Comparing survey data on functional disability: the impact of some methodological differences

H S J Picavet, G A M van den Bos

Abstract

Study objective – To examine the impact of some differences in survey methodology on the prevalence of functional disability in population based surveys of the elderly. Design and methods – Nine surveys of Dutch people aged 55 years and older were compared to investigate the differences in the methods of data collection (proxy questioning, yes/no; interview versus self administered questionnaire) and construction of the questionnaire (wording of introductory text, activities, and response categories). The effect of these differences on prevalences in three domains of functional disability – activities of daily living, mobility, and communication – were studied. Both univariate analyses and multivariate logistic regression were used to quantify the methodological influences. Results – No effect of proxy questioning could be shown. Self administered questionnaires yielded higher prevalences of disability than interviewer administered questionnaire – in particular for mobility (odds ratio (OR) 1·4, 95% confidence interval (95% CI) 1·3, 1·6) and communication (OR = 1·7, 95% CI 1·5, 1·9), resulting in prevalence differences of 9 and 11 percentage points respectively. Seemingly minor differences in the structure and wording of the questionnaires resulted in major differences (up to 15·6 percentage points) in prevalence estimates of functional disability. These differences were associated with the severity level of the disability indicated by the wording of the questions. Conclusions – Differences in survey methodology have a substantial effect on the prevalence estimates of disability in the elderly. These differences should be taken into account when making international comparisons and studying time trends based on survey data.
Survey methodology and disability prevalences

Juan G. Pérez-Rossello

Institutes such as nursing homes, which implies the exclusion of a group of probably severely disabled elderly. This should be taken into account when data from a survey which excludes institutionalised people is compared with a survey that includes people in institutions. The same is true when comparing surveys that provide data on target population for which very different criteria for nursing home admission are present, particularly when data from different countries are compared.

The method of data collection refers to the choice of using either interviewer administered techniques, self completion, or both, and whether to use proxies. The validity of proxy responses has been investigated before. For the assessment of functional disability a proxy effect for minor disabilities (under-reporting by proxy) might be expected but not for obvious and long standing disabilities. There is some evidence that self administered (written) questionnaires show systematically higher prevalences of disability than interviewer administered questionnaires. It is indisputable that different instruments may yield different results. Often, however, there is a high similarity between instruments designed to measure the same concept of "functional disability", with only slightly different wordings of the actual questions or a different number of activities, or both. It is important to know the extent to which these seemingly unimportant differences between instruments affect the estimation of prevalences of functional disability, even more so because it has already been shown that activities of daily living (ADL) prevalences differ substantially between surveys in the USA.

In this paper, we will describe the impact of some differences in the construction of the instruments and differences in the method of data collection on the prevalences of functional disability.

Methods

THE MATERIAL

Nine different surveys concerning elderly Dutch people provide unique material with which to study the effect of methodological differences. The data consist of a number of years of the continuous Netherlands health interview survey (HIS: 1983, 1984, 1985, 1989, 1990) and three multipurpose surveys: the continuous Netherlands quality of life survey (QLS: 1989, 1990), the quality of life survey of the elderly of 1982 (QLE82), and the public service survey of 1987 (PSS87). Like the HIS, these multipurpose studies collect information on health characteristics.

The general characteristics of the surveys are summarised in table 1. All the surveys provided data on a probability sample of the non-institutionalised population aged 55 years and older. The exclusion of people living in institutions – mainly homes for the elderly and nursing homes, but also prisons – is a common procedure in health interview surveys. Because the prevalence of functional disability among the institutionalised population will be relatively large, the prevalences based on these surveys will be an underestimation of the total absolute prevalences in the population. Because all the surveys excluded the institutionalised population of the same target population, however, it will not distort the comparisons.

All surveys were carried out by trained interviewers/pollsters who visited people at home, and used questionnaires, part of which, in some cases, was left behind and collected later. Thus, part of the data collection was oral and some written. None of the surveys used telephone interviews or self-administered postal questionnaires.

