SF-36v2 norms for New Zealanders aged 55–69 years
Christine Stephens, Fiona Alpass, Merel Baars, Andy Towers, Brendan Stevenson

Abstract
Owing to an ageing population there is growing interest in research to improve the health of older New Zealanders. To facilitate the use of the internationally used SF-36 (version 2) measure of health and quality of life for this work in New Zealand we provide norms and comparative data from the first wave in a longitudinal study of a representative sample of New Zealanders aged 55–69 years. The use of the normative data from this study will facilitate comparisons of results from small clinical samples of older people with the general New Zealand population and international populations. The norms are also available for use in calculating summary physical and mental health summary scores for data from clinical trials and national surveys.

The first ‘baby-boomers’ (a demographic bulge born between 1946 and 1960) will reach 65 in 2011 and the movement of this group into retirement and older age will present new issues in health and health care research. Because of this ageing population there is growing interest in studies of the health and well-being of older New Zealanders.

To facilitate the use of the Medical Outcomes Study Short Form questionnaire version 2 (SF-36v2) for this work in New Zealand we provide norms and comparative data from the first data wave in a longitudinal study of a representative sample of New Zealanders aged 55–69 years.

The SF-36 Health Survey is a widely used reliable self-reported measure of generic health status. Designed for use in clinical practice and research, health policy evaluations, and general population surveys, it has been widely applied to assess changes and differences in mental and physical health in medical intervention studies and has been shown to be valid for use in older people. Because of improved wording and lay-out, and the changed number of responses in some questions, the second version of this measure minimises ambiguity and allows for greater comparability between cultural adaptations and translations.

At the same time scores can still be directly compared to SF-36 version 1. Scores from eight sub-scales measuring different aspects of health may be standardised, and also summarised to provide scores for two underlying factors of physical and mental health. When used as a baseline and outcome measure in smaller samples Ware and colleagues strongly recommend that these measures are standardised using population norms to provide the basis for national and international comparisons, and also improve the validity of observations within the sample under study.

The authors of the scale also stress the importance of the norms used being as up to date as possible, and Scott et al, point to the variations in scores across cultures so that local norms are required.
In New Zealand, norms on the SF-36 (version 1) sub-scales for the general population from the New Zealand national health survey (1996/7) have been published by Scott et al., who also assessed the appropriateness of the SF-36 for the New Zealand population. However, more recent normative data using SF-36v2, and norms for particular population groups, such as older people, are now required.

The Health Work and Retirement study (HWR) is a longitudinal study of older New Zealanders funded by the Health Research Council of New Zealand. The first wave of data collection for the HWR involved a representative sample of 6662 New Zealanders aged between 55 and 70 years. In this paper we provide normative data from this sample on SF-36v2 sub-scales and Physical and Mental Health summary scores for the whole sample as well as those for 5 year age groups, ethnicity and gender. Sub-group comparisons are given for physical health summary scores and mental health summary scores.

Method

Sampling method—The population of interest for the study was New Zealanders aged 55 to 70 who are generally in the later stages of work life or early stages of retirement (there is no legal age of retirement in New Zealand, although a universal superannuation scheme provides a pension from age 65). There are approximately 814,464 New Zealanders aged 50 to 69, with 65790 (8%) of those identifying as Māori, the indigenous people of Aotearoa/New Zealand. The New Zealand Electoral Roll was the source for sample selection. Registration on the roll is mandatory for all citizens eligible to vote in government elections and in 2007, 96% of all eligible New Zealanders were registered. Equal probability random sampling procedures were used to select two independent samples to represent the general population (N=5264) and the Māori population (N=7781).

Māori were over-sampled for this study using the Māori descent indicator on the general and Māori electoral rolls. This was done to maximise participant recruitment and provide sufficient numbers for statistical analysis in later data collection waves. In total 13,045 55–70 year olds were surveyed. The total response rate (after exclusions, e.g. unable to be contacted, deceased, or institutionalised, N=551) was 53.32% (N=6662). The response rate for the general population was 61% and for the Māori sample, 48%. Both samples have similar gender and age group proportions to the general and Māori populations respectively when compared to the 55–69 year-old population from the 2006 Census.

