Development and validation of the Observation List for early signs of Dementia (OLD)

Marijke Hopman-Rock, Erwin C. P. M. Tak and Patricia G. M. Staats

1Research Co-ordinator, Epidemiologist, Biologist, Psychologist, TNO Prevention and Health, Public Health Division, Leiden, The Netherlands
2Research Assistant, Psychologist, TNO Prevention and Health, Public Health Division, Leiden, The Netherlands
3Researcher, Social Gerontologist, TNO Prevention and Health, Public Health Division, Leiden, The Netherlands

SUMMARY
Objective Development and validation of a short Observation List of possible early signs of Dementia (OLD) for use in general practice.
Design Stepwise development using reviews of publications and expert consensus. Field study for evaluation of reliability. Validation study (interviews, family forms) using existing valid and reliable measures. Use of data reduction techniques to construct a short version.
Setting of field study Twenty-two GPs in 19 Dutch practices.
Participants The first two patients seen on 15 working days (n = 470) were observed. Inclusion: age > 75, without a known diagnosis of dementia. Exclusion: psychiatric treatment, severe depression, acute illness with confusion. Division of patients into three groups with no, intermediate, and the most signs (total of interviewed patients, n = 60; family forms, n = 39).
Outcome measures Reliability (Cronbach’s alpha and factor-analysis). Convergent validity using the Cognitive Screening Test (CST), the Word Learning Test (WLT, total and retention), the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE), the Groningen Activities Restriction Scale (GARS), and an IADL scale. Discriminant validity using the geriatric depression scale (GDS). Construct validity using a Principal Component Analysis (PRINCALS). Incremental validity using the intuitive opinion of the GP (McNemar test).
Results Reliability in the total group 0.88, first factor explained variance 42.5%. Convergent validity (two-way ANOVA) results: CST (p = 0.00), WLT-total (p = 0.001), WLT retention (p = 0.00), IQCODE (p = 0.09). No statistically significant differences for GARS and IADL, GDS (p = 0.30) not different. PRINCALS first factor explained 48% of variance. The OLD added to the GP opinion (McNemar p = 0.00). Reliability short version 0.89 (interviewed group), 0.86 (total group).
Conclusion The OLD is a valid and reliable method to detect early signs of dementia in general practice that can indicate when it may be useful to employ existing screening instruments. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS — activities of daily living; Alzheimer disease; behaviour; cognition; dementia; family physician; observation; signs and symptoms

INTRODUCTION
Dementia, which is highly prevalent among aged populations, is associated with a diminished quality of life and an increased morbidity and mortality (Jorm et al., 1987; Jacoby and Oppenheimer, 1991; Howard and Rockwood, 1995). A meta-analysis has shown that the prevalence increases from 1% in people aged 65–69 years, to 22% in people aged 85–89 years, and 35% in people aged 95 years and older (Hofman et al., 1991). While dementia is difficult to diagnose (O’Connor et al., 1988; Bowers et al., 1990; Wind et al., 1994, 1997), early detection is important in order to provide timely support for the patient and his or her relatives (Wilcock et al., 1994) and to start...
potential drug treatment as soon as possible. However, the most frequently used dementia screening instruments have a relatively low predictive positive value for the general aged population compared with that for clinical populations with dementia (Flicker et al., 1997). This means that the use of such screening tests in unselected elderly populations will probably produce many false positive diagnoses of dementia. This also raises the ethical question of whether older people should be confronted with the suspicion of dementia, which may cause them unnecessary anxiety.

General practitioners need to be able to detect possible dementia without unduly alarming their patients and their relatives. The detection of certain signs that are characteristic of early dementia (and not of other disorders, such as depression) and monitoring their progression will enable general practitioners to take timely action. In the Netherlands, the Dutch College of General Practitioners published a standard to help general practitioners to detect signs of dementia (Wind et al., 1999). The standard provides information about the anamnesis and hetero-anamnesis of dementia, using some examples of signs that may be observed (without mentioning their reliability or validity).