Functional Disability

Using a list of activities (or functions), functional disability was measured by asking the respondents to state whether they had difficulty carrying out the activity, needed help, or were unable to carry out the activity. The list of activities was adapted from the OECD indicator for long term disability to which some ADL items had been added. The original OECD indicator as well as the ICIDH had been constructed on the basis of the concepts described by Wood. Questions on mobility, eg walking, communication, eg hearing, and seeing, and ADL activities are part of the most essential questions on disability to be used in survey research. The number of activities differed per survey (see table 2).

Methodological Differences

In general, the survey designs were very similar so we had the unique possibility of quantifying the effects of the seemingly minor differences in survey characteristics. The methodological differences that could be studied were: differences in the methods of data collection and in the construction of the questionnaires.

Data Collection

Two aspects of the method of data collection could be studied: allowing a proxy to be interviewed and the use of interviewer administered
versus self administered questionnaires. In some data years of the HIS, the use of proxies for people who were not at home or were unable to respond was a standard procedure for the interview administered questionnaire. In 1989 and 1990 a self administered questionnaire was used as well as the interview, and this was left behind by the interviewer and collected later. Because the ADL questions were part of the interview in those years (and therefore possibly answered by proxies) and the mobility questions were part of the written questionnaire (and thus answered by the respondent) it was possible to compare the prevalence of disability reported by proxies with that reported by the people themselves.

**Questionnaire construction**

With regard to the construction of the instruments, three aspects were studied. The first concerns the possible effect of using an introductory text to the questions on disability. Some surveys included an introductory text to the questions on mobility and communication which emphasised that the questions referred to *longstanding disability*. Although the original OECD instrument aimed to assess long term disability, temporary problems were not necessarily excluded. The introductory text aimed therefore to prevent the reporting of temporary activity limitations. Comparison of the estimated prevalences in the surveys concerned should give an indication of the effect of the introductory text. If temporary disability is reported, the prevalence is likely to be higher.

The second aspect concern the effect of the wording of the activities, how they are described. Within all domains of disability some items are phrased slightly differently across the surveys. These differences in wording will be described in detail in the results section. The original questions were, of course, phrased in Dutch. The translations used in this article are as literal as possible and do not necessarily represent normal phrasings in the English language or should give the same results when used in these translations. The results should be seen as an illustration of the potential effect of differences in wording of the activities.

**Table 2. Methodological characteristics and adjusted prevalences by domain of disability, per survey. All surveys are conducted in The Netherlands among the non-institutional population of 55 years and older. Prevalences include the least severe severity level that was assessed in the survey**

<table>
<thead>
<tr>
<th>Survey</th>
<th>HIS83</th>
<th>HIS84</th>
<th>HIS85</th>
<th>HIS89</th>
<th>HIS90</th>
<th>QLE82</th>
<th>QLS89</th>
<th>QLS90</th>
<th>PSS87</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of respondents aged 55+</td>
<td>2060</td>
<td>2119</td>
<td>2017</td>
<td>1924</td>
<td>1964</td>
<td>4283</td>
<td>848</td>
<td>1283</td>
<td>3840</td>
</tr>
</tbody>
</table>

**Activities of daily living (ADL)**

Methodological differences:

- Method of data collection
  - (i) interview, s = self administered
  - Construction of response categories
    - Distinction between minor and major difficulty: Yes Yes Yes Yes No Yes No No
  - Including “not able to”: Yes Yes Yes No No Yes No Yes
  - Including “only with help”: No No No Yes Yes Yes No Yes