Because of the over-sampling, the total sample was weighted to represent the New Zealand population. A post-stratified weighting variable according to primary ethnicity was applied to the present analyses based on the population estimates from the 2006 Census for the 55–69 year old age group. In the weighted sample age was well distributed with 2422 (40.5%) aged 55–59, 1905 (31.9%) aged 60–64, and 1651 (27.6%) aged 65–69. There were 219 missing and 513 who were outside the 55–69 age range and accordingly not included in the present analysis (N=5978). The numbers (rounded) for ethnicity were: European descent (4153), Māori (461), Pasifika (199), Asian (265), other (864), and missing (i.e. did not report ethnicity; 40). In regard to gender, 2856 (47.8%) were males and 3103 (51.9%) females, (22 missing).

Data collection method—The postal survey used multiple contact points to maximise participation.

1. A brief pre-notice letter was sent to inform potential participants about their selection and the questionnaire study.
2. One week later, a questionnaire which included the SF-36 items and other measures for the Health Work and Retirement Longitudinal Study (see http://hwr.massey.ac.nz), a detailed information sheet and a free-post return envelope were sent.
3. At 3 weeks a reminder postcard was sent to the whole sample.
4. At 6 weeks a replacement questionnaire was sent to all non-respondents.
5. At 11 weeks a final postcard was sent to all non-respondents.
These procedures were approved by the Massey University Human Ethics Committee.

Measures and analysis—The SF-36v2 comprises 36 items grouped into eight sub-scales each examining a different dimension of health (physical function, role limitations for physical problems, bodily pain, general health perception, vitality, social functioning, role limitations for emotional problems and general mental health). The eight sub-scales of the SF-36v2 were calculated using algorithms provided by the developers. For all measures of the eight SF-36v2 sub-scales, scores were transformed to scales of 0 to 100, with higher scores reflecting better health and fewer role limitations.

In addition, the 8 sub-scales were combined using principle components (orthogonally rotated) derived coefficients to form two components assessing physical (PCS) and mental health (MCS). Each physical and mental component summary score was standardised (using standard deviations from the present representative study) with lower scores implying poorer health.

Cronbach’s alpha coefficients were used to test internal consistency of the 8 SF-36v2 subscales and these are reported in Table 1. Cronbach’s alpha exceeded .80 in all but one subscale (social functioning with alpha .67).

Table 1. Cronbach’s alpha internal consistency statistics for the eight sub-scales of the SF-36v2

<table>
<thead>
<tr>
<th>Sub-scale</th>
<th>No. of items</th>
<th>N</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical function</td>
<td>10</td>
<td>6478</td>
<td>0.92</td>
</tr>
<tr>
<td>Role – physical</td>
<td>4</td>
<td>6491</td>
<td>0.95</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>2</td>
<td>6521</td>
<td>0.90</td>
</tr>
<tr>
<td>General health</td>
<td>5</td>
<td>6457</td>
<td>0.83</td>
</tr>
<tr>
<td>Vitality</td>
<td>4</td>
<td>6498</td>
<td>0.85</td>
</tr>
<tr>
<td>Social function</td>
<td>2</td>
<td>6653</td>
<td>0.67</td>
</tr>
<tr>
<td>Role emotional</td>
<td>3</td>
<td>6424</td>
<td>0.92</td>
</tr>
<tr>
<td>Mental health</td>
<td>5</td>
<td>6491</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Mean differences between groups were tested using one-way ANOVA. Owing to different group sizes and unequal variances the Welch statistic was used to test equality of means, and if significant differences were found, post hoc Tamhane’s t tests (p<0.05), assuming a non-normal distribution and unequal variances, were performed.

Each variable had some missing data so Ns for every analysis vary and are reported in the results.

Results

Table 2 provides means and standard deviations for each of the eight sub-scales of the SF-36v2 in the whole sample (55–69 years of age). Means and standard deviations are also provided across three different age groups: 55–59 years, 60–64 year and 65–69 years.