Our aim was to develop and validate a short observation list (the so-called observation list for early signs of dementia, OLD) of possible early signs of dementia for use by general practitioners in a general older population. Because dementia of the Alzheimer type is the most common form, the OLD was restricted to this disorder. We give the results of a field study involving 22 general practitioners who observed 470 people aged 75 years and older who did not have a known diagnosis of dementia.

METHODS

Initial development

OLD was developed in several stages (see also Figure 1): (1) review of 32 publications about early signs of Alzheimer dementia and signs of depression (because these two are frequently mixed up); (2) submission of 187 signs to 10 experts (six disciplines) in the field, asking them to judge the frequency and prevalence of the signs in patients with early and progressive dementia, and their specificity and observability; (3) reduction of numbers of signs by previously set criteria; (4) submission of 98 signs (75 dementia, 23 depression) to 27 experts (six disciplines; four general practitioners) in the field; (5) reduction of numbers by previously set criteria; (6) submission of 85 signs of dementia and depression to the four general practitioners, asking them to judge the observability in general practice; (7) reduction to 23 signs of dementia and eight of depression; (8) submission of the 23 signs of dementia to two experts (one general practitioner with special dementia expertise, one neurologist), asking them to add signs that are important but which were ‘lost’ in the reduction process; (9) initial version of the OLD with 29 early signs of Alzheimer dementia (the eight depression signs were removed).

The initial 29-signs version of the OLD included early signs of dementia in the following areas: forgetting 5, repeating 3, language 4, understanding 3, orientation in time and place 4, daily living 5, social withdrawal 1, confabulation 3, dependence (head-turning sign) 1. The initial development phase has been described in detail elsewhere (Hopman-Rock et al., 1998). Signs could be scored ‘present’, ‘doubt about presence’, ‘not present’, or ‘not observable’. The general practitioners were encouraged to develop their own methods to make signs observable, for instance, by asking certain relevant key questions like ‘how are things going these days?’ or ‘when was the last time that you visited me?’ or ‘how old are you now?’. The initial version was used to study the usefulness of the OLD in general practice before the definitive study of the reliability and validation of the OLD was performed.

Recruitment and participation of general practitioners

All participating general practitioners (19 practices, 22 general practitioners) were recruited by networking by the researchers and the committee of experts, using an information leaflet. The response was 67%; non-participating general practitioners could mostly not participate because of lack of time. None of the participating general practitioners was specialised in dementia. Their practices were in the western and southern part of the Netherlands. All general practitioners were given a short oral explanation of the study by one of the researchers. Thirteen of the recruited general practitioners also participated in a preliminary study to assess the usefulness of the OLD, especially the observability of the signs. Signs that could not be observed by more than 40% of the general practitioner’s had to be removed before the main study started.

The TNO Medical Ethics Committee approved the study protocol. Patients were informed about the
study by a poster in the waiting room, or by receiving an information leaflet after the observation took place during a home visit.

Procedure, inclusion and exclusion criteria

On 15 working days, the first two patients seen by the general practitioner were observed. The inclusion criterion was age 75 years and older. Data from the national Dutch registration system (Bakker et al., 1998) indicate a general practitioner has, on average, nine contacts per day with patients aged 75 or more. Exclusion criteria were psychiatric treatment, known diagnosis of dementia or severe depression, and acute illness with confusion. Thirty patients were observed per practice. Thus, in theory there would be almost 600 observations. The general practitioners received a reminder (desk standard) to help them to remember their tasks in the study and were monitored following the study protocol. They returned the completed OLDs to the research institute every 2 weeks. On all observation forms general practitioners were asked to indicate whether they thought that the patient was suspected
or not (‘would you have used the OLD spontaneously?’).

Reliability

The reliability of the OLD was measured using internal consistency (Cronbach’s alpha) and factor analysis. Test–retest was not possible because patients were observed during a real-life consultation. For the same reason, it was not possible to determine inter-observer reliability. The study protocol required a reliability analysis to be performed to identify less reliable signs, which had then eventually to removed from the OLD.