<table>
<thead>
<tr>
<th>Prevalences (55+):</th>
<th>Eating and drinking</th>
<th>Getting in/out of a chair</th>
<th>Getting in/out of bed</th>
<th>(Un)dressing</th>
<th>Transferring one level</th>
<th>Walking stairs</th>
<th>Entering/leaving the house</th>
<th>Transferring outdoors</th>
<th>Washing face and hands</th>
<th>Washing entirely</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-4</td>
<td>15-7</td>
<td>3-8</td>
<td>4-8</td>
<td>3-5</td>
<td>26-2</td>
<td>7-9</td>
<td>15-2</td>
<td>2-4</td>
<td>6-8</td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>16-3</td>
<td>4-9</td>
<td>6-0</td>
<td>3-8</td>
<td>25-4</td>
<td>9-4</td>
<td>17-0</td>
<td>2-8</td>
<td>8-5</td>
</tr>
<tr>
<td></td>
<td>0-5</td>
<td>6-8</td>
<td>4-9</td>
<td>7-5</td>
<td>6-2</td>
<td>24-4</td>
<td>4-6</td>
<td>11-4</td>
<td>1-0</td>
<td>8-4</td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>16-5</td>
<td>7-5</td>
<td>5-2</td>
<td>5-8</td>
<td>27-8</td>
<td>8-5</td>
<td>20-8</td>
<td>3-8</td>
<td>7-6</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>11-2</td>
<td>9</td>
<td>9-7</td>
<td>6-8</td>
<td>29-6</td>
<td>8-8</td>
<td>15-5</td>
<td>3-1</td>
<td>6-9</td>
</tr>
</tbody>
</table>

One ADL or more

| 26-7 | 30-2 | 29-9 | 28-3 | 32-0 | 21-9 | 34-3 | 34-5 | 24-7 |

**Mobility**

Methodological differences:

- Method of data collection:
  - (i) interview, s = self administered
  - Emphasis on long duration of disability in introduction: No No No Yes Yes No

<table>
<thead>
<tr>
<th>Prevalences (55+):</th>
<th>Carrying 5 kg for 10 m</th>
<th>Bending and picking up</th>
<th>Walking 400 m†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30-9</td>
<td>24-2</td>
<td>27-1</td>
</tr>
<tr>
<td></td>
<td>34-3</td>
<td>32-0</td>
<td>34-4</td>
</tr>
<tr>
<td></td>
<td>33-0</td>
<td>29-8</td>
<td>33-4</td>
</tr>
<tr>
<td></td>
<td>33-1</td>
<td>26-4</td>
<td>23-4</td>
</tr>
<tr>
<td></td>
<td>32-1</td>
<td>25-2</td>
<td>22-5</td>
</tr>
<tr>
<td></td>
<td>22-3</td>
<td>18-1</td>
<td>18-2</td>
</tr>
</tbody>
</table>

Problems with mobility‡

| 44-7 | 54-8 | 52-6 | 41-4 | 38-4 |

**Communication**

Methodological differences:

- Method of data collection:
  - (i) interview, s = self administered:
  - Emphasis on long duration of disability in introduction: No No No Yes No No

| Construction of response categories (D = descriptive, E = evaluative): § |
|-------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                   | D                   | D                   | D                   | D                   | E                   | E                   | E                   |

<table>
<thead>
<tr>
<th>Prevalences (55+):</th>
<th>Conversation 3/4 persons†</th>
<th>Conversation 1 person</th>
<th>Seeing on a distance†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16-7</td>
<td>5-2</td>
<td>6-9</td>
</tr>
<tr>
<td></td>
<td>21-9</td>
<td>6-6</td>
<td>11-2</td>
</tr>
<tr>
<td></td>
<td>21-1</td>
<td>7-2</td>
<td>10-0</td>
</tr>
<tr>
<td></td>
<td>20-3</td>
<td>6-2</td>
<td>10-6</td>
</tr>
<tr>
<td></td>
<td>20-2</td>
<td>6-2</td>
<td>10-7</td>
</tr>
<tr>
<td></td>
<td>19-5</td>
<td>5-5</td>
<td>13-2</td>
</tr>
<tr>
<td></td>
<td>12-2</td>
<td>4-6</td>
<td>14-7</td>
</tr>
<tr>
<td></td>
<td>10-8</td>
<td>12-0</td>
<td>13-2</td>
</tr>
<tr>
<td></td>
<td>24-8</td>
<td>16-3</td>
<td>7-7</td>
</tr>
</tbody>
</table>