PCS and MCS scores for this group are also provided in Table 2. The Welch test of equality of means revealed that the three age groups differed significantly on PCS scores ($F(2, 3372)=44.83, p<.001$). Post hoc comparisons using the Tamhane test indicated significant differences between all three group means, with PCS scores decreasing as age increased. One-way ANOVA also showed a significant difference between age groups for MCS scores ($F(2, 3494)=15.16, p<0.001$). Post hoc analysis revealed a significant difference between the 55–59, and 60–64, and the oldest (65–69) sub-group, with the oldest sub-group scoring higher than the two younger groups.
Thus, mean PCS and MCS scores across the three age groups showed a decrease in physical health and an increase in mental health scores as people age (see Figure 1).

**Figure 1. PCS and MCS mean scores across three age groups**

![Graph showing PCS and MCS mean scores across three age groups](image)

*Note:* Cases weighted for ethnicity.

Mean scores and standard deviations broken down by three ethnic groups (i.e. New Zealand European, Māori, and Pasifika) and across the three different age groups are shown in Table 3. The three ethnic groups were found to have significantly different scores on both PCS, $F(2, 331.50)=67.22$, $p<.001$, and MCS $F(2, 325.70)=30.17$, $p<.001$. Post hoc comparisons indicated that the mean PCS score for New Zealand Europeans was significantly different from Māori and Pasifika Peoples. Māori also differed significantly from Pasifika Peoples. The same pattern was observed for the MCS summary scores. Here, the mean MCS score for New Zealand Europeans significantly differed from Māori and Pasifika Peoples. However, Māori and Pasifika Peoples did not differ significantly on the Mental Health summary score.
Table 2. SF-36v2 Mean scores (SD) on eight sub-scales for whole sample (with 95% confidence intervals [CI]) and across three age groups

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>55–69</td>
<td>5536</td>
<td>79.32</td>
<td>81.49</td>
<td>71.12</td>
<td>71.12</td>
<td>65.57</td>
<td>85.22</td>
<td>87.37</td>
<td>80.71</td>
<td>50.00</td>
<td>50.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(22.81)</td>
<td>(24.17)</td>
<td>(23.75)</td>
<td>(21.63)</td>
<td>(19.64)</td>
<td>(22.48)</td>
<td>(20.63)</td>
<td>(15.78)</td>
<td>(9.56)</td>
<td>(9.44)</td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td>78.73–79.91</td>
<td>80.87–82.11</td>
<td>70.51–71.73</td>
<td>70.56–71.68</td>
<td>65.07–66.07</td>
<td>84.65–85.80</td>
<td>86.84–87.90</td>
<td>80.31–81.11</td>
<td>49.75–49.85</td>
<td></td>
</tr>
<tr>
<td>55–59</td>
<td>2296</td>
<td>81.50</td>
<td>84.48</td>
<td>72.10</td>
<td>72.02</td>
<td>65.41</td>
<td>85.33</td>
<td>87.77</td>
<td>79.93</td>
<td>51.22</td>
<td>49.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(22.52)</td>
<td>(22.94)</td>
<td>(23.33)</td>
<td>(21.66)</td>
<td>(19.69)</td>
<td>(22.32)</td>
<td>(20.30)</td>
<td>(16.13)</td>
<td>(9.15)</td>
<td>(9.65)</td>
</tr>
<tr>
<td>60–64</td>
<td>1743</td>
<td>79.93</td>
<td>81.69</td>
<td>70.76</td>
<td>70.52</td>
<td>65.32</td>
<td>84.97</td>
<td>87.74</td>
<td>80.23</td>
<td>50.00</td>
<td>49.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(21.78)</td>
<td>(23.87)</td>
<td>(23.92)</td>
<td>(21.60)</td>
<td>(19.65)</td>
<td>(22.55)</td>
<td>(20.54)</td>
<td>(16.16)</td>
<td>(9.22)</td>
<td>(9.55)</td>
</tr>
<tr>
<td>65–69</td>
<td>1497</td>
<td>75.33</td>
<td>76.70</td>
<td>70.05</td>
<td>70.42</td>
<td>66.10</td>
<td>85.36</td>
<td>86.33</td>
<td>82.42</td>
<td>48.14</td>
<td>51.17</td>
</tr>
</tbody>
</table>

