generated similar trajectories of MUNSH and SWLS scores over 12 months in sedentary adults aged 60 to 75; these trajectories showed significant growth in happiness and satisfaction over the six-month exercise period, followed by a significant decrease at 12 months (McAuley et al., 2000, RCT**+**). The findings are likely to be broadly applicable to similar populations in the UK. |
| 4 | **Walking interventions** A walking programme delivered to older people in 28 heterogeneous neighbourhoods in Portland, Oregon by trained leaders three times a week over six months improved SF-12 mental health and SWLS life satisfaction scores relative to control neighbourhoods (Fisher & Li, 2004, Cluster RCT**+**). This cluster randomised trial recruited 279 people to the intervention group (of whom 156 completed the intervention) and compared them with 303 controls who received education only. Though recruitment and retention of participants is important for such programmes, the results are likely to be broadly applicable to similar populations in the UK. |
| 5 | **Tai Chi** Two out of three rigorous evaluations in the US showed that 3 to 6-month community-based Tai Chi programmes delivered by professionals improve differing mental health measures in older people (Li et al., 2002 & 2004; RCT**+** but not Kutner et al. 1997; NCT**+**). However there was little difference between Tai Chi and less specific exercise programmes. Hence there is no evidence that the distinctive element of Tai Chi confers any benefit. |
| 6 | **Yoga** |
| | One good quality study (Oken et al., 2006; RCT+) comparing the effects of yoga with walking exercise and wait list controls, undertaken in the US with relatively healthy adults between 65-85 finds improvements in some aspects of health related quality of life (SF-36) but no improvement in mental well-being. |
| | |
| 7 | **Other exercise** A US pilot study (Williams et al CBAS-) found that home-based balance-training for 13 older females had no effect on the self-reported SWLS. Another very small study (Tanaka et al., 2002, UBAS-) tested a four-week programme of exercise and short naps on 11 older people in Japan. As |

| | |
|---|---|
| | only these weak studies were found in this category, the conclusion is that there is no robust evidence that these forms of exercise improve mental well-being (life satisfaction and GHQ). |

| | |
|---|---|
| 8 | **Group-based health promotion**<br>There is evidence from one well-designed longitudinal trial [Clark et al., 1997, RCT**++**; Clark et al., 2001, RCT**(++)**] that weekly educational sessions led by occupational therapists promoted and maintained positive changes in the SF-36 mental health summary score in participants recruited from two federally-subsidised apartment complexes for older adults in the US. Though the findings are likely to be broadly applicable to a similar population in the UK, the findings may not generalise to those in other circumstances (e.g. owner-occupiers & nursing home residents). A small pilot study adapted the intervention for the UK context (Mountain et al., 2006; Q+). The findings indicate that the intervention 'Lifestyle Matters' is acceptable to older people with diverse health status living in private housing, and a range of positive benefits were reported. |
| | |

| | |
|---|---|
| 9 | **Mixed health promotion programmes** There is no quantitative evidence on the effectiveness of mixed health promotion in improving mental well-being. Four qualitative studies suggest that comprehensive health promotion programmes delivered by professionals to homeless, poor or socially isolated older people are acceptable to users and perceived to improve mental wellbeing markedly (Buijs et al., 2003, Q**+**; Greaves & Farbus, 2006, MM**+**; Wilcock, 2006a & 2006b, Q**+**). As three of these studies are British and the fourth Canadian, they are applicable to the UK. |
| 10 | **Individually targeted health promotion**<br>There is conflicting evidence from four randomised trials (Halbert et al., 2000, RCT**+**; Kerse, 2005, RCT**+**; Frieswijk et al., 2006, RCT**–**; Markle-Reid et al., 2006, RCT**++**) on the effects on mental well being.of differing health promotion interventions delivered to individuals by professionals<br><br>In Canada Markle-Reid et al. found that monthly home visits of 1 hour by health |

| | |
|---|---|
| | promotion nurses significantly improved SF-36 mental health summary scores in the intervention group compared with usual care. The intervention also significantly reduced the costs of prescription drugs in the health promotion group, by enough to offset the costs of the scheme.<br><br>In the Netherlands Frieswijk et al.. found that a five-part bibliotherapy correspondence course to aid self management in slightly to moderately frail older volunteers living at home, resulted in significant improvements on the Sense of Mastery Scale in the short term, but not at six months..<br><br>In New Zealand Kerse found that a primary care intervention in which independent sedentary older patients received monthly phone calls from exercise specialists improved SF-36 vitality subscale scores, but had no effect on the mental health scores.<br><br>In contrast in Australia Halbert et al found that both the provision of 20 minutes of advice on physical activity by an exercise specialist to older patients, and no treatment for controls, in two general practices significantly reduced mental well-being in two SF-36 dimensions – vitality and emotional limitations on role. |
| 11 | **Psychological interventions**<br>A meta-analysis (Pinquart & Sörensen, 2001; MA**+**) covering a total of 84 studies from many developed countries provides strong evidence for the effectiveness of cognitive training, control-enhancing interventions, psycho-education, relaxation and supportive interventions in improving the subjective well-being of older people. The meta analysis draws on the international literature and is likely to be applicable to similar populations and settings in the UK It also reports that psychosocial interventions worked better in nursing homes that in the community.<br><br>There is little robust evidence on the effectiveness of more specific psychological interventions – dream telling, memory tapping, mental fitness training, resourcefulness training and visual stimulation. |
| 12 | **Computer use** Four trials examined the effect of computer training or use |

| | |
|---|---|
| | (Shrerer et al., 1996, NCT–; White et al., 1999, NCT–; Billipp et al., 2001, NCT-; White et al, 2002, RCT–). As all were of poor quality, there is no robust evidence on the effectiveness of computer use in improving mental well-being. |
| 13 | **Gardening interventions** Three studies examined the role of gardening in the mental wellbeing of older people (Barnicle & Midden 2003, CBAS-, Milligan et al, 2004, MM-, Heliker et al., 2000; UBAS-). As there were critical flaws in each study, there is no robust evidence on the effectiveness of gardening interventions in improving mental well-being. |

| | |
|---|---|
| 14 | **Support groups** Three studies reported the effect of support groups on mental well-being (Barnes & Bennett; 1998, Q-; Stewart et al., 2001, UBAS–; Powers & Wisocki, 2006, UBAS–). As each was of poor quality, there is no robust evidence that support groups improve mental well-being. |
| 15 | **Volunteering interventions**<br>Three studies reported the effect of volunteering interventions on older people (Wheeler et al., 1998, MA–; Rabiner et al., 2003, CBAS-; Butler, 2006, MM–). As all were of poor quality, there is no robust evidence on the effectiveness of volunteering in improving the mental well-being of older volunteers or older clients. |

| | |
|---|---|
| 16 | **Other interventions**<br>A range of other interventions have been evaluated in poor quality studies. It is concluded that there is no robust evidence on the effectiveness of altruistic activity, art therapy, catering redesign in long-term care, home massage, occupational therapy, pet therapy, sleep management, video games and wheelchair modification. |
| 17 | **Published studies of cost-effectiveness**<br>Two studies provided good evidence about the cost-effectiveness of interventions to improve the mental well-being of older people. First Hay et al. (2002; RCT**+**) showed that a two-hour group session of preventive advice from an occupational therapist per week is cost-effective in the US with an incremental cost per QALY of $10,700 (95% CI $6,700 to $25,400). Secondly Munro et al (2004; RCT**+**) showed that twice-weekly exercise classes led by |

| | |
|---|---|
| | qualified instructors are probably cost-effective in the UK with an incremental cost per QALY of £12,100 (95% CI = £5,800 to £61,400). While both studies are sound, one cannot be confident that such sparse findings will apply to similar populations (relatively healthy, living independently, and motivated to take advice and exercise) in similar community-based settings in the UK. |

| | |
|---|---|
| 18 | **Economic modelling of cost-effectiveness**<br>There are only two published economic analyses of interventions to improve the mental well-being of older people (evidence statement 16). To complement these sparse data needs economic modelling based on the integration of existing studies of effectiveness and existing sources of data about patient utilities and resource costs. The most cost-effective intervention was a thrice-weekly community-based walking programme, delivered to sedentary older people who are able to walk without assistance (Fisher & Li, 2004; Cluster RCT**+**). Modelling yielded an incremental cost per QALY of £7,400 after six months, which is comparable with the two published economic analyses.<br><br>Modelling was also used to enhance three RCTs of advice about physical activity. Such advice had an estimated incremental cost per QALY of £26,200 when modelled from Kerse et al. (2005; NCT**+**), who estimated the effects of the primary care 'green prescription' counselling programme in New Zealand. The estimated incremental cost per QALY rose to £45,600 when modelled from Markle-Reid et al. (2006; RCT**++**), who evaluated proactive health promotion by nurses in Canada in addition to usual home care for people over 75. However Halbert et al. (2000; RCT**+**) reported decreased mental well-being in response to 20 minutes of individual advice on physical activity by an exercise specialist in general practice in Australia. Thus the advice was dominated by the control group to whom no advice was given. |

# Glossary of specialist terms

| Term | Definition |
|---|---|
| **Bias** | Any process in the collection, analysis, interpretation, publication or review of data or studies that can lead to conclusions that deviate systematically from the truth. |
| **Cluster RCT** | RCT in which the unit of randomisation is a cluster of participants, e.g. a class, practice or Primary Care Trust |
| **Concurrent validity** | Concurrent validity is demonstrated where a test correlates well with a measure that has previously been validated. The two measures may be for the same or related constructs. |
| **Confidence interval** | An interval around a statistical estimate to show where the true parameter lies with specified probability or 'confidence'. |
| **Controlled Before and After study (CBAS)** | Intervention & control groups are defined and data collected before & after the intervention is administered. CBASs differ from controlled trials in that participants are not allocated to intervention or control groups, but an opportunistic control group is used. |
| **Cost-effectiveness analysis** | The consequences of the alternatives to particular interventions are measured in natural units, such as years of life gained. The consequences are not given a monetary value. |
| **Effect size** | Magnitude of the effect of an intervention or a relationship between variables, calculated as the ratio of the net effect divided by the population standard deviation of the relevant outcome measure. Since this index is independent of sample size, unlike statistical tests of significance, it is useful in meta-analysis. In this review we describe effect sizes less than 0.2 as trivial, those between 0.2 & 0.4 as small, those between 0.4 & 0.6 as 'small to moderate', those between 0.6 & 0.8 as moderate, those between 0.8 & 1.0 as 'moderate to large' and those greater than 1.0 as large. |
| **External validity (also known as generalisability)** | A study is externally valid, or generalisable, if it yields unbiased inferences about a specified target population beyond the subjects in the study (Last, 2001) |
| **Forest plot** | Common method of displaying the results from a meta-analysis. The results of each study are displayed graphically as squares centred on each study's point estimate of the intervention effect with horizontal lines representing confidence intervals (usually 95%) for that effect. |