Communication disability‡

| 26-4 | 36-6 | 38-3 | 35-1 | 34-1 | 31-3 | 21-3 | 18-9 | 34-0 |

* Proxy questioning allowed; † wording of function can differ per survey; ‡ the aggregated prevalence rates are based on all available items per survey; § Descriptive = <Can do activity> without difficulty, with minor difficulty, with major difficulty, not able to; evaluative = <<Can do activity>> good, moderate, bad, not at all. HIS = Netherlands health interview survey; QLE = quality of life survey of the elderly; QLS = quality of life survey; PSS = public service surveys.
The third aspect of questionnaire construction concerns the response categories. All surveys used questions on disability with response categories which included levels of severity – for example, “with difficulty”, “needing help”, or “unable to carry out the activity” (the D variant (of Difficulty)). The effects of differences within this D variant were studied. The response categories for the disability for hearing and seeing are either phrased as a D variant or in terms of the evaluation of the ability to hear or see: <can see/hear>, well, moderately, badly, not at all (the E variant (of Evaluation)).

**ANALYSIS**

Univariate and multivariate statistical analyses were performed as was visual inspection of the data. For each survey, the estimated prevalences of functional disability, per item as well as aggregated per domain of disability, are presented in table 2. These prevalences were adjusted for differences in the age-sex composition of the respondents and for differences in sample design. In such a table it is not feasible to present the statistical differences between every two surveys. In all cases a difference of 5 percentage points was statistically significant (p<0.01). In most cases a lesser difference approaches statistical significance. Differences between prevalences were described as the absolute difference in percentage points (prevalence difference) and as the ratio between the two prevalences. Multiple logistic regression was used to calculate the relative effects of some of the methodological differences.

**Results**

The methodological differences and prevalences per disability item and aggregated per domain of disability are presented in table 2. For the three different domains of disability, the aggregated prevalences differed significantly per survey. The percentage of people of 55 years and older who reported at least minor difficulty with at least one ADL item ranged from 21.9% to 34.5%. These figures were 38.4% and 54.8% for disability in terms of mobility, and 18.9% and 38.3% for problems with hearing and/or seeing.

When surveys used identical methods, however, the estimated prevalences were very similar. This is true for the survey pairs HIS84 and HIS85, HIS89 and HIS90, QLS89 and QLS90, and for all domains of disability.

The great variations in disability prevalences across surveys that used different methods and similar prevalences for surveys with identical methodology reflected the impact of methodological differences.

**PROXY QUESTIONING**

Interviews involving proxies are a common feature of the Netherlands health interview survey. For the elderly population, the percentage of proxies is about 21% for men and 5% for women, which is much lower than the percentage for the entire population, which is 30%. The main reasons for proxy interviews were “not at home” (because of work, shopping, hobbies, or other activities) (more than 70%) and “unable to” because of illness (at home or in hospital). Because the proxy group consisted of apparently healthy persons on the one hand and severely disabled on the other, it is not known what the expected mean health status of this group is compared with the non-proxy group. Table 3 presents the prevalences of disability in terms of ADL and mobility for all persons for whom a proxy was interviewed and for all persons who were interviewed themselves. The mean health status seemed to be better in the proxy than in the non-proxy group. This held true for disability in ADL, which can represent a proxy effect, and it also held true for disability in mobility, which cannot represent a proxy effect. At an aggregated level, there was no indication of under-reporting or over-reporting of disabilities by the use of proxies.

**INTERVIEWER VERSUS SELF ADMINISTERED QUESTIONNAIRES**

To measure the effect of interviewer versus self administered questionnaires, HIS83 and HIS84/85, and QLE82 and PSS87 were compared. Separate comparisons were necessary because HIS83–85 and QLE82/PSS87 used different response categories. The differences are summarised in table 4. For all domains of disability, self administered questionnaires showed systematically higher prevalences, with
disability in mobility and communication showing higher differences than disability in ADL. In most cases the prevalence difference of the aggregated prevalences was higher than the mean prevalence difference of the single items. The effect of the method of data collection seemed to accumulate in the aggregated prevalences. Using self administered questionnaires increased the measured aggregated prevalence of disability in communication by an absolute difference of 11.0 percentage points (that is, a relative increase of 42%) and for the prevalence of disability in mobility by an absolute difference of 9.0 percentage points (that is, a relative increase of 20%). For disability in ADL these figures were 3.4 (13%) and 2.6 (12%) respectively for the two methods (HIS83 versus HIS84/85 and QLE82 versus PSS87 respectively).