**Note:** Range 0–100, sample size varies due to missing data; Cases weighted for ethnicity. Group Ns are different from total count because weighted Ns are rounded.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ Eur</td>
<td>55–69</td>
<td>3852</td>
<td>80.42 (22.16)</td>
<td>82.38 (23.80)</td>
<td>71.97 (23.55)</td>
<td>71.98 (21.36)</td>
<td>65.92 (19.65)</td>
<td>86.21 (22.21)</td>
<td>88.79 (19.80)</td>
<td>81.45 (15.39)</td>
<td>50.22 (9.60)</td>
</tr>
<tr>
<td></td>
<td>55–59</td>
<td>1491</td>
<td>83.44 (21.15)</td>
<td>85.79 (22.33)</td>
<td>73.39 (23.24)</td>
<td>73.14 (21.71)</td>
<td>66.32 (19.77)</td>
<td>86.32 (22.17)</td>
<td>89.69 (19.07)</td>
<td>80.94 (15.57)</td>
<td>51.55 (9.18)</td>
</tr>
<tr>
<td></td>
<td>60–64</td>
<td>1293</td>
<td>81.15</td>
<td>83.07 (23.48)</td>
<td>71.84 (23.48)</td>
<td>72.12 (21.00)</td>
<td>65.88 (19.59)</td>
<td>86.54 (22.00)</td>
<td>88.84 (20.07)</td>
<td>81.34 (15.83)</td>
<td>50.44 (9.20)</td>
</tr>
<tr>
<td>65–69</td>
<td>1067</td>
<td>75.42 (23.66)</td>
<td>76.82 (25.12)</td>
<td>70.19 (23.95)</td>
<td>70.22 (21.21)</td>
<td>65.42 (19.56)</td>
<td>85.66 (22.53)</td>
<td>87.46 (20.40)</td>
<td>82.99 (14.58)</td>
<td>48.07 (10.25)</td>
<td>51.33 (8.64)</td>
</tr>
<tr>
<td>Māori</td>
<td>55–69</td>
<td>410</td>
<td>71.91 (26.98)</td>
<td>72.54 (27.64)</td>
<td>66.61 (26.61)</td>
<td>66.74 (23.53)</td>
<td>62.39 (20.54)</td>
<td>75.60 (25.14)</td>
<td>79.03 (25.22)</td>
<td>77.69 (17.22)</td>
<td>47.39 (10.60)</td>
</tr>
<tr>
<td></td>
<td>55–59</td>
<td>174</td>
<td>76.07 (25.89)</td>
<td>77.22 (26.12)</td>
<td>68.03 (26.03)</td>
<td>68.22 (23.29)</td>
<td>62.23 (20.89)</td>
<td>79.74 (25.62)</td>
<td>81.50 (24.02)</td>
<td>77.38 (17.35)</td>
<td>49.00 (10.27)</td>
</tr>
<tr>
<td></td>
<td>60–64</td>
<td>127</td>
<td>70.61 (26.80)</td>
<td>70.89 (28.80)</td>
<td>65.43 (27.40)</td>
<td>65.95 (24.44)</td>
<td>61.82 (20.77)</td>
<td>76.81 (27.15)</td>
<td>77.63 (26.56)</td>
<td>77.24 (17.73)</td>
<td>46.87 (10.79)</td>
</tr>
<tr>
<td></td>
<td>65–69</td>
<td>108</td>
<td>66.95 (28.06)</td>
<td>67.15 (27.60)</td>
<td>65.74 (26.70)</td>
<td>65.33 (22.90)</td>
<td>63.29 (19.83)</td>
<td>78.87 (25.88)</td>
<td>76.77 (25.36)</td>
<td>78.67 (16.52)</td>
<td>45.42 (10.61)</td>
</tr>
<tr>
<td>Pasifika</td>
<td>55–69</td>
<td>155</td>
<td>61.69 (30.05)</td>
<td>61.26 (28.57)</td>
<td>57.83 (25.79)</td>
<td>59.26 (21.14)</td>
<td>63.74 (18.16)</td>
<td>70.75 (26.12)</td>
<td>70.83 (25.79)</td>
<td>73.71 (17.90)</td>
<td>42.51 (8.78)</td>
</tr>
<tr>
<td></td>
<td>55–59</td>
<td>64</td>
<td>58.33 (31.13)</td>
<td>61.84 (27.94)</td>
<td>57.62 (28.42)</td>
<td>64.34 (19.18)</td>
<td>62.37 (18.14)</td>
<td>73.21 (23.05)</td>
<td>66.42 (24.11)</td>
<td>72.63 (20.29)</td>
<td>43.02 (8.88)</td>
</tr>
<tr>
<td></td>
<td>60–64</td>
<td>56</td>
<td>63.44 (31.41)</td>
<td>67.16 (27.67)</td>
<td>56.25 (18.29)</td>
<td>56.80 (17.45)</td>
<td>66.22 (14.80)</td>
<td>68.75 (27.27)</td>
<td>75.49 (22.02)</td>
<td>71.87 (14.02)</td>
<td>43.32 (7.06)</td>
</tr>
<tr>
<td></td>
<td>65–69</td>
<td>36</td>
<td>64.65 (26.09)</td>
<td>51.14 (28.89)</td>
<td>60.55 (29.89)</td>
<td>53.30 (27.24)</td>
<td>62.73 (22.03)</td>
<td>69.32 (29.71)</td>
<td>70.37 (33.60)</td>
<td>78.09 (17.89)</td>
<td>40.35 (11.13)</td>
</tr>
</tbody>
</table>