Validation

The reliable OLD was then used to divide the observed patients into three groups of approximately (because of research logistic reasons) 20 persons (group 1, no signs; group 2, intermediary number of signs; group 3, with the most signs). These three groups of patients were then invited to participate in the final validation study using interviews including existing valid and reliable measures.

To check the validity of the OLD, we looked for reliable and valid tests of dementia (convergent validity) and depression (discriminant validity) (Campbell and Fiske, 1959). We also expected some association with problems in the areas of activities of daily living (ADL) and Instrumental ADL (I-ADL) such as using the telephone and paying bills. Because of the burden on the patients, these tests had to be as short as possible and administered by trained interviewers (psychologists). The delay between the observation date and the test date was minimally 3 months (8 months maximally). All interviews were held in the patients’ homes during morning hours. The following tests were used: (1) The Cognitive Screening Test (CST) (Graaf and Deelman, 1991), the CST-20 version (20 points maximally). Scores for cognitive impairment are <12 if age <82 years, and <10 if age >81 years. (2) The Word Learning Test (WLT) (Deelman et al., 1980; Houx, 1991) which requires subjects to remember words written on sheets of paper. The WLT total score (total of words remembered in five trials of 15 words) and the WLT-retention score: (total of words remembered after 15–25 min/number of words remembered at fifth trial)×100 are mentioned. (3) The Geriatric Depression Scale (GDS) (Yesavage and Brink, 1983). The maximum score is 30. A score >9 is indicative of depression. (4) The Groningen Activities Restriction Scale (GARS) (Kempen et al., 1996). The maximum score 72 is indicative of disability. (5) The Fillenbaum IADL scale (Fillenbaum and Smyer, 1981). The maximum score 14 is indicative of no disability.

The GARS and IADL tests were administered, while the words of the WLT were being memorised. The GDS was the last test to be administered because of possible emotional reactions. Because some signs of starting dementia are best observed by relatives or friends, we asked relatives or friends to complete additional questionnaires, namely the short version of the Informant Questionnaire on Cognitive Decline in the Elderly (IQQCODE) (Jorm and Jacob, 1989; de Jonghe et al., 1991), which measures observed change in the past 10 years. Higher (>3) mean scores indicate more negative changes; the norm for cognitive decline is a mean score >3.6. The GARS and IADL scores were used as measures to detect eventual confabulation by analysing the differences between patients and their relatives (or caregivers). Difference scores were calculated as the score of the patient minus the score of the relative.

Construct validity was checked with a principal component analysis. Analysis of the opinions of the general practitioners (‘would you have used the OLD spontaneously?’) was used to check the incremental validity. Correlations with general practitioner–patient relationship characteristics and knowledge of the general practitioner about dementia (Buijsen, 1998) are given. Because OLD is not intended to be used as a screening instrument, no specificity or sensitivity can be given.

Statistical analysis

All analyses were carried out using SPSS 8.0 for Windows. The number of observations was based on the expected prevalence of diagnosed dementia in an older population (4% at age 70–75 years; power 80%). Analysis of (co)variance with Scheffe tests for multiple comparisons was used to detect statistically significant differences between the three groups. An alpha level of 0.05 was used. If variances were non-homogeneous, a non-parametric test was carried out (Kruskall-Wallis). Differences in nominal variables were analysed with χ²-tests. Data reduction in the validation phase was carried out with a principal component analysis on ordinal variables, using an alternating least-squares method (PRINCALS). Incremental validity was tested with a McNemar test (difference between opinion of general practitioner and presence of minimally one OLD sign). Correlations are given as Pearson r-values.
RESULTS

Test phase for observability of the initial OLD

Before the main study started, 13 general practitioners co-operated in three test observations with the initial OLD (with 29 signs to observe). These general practitioners completed an evaluation form about the observability of the signs in real practice, as a way of prevalidation. The initial 29-signs OLD took approximately 5 min to complete (range 2–10 min). These observations were not used for further analyses. Seven signs that were evaluated as not observable by more than 40% of the general practitioners were removed (these were all signs in the area of orientation and daily living). A list with 22-signs remained.

Reliability of the 22-signs OLD

Cronbach’s alpha for the 22-signs OLD (n = 470) was 0.88. Factor analysis showed a strong first factor (42.5% explained variance) and all signs had factor loadings > 0.40. Because no strong clues were found for reduction based on less reliability, the scores for the 22-signs OLD were used to divide the patients arbitrary into three groups, based on the frequencies of signs that were found. The characteristics of the study participants are given in Table 1. The difference in mean age between the three groups, within the total group, was significant (K-W: \( \chi^2 = 9.02 \), d.f. = 2, \( p = 0.011 \)), the groups with signs were older. Table 2 shows the response of the patients and the number of forms completed by relatives. \( \chi^2 \)-test showed no differences between non-response/dropout versus interview groups (\( \chi^2 = 6.25 \), d.f. = 4, \( p = 0.18 \)).

Convergent validity

The convergent validity results of the cognitive tests are shown in Table 3. Statistically significant results were found for CST, WLT, and IQCODE, indicating a good validity.

Because the scores on the CST-20 were non-homogeneously distributed, a non-parametric test was also performed, which showed the same significant result (K-W: \( \chi^2 = 10.6 \), d.f. = 2, \( p = 0.005 \)). Three people in group 3 had relatively high scores on the CST-20 and two had this on the WLT. One person suffered from pulmonary oedema when observed by his doctor some months before the interview, but his cognitive problems had resolved by the time of the interview. One person scored relatively high on the CST-20 because it was her husband’s birthday (easy to remember date, month, and year), but scored lower on the WLT. The third person had hearing problems and was very unsure and repetitive (in this case the OLD showed other signs than forgetfulness). One person from group 1 had a low score on the CST-20 (12 points): she had recently bruised her ribs and was very tired during the interview. Few patients had scores below 12 points (two in group 3 and one in group 2) because of the exclusion of patients with a known diagnosis of dementia. Covariate analysis showed that the time between observation and interview did not contribute significantly to score on the CST-20 and the WLT. The WLT retention score was only significantly lower in group 3. The IQCODE showed that group 1 differed from group 3, with the latter group showing cognitive decline.

No statistically significant differences were found for GARS and IADL (see Table 4). The family mean scores in the three groups were all in the direction of more GARS and IADL problems than

<table>
<thead>
<tr>
<th>Table 1. Characteristics of study participants*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean/SD)</td>
</tr>
<tr>
<td>Total group (n = 470)</td>
</tr>
<tr>
<td>Group 1 (n = 363)</td>
</tr>
<tr>
<td>Group 2 (n = 82)</td>
</tr>
<tr>
<td>Group 3 (n = 25)</td>
</tr>
<tr>
<td>Interview group (n = 60)</td>
</tr>
<tr>
<td>Group 1 (n = 27)</td>
</tr>
<tr>
<td>Group 2 (n = 22)</td>
</tr>
<tr>
<td>Group 3 (n = 11)</td>
</tr>
</tbody>
</table>

*Group 1: without any observed signs of dementia; Group 2: 1–5 signs of dementia; Group 3: > 5 signs of dementia.
Table 2. Response and the number of completed family forms

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of cases</th>
<th>Number observed</th>
<th>Number approached</th>
<th>Non-response</th>
<th>Dropout</th>
<th>Number with interview</th>
<th>Family forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>363</td>
<td>51</td>
<td>21</td>
<td>3</td>
<td>27</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>82</td>
<td>61</td>
<td>32</td>
<td>7</td>
<td>22</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>25</td>
<td>25</td>
<td>9</td>
<td>5</td>
<td>11</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>470</td>
<td>137</td>
<td>62</td>
<td>15</td>
<td>60</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

*Group 1: without any observed signs of dementia; Group 2: 1–5 signs of dementia; Group 3: > 5 signs of dementia.

Table 3. Results of convergent validity analyses on cognitive tests

<table>
<thead>
<tr>
<th>Group</th>
<th>CST-20 (mean, SD)</th>
<th>WLT-total (mean, SD)</th>
<th>WLT-retention (mean, SD)</th>
<th>ICODE (mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>18.7 (1.7)</td>
<td>38.6 (9.9)</td>
<td>81.1 (25.3)</td>
<td>n = 17 3.2 (0.29)</td>
</tr>
<tr>
<td>Group 2</td>
<td>17.4 (2.5)</td>
<td>30.0 (9.2)</td>
<td>72.9 (26.5)</td>
<td>n = 12 3.4 (0.55)</td>
</tr>
<tr>
<td>Group 3</td>
<td>14.4 (3.7)</td>
<td>24.3 (14.5)</td>
<td>35.0 (38.3)</td>
<td>n = 8 3.7 (0.59)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>F, p</td>
<td></td>
<td></td>
<td>2.6 (0.089)</td>
</tr>
</tbody>
</table>

*Group 1: without any observed signs of dementia; Group 2: 1–5 signs of dementia; Group 3: > 5 signs of dementia.

Table 4. Scores on the GARS, IADL and GDS

<table>
<thead>
<tr>
<th>Group</th>
<th>GARS (mean, SD)</th>
<th>IADL (mean, SD)</th>
<th>GARS-difference* (mean, SD)</th>
<th>IADL-difference* (mean, SD)</th>
<th>GDS (mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>25.1 (8.6)</td>
<td>12.4 (2.4)</td>
<td>-4.4 (10.6); n = 18</td>
<td>0.0 (1.2); n = 18</td>
<td>6.9 (4.9)</td>
</tr>
<tr>
<td>Group 2</td>
<td>27.8 (8.8)</td>
<td>11.9 (2.6)</td>
<td>-1.81 (9.4); n = 11</td>
<td>0.9 (2.7); n = 12</td>
<td>8.7 (5.6)</td>
</tr>
<tr>
<td>Group 3</td>
<td>24.4 (8.7)</td>
<td>11.6 (3.5)</td>
<td>-4.6 (7.4); n = 8</td>
<td>0.7 (2.4); n = 7</td>
<td>6.0 (4.8)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>0.77 (0.47)</td>
<td>0.35 (0.70)</td>
<td>0.29 (0.75); n = 8</td>
<td>0.72 (0.49)</td>
<td>1.23 (0.30)</td>
</tr>
</tbody>
</table>

*Group 1: without any observed signs of dementia; Group 2: 1–5 signs of dementia; Group 3: > 5 signs of dementia.

Discriminant validity

Because the signs of depression were removed from the initial OLD and people with severe depression were excluded, we expected, and found, there to be no significant relationship between the OLD score and the GDS score (see Table 4). There were no

the patients indicated themselves. A significant interaction effect for the difference scores on the IADL was found (group X gender, $F = 3.6$, $p = 0.04$), indicating that more difference was found for women in group 2 and for men in group 3. No relationship was found with confabulation signs in the OLD ($t$-test).
differences between the three groups \((F=1.2, p=0.30)\), the mean score for all groups being \(< 9\) (indicating no depression).

**Construct validity**

A PRINCALS with ordinal scores on the OLD (higher score = sign observed) was carried out in two dimensions. The quartile scores on the CST-20, WLT total, and WLT retention were added in the analyses (higher quartile = fewer cognitive problems). The first component explained 48% of the total variance, the second 14%. All signs except the sign ‘the patient does not start a conversation’ (loading 0.38) loaded \(> 0.50\) on the first dimension. Component loadings: CST-20 \((-0.7; -0.38)\), WLT total \((-0.69; -0.38)\), WLT retention \((-0.60; -0.23)\). Highest component loadings were found on the second dimension for ‘keeps forgetting the current day’ (0.78; 0.62), and ‘regularly repeats the same story during conversation’ (0.76; -0.61).

**Incremental validity**

The general practitioners’ opinion was tested against the 22-signs OLD \((n=59)\); group 1 versus groups 2 and 3 together). All 17 patients whom the general practitioners considered to be in an early stage of dementia showed one or more of the OLD signs. However, of the other 42 non-suspect patients, 16 (38%) exhibited signs. The difference between the opinion of the OLD and the general practitioner was significant (McNemar, \(p=0.00\)). Of these 16 patients, 14 were from group 2 \((n=22, 1–5\) signs) and two from group 3 \((n=11, > 5\) signs).

In the total group of 470 patients, the general practitioners considered 55 patients as suspect, of whom 16 (29%) did not show any of dementia signs. The general practitioners considered 413 patients as non-suspect, of whom 67 (16%; five of whom with more than five signs) showed minimally one sign (McNemar, \(p=0.00\)). No relationships were found between the number of positive signs and general practitioner–patient characteristics such as the frequency of visits \((r=0.08)\), years of relationship with general practitioner \((r=0.03)\), and familiarity with the patient \((r=0.002)\). A small correlation was found between the number of observed signs and age \((n=470, r=0.09, p<0.05)\). Interestingly, in the group with more than five signs \((n=25)\) the correlation of the number of signs with the number of years of the doctor–patient relationship was \(-0.36 (p<0.10)\) and with age \(-0.40 (p<0.05)\). No relationship has been found with the general practitioners knowledge about dementia \((r=0.05)\). The time needed for completion of the 22-signs version of the OLD was approximately 4 min.

**Reduction of number of signs**

It was decided to reduce the number of signs further to 12 signs (the number was chosen arbitrarily) to make OLD as useful as possible in general practice, by removing signs in subgroups with the lowest component loadings on the first dimension. In this way, a short version of the OLD with 12 signs was created with the following subgroups of signs: forgetting (3); repeating (2); language (2); understanding (2); orientation in time (1); confabulation (1); dependence (1). The reliability of this short version of the OLD was 0.89 in the validation sample \((n=48)\) and 0.86 in the total sample \((n=470)\). Pearson correlations to give an indication of validity were: \(-0.49\) (CST-20), \(-0.40\) (WLT-total), \(-0.43\) (WLT-retention), 0.38 (IQCODE), \(-0.17\) (GARS), \(-0.06\) (IADL), and \(-0.15\) (GDS). The correlations with GDS, GARS, and IADL were statistically not different from zero. A McNemar test on the difference between the OLD score (one sign minimally) and the opinion of the general practitioner for the short version of the OLD, was also significant \((p=0.019)\), indicating good incremental validity.

**DISCUSSION**

This paper describes a new, patient-friendly approach to the early detection of possible Alzheimer disease by general practitioners. This reliable and valid observation list was designed on the basis of expert opinion about early signs and symptoms of Alzheimer’s disease. Both long (22 signs) and short (12 signs) versions of the OLD are available. The usability in general practice is very promising, because the OLD is easy to administer and takes only a few minutes to complete. Following the conclusion of the study, 15 of the 22 general practitioners who had participated stated that they intended to use OLD in the future, clearly indicating that OLD can play a useful role in practice settings.
General practitioners could use the OLD as an aid for observing and monitoring patients (‘watchful waiting’) and as a tool for determining whether a patient should be referred to a specialist or whether they have to use further assessment of possible dementia with known psychometric (screening) instruments (Heun et al., 1998; Brooke and Bullock, 1999). An important advantage of OLD is that patients can be assessed at multiple time points. As stated by Morris et al. (1999), change in performance is a superior outcome measure in studies of cognitive decline because many factors other than the disease can influence test performance. In our study, this was clearly the case in one patient who had relatively many early signs of Alzheimer but who did not show cognitive impairment several months later. In this case the OLD scores could be explained by pulmonary oedema.

A problem with this study was that many observations had to be made to obtain a sufficient number of cases of starting dementia. Although the number of cases was based on the prevalence of dementia of the Alzheimer type in an older population, the patients in the small group with relatively many signs of dementia (response 64%, which is high) were as weak as expected on that age (death, illnesses). This meant that fewer patients than expected participated in the validation study (11 versus 20). Nevertheless, the expected significant effects between the study groups were found and thus the power of the study was sufficient.

An important aspect of the OLD is its incremental validity. If the general practitioner had a valid intuitive opinion about signs of starting dementia, he or she did not need to use the OLD at all. Apparently, this was not the case: the OLD clearly adds to the general practitioner’s opinion.

Another important aspect is the construct validity. Because all signs were selected on the basis of the opinion of experts, we expected the first dimension of the principal component analysis to be strong, indicating possible starting dementia. This was definitely the case. Besides signs of forgetfulness, other aspects of cognitive impairment (such as problems with language and with understanding) and behaviour added to the first dimension. Unfortunately, there are currently no instruments available to validate these other aspects. The results of the principal component analysis were used to reduce the number of signs in the 22-signs (long-form OLD) into a short form with 12 signs, including all areas of impairment.

The convergent (cognitive impairment and decline) validity was very good and was evaluated with three different instruments (CST, WLT, and IQCODE). The difference in the WLT scores between the two groups with fewer or more than five signs on the OLD was interesting. It indicated that patients with relatively many signs had poorer memory function when it came to retrieving information after 15–25 min. The expected relationship between the OLD score and problems with activities of daily living was not found. It is unclear whether this is because of measurement problems (sensitivity of this kind of tests in these groups, and differences between men and women) or because people with starting dementia experience few problems in this area. Most general practitioners find it difficult to detect problems in ADL. Interestingly, the relatives or friends completed the ADL lists differently from the patients, irrespective of whether the patient showed many or none of the OLD signs.

As expected and designed, the OLD score did not correlate with the GDS score for depression. This indicated a good discriminant validity. Other tools are available for the general practitioner to measure depression in clinical practice (Goldberg et al., 1988; Tiemens et al., 1995).

In group 3 (patients with more than five early signs of dementia), the number of signs was negatively correlated with age. A possible explanation is that men with dementia have higher mortality risks by atherosclerosis and cerebrovascular disease (Gusseklo et al., 1999, submitted manuscript).

The reliability and validity of both the long and short versions of the OLD are very good. However, the final validation of an instrument is its predictive value: which patient develops Alzheimer dementia and which patient does not. This question cannot be answered now (because the convergent measures that we used are screening instruments and do not give a valid diagnosis) and must await 5-year follow-up data for the subjects included in this study. In the implementation phase of the OLD, more recommendations for use in general practice can be developed.

CONCLUSION

The newly developed OLD offers a good opportunity for general practitioners to observe and monitor their older patients for early signs of Alzheimer dementia. The method is reliable, valid, patient-friendly, and easy to use.
KEY POINTS
- Development and validation of the observation list for early signs of dementia (OLD).
- Reviews of publications and expert consensus (development); field study with 22 general practitioners and interviews with standard cognitive tests (validation).
- Observation of 470 patients with OLD by general practitioner; validation by interviews with 60 selected patients and information of 39 family members.
- OLD is a valid and reliable method to detect early signs of dementia in general practice.

ACKNOWLEDGEMENTS
This study was financially supported by Janssen-Cilag B.V., Tilburg, The Netherlands. We thank M. Blom, T. Bosch, W. Davidsen, C. Jonker and AW. Wind of the expert committee in the validation phase for their helpful comments. We thank C. ten Kate for her work in the initial development phase. P. van der Heijden is gratefully acknowledged for his statistical support. The OLD is available at TNO Prevention and Health (request to the first author), PO Box 2215, 2301 CE Leiden, The Netherlands.

REFERENCES

Copyright © 2001 John Wiley & Sons, Ltd.