**EMPHASIS ON LONG DURATION OF DISABILITY IN AN INTRODUCTORY TEXT**

The effect of mentioning explicitly a long duration of disability on prevalences was studied for disability in mobility and in communication by comparing HIS89/90 (duration emphasised) with HIS84/85 (not emphasised) (see table 5). Because of the five year gap between the two surveys, real changes in prevalences cannot be ruled out. Since very large changes in prevalences are unlikely over such a period, however, the dramatic decrease in prevalence was mainly due to the difference in methodology, in particular for disability in mobility. The use of the introduction lowered the estimated prevalence of disability in mobility by 13.7 percentage points. Where there was no emphasis on the long duration of disability, up to 26% of the reported disability in mobility could be attributed to temporary disability. This finding held true for the separate mobility items, although the comparison for the item “walking” was hampered by differences in wording of the function (see below). When the duration of the disability is not emphasised, temporary problems with mobility can also be reported. The lack of effect on the prevalences of problems in hearing and eyesight of emphasising the long duration of disability, indicated that these problems are of a permanent nature when they are reported.

**DIFFERENCES IN WORDING OF THE FUNCTIONS**

The effect of differences in the wording of the function was studied for five items – walking stairs, dressing, seeing, hearing, and walking a certain distance. For the item “walking stairs” the wording in HIS83 and HIS84/85 was: “Can you walk up and down a staircase of 15 steps without standing still?”. In the other surveys, “walking stairs” was part of the ADL list, which is more generally phrased. HIS83 and HIS89/90 were comparable because they both assessed this disability by self administered questionnaire. The phrasing in HIS83 resulted in a slightly higher estimation of disabilities than the general wording. This was true, in particular, for severe disabilities: the percentage of the elderly who reported inability to climb the stairs was 7.3% in HIS83, while in HIS89/90 the mean percentage of elderly who reported that they were able to climb stairs only with help was 3.4% (prevalence difference: 3.9, 95% CI 2.6, 5.2). The addition of the phrase “without standing still” in HIS83 probably made the function more difficult to perform.

Unlike all other surveys, QLE and PSS included the phrase “putting on shoes” in the item “(un)dressing”, which might have made it a more difficult activity to perform. For these surveys the mean prevalence of disability with dressing was 7.7%, for all other surveys it was 7.2%. The main problem in this comparison was that QLE and PSS also used different response categories which probably resulted in lower prevalences (see next paragraph), indicating that the extended phrasing of the item resulted in (relative) higher prevalences.

To investigate problems with eyesight, the two questions most commonly used were: “Is your eyesight good enough to read ordinary newspapers print?” and “Is your eyesight good enough to recognise a face at a certain distance?”. The wording of a face at a certain distance there were two main ways of asking – seeing at a specific distance (4 metres, on the other side of the room) and just “seeing on the street”. The mean prevalences for these two variants were 9.2% (HIS and PSS) and 13.4% (QLE and QLS) respectively (prevalence difference 4.2, 95% CI 3.2, 5.2). Within the elderly, in particular, problems with eyesight are very diverse. This diversity can lead to serious problems when disability in seeing is to be assessed with such global questions. The observed differences in prevalences when questions have different wording illustrate this.

Three methods were used to assess problems with hearing. PSS and QLS simply asked whether the hearing of a person was good, moderate, or bad, on the basis of which a mean of 22% of the respondents reported problems. The HIS questionnaires included two questions on the degree of difficulty, that is, “Can you hear what is said in a normal conversation.
DIFFERENCES IN RESPONSE CATEGORIES

The greatest variation in the phrasing of questions on functional disability between the surveys concerned the function of walking. Differences included the wording of the time of walking, the distance (400 metres or one block), the use of an aid such as a cane, and the level of difficulty. Comparability was quite problematic because apart from these differences the method of data collection and the duration of the disability also played a part. As an illustration, the five versions of the question on walking disability and the prevalences are presented in Table 6. The absolute difference can be as big as 15.5 percentage points when two self-administered surveys are compared: HIS85 versus PSS87 (95% CI 12.9–17.7). Even for the estimation of the elderly population with the most severe disability in mobility (“not able to walk”) the prevalences varied from 4.2% to 15.9%.