**Note:** Range 0–100, sample size varies due to missing data. Cases weighted for ethnicity. Group Ns are different from total count because weighted Ns are rounded.
Table 4. SF-36v2 mean scores (SD) for the eight sub-scales, PCS and MCS across gender whole sample (with 95% confidence intervals [CI]) and across three age groups

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55–69</td>
<td>2658</td>
<td>80.93</td>
<td>81.61</td>
<td>71.88</td>
<td>70.48</td>
<td>66.65</td>
<td>85.17</td>
<td>87.52</td>
<td>81.32</td>
<td>50.24 (9.23)</td>
<td>50.30 (9.08)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(22.16)</td>
<td>(23.97)</td>
<td>(23.48)</td>
<td>(21.21)</td>
<td>(19.65)</td>
<td>(21.89)</td>
<td>(20.52)</td>
<td>(15.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55–59</td>
<td>1116</td>
<td>83.59</td>
<td>85.69</td>
<td>72.98</td>
<td>71.01</td>
<td>69.69</td>
<td>65.92</td>
<td>84.36</td>
<td>86.76</td>
<td>80.74</td>
<td>49.89–</td>
</tr>
<tr>
<td>55–69</td>
<td>2859</td>
<td>77.83</td>
<td>81.35</td>
<td>70.43</td>
<td>71.69</td>
<td>64.59</td>
<td>85.27</td>
<td>87.27</td>
<td>80.17</td>
<td>49.76 (9.86)</td>
<td>49.93 (9.75)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(23.31)</td>
<td>(24.36)</td>
<td>(23.96)</td>
<td>(22.00)</td>
<td>(19.58)</td>
<td>(23.00)</td>
<td>(20.72)</td>
<td>(16.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55–69</td>
<td>2859</td>
<td>77.83</td>
<td>81.35</td>
<td>70.43</td>
<td>71.69</td>
<td>64.59</td>
<td>85.27</td>
<td>87.27</td>
<td>80.17</td>
<td>49.76 (9.86)</td>
<td>49.93 (9.75)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(23.31)</td>
<td>(24.36)</td>
<td>(23.96)</td>
<td>(22.00)</td>
<td>(19.58)</td>
<td>(23.00)</td>
<td>(20.72)</td>
<td>(16.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55–59</td>
<td>1172</td>
<td>79.56</td>
<td>83.88</td>
<td>71.33</td>
<td>72.36</td>
<td>64.23</td>
<td>85.43</td>
<td>87.26</td>
<td>79.36</td>
<td>50.87 (9.85)</td>
<td>49.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(23.75)</td>
<td>(24.37)</td>
<td>(24.44)</td>
<td>(22.87)</td>
<td>(20.35)</td>
<td>(23.02)</td>
<td>(20.86)</td>
<td>(16.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60–64</td>
<td>946</td>
<td>79.31</td>
<td>82.24</td>
<td>71.28</td>
<td>72.47</td>
<td>65.28</td>
<td>86.08</td>
<td>88.01</td>
<td>80.74</td>
<td>50.12</td>
<td>50.29</td>
</tr>
<tr>
<td>65–69</td>
<td>740</td>
<td>73.28</td>
<td>77.01</td>
<td>68.04</td>
<td>70.38</td>
<td>64.43</td>
<td>84.88</td>
<td>86.73</td>
<td>81.55</td>
<td>47.48</td>
<td>51.07 (8.92)</td>
</tr>
</tbody>
</table>