| | |
|---|---|
| **Heterogeneity** | Differences in study design or methods, or in the characteristics or distributions of populations to be compared, rendering comparison invalid. |
| **Homogeneity** | In contrast to the previous term, this describes consistency in study design or methods, or in the characteristics or distributions of populations to be compared, rendering comparison valid. |
| **Intention-to-treat analysis** | Method of data analysis in which participants are analysed in the group to which they were allocated regardless of whether they complied with their allocated intervention or treatment. |
| **Internal consistency** | Internal consistency is an estimate of how much a measure is based on systematic experimental technique, so that reliable inferences about cause-consequence relations may be made |
| **Internal validity** | A study is internally valid if it yields unbiased comparisons of cases and controls within the study (or intervention and control groups) apart from sampling error (after Last, 2001) |
| **Meta-analysis (MA)** | A mathematical procedure that combines quantitative evidence from a number of different studies, facilitating combination of these studies and comparison between them. |
| **Mixed methods (MM)** | The combination of two or more distinct research methods, typically quantitative & qualitative, to validate findings through triangulation, i.e. showing consistent results from the independent sources. |
| **Non-Randomised Controlled Trial (NCT)** | Trials in which individuals are allocated between intervention and control groups but the allocation is not randomised (for example alternate allocation). |
| **Qigong** | The form of traditional Chinese medicine that coordinates breathing patterns with physical postures and body movements. Though it is mostly taught for health maintenance, some teach it as a therapy. |
| **Randomised Controlled Trial (RCT)** | Individuals or defined groups of individuals are randomised to either an intervention or a control group. If well implemented randomisation should ensure that intervention and control groups only differ in their exposure to treatment. |
| **Reliability** | The consistency of measurements or measuring instruments. Reliability does not imply validity, because a reliable measure can be consistent without measuring what it is supposed to measure. |
| **Systematic review (SR)** | Method of finding & selecting primary studies relating to a defined topic, and appraising & synthesising the resulting evidence. |

| | |
|---|---|
| **Tai Chi** | The slow motion routines practiced every morning in parks around the world, notably in China, to promote health and longevity. |
| **Uncontrolled Before and After study (UBAS)** | Intervention group is defined and data collected before & after the intervention is administered. UBASs differ from CBASs in that there is no control group. |
| **Validity** | Achieved when the measuring instrument (e.g. questionnaire) measures what it is designed to measure. |

## Measures of mental well-being used by studies reported in this review

| Measure | Description |
|---|---|
| Affect Balance Scale (Bradburn, 1969) | A 10-item rating scale containing five statements reflecting positive feelings and five statements reflecting negative feelings. Administered to determine overall psychological well-being at a given point in time. Positive affect questions receive a rating of 1 for yes and 0 for no. A Positive Affect Scale score is obtained by summing ratings for the five positive affect questions. Scores range from 0 to 5. A Negative Affect Scale score is obtained by summing the ratings for the five negative affect questions. Scores range from 0 to 5. The Affect Balance Scale score is computed by subtracting Negative Affect Scale scores from Positive Affect Scale scores and adding a constant of 5 to avoid negative scores. Scores range from 0 (lowest affect balance) to 10 (highest affect balance). Satisfactory levels of reliability have been reported ranging from 0.47 to 0.73 for the positive scale and 0.48 to 0.73 for the negative scale (Cherlin & Reeder, 1975; Warr, 1978).<br>Barnicle & Midden (2003) used a derived five-point scale (strongly disagree, disagree, neutral, agree and strongly agree). This gave a score range from -20 (lowest level of psychological well-being) to +20 (highest psychological well-being). A score of 0 indicates neutral psychological well-being. This scale was found to be reliable in this population over a 5-7 day test-retest period with Pearson correlation = 0.72. Thus the adjustment of the measure did not compromise reliability. |
| Apathy Scale (Starkstein et al., 1995) | The scale consisted of 14 headings, with points between 0 & 3 allotted to each. Higher scores reflect apathetic mood. The scale was validated for use in stroke, Parkinson's and Alzheimer's disease. |
| Campbell's index of well-being (Campbell et al., 1976) | Self-report measure with multiple choice items ranging from 1 to 7. The index is the sum of 2 measurements: the average score on an Index of General Affect (8 items on semantic differential scales); and a single-item assessment of life satisfaction. The measure is reliable and valid: the index has a Cronbach's alpha of 0.89, and a correlation of 0.55 with the life satisfaction question (Robinson, 1981). |

| | |
|---|---|
| Coopersmith Self-Esteem Inventory (1981) | The scale ranges from 25 to 50. Reliability and validity have been reported as adequate in several studies. |
| Emotional well-being scale (Campbell & Aday, 2001). | Campbell & Aday (2001) developed several measures to explore health and well-being, including a 7-item emotional well-being subscale that asked participants whether they felt the intervention under evaluation had brought about any changes in this dimension. Life satisfaction and mental health were also assessed in this study, but there is no reference to how scales were constructed. Cronbach's alpha of 0.86 indicated strong internal consistency. The scale appears to require further validation. |
| Emotional well-being scale (Hermans & Tak-van der ven, 1973) | A measure of the relative proportion of positive to negative feelings experienced by an individual. Positive and negative feelings are treated as independent of each other. Respondents agree or disagree with positive and negative statements. Goldstein et al. (1997) selected 10 items, 5 positive and 5 negative, from the 36-item scale by Hermans & Tak-van der Ven (1973). There is no mention of validating this alternative version. |
| Exercise induced feeling inventory (Gauvin & Rejeski, 1993) | A 12-item inventory that assesses 4 dimensions: positive engagement (enthusiastic, happy, and upbeat); revitalisation (refreshed, energetic, and revived); tranquillity (calm, relaxed and peaceful); and physical exhaustion (fatigued, tired and worn-out). On the 5-point scale subjects indicate how strongly they experience each feeling state immediately after one period of exercise. Anchors were: 0 = do not feel and 4 = feel very strongly. Internal consistency exceeded 0.70 for each subscale (McAuley & Courneya, 1994). In Matsouka et al. (2003) Cronbach's alpha exceeded 0.85. The subscales have good internal consistency, share expected variance with related constructs, are sensitive to different exercise settings and responsive to different social contexts. |
| General Health Questionnaire (GHQ-28; Goldberg and Hillier, 1979) | A 28-item version of the GHQ commonly used as a measure of psychological well-being (low scores, while poor health scores high). There has been extensive testing of validity, reliability, and sensitivity (Bowling, 1995). It has been widely used and found to be acceptable by participants over 65 years. The scale has four subscales: A – somatic symptoms; B – anxiety & insomnia; C – social dysfunction; D – severe depression. |

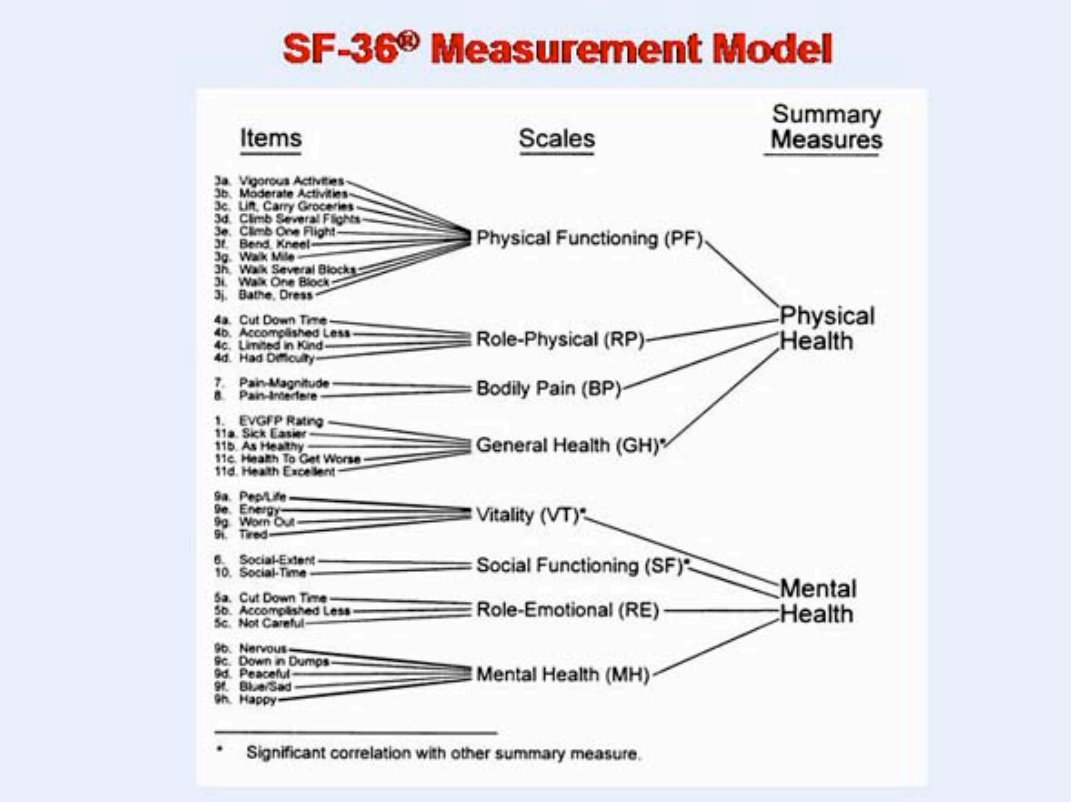
| | |
|---|---|
| General self efficacy (Sherer et al., 1982) | A 16-item scale that measures self-efficacy expectations across a variety of situations. The scale consists of two components – initiation & persistence of behaviour, and efficacy in the face of adversity. |
| LEIPAD II Short version (LEIPAD SV; De Leo et al., 1998) | The measure is designed to gauge subjective perceptions of quality of life in the elderly. It comprises 25 items in 6 scales – cognitive functioning scale (CFS), depression & anxiety scale (DAS), life satisfaction scale (LSS), physical function scale (PFS), self-care scale (SCS) and social functioning scale (SFS). Each item is scored from 0 (best) to 3 (worst). Validity and reliability of this short version of the scale are not reported. |
| Life Attitude Profile (LAP-R; Reker & Peacock, 1981) | A measure comprising 6 sub-scales (choice responsibility; coherence; death acceptance; existential vacuum; goal seeking, life purpose) and two composite scales (personal meaning index [PMI) and existential transcendence [ET]). Each of the 8 items per subscale is rated on a 7-point Likert-type scale (1-7) from “strongly disagree” (1) to “strongly agree” (7). |
| Life Satisfaction Index-A (LSI-A; Neugarten et al., 1961). | A 20 item self-report instrument used to measure subjective well-being (reflecting satisfaction with life) among individuals aged 65 or over. Respondents are asked to either agree, disagree or express neutrality on each items. Twelve items are positively worded, and eight are negatively worded. Each agreement with a positively worded item receives 1 point and each disagreement with a negatively worded item also receives 1 point. Other responses are scored 0. Totalling the number of points creates a score ranging from 0 to 20, with higher values indicating greater life satisfaction. Neugarten et al. (1961) report a mean score of 12.4 (SD 4.4). Reported internal consistency ranged from 0.73 (Hooker & Ventis, 1984) to 0.84 (Wolk & Kurtz, 1975). Concurrent validity is also reported. This measure has been extensively used in gerontological research. |
| Life Satisfaction Index-Z (Wood et al, 1969) | A global measure of past, present and future states, this scale was developed for use with older populations in different ethnic groups. It is commonly used to measure well-being in gerontology research and is considered to indicate successful aging (Bowling, 1991). Respondents agree or disagree with each of 13 items about satisfaction with life scoring 0, 1 or 2, yielding a total score ranging between 0 & 26. In initial scale development, split-half reliability was 0.79. |

| | |
|---|---|
| Life Satisfaction Ladder Scale (Cantril, 1965) | A 10-item vertical self-report scale with item responses from 1 = very dissatisfied to 10 = very satisfied. Scale scores therefore range from 10 to 100, with higher scores indicating higher satisfaction. Adequate reliability and validity have been reported. |
| Life Satisfaction Questionnaire (LSQ; original ref. not cited) | Carlsson et al. (1999) concluded that the LSQ, developed for women with breast cancer, has acceptable validity and reliability. Respondents indicate their degree of satisfaction with their finances, health, religious experience and social relationships. Neither users, Powers & Wisocki (1997) or Carlsson, present information on validity or reliability. |
| Life Satisfaction Visual Analogue Scale | Apparently created by Dungan et al. (1996) for their own use. Unfortunately they give no details of content, scoring or use. |
| Locus of Control Scale (Levenson, 1974). | A 24-item self-report measure of internal orientation and two types of external orientation – belief in chance and belief in control by others. The scale shows adequate reliability and validity, and has been used effectively with older adults (Shewchuk et al. 1990). |
| Loneliness Scale (De Jong et al., 1999). | The self-report scale has 5 positive and 6 negative items assessing sense of belonging and discrepancies in desired relationships. Item scores range from 0 (not lonely) to 11 (extremely lonely); scores above 3 show loneliness, while those above 9 show extreme loneliness. The scale has been used in several surveys and is reliable and valid with older persons. |
| Memorial University of Newfoundland Scale of Happiness (MUNSH; Kozma & Stones, 1980) | Self-report measure with 4 subscales – positive affect (PA), negative affect (NA), positive experiences (PE) and negative experiences (NE). There are 5 binary items for PA, 5 for NA, 7 for PE, and 7 for NE. The total score comes from the formula (PA-NA) + (PE-NE). As this can give a negative score, Elavsky et al. (2005) added 24 to the total score, thus giving a range from 0 to 48. The scale has been validated in several settings. Elavsky et al. (2005) reported internal consistency as more than 0.75. |
| Mental Health Index (MHI-5; McCabe et al., 1996) | This is derived adapted from five mental health items of the SF-36, leading to a summary score between from 0 & 25. Each item asks respondents about mood over the last 4 weeks and how long they have felt that way (from none of the time to all). There is no information on reliability or validity in the study which employed this measure (Clark et al., 2003). |

| | |
|---|---|
| Morale & Life Satisfaction Scale (Clark & Anderson, 1967) | A 45-item measure covering 8 dimensions: depression & satisfaction; equanimity; negative aspects of age; physical condition; positive aspects of age; social accessibility; social alienation; & will to live. Evidence of validity & reliability reported by Clark & Anderson (1967) and Pierce & Clark (1973). |
| Perceived Control of Life Situations (Eizenman et al., 1997). | Self-report measure with eight items answered on a 4-point scale from 1 = agree strongly to 4 = disagree strongly. Scores therefore range from 8-32 with higher scores indicating greater control. White et al. (2002) reported adequate reliability and validity. |
| Perceived well-being scale (Reker & Wong, 1984) | The scale includes 14 items about physical and mental function. No information available about validation. |
| Perceived well being scale – revised (PWB-R) | Measures perceived physical & emotional well being, using 16 items – 8 psychological & 8 physical, each measured on 7-point Likert scale. Items are randomly ordered and varied between positive and negative to control for response set bias. Cronbach's alpha reported between 0.79 & 0.85. |
| Philadelphia Geriatric Center (PGC) Morale Scale (Lawton, 1975) | This 17-item revised scale is a measure of general well-being and positive future outlook. There are two widely used alternative versions – interview, or self-report questionnaire. High morale responses are scored as a 1; low morale responses are scored as 0, so scores range from 0-17, 0 being low morale, 17 being high morale. Three factors represent morale – agitation, attitude toward own aging, and lonely dissatisfaction. Valid and reliable |
| Philadelphia Geriatric Center (PGC) Morale Scale -- modified | A measure of life satisfaction. The resulting scale ranges from 0 to 11, with higher values indicating greater satisfaction with life. The alpha reliability of this measure is reported as 0.78 by Rabiner et al. (2003). |
| Positive and negative affect schedule (PANAS; Watson et al., 1988). | A 20 item self-report measure with 10 items reflecting positive affect (PA), and 10 negative affect (NA). High PA reflects high energy, full concentration and pleasurable engagement. Low PA is characterised by sadness and lethargy. Low NA reveals a state of calmness and serenity. Internal consistency is 0.80 for PA and 0.84 for NA (Martina & Stevens, 2006). |

| | |
|---|---|
| Profile of Mood States (POMS; McNair et al., 1971) | Self-report measure assessing general mood over "the past week including today". Respondents, provided with a list of 65 adjectives describing mood states, indicate their agreement with adjectives on a 5-point Likert scale ranging from 0 (not at all) to 4 (extremely). This yields 6 sub-scores – anger & hostility, confusion & bewilderment, depression & dejection, fatigue & inertia, tension & anxiety and vigour & activity (with differing numbers of items per subscale). Total mood disturbance is calculated by summing the 5 negative scores and subtracting the one positive score – vigour & activity (McLafferty et al., 2004). Most respondents complete the POMS in 3 to 5 minutes (McNair et al., 1992). Across subscales internal consistency ranges from 0.87 to 0.95, and test-retest coefficients from 0.65 to 0.74. Concurrent validity has also been shown. |
| POMS short form (POMS-SF; Curren et al., 1995) | Curren et al (1995) reduced POMS (previous row) to 30 items – 5 for each of the 6 subscales. All internal consistencies remained above 0.7. |
| POMS modified version (Jette et al., 1996) | Another shortened version of the POMS, similar to the 30-item POMS-SF, but with the 5 items with highest item-total correlations (as reported in the test manual) selected for 5 of the subscales and 8 such 'best' items for the depression subscale. Though methodologically slightly better than POMS-SF, internal consistencies were similar, i.e. greater than 0.7. |
| POMS – Korean version (Shin, 1999) | The POMS was adapted for Korean elders through cultural verification and psychometric evaluation (Shin, 1996). The new instrument comprises 3 factors – anxiety & depression (21 items), vigour (8 items) and anger (5 items). Cronbach's alpha has risen to 0.95 – very high if not too high. |
| Psychological well-being (Becker, 1989) | Self-report personality questionnaire with 3 subscales – meaning of life, self-attentiveness & preoccupation (having self-centred thoughts and anxiety & concern about self & the future), and 'complaintlessness'. Validity and reliability are not reported. |
| Psychological well-being (Ryff,1989) | Scale to measure perceived psychological well-being comprising 14 items about personal growth, 3 items about positive relations with others, and 3 about self-acceptance. High internal & test-re-test reliability, and convergent & discriminant validity reported for various age-groups including older adults. |

| | |
|---|---|
| Quality of Life Profile: Senior Version (Renwick et al., 1996) | According to Raphael (1996) the QOLP-SV rates highly in importance, satisfaction and quality of life scores. Nevertheless neither paper reports on content or scoring! |
| Rosenberg Self-Esteem Scale (Rosenberg, 1965) | This scale comprises 10 self-reported 4-point Likert scales labelled "strongly disagree", "agree", "disagree" & "strongly disagree". High total score shows high self-esteem. Reliability over time > 0.83; reproducibility coefficient = 0.92; scaleability coefficient = 0.72. Validity shown by significant negative correlations between self-esteem and clinical ratings of depression. |
| Satisfaction With Life Scale (SWLS) (Deiner et al., 1985) | Measure with 1 item for each of 5 life domains, all rated on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). Higher scores show greater level of satisfaction with life. Reports of good psychometric properties including internal consistency include those from Diener et al. (1985) [also test-retest reliability], McAuley et al. (2000), Macfarlane et al. 2005 [also test-re-test reliability and construct validity] and Martina & Stevens (2006). |
| Self-rated mood (Tamake et al., 1999) | Single-item self-report measure of mood. No other details are provided. |
| Sense of Mastery Scale (Pearlin & Schooler, 1978) | Self–report scale with 7 items assessing sense of control over one's life. (Froelicher et al. 2004) reported construct validity and internal consistency in a sample of women with mean age = 61. |
| Short Form-12 (SF-12) (Ware et al., 1995) | Widely used self-report measure comprising 12 items from SF-36 (following row) yielding separate scores for physical and mental health. One can also derive a single health utility score for economic analysis. Scores are transformed to lie between 0 & 100, with higher scores showing better health. Reliability & validity have been established in numerous studies, including Ware et al. (1996) and Ware et al. (1998). |

| | |
|---|---|
| Short Form 36 (SF-36; Ware & Sherbourne,1992), previously known as Medical Outcomes Study Instrument (MOSI) | Ubiquitous, multi-purpose, self-reported, generic health survey, comprising 8 scales derived from 36 questions about functional health & well-being, and yielding psychometrically-based physical & mental health summary scores and a preference-based health utility index. Scores are transformed to lie between 0 & 100 with high scores showing better health. There are many reports of good psychometric properties including internal consistency, test-retest reliability, and all types of validity (content, concurrent, predictive, criterion & construct) including Ware et al. (1993), McHorney et al. (1994), Ware et al. (1994), Tsai et al. (1997) and Schechtman & Ory (2001).<br>  |
| Social Production Function Index Level Scale (SPF-IL; Nieboer et al, 2005) | Self-report measure of affective and cognitive components of well-being. There are 15 items – 3 in each of the 5 subscales representing the dimensions of well-being from the SPF theory – affection, behavioural confirmation, comfort, status and stimulation. Nieboer et al. (2005) report that validity and reliability have been extensively tested and that SPF-IL has other good psychometric properties. |

| | |
|---|---|
| Sources of Meaning Profile-Modified (Reker, 1988) | Measures the sources and degree of personal meaning in one's life, by using 16 7-point items to tap activities, commitments, and pursuits. |
| Subjective Quality of Life Profile (Gerin et al. 1992). | This tool has 4 categories from which investigators can choose items relevant to their study, including functional life (concerning both physical and mental health), social life (relationships, social roles & interest in the exterior world) & spiritual life (capacity to have aesthetic or religious experiences, to meditate and to reflect). The last category asks subjects to evaluate their personal programme, to rate its relevance, and describe how it was experienced. Item scores lie between -2 (perceived to be highly unsatisfactory) and a maximum of 2 (perceived to be highly satisfactory). Importance attached to items has 3 levels – without importance (0) up to very important (2). |
| Subjective Satisfaction & Refreshment Scale (Hirawayka et al., 2001) | Responses to variations on the question "To what extent do you feel satisfied and refreshed in daily life?" are assessed on a 4 point scale (3 = strongly; 2 = moderately; 1= slightly, 0 = not at all). Details of validity & reliability are not available. |
| Subjective scale of well-being for older persons (SSWO; Tempelman, 1987) | Paw et al. (2007) used the Dutch version of the SSWO, a self-report measure with 30 items divided into 5 subscales – health (5 items), contacts (5 items), morale (6 items), optimism (7 items) and self-respect (7 items). The total score measures general well-being. The test-retest reliability coefficient was 0.85. |
| UCLA Loneliness | This scale is often considered the gold standard of loneliness scales. It assesses subjective feelings of loneliness or social isolation on a |

| | |
|---|---|
| Scale (Russell et al., 1980). | 20-item scale with scores ranging from 20 to 80. McAuley et al. (2000) & Brown et al. (2004) both reported internal consistency of more than 0.9, & Brown added test-retest reliability of 0.73. Russell (1996) established convergent validity through highly significant correlations of 0.65 with the NYU Loneliness Scale and 0.72 with the Differential Loneliness Scale |

| | |
|---|---|
| World Health Organisation Quality of Life questionnaire (WHO-QOL-100; WHO Quality of Life Group, 1998) | The WHO-QOL-100 was developed simultaneously in 15 international centres through item creation, focus groups, pilot tests and field tests. An initial pool of 1000 questions was reduced to 100 items, grouped into one generic facet on quality of life & health perceptions, and 24 specific facets, originally grouped into 6 domains – Environment, Independence, Physical, Psychological, Social Relationships & Spirituality. Most participants complete the survey themselves, though a small number with literacy problems get a structured interview. Items are scored on a 5-point Likert scale specifying only anchor points (e.g. 'never' & 'always'). Test-retest reliability & internal consistency are both good, especially in Britain (Skevington, 1999). The WHO-QOL-100 is adept at identifying facets of quality of life which are cross-culturally important (WHO-QOL Group 1998, Power et al. 1999). Confirmatory factor analysis showed that the 6-domain model was not as good a fit as a 4-domain model combining Independence with Physical, and Spirituality with Psychological. Scores discriminate well between sick and well people and concur with reported health status. |
| Worry Questionnaire (Wisocki, 1988) | Respondents estimate the amount of worry & associated physical feelings they experience in the domains of health, finance and social relationships. Wisocki (1988) presents no information on reliability or validity. |

## Abbreviations

| | |
|---|---|
| ABS | Affect Balance Scale |
| ADL | Activity of Daily Living |
| ANOVA | Analysis of Variance |
| ANCOVA | Analysis of Covariance |
| AOR | Adjusted Odds Ratio |
| APV | Analysis of Partial Variance |
| C | Control |
| CBAS | Controlled Before-&-After study (two groups of participants) |
| CI | Confidence Interval |
| CFI | Comparative Fit Index |
| GHQ | General Health Questionnaire |
| HUI | Health Utilities Index |
| HR | Heart rate |
| HRQoL | Health-related quality of life |
| I or IV | Intervention |
| LSI | Life Satisfaction Index |
| M | Mean |
| MA | Meta-analysis |
| MANOVA | Multivariate Analysis of Variance |
| MANCOVA | Multivariate Analysis of Covariance |
| MCS | Mental Component Score (of SF-36) |
| MI | Mental Illness |
| MM | Mixed methods |
| MMSE | Mini Mental Status Examination |
| MOS | Medical Outcome Study |
| N | Number |
| NCT | Non-randomised controlled trial |
| NSF | National Service Framework |
| OR | Odds Ratio |
| OT | Occupational Therapy ***or*** Occupational Therapist |
| PAQ | Physical Activity Questionnaire |

| | |
|---|---|
| PCS | Physical Component Score (of SF-36) |
| POMS | Profile of Mood States |
| PWB | Psychological Well-being |
| Q | Qualitative study |
| QALY | Quality adjusted life year |
| QoL | Quality of life |
| QWB | Quality of Well-being |
| RCT | Randomised controlled trial |
| RPE | Rating of Perceived Exertion |
| SD | Standard Deviation |
| SE | Standard Error |
| SEIQoL | Schedule for Evaluating Individual Quality of Life (validated questionnaire) |
| SES | Socio-economic Status |
| SWB | Subjective Well-being |
| T | Time |
| UBAS | Uncontrolled Before-&-After Study (single group of participants) |
| wk | Week |

# 1 Introduction

## 1.1 Aims of the review

1.1.1 To identify and review all relevant evidence about public health interventions to promote mental well-being in older people aged 65 and over.

1.1.2 To identify and review data on the costs and cost-effectiveness of public health interventions to promote mental well-being in older people aged 65 and over.

1.1.3 To highlight gaps in the evidence base and make recommendations for further research.

## 1.2 Target audience and structure

The work is aimed at professionals and practitioners working in the NHS, other public sector organisations, the private sector and the voluntary and community sectors. It is also relevant to carers and family members who have direct or indirect responsibility for the care and support of older people.

The report therefore adopts the traditional structure – introduction (Chapter 1), methods (Chapter 2), results and discussion (Chapter 6). To expedite the work we divided it into three tasks – review of effectiveness (Chapter 3), review of cost-effectiveness (Chapter 4) and economic modelling of cost-effectiveness (Chapter 5). As Chapter 3 reports on 97 included papers, while Chapter 4 reports on only two, Chapter 5 shows how studies of effectiveness can be extended after publication to throw light on cost-effectiveness, provided they are reported assiduously.

## 1.3 Research Questions

The review addresses one main research question:
What are the most effective and cost-effective ways for primary and residential care services to promote the mental wellbeing of older people?

Table 1 below uses the Population Intervention Comparison Outcome (PICO) format to show how the main research question addresses specific issues within the types of interventions – for all older people and for sub-groups of this population.
Within the main question the review addresses six sub-questions:

1.3.1 What is the frequency and duration of an effective intervention?

1.3.2 What are the significant features of an effective intervener?

1.3.3 Are interventions that engage older people in their design and delivery more effective than those that do not?

1.3.4 Are interventions that engage immediate family members or carers more effective than those without such engagement?

1.3.5 Does the intervention lead to any adverse or unintended effects?

1.3.6 What are the barriers to and facilitators of effective implementation?

### 1.4 Background – setting the context

Population ageing is emerging as a worldwide trend, reflecting economic development, improvements in education and health care, increases in life expectancy and falls in fertility. The oldest old (80 and over) are the fastest growing group in many nations (Kinsella & Velcoff, 2001). This demographic change emerged first in Europe (Scharf et al., 2003). By 2021 this oldest age group will constitute almost 5 per cent of the population of the UK (Office for National Statistics, 1999). The UK is one of the world's 25 oldest countries, with 20.4% being aged 60 or over (Kinsella & Velcoff, 2001). By 2025 the number of people over 65 in the UK will exceed the number under 16 by 1.6 million (Office for National Statistics, 2003).

Population ageing presents many challenges for government policies and the health and social services, particularly the perceived increasing burden of pensions, and health and social care provision. Ageing can be accompanied by biological changes that increase the risk of illness, disability and death (Office for National Statistics, 1999). Although life expectancy has increased and mortality decreased, it is not clear whether there have been concomitant improvements in morbidity in older age (Office for National Statistics, 1999).