In the HIS (all years) and the QLS, a distinction between minor and major difficulty was made, whereas no such distinction was drawn in QLE82 and PSS87. Where this distinction was made, higher prevalences were found for almost all ADL items, comparing the percentage reporting at least “minor difficulty” with the percentage reporting at least “difficulty”. Furthermore, it can be shown that the percentage who reported “major difficulty” was systematically lower (see figure 1). This indicates that the response category “with difficulty” does not cover “minor and major” difficulty but represents something in between. As a consequence these absolute prevalences have become incomparable.

The other difference in the response categories for the ADL could be found in the wording of the most severe response category. HIS83–85 used “unable to”, HIS89/90, QLE82, and PSS87 used “only with help”, and QLS89/90 used both. There was considerable variation in the estimates of the size of the group of elderly suffering from the most severe ADL disabilities (not shown). These different phrases used to describe the most severe ADL disabilities were not interchangeable, which was also illustrated by the observation in QLS89/90 that both categories were “filled”. Exclusion of one of these responses might have hampered the classification of severity and it might also have affected the interpretation of the other response categories.

The D variant and E variant can be compared for the questions on disability in seeing. QLE82, QLS89/90 and PSS87 used response categories in evaluative terms. A comparison

### Table 6: Prevalences of walking disability: five variants in wording of the question, asked of sample of the non-institutional population of elderly, The Netherlands

<table>
<thead>
<tr>
<th>QLE82, interview</th>
<th>How long can you walk without help?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;40 minutes</td>
</tr>
<tr>
<td></td>
<td>30–39 min</td>
</tr>
<tr>
<td></td>
<td>20–29 min</td>
</tr>
<tr>
<td></td>
<td>10–19 min</td>
</tr>
<tr>
<td></td>
<td>&lt;10</td>
</tr>
<tr>
<td></td>
<td>Not able</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PSS87, self administered</th>
<th>Walking for ten minutes without standing still</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without difficulty</td>
</tr>
<tr>
<td></td>
<td>With difficulty</td>
</tr>
<tr>
<td></td>
<td>Only with help</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIS83, interview</th>
<th>Can you walk 400 metres without standing still?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without difficulty</td>
</tr>
<tr>
<td></td>
<td>Minor difficulty</td>
</tr>
<tr>
<td></td>
<td>Major difficulty</td>
</tr>
<tr>
<td></td>
<td>Not able</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIS85, self administered</th>
<th>Can you walk 400 metres without standing still?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without difficulty</td>
</tr>
<tr>
<td></td>
<td>Minor difficulty</td>
</tr>
<tr>
<td></td>
<td>Major difficulty</td>
</tr>
<tr>
<td></td>
<td>Not able</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIS89, self administered</th>
<th>&lt;Emphasising long duration of disability&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without difficulty</td>
</tr>
<tr>
<td></td>
<td>Minor difficulty</td>
</tr>
<tr>
<td></td>
<td>Major difficulty</td>
</tr>
<tr>
<td></td>
<td>Not able</td>
</tr>
</tbody>
</table>

QLE = quality of life survey among the elderly; PSS = public services survey; HIS = Netherlands health interview survey.

with three or four other persons?” and the same question with one other person (both despite hearing aid, if usually worn). Of the elderly, 19.9%—9% reported problems with one or both of these hearing functions. Whether or not the wording of the function or the construction of the response categories had any effect on prevalence is unknown because both differed and could have had an (adverse) effect.

Prevalences of activities of daily living (ADL) disability in relation to different response categories for each ADL item. The prevalences for the disability levels “including minor difficulty” and “including major difficulty” are based on The Netherlands health interview survey 1989/90: data for the disability level “with difficulty” are derived from the quality of life survey of the elderly of 1982 and the public services survey of 1987.
of the prevalences on the basis of PSS87 and HIS89/90 (both self administered questionnaires) showed that the E variant lead to lower prevalences than the D variant. This is particularly true for the most severe category: the estimated prevalence among those whose eyesight was too bad to read ordinary newspaper print (despite glasses) was 5·1% for HIS89 and 1·4% for PSS87 (p<0·01). The figure for recognising a face were 2·9% and 0·4% respectively (p<0·01). These different response categories yielded different, and therefore incomparable, prevalences.

MULTIVARIATE ANALYSES

The results of the multivariate logistic regression are presented in table 7. The main methodological differences were included in the model as dummy variables. The methodological effects that were identified by the univariate analyses were still present in the multivariate analyses. The results can be summarised as follows: written self administered surveys provided consistently higher aggregated disability prevalences than interviewer administered surveys and differences in the actual instrument (introductory text and response categories) had significant effects on the estimated prevalences.

In addition to this we observe that for disability in ADL and disability in mobility, the effects of differences in the construction of the instrument were more important than those related to the method of data collection. For disability in communication, the effect of differences in the method of data collection was more pronounced.

Discussion

The finding that different surveys yield different estimates of the size of the (elderly) population with disabilities had already been reported for ADL disabilities on the basis of surveys carried out in the USA. Our study shows that the prevalences of disability in mobility and communication can be added to the list of disability prevalences that differ across surveys, although this held true only for surveys that differ in methodology. Identical surveys yielded the same prevalences for functional disability, both for the aggregated prevalences and the prevalences per item. All methodological differences studied in this study were shown to affect appreciably the prevalences of disability, with exception of the proxy effect. At the aggregated level we could not show that the use of proxies yielded higher disability prevalences a finding which had previously been reported.9 12 The finding that self administered questionnaires yielded higher prevalences than interview questionnaires is consistent with published reports.13 14 Prevalences for disability in ADL, mobility, and communication based on self administered surveys are 13%, 20%, and 42% respectively higher than those in interview based surveys, resulting in prevalence differences of up to 11 percentage points. Self administered instruments and interviewer instruments are not therefore equivalent, although more research is needed to determine which of these is the most valid.

The introductory text which emphasised the long duration of the disability reduced the prevalences of some of the disabilities, as would be expected if temporary disability were reported when the introductory text was omitted. More than 25% of the estimated elderly population with disability in mobility can be due to reporting of temporary disability.

The most striking result of this study was the major difference in estimated prevalences due to seemingly minor differences in instruments (wording of activities and of response categories). This was mainly due to the assessment of the degree of difficulty in performing the activity, which was defined by the exact wording of the activity and the wording of the response categories. The actual wordings of the disability questions define the threshold for the disability that will be identified. When there are differences in the wording of these questions, even seemingly minor ones, different prevalences of disability result.

These results were based on comparing data that were already available and were not collected for the purpose of investigating methodological differences. The design is not therefore ideal for the study of methodological influences on the outcome of surveys. Alternative factors that could have contributed to the differences between the surveys are differences in the selectivity of the relative high non-response and/or large shifts in the real prevalences of disability per year. However, the systematic similarity of the results of surveys that used identical methodology makes these hypotheses very unlikely.

To quantify health differences between groups, we need comparable data of the prevalences of health problems. For the quantification and monitoring of health problems in an ageing population, data on functional disability and chronic conditions based on population surveys are increasingly used. Our study illustrates the size of the influences that methodological differences in surveys have on the estimated absolute prevalences of functional disability in the elderly, which we con-
sider as alarmingly large. No true disability prevalence exists: the prevalence is very sen-
sitive to characteristics of the survey and de-
pends, in particular, on the severity of disability as
is indicated by the actual wording of the
questions and response categories. For in-
ternational comparisons, agreement on in-
struments in population surveys such as health
interview surveys are a prerequisite. Our find-
ings also imply that consistency in methodology
is necessary if the assessment of time trends
on the basis of series of surveys is required.
These requirements seem to conflict with one
another because adaptation of survey metho-
dology to meet international agreements implies
interruption of time trend assessment in those
countries which already have series of surveys.
For the interpretation of the results from
regional surveys or surveys within a specific
patient population it may be necessary to com-
pare these with reference values from other
surveys, surveys concerning other regions, or
other patient categories or even national popu-
lation based surveys. Such comparisons are
only possible when surveys use the same meth-
ological procedures as the "comparison standard".
The design and execution of a popu-
lation survey or the assessment of health char-
acteristics require many decisions on the
instruments and methodological procedures.
The potential effects of methodological char-
acteristics on the outcome of surveys and the
implication of these for comparisons with other
surveys should be taken into account during
this decision phase.

The authors wish to thank Jan L A van Sonsbeek and the
Netherlands Central Bureau of Statistics for their expert advice
and provision of the data for this study. We also wish to thank
Louise Gunning-Schepers, Rob Spruit, and Wien Limburg for
their comments on earlier versions of this manuscript.

  Report on a WHO meeting. Copenhagen: WHO Regional
  Office for Europe (EUR/ICP/RHB 023); 1991; 12.
  Hyattsville: US Department of Health and Human
  Services (DHHS Publication no (PHS) 94–1007),1994;
  118.
3 Evers, SMAA. Health for All indicators in health interview
  surveys. Health Policy 1993;23:205–18
4 Boshuizen HC, van de Water HPA. An international com-
parison of health expectancies. Leiden: TNO Health Report
  (TN0 PG 94/046), 1994;145.
5 EUROSTAT. Disabled persons statistical data. Vol 1. Ger-
  many, Greece, France, Luxembourg, The Netherlands
  and Portugal, 1991. Volume 2: Belgium, Denmark, Spain,
  Ireland, Italy and the United Kingdom. Luxembourg:
6 Heikinänen E, Waters WE, Breenainski ZJ, eds. The elderly in
deleven countries, a sociomedical survey. Copenhagen: WHO
  Regional Office for Europe (Public Health in Europe 21);
  1983: 231.
7 Robine JM, Mathers CD, Bone MR, Romieu I, eds. Cal-
culation of health expectancies: harmonization, consensus
  achieved and future perspective. (Colloque intern, vol 226).
  Montreouge: John Libbey, 1993;388.
8 Locker D, Wiggins R, Sirtampalam Y, Patrick DL. Esti-
mating the prevalence of disability in the community: the
  influence of sample design and response bias. J Epidemiol
9 Mosley II RR, Wolinsky FD. The use of proxies in health
  surveys, substantive and policy implications. Med Care
10 Nelson LM, Longstreth WT, Koepsell TD, van Belle G.
  Proxy respondents in epidemiological research. Epidemiol
11 Rothman ML, Hedrick SC, Bulcroft KA, Hickam DH,
  Rubenstein LZ. The validity of proxy-generated scores
  as measure of patient health status. Med Care 1991;29:
  165–24.
12 Tennant A, Badley EM, Sullivan M. Investigating the proxy
  effect and the saliency principle in household based postal
  questionnaires. J Epidemiol Community Health 1991;45:
  312–6.
13 Anderson JP, Bush JW, Berry CC. Classifying function for
  health outcome and quality of life evaluation. Self-versus
  interview modes. Med Care 1986;24:54–69.
14 Cook DJ, Guyatt GH, Juniper E, et al. Interviewer versus
  self-administered questionnaires in developing a disease-
specific, health-related quality of life instrument for
15 Wiener JM, Hanley RJ, Clark R, van Nostrand JF. Measuring
  the activities in daily living: comparisons across national
16 McWinnie JR. Disability assessment in population surveys:
  results of the OECD common development effort. Rev
17 McWinnie JR. Measuring disability. OECD Social Indicators
  Programme (Special Studies no 5). Paris: OECD, 1982;
  44.
18 World Health Organization. International classification of im-
  pairments, disabilities, and handicap – a manual relating to
19 Wood PHN. Classification of impairments and handicap.
  Reviews/Conference Series No 75/13. Geneva: WHO,
  1975.
20 Chamie M. Survey design strategies for the study of dis-