Note: Range 0–100, sample size varies due to missing data; Cases weighted for ethnicity; Group Ns are different from total count because weighted Ns are rounded.
Table 5. Comparison of the SF-36v2 mean scores (SD) for the eight sub-scales between 55–64 year olds in the HWR survey with 55–64 year olds in the New Zealand Health Survey (2006/7)

<table>
<thead>
<tr>
<th>Survey</th>
<th>Physical Function</th>
<th>Role Physical</th>
<th>Bodily Pain</th>
<th>General Health</th>
<th>Vitality</th>
<th>Social Function</th>
<th>Role Emotional</th>
<th>Mental Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWR</td>
<td>79.3 (78.7–79.9)</td>
<td>81.5 (80.8–82.1)</td>
<td>71.1 (70.5–71.7)</td>
<td>71.1 (70.5–71.6)</td>
<td>65.6 (65.0–66.1)</td>
<td>85.2 (84.7–85.8)</td>
<td>87.4 (86.8–87.9)</td>
<td>80.7 (80.3–81.1)</td>
</tr>
<tr>
<td>New Zealand Health Survey</td>
<td>81.3 (80.1–82.6)</td>
<td>83.2 (81.9–84.5)</td>
<td>73.0 (71.6–74.5)</td>
<td>73.6 (72.5–74.6)</td>
<td>65.0 (64.1–66.0)</td>
<td>89.0 (88.0–90.1)</td>
<td>94.6 (93.9–95.4)</td>
<td>84.6 (83.9–85.2)</td>
</tr>
</tbody>
</table>
In Table 4, normative data is provided for males and females for the sample as a whole and across three age groups. When examining the sample as a whole, no significant difference was found in scores on PCS and MCS, $F(1, 5514.53)=3.47$, ns, and $F(1,5514.75)=2.04$, ns, respectively. Only PCS scores differed significantly between males and females aged 55–59, $F(1, 2559)=4.39$, p<.05, with males reporting better physical health. No significant gender differences were found for the other age groups on PCS or MCS scores.

**Discussion**

Reliability analysis indicates good internal consistency of the eight sub-scales of the SF-36v2. Reliability statistics for the two role functioning sub-scales and vitality are higher than those previously reported for the general New Zealand population $^6$, whereas the social functioning scale shows lower reliability. Although social functioning has only two items, the items are very similar in wording and it is not immediately apparent why this group was not responding consistently. Given past reliable findings this subscale may be used with confidence while this anomalous finding is tested in future surveys of this population.

Validity of the scales for use with this population is also supported by the results for health outcomes and for ethnic differences. For the sample as a whole, physical health summary scores decreased and mental health summary scores increased with age. The observation that physical health deteriorates across these age groups is not surprising. The observation that mental health initially improves with age after 55 may not be expected by those unfamiliar with ageing populations, yet is well supported by previous findings for the general New Zealand population $^6$ and in international samples $^9$.

The SF-36v2 was used in the 2006/7 New Zealand Health Survey $^{10}$; a survey of the whole population. Comparison of the 55–64 year old age group with the HWR 55–64 year old age group (see Table 5) shows that physical health scores are comparable, although slightly higher, while mean mental health sub-scale scores (social functioning, role emotional, and mental health) are higher in the New Zealand Health survey. The higher mental health scores in this smaller sample of the older population support the trends shown in the present data.

Comparison of three ethnic sub-groups, New Zealanders of European descent, Māori and Pasifika Peoples, revealed significant differences in both physical and mental health. Overall, New Zealand Europeans reported better physical and mental health than Māori and Pasifika Peoples. These results are consistent with those from earlier studies of the SF36 in New Zealand $^{6,11,12}$.

These differences have also been reliably observed in other studies in New Zealand using different indicators of health outcomes, including mortality $^{14,15}$. However, Scott et al $^{16}$ have questioned the validity of the factor structure of these summary scores (which attempt to separate mental and physical health as separate constructs) for Māori and Pasifika People. They found that the expected factor structure was not reproduced for older Māori and this suggests that comparisons using the summary scores as presently constructed should be treated with caution, and other ways of...
using the responses to SF-36v2 items must be investigated. Scott et al. also showed that these scores are not valid for Pasifika peoples.

In addition, the SF-36v2 mean scores for Pasifika peoples in these data are not appropriate for use as normative data since there are low numbers, no indication of particular ethnicity, and wide confidence intervals in the results. However these results do suggest areas for ongoing enquiry using more focussed samples.

An apparent difference for this older age group compared with general populations, is less marked gender differences. Unlike results reported previously for the general New Zealand population and similar U.K. samples, data from our sample of New Zealanders aged 55–69 did not indicate significant sex differences on either physical or mental health summary scores. Differences were found only for the youngest age group (aged 55–59) and only for the physical health summary score. This finding indicates the importance of examining different population groups, as norms for certain groups can differ from the general population and patterns of change (such as improving mental health in general) may be revealed.

The present indication of a narrowing of the gap between male and female self-rated health may also be reflected in United Nations estimates of a difference in male and female mortality in New Zealand of less than 3 years by 2020.

The representativeness of the present data may be limited because of a low response rate from the selected population. Although the response rate of 61% for New Zealand Europeans was high for a postal survey, questions remain regarding the characteristics of those who did not respond. Because the sample in this study reflects census data on key variables such as age groups and gender, and owing to the oversampling of Māori, we have confidence that this large sample can indicate trends and differences in health in the population.

The SF-36v2 is now one of the most well used measures of health and quality of life in survey and clinical research. Although the measure has its critics and limitations, an advantage of its widespread use is the ability to compare results internationally and within local populations. The use of the norms from this study will facilitate the comparison of sub-scale data collected from small clinical samples of older Māori and European New Zealanders with the general New Zealand population and international populations of those aged 55 to 69 years. The norms are also available for use in calculating summary physical and mental health summary scores for data from national surveys (as opposed to using norms generated in the US or other countries). These norms may be used in any research with those aged 55–69 years and within three different age groups.

**Competing interests:** None.

**Author information:** Christine Stephens, Associate Professor, School of Psychology, Massey University, Palmerston North; Fiona Alpass, Associate Professor, School of Psychology, Massey University, Palmerston North; Merel Baars, Research Assistant, School of Psychology, Massey University, Palmerston North; Andy Towers, Research Officer, School of Psychology, Massey University, Palmerston North; Brendan Stevenson, Research Officer, Research Centre for Māori Health and Development, Massey University, Palmerston North
Acknowledgements: These data were collected with the assistance of research funding from the Health Research Council of New Zealand. The authors also acknowledge the very helpful contributions of the two anonymous NZMJ reviewers.

Correspondence: Christine Stephens, School of Psychology, Massey University, Private Bag 11 222, Palmerston North, New Zealand. Fax: + 64 (0)6 3505673; email: c.v.stephens@massey.ac.nz

References: