

Cognitive & Behavioral Assessment

# Scientific and ethical features of English-language online tests for Alzheimer's disease

Julie M. Robillard<sup>a,\*</sup>, Judy Illes<sup>b</sup>, Marcel Arcand<sup>c</sup>, B. Lynn Beattie<sup>d</sup>, Sherri Hayden<sup>e</sup>, Peter Lawrence<sup>f</sup>, Joanna McGrenere<sup>g</sup>, Peter B. Reiner<sup>h</sup>, Dana Wittenberg<sup>i</sup>, Claudia Jacova<sup>j,1</sup>

<sup>a</sup>National Core for Neuroethics, Division of Neurology, Department of Medicine, Faculty of Medicine, University of British Columbia, Djavad Mowafaghian Centre for Brain Health, Vancouver, BC, Canada

<sup>b</sup>National Core for Neuroethics, Division of Neurology, Department of Medicine, Faculty of Medicine, University of British Columbia, UBC Hospital, Vancouver, BC, Canada

<sup>c</sup>Department of Family Medicine, Faculty of Medicine and Health Sciences, University of Sherbrooke, Sherbrooke, QC, Canada

<sup>d</sup>Division of Geriatric Medicine, Department of Medicine, Faculty of Medicine, University of British Columbia, Gordon and Leslie Diamond Health Care Centre, Vancouver, BC, Canada

<sup>e</sup>Division of Neurology, Department of Medicine, Faculty of Medicine, University of British Columbia, UBC Hospital, Vancouver, BC, Canada

<sup>f</sup>Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada

<sup>g</sup>Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

<sup>h</sup>National Core for Neuroethics, Department of Psychiatry, Faculty of Medicine, University of British Columbia, UBC Hospital, Vancouver, BC, Canada

<sup>i</sup>Division of Neurology, Department of Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

<sup>j</sup>Division of Neurology, Department of Medicine, Faculty of Medicine, University of British Columbia Hospital, Vancouver, BC, Canada

## Abstract

**Introduction:** Freely accessible online tests for the diagnosis of Alzheimer's disease (AD) are widely available. The objective of this study was to evaluate these tests along three dimensions as follows: (1) scientific validity; (2) human-computer interaction (HCI) features; and (3) ethics features.

**Methods:** A sample of 16 online tests was identified through a keyword search. A rating grid for the tests was developed, and all tests were evaluated by two expert panels.

**Results:** Expert analysis revealed that (1) the validity of freely accessible online tests for AD is insufficient to provide useful diagnostic information; (2) HCI features of the tests are adequate for target users, and (3) the tests do not adhere to accepted ethical norms for medical interventions.

**Discussion:** The most urgent concerns raised center on the ethics of collecting and evaluating responses from users. Physicians and other professionals will benefit from a heightened awareness of these tools and their limitations today.

© 2015 The Alzheimer's Association. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Keywords:

Alzheimer's disease; Online health; Diagnosis; Online tests; Aging; Self-assessment; Ethics

## 1. Background

Online health information and services for a variety of medical conditions have been available since the early days of the Internet but have become increasingly popular

J.M.R., J.I., M.A., P.L., P.B.R., and D.W. have no conflicts of interest to disclose.

<sup>1</sup>Present address: School of Psychology, Pacific University, Creighton Hall 295, 222 SE 8th Ave., Hillsboro, OR 97123, USA.

\*Corresponding author. Tel.: 1-604-827-0642; Fax: 1-604-827-5229.

E-mail address: [jrobilla@mail.ubc.ca](mailto:jrobilla@mail.ubc.ca)

<http://dx.doi.org/10.1016/j.dadm.2015.03.004>

2352-8729/© 2015 The Alzheimer's Association. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

since the rise of eHealth in the late 1990s [1]. In 2012, 72% of Internet users reported looking online for health information [2]. Online self-diagnosis is an increasingly prevalent behavior associated with Internet use: 35% of US adults report turning to the Internet specifically to establish whether they or someone they know suffer from a medical condition, and 70% of adults track at least one health indicator online [3]. A significant portion of people say they are influenced by online health information [2] and believe that this information can benefit their health [4].

The older adult population also participates in online information-seeking: over half of adults aged  $\geq 65$  years use the Internet and this figure rises to over three-quarters for adults aged 50–64 [3], with a majority of these users turning to the Internet for health information [2]. As a result, online resources are emerging to meet the demands of this demographic group and many self-diagnostic or screening tests are now available for a variety of age-associated conditions. Alzheimer's disease (AD) and other dementias are among the health issues most feared by older adults and their families [5], and freely accessible online tests for mild cognitive impairment (MCI), AD and unspecified dementia are currently available on high-traffic Web sites reaching up to several million users each month. However, little is currently known about the quality, the content, and the potential impact of self-tests for these conditions.

No doubt, online testing for MCI and AD/dementia has benefits: users feel empowered, they can access information when traditional health services are difficult to reach, they may become motivated to seek medical advice [6,7], or self-monitor for preventive purposes [8]. However, the use of online health services is not without risks. The limited oversight of online tests in particular may lead to inaccurate self-diagnosis and treatment [9]. Users may misinterpret graphical displays of risk and associated terminology [10], which can lead to inappropriate actions with negative health consequences [9,11]. Further confusion may arise when unregulated and potentially commercially driven Web site content is represented as educational [12]. The adverse psychological impact of poor cognitive test results could be magnified by the lack of inperson support and counseling. These risks are particularly relevant when the users of test resources are older adults with any degree of cognitive impairment.

The assessment of MCI and AD/dementia routinely includes self or third party reports on current functioning and performance tests based on standardized cognitive tasks [13]. Computerized versions of cognitive tests are an accepted assessment method in clinical settings [14,15]. There remains, however, a number of issues that limits their utility in, and particularly outside the clinic setting. Computerized adaptations of cognitive tests cannot be assumed to have the same psychometric properties as their paper-and-pencil counterparts and require their own validation studies [16]. The computerized interface must have demonstrated usability for older adults with variable cognitive and computer skills [14]. For tests that are self-administered in an online format, uncontrolled person and environmental variables including inadequate understanding of test instructions, distractibility, and interruptions pose further challenges [17]. There are also issues specific to the noninteractive and impersonal online test context including information on the test, consent, confidentiality, and the nature of the disclosed test results. Beyond specific considerations of the tests themselves,

self-diagnostic tools in the unregulated online environment may not take into consideration the evolving definition of dementia [18] and the ethical conflict of diagnosing a condition for which no broadly effective treatments exist [19,20]. Finally, the emphasis on early detection and diagnosis promoted in online self-assessments often ignores the broader cultural and social processes that contribute to aging and dementia [19,21]. There is a dearth of research on the psychological and health effects of cognitive screening on older adults, and to our knowledge, no research on such effects associated with online cognitive test information.

The aim of the current work was to evaluate English-language online self-assessment tools for MCI and AD/dementia through expert panel reviews. We identified a sample of existing online tests and evaluated them on three dimensions of interest as follows: (1) scientific validity, (2) human-computer interaction (HCI) features; and (3) ethics variables.

## 2. Methods

### 2.1. Identification and characterization of tests

We identified publicly available online tests for AD through a keyword search on Google, the leading search engine on the Internet (comScore, June 2013), using a location-independent search with a clear search history. To capture the largest possible sample, we used combinations of the following keywords: "online," "Alzheimer," "memory," "test," "evaluation," and "free." As evidence suggests that most online users do not look beyond two pages of search results when using a search engine [22], we limited our search to the first two pages of results (total of 20 returns per combination of keywords). Initial keyword searches were conducted in July 2012 and repeated in January 2013. Inclusion criteria were that the test (1) is in the English language; (2) contains an "Alzheimer" keyword in the test title, header, claim, or outcome; (3) is freely accessible (e.g., does not require a special password or a login from a physician); (4) can be administered either online or in print through materials downloaded online; and (5) is designed to yield outcomes that are available to the test-taker. Exact duplicate search returns were excluded, and 16 unique tests were retained for analysis.

After retrieval, each test was characterized by the lead author based on the following criteria: (1) type of site hosting the test; (2) monthly traffic to parent site as determined by analytics provider Compete; (3) test claim(s); (4) test length or number of question; (5) administration (self or through third party); (6) types of questions (questionnaire-based or performance-based); and (7) possible outcomes. Each test was taken several times by the lead investigator, varying performance and answers, to elicit all possible outcomes and screen captures were taken for each of the outcomes (e.g., pass or fail).

2.2. Development of the rating grid

A rating grid was constructed based on issues highlighted in reviews on computerized cognitive assessments of older adults [14,23] and on author expertise. The rating grid included 17 items to evaluate each of the three dimensions of interest (validity-4; HCI-5, and ethics-8) and an overall quality item (Fig. 1). All items are described in full in Fig. 1. Reviewers were asked to rate each test on the 18 items using a 10-point Likert scale (1,2: very poor, 3,4: poor, 5,6: acceptable, 7,8: good, and 9,10: excellent).

2.3. Expert panel review

In line with other assessments of the quality of health information on the Internet, we recruited two panels of four members each to review the tests [24]. Members of both panels had expertise in aging and dementia, and members of panel 1 additionally had specific expertise in the field of computerized testing for the clinical setting. Each panel consisted of a geriatrician, neuropsychologist, HCI expert, and ethicist. Expertise in aging and dementia was defined as follows: for clinicians, weekly interactions with older adults with or at risk for dementia; for researchers, at least one active research project related to dementia.

Each expert reviewer rated the tests over two 3-hour sessions. Sessions were conducted with one or two experts in the room but they did not interact about the evaluation when more than one was present. All sessions were moderated by the lead investigator and included an introduction to the project, a review of the rating grid and rating system, and a review of the rating procedure. The rest of the sessions unfolded in the following pattern:

- 1) Moderator shares computer screen through a projector or through screen sharing and takes a test online in real-time for the expert reviewer(s) to view.
- 2) All possible outcomes for the test are presented as screen captures for expert evaluation.
- 3) Additional information (if applicable, e.g., frequently asked questions rubrics, guidelines for the interpretation of results) is shown to the experts in the form of screen captures.
- 4) Experts are invited to evaluate the test on the 18 rating items.

2.4. Data analysis

Arbitrary numbers were assigned to each test so as to maintain its anonymity.

To test for overall expert impressions of the online tests, we computed interrater agreement with intraclass

| Online tests for Alzheimer disease and mild cognitive impairment<br><i>Please rate each characteristic with a number from 1 (extremely poor) to 10 (excellent)</i> | Expert panel review |   |   |   |   |   |   |   |
|--|---------------------|---|---|---|---|---|---|---|
|  | TEST                |   |   |   |   |   |   |   |
| VALIDITY AND RELIABILITY   | 1                   | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Is the content test appropriate to achieve its claim?  |                     |   |   |   |   |   |   |   |
| Is the breadth of the test appropriate to achieve its claims?  |                     |   |   |   |   |   |   |   |
| Are the test questions based on current, peer-reviewed evidence?   |                     |   |   |   |   |   |   |   |
| Do you think this test would have test-retest reliability?   |                     |   |   |   |   |   |   |   |
| INTERFACE  |                     |   |   |   |   |   |   |   |
| Are the test instructions clear and easy to understand?  |                     |   |   |   |   |   |   |   |
| Is the test visually adequate (font size, etc.)?   |                     |   |   |   |   |   |   |   |
| Is the test interface easy to use?   |                     |   |   |   |   |   |   |   |
| Is there a mechanism to evaluate whether performance is valid?   |                     |   |   |   |   |   |   |   |
| Does the test take varying levels of computer knowledge into consideration?  |                     |   |   |   |   |   |   |   |
| ETHICS   |                     |   |   |   |   |   |   |   |
| Is there sufficient and clear information prior to starting the test?  |                     |   |   |   |   |   |   |   |
| Is there an informed consent/comprehension check process before the start?   |                     |   |   |   |   |   |   |   |
| Are issues of privacy, data collection and confidentiality discussed?  |                     |   |   |   |   |   |   |   |
| Are the scope and the limitations of the test clearly stated?  |                     |   |   |   |   |   |   |   |
| Is the wording of the outcomes ethically appropriate?  |                     |   |   |   |   |   |   |   |
| Is the interpretation of the outcomes clear?   |                     |   |   |   |   |   |   |   |
| Are conflicts of interest clearly stated?  |                     |   |   |   |   |   |   |   |
| Do you receive appropriate advice on follow-up after taking the test?  |                     |   |   |   |   |   |   |   |
| OVERALL  |                     |   |   |   |   |   |   |   |
| How would you rate the overall quality of the test?  |                     |   |   |   |   |   |   |   |

Fig. 1. Rating grid. All items to be rated on Likert scale ranging from 1 (very poor) to 10 (excellent).

correlation coefficients (ICC). We calculated absolute agreement with a two-way random effects model. For ICCs <0.60 (where <0.40 poor, 0.40–0.59 fair, 0.60–0.74 good, and >0.74 excellent), we performed follow-up comparisons of panel (1 or 2) and expertise (clinician, neuropsychologist, HCI expert, and ethicist) with nonparametric tests (Mann-Whitney for panels; Kruskal-Wallis for experts). We computed global scores for each expert by averaging the ratings he or she assigned to each of the tests evaluated (global item score =  $M_{\text{Test } 1-n}$ ). As such, global scores for each expert had the same metric as the original rating scale composed of the 17 evaluation criteria, with a range of 1–10, and reflected experts' overall impression of the test sample. We evaluated descriptive data for these scores with experts as cases.

To further characterize the identified tests themselves, we averaged ratings across experts for each test on each item (global expert score =  $M_{\text{Expert } 1-k}$ ) and evaluated descriptive data with tests as cases.

### 3. Results

#### 3.1. Test characteristics

The keyword search strategy yielded 16 unique tests for which names, references, and detailed characteristics are reported in Table 1. All tests retrieved were in the English language. We found 11 different types of Web sites hosting online tests for AD, including news organizations, not-for-profit organizations, commercial entities, and entertainment platforms. Monthly unique visitors to the sites, as assessed by Web site analytics provider Compete, ranged from 200 to 8.8 million. Of the 16, 12 were designed to be self-administered and 4 were designed to

be administered by a friend or relative, according to the test instructions.

Time to completion ranged between 2 minutes to approximately 30 minutes based on experimenter testing over three trials. Tests were questionnaire-based (6/16) or performance-based (10/16). Questionnaire-based tests typically required responses about behaviors (e.g., “Do you forget the names of loved ones?”), risk factors (e.g., “Do you use underarm antiperspirant?”), or both. Performance-based tests measured various aspects of cognitive function including digit recall, arithmetic, and memory for words or faces.

Nine tests provided outcome information in two or three categories (example for a test with two categories of outcomes: “Not at risk for AD/At risk for AD”). Two tests used a pass/fail scheme. Three described user results by providing a unique score or result on a continuum of possibilities. One test yielded only one possible outcome regardless of input data: severe risk of AD. One test did not result in any outcomes.

#### 3.2. Overall quality

Table 2 shows descriptive data for the test-averaged item ratings. Overall quality averaged across the 16 tests ranged from very poor to poor, with a mean rating of 2.8 and a narrow 1.7-point range. Interrater agreement was 0.87 [41].

#### 3.3. Scientific validity and reliability

Experts rated the appropriateness of test content and test breadth to achieve the goal described in the test claims (e.g., “Determine your risk for AD”) as very poor to poor (interrater agreement: 0.86 [content], 0.90 [breadth]). Test grounding in peer-reviewed literature was also rated as

Table 1  
Test characteristics

| Web site name                     | Type of host       | Monthly traffic* | Length | Administration | Types of questions | Possible outcomes |
|-----------------------------------|--------------------|------------------|--------|----------------|--------------------|-------------------|
| Nutritional test [25]             | Commercial 1       | 1900             | 20 Q   | Self           | Questionnaire      | Fail              |
| CogniCheck [26]                   | Commercial 2       | 200              | 30 min | Self           | Performance        | Continuum         |
| PreventAD [27]                    | Commercial 3       | Not available    | 15 Q   | Self           | Questionnaire      | Three categories  |
| Dr Dharma [28]                    | Commercial 4       | Not available    | 15 min | Self           | Questionnaire      | Two categories    |
| Dr Oz [29], <sup>†</sup>          | Entertainment 1    | 3.0 M            | 17 Q   | Self           | Performance        | Three categories  |
| Memozor [30]                      | Entertainment 2    | 1100             | 20 Q   | Self           | Questionnaire      | Two categories    |
| Cognitive Labs [31]               | Research 1         | 1400             | 3 min  | Self           | Performance        | Continuum         |
| MemTrax [32]                      | Research 2         | 1800             | 3 min  | Self           | Performance        | Three categories  |
| My Brain Test [33]                | Market research    | 900              | 3 Q    | Self           | Performance        | Pass/fail         |
| Daily Mail Online [34]            | News               | 8.8.M            | 21 Q   | Proxy          | Questionnaire      | Three categories  |
| Health24 [35]                     | Health news        | 64,000           | 6 Q    | Proxy          | Performance        | None              |
| Food for the Brain [36]           | Not-for-profit     | 8000             | 15 min | Self           | Performance        | Continuum         |
| Free Online Alzheimer's Test [37] | Aggregator         | 5000             | 4 Q    | Self           | Performance        | Pass/fail         |
| On Memory [38]                    | Advocacy           | 1200             | 11 Q   | Proxy          | Questionnaire      | Two categories    |
| SAGE [39]                         | Academic           | 800              | 15 min | Self           | Performance        | Three categories  |
| Way of the Mind [40]              | Health information | Not available    | 20 Q   | Proxy or self  | Performance        | Two categories    |

\*Traffic to parent site as per analytics provider Compete.

<sup>†</sup>Page no longer available as of January 9, 2015.

Table 2  
Descriptives for test-averaged item ratings with experts as cases

| Item                   | Mean | Median | SD  | Minimum | Maximum | ICC average measures |
|------------------------|------|--------|-----|---------|---------|----------------------|
| Content                | 3.5  | 3.4    | 0.7 | 2.4     | 4.6     | 0.86                 |
| Breadth                | 3.3  | 3.3    | 0.6 | 2.4     | 4.4     | 0.90                 |
| Peer review            | 3.6  | 3.7    | 0.7 | 2.9     | 4.9     | 0.88                 |
| Reliability            | 4.2  | 4.3    | 1.0 | 2.8     | 5.7     | 0.54                 |
| Instructions           | 5.0  | 4.8    | 1.4 | 2.6     | 7.3     | 0.69                 |
| Visual quality         | 5.3  | 5.3    | 1.7 | 3.0     | 8.0     | 0.76                 |
| Interface quality      | 5.6  | 5.5    | 1.3 | 3.7     | 8.3     | 0.75                 |
| Performance check      | 2.9  | 2.0    | 2.3 | 1.5     | 8.2     | 0.50                 |
| Computer knowledge     | 4.1  | 3.2    | 2.5 | 1.6     | 8.7     | 0.34                 |
| Information            | 3.8  | 3.8    | 0.9 | 2.2     | 4.8     | 0.71                 |
| Consent                | 1.9  | 1.9    | 0.3 | 1.6     | 2.2     | 0.94                 |
| Privacy                | 1.8  | 1.8    | 0.4 | 1.3     | 2.3     | 0.93                 |
| Scope                  | 2.7  | 2.4    | 1.0 | 1.6     | 4.1     | 0.74                 |
| Outcome wording        | 3.1  | 3.1    | 0.8 | 1.6     | 4.3     | 0.83                 |
| Outcome interpretation | 3.0  | 3.0    | 0.7 | 1.7     | 3.8     | 0.81                 |
| Conflict of interest   | 1.4  | 1.3    | 0.3 | 1.1     | 3.5     | 0.48                 |
| Advice                 | 2.8  | 3.1    | 0.8 | 1.1     | 3.5     | 0.75                 |
| Overall quality        | 2.8  | 2.7    | 0.6 | 1.9     | 3.6     | 0.88                 |

Abbreviations: SD, standard deviation; ICC, intraclass correlation coefficients.

very poor to poor (interrater agreement: 0.88). Ratings of test reliability were in the poor to acceptable range (interrater agreement: 0.54), with neuropsychologists rating this criterion lower than other types of experts. Experts and panels did not differ significantly in their ratings.

### 3.4. Human-computer interaction

Ratings for HCI items were variable, with a range from very poor to good. Clarity of the instructions was rated as acceptable (interrater agreement: 0.69). Panel and expert rankings did not differ significantly. Quality of the visual display (e.g., font size, color contrast) and interface (e.g., size of clickable areas) was also rated as acceptable, with good to excellent interrater agreement (0.75, visual display, 0.76, interface). Experts found that the tests lacked a mechanism to detect invalid performance such as simply clicking through the test (interrater agreement: 0.50) and that the tests did not accommodate different levels of computer knowledge (interrater agreement: 0.34). Differences in ratings between experts and panels were not statistically significant but we noted a trend toward HCI experts providing higher ratings than the other experts.

### 3.5. Ethics items

Expert ratings of ethics items were generally low, with most ethics items scoring within a very narrow range and with high interrater agreement. They rated the introductory information provided by the tests as very poor to

poor (interrater agreement: 0.71), and ratings concerning the process of informed consent and the disclosure of measures for privacy and confidentiality were also very poor to poor (interrater agreements: 0.93 and 0.94). Experts also rated posttest debriefing process, including wording and interpretability of results, as very poor to poor (interrater agreement: 0.81 and 0.83). There was insufficient acknowledgement of the limitations of the test (interrater agreement: 0.74) and advice on follow-up (interrater agreement: 0.75). Expert ratings for disclosure of conflicts of interests ranged from very poor to poor (interrater agreement: 0.48 with some tests scoring low across all experts and others scoring from low to high). For all ethics items for which interrater agreement was <0.60, there were no significant differences associated with expertise or panel. We further scrutinized the disclosure of conflict of interest item data and noted that a handful of tests received either low or moderately high ratings from experts, with no discernible relationship to panel and expertise.

### 3.6. Ranking of tests

In addition to the analysis of scores with experts as cases, we also analyzed scores averaged across experts with tests as cases. For this analysis, each test has one score averaged across experts for each item of the scoring grid. Using this analytic method, we found that expert ratings portrayed the sample of 16 tests as having a quality range from very poor to good on most validity and HCI items. Overall, the quality of individual tests ranged from very poor to acceptable, with a 6-point range. The ability to ensure valid performance, provision of advance and debriefing information, and protection of privacy and confidentiality of the test data were found at best only acceptable. Disclosure of conflict of interest was found very poor for all tests. Test rankings were highly variable, with no test meeting good quality standards on all items. Conversely, no test was consistently poor on all items, suggesting that each of the sampled tests presents a mix of acceptable and problematic qualities.

## 4. Discussion

The results of this study show that freely accessible online tests for AD, as a whole, are of limited value: (1) they have insufficient scientific validity and reliability to provide accurate information about cognitive functioning and (2) they lack adherence to ethical norms that conventionally safeguard people, patients, and other users from disclosures of confidential information and promote transparency and informed decision making.

Various tools such as LIDA, DISCERN, the HONcode and the Michigan consumer health Web site checklist [42,43] have been developed to evaluate the quality of

online health information for specific diseases or conditions. The rating grid we developed for the present study built on past work and takes into consideration the specific issues and challenges of a self-assessment. Regardless of which tool is used, the findings are consistent across studies: variability in quality and reliability [9,44,45] and limited attention to ethics features such as disclosures of conflicts of interest [46–49].

Poor adherence to ethical norms was the greatest concern regarding online testing for MCI and AD/dementia. In the online setting where there is lack of oversight and regulation, users of test resources may not be able to discern predatory organizations with conflict of interest. In the present study, we found that the conflict of interest item as pertaining to online tests appears to have been interpreted differently by experts. This finding supports the need to develop a consensus around what constitutes a conflict of interest for a Web site hosting online assessments for AD and more broadly for other medical conditions. Aside from issues of conflict of interest, older adults have more difficulty detecting untrustworthy information than younger adults [50] and this difficulty may be exacerbated by cognitive impairment. Scientifically invalid testing procedures may lead to incorrect self-assessments, with potentially serious consequences for health: false positives can lead to anxiety and to an increased burden for the health care system [51], whereas false negatives may contribute to a false sense of well-being when further testing should be carried out, and delay critical services and interventions. Regardless of their validity and adherence to ethical norms, online tests may also impact health through the patient-physician relationship, as online health information in general is increasingly shown to alter the ways in which patients and their families interact with health care providers [52,53]. Beyond the self-assessments themselves, the Web sites hosting online tests are far-reaching, easily accessible platforms that may lead to the wide dissemination of inappropriate messages and misinformation about the diagnostic process for AD [54]. Dementia, type of dementia (AD), and MCI are clinical diagnoses for which cognitive impairment is a necessary but a not sufficient criterion. As such, irrespective of the validity or ethical context of an online test, no online cognitive test should claim to diagnose dementia, AD, or MCI. On this basis, and despite the fact that online tests may serve to open a dialogue about dementia, we suggest that currently the risks associated with online testing outweigh the potential benefits at this time.

We expect that improved tools that meet target scientific and ethics benchmarks will become available to Internet users in the future but an important concern remains: new tools will exist amid a pool of invalid and unethically developed ones. In our sample, we did not observe a relationship between the traffic to the Web site host and the quality of the test and this is further supported by evidence

that demonstrates that popular Web sites or those that rank high in search returns are not necessarily of the highest quality [44,55]. In the current online environment, the burden of selection falls on the user. This dilemma is exacerbated in the case of vulnerable older adults whose ability to discern trustworthy information may be compromised [50]. Moving forward, a useful starting point for the development of future computerized tools would be to consider the eight ethics items listed in the rating grid we developed for this study (Fig. 1) and to make these tools available only within the clinical context until further studies have established the net impact of freely accessible self-diagnostic tests for AD on health decision making in older adults.

There are several limitations of the present report. The online environment is dynamic and fluid, and the online tests evaluated here were retrieved over two discrete periods. In addition, although interaction between experts was discouraged before the ratings of each test, some interdependence in the setting was inevitable and may have biased ratings but the variable interrater agreement suggests that such bias is minimal. We did not measure intrarater reliability and cannot confirm the reproducibility of the ratings. More broadly, our evaluation did not incorporate the experience of the self-assessment from the perspective of target users. Future work looking at the impact of online self-assessments will include an evaluation of the perspectives of test-takers.

Our results highlight important scientific and ethical considerations of online self-assessments for AD with significant implications for health care providers, who encounter patients. As the number of adults who seek health information online continues to rise and online self-assessments become more popular, physicians and allied health professional will benefit from a heightened awareness of these tools, the benefits they afford, and the risks they raise.

### Acknowledgments

B.L.B., S.L.H., J.M., and C.J. were involved in the development of C-TOC (cognitive testing on computer), a computerized cognitive test. C-TOC is being developed for use in clinic settings, with the oversight of a neuropsychologist. C-TOC is not freely accessible online or are there plans to provide such access in the future.

Funding for this work was provided by the Canadian Institutes for Health Research, the Foundation for Ethics and Technology, the Vancouver Coastal Health Research Institute, the Canada Research Chairs Program (J.I.), and the Ralph Fisher Professorship in Alzheimer's Research (Alzheimer Society of British Columbia) (C.J.). None of the sponsors had a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the article. The researchers operated independently from funders.

## RESEARCH IN CONTEXT

1. Systematic review: The authors reviewed the literature using academic and medical databases (PubMed, Google Scholar, Cochrane Library, and Medline) and identified that although freely accessible online tools for the self-assessment of Alzheimer disease (AD) are available, no study has evaluated the validity and the ethics of these tools.
2. Interpretation: Our study suggests that freely accessible online tests for AD do not provide useful diagnostic information about the disease, and we identified important ethical lapses that may lead to potential harms.
3. Future directions: Future work should address the perspectives and experiences of target users of online self-assessments for AD, as well as the benefits and harms of these tests.

## References

- [1] Della Mea V. What is e-Health (2): The death of telemedicine? *J Med Internet Res* 2001;3:e22.
- [2] Fox S. Health Online 2013. Pew Res Cent Internet Am Life Proj. Available at: <http://www.pewinternet.org/2013/01/15/health-online-2013/>; 2013. Accessed August 19, 2014.
- [3] Zickuhr K. Older adults and internet use. Pew Res Cent Internet Am Life Proj. Available at: <http://www.pewinternet.org/2012/06/06/older-adults-and-internet-use/>; 2012. Accessed August 19, 2014.
- [4] Mead N, Varnam R, Rogers A, Roland M. What predicts patients' interest in the Internet as a health resource in primary care in England? *J Health Serv Res Policy* 2003;8:33–9.
- [5] Corner L, Bond J. Being at risk of dementia: Fears and anxieties of older adults. *J Aging Stud* 2004;18:143–55.
- [6] Suziedelyte A. How does searching for health information on the Internet affect individuals' demand for health care services? *Soc Sci Med* 1982;2012:1828–35.
- [7] Giovanni MA, Fickie MR, Lehmann LS, Green RC, Meckley LM, Veenstra D, et al. Health-care referrals from direct-to-consumer genetic testing. *Genet Test Mol Biomarkers* 2010;14:817–9.
- [8] Trustram Eve C, de Jager CA. Piloting and validation of a novel self-administered online cognitive screening tool in normal older persons: The cognitive function test. *Int J Geriatr Psychiatry* 2014; 29:198–206.
- [9] Lovett KM, Mackey TK, Liang BA. Evaluating the evidence: Direct-to-consumer screening tests advertised online. *J Med Screen* 2012; 19:141–53.
- [10] Johnson CM, Shaw RJ. A usability problem: Conveying health risks to consumers on the Internet. *AMIA Annu Symp Proc* 2012; 2012:427–35.
- [11] Waters EA, Sullivan HW, Nelson W, Hesse BW. What is my cancer risk? How internet-based cancer risk assessment tools communicate individualized risk estimates to the public: Content analysis. *J Med Internet Res* 2009;11:e33.
- [12] Wolfe SM. Direct-to-consumer advertising—Education or emotion promotion? *N Engl J Med* 2002;346:524–6.
- [13] Jacova C, Kertesz A, Blair M, Fisk JD, Feldman HH. Neuropsychological testing and assessment for dementia. *Alzheimers Dement* 2007; 3:299–317.
- [14] Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: A systematic review. *Alzheimers Dement* 2008;4:428–37.
- [15] Snyder PJ, Jackson CE, Petersen RC, Khachaturian AS, Kaye J, Albert MS, et al. Assessment of cognition in mild cognitive impairment: A comparative study. *Alzheimers Dement* 2011; 7:338–55.
- [16] Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Clin Neuropsychol* 2012;26:177–96.
- [17] Jacova C, McGrenere J, Lee HS, Wang WW, Le Huray S, Corenblith EF, et al. C-TOC (cognitive testing on computer): Investigating the usability and validity of a novel self-administered cognitive assessment tool in aging and early dementia. *Alzheimer Dis Assoc Disord* 2014; <http://dx.doi.org/10.1097/WAD.0000000000000055>.
- [18] George DR, Qualls SH, Camp CJ, Whitehouse PJ. Renovating Alzheimer's: "Constructive" reflections on the new clinical and research diagnostic guidelines. *Gerontologist* 2013;53:378–87.
- [19] Lock M. *The Alzheimer conundrum: Entanglements of dementia and aging*. Princeton, New Jersey: Princeton University Press; 2013.
- [20] Portacolone E, Berridge C, Johnson JK, Schickel S. Time to reinvent the science of dementia: The need for care and social integration. *Aging Ment Health* 2014;18:269–75.
- [21] Katz S. *Cultural aging: Life course, lifestyle, and senior worlds*. Orchard Park, NY: Broadview Press; 2005.
- [22] Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ* 2002; 324:573–7.
- [23] Tierney MC, Lerner MA. Computerized cognitive assessment in primary care to identify patients with suspected cognitive impairment. *J Alzheimers Dis* 2010;20:823–32.
- [24] Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: Accessibility, quality, and readability in English and Spanish. *JAMA* 2001; 285:2612–21.
- [25] Free Alzheimer's Health Test n.d. Available at: <http://www.nutritionaltest.com/fre Alzheimer.html>. Accessed January 9, 2015.
- [26] Online Short-term Memory Tests & Memory Screenings n.d. Available at: <http://www.cognitichack.com/>. Accessed January 9, 2015.
- [27] PreventAD.com n.d. Available at: <http://www.preventad.com/>. Accessed January 9, 2015.
- [28] Free Memory Test Questions - Dr Dharma Internet Supplement Sales, Inc. n.d. Available at: <http://www.drDharma.com/Public/FreeStuff/FreeMemoryTestQuestions/index.cfm>. Accessed January 9, 2015.
- [29] Page not found | The Dr. Oz Show n.d. Available at: <http://www.doctoroz.com/quiz/memory-quiz>. Accessed January 9, 2015.
- [30] Test for Alzheimer online n.d. Available at: <http://www.memozor.com/memory-tests/test-for-alzheimer-online>. Accessed January 9, 2015.
- [31] Cognitive Labs online Test Center Memory Loss, Alzheimer's, dementia, cognitive fitness n.d. Available at: [http://cognitivelabs.com/global\\_cog\\_map1.htm](http://cognitivelabs.com/global_cog_map1.htm). Accessed January 9, 2015.
- [32] Online Memory Test - MemTrax n.d. Available at: <http://memtrax.com/test/>. Accessed January 9, 2015.
- [33] Free Online Memory, Alzheimers, dementia test | MyBrainTest n.d. Available at: <http://www.mybraintest.org/online-memory-screening-tests/>. Accessed January 9, 2015.
- [34] Alzheimer's questionnaire: Test that can reveal if YOU are at risk | Daily Mail Online n.d. Available at: <http://www.dailymail.co.uk/health/article-2095705/Alzheimers-questionnaire-Test-reveal-YOU-risk.html>. Accessed January 9, 2015.

- [35] Short memory test | Health24 n.d. Available at: <http://www.health24.com/Medical/Alzheimers/Memory-tests/Short-memory-test-20120721>. Accessed January 9, 2015.
- [36] Home n.d. Available at: <http://www.foodforthebrain.org/>. Accessed January 9, 2015.
- [37] Science: Free online Alzheimer's test n.d. Available at: <http://www.episodeseason.com/science/online-test-for-alzheimer.html>. Accessed January 9, 2015.
- [38] Alzheimer's disease | Memory Test | onmemory.ca n.d. Available at: <http://www.onmemory.ca/en/signs-symptoms/memory-test>. Accessed January 9, 2015.
- [39] SAGE: A test to measure thinking abilities | OSU Wexner Medical Center n.d. Available at: <http://wexnermedical.osu.edu/patient-care/healthcare-services/brain-spine-neuro/memory-disorders/sage>. Accessed January 9, 2015.
- [40] Alzheimers Test n.d. Available at: <http://www.way-of-the-mind.com/alzheimers-test.html>. Accessed January 10, 2015.
- [41] Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284–90.
- [42] Soobrah R, Clark SK. Your patient information website: How good is it? *Colorectal Dis* 2012;14:e90–4.
- [43] Jadad AR, Gagliardi A. Rating health information on the internet: Navigating to knowledge or to Babel? *JAMA* 1998;279:611–4.
- [44] Patel U, Cobourne MT. Orthodontic extractions and the Internet: Quality of online information available to the public. *Am J Orthod Dentofacial Orthop* 2011;139:e103–9.
- [45] McLean T, Delbridge L. Comparison of consumer information on the internet to the current evidence base for minimally invasive parathyroidectomy. *World J Surg* 2010;34:1304–11.
- [46] Ward JB, Leach P. Evaluation of internet derived patient information. *Ann R Coll Surg Engl* 2012;94:300–1.
- [47] Coleman JJ, McDowell SE. The potential of the internet. *Br J Clin Pharmacol* 2012;73:953–8.
- [48] Yeung TM, Mortensen NJ. Assessment of the quality of patient-orientated Internet information on surgery for diverticular disease. *Dis Colon Rectum* 2012;55:85–9.
- [49] Qureshi SA, Koehler SM, Lin JD, Bird J, Garcia RM, Hecht AC. An evaluation of information on the internet about a new device: The cervical artificial disc replacement. *Spine* 2012;37:881–3.
- [50] Castle E, Eisenberger NI, Seeman TE, Moons WG, Boggero IA, Grinblatt MS, et al. Neural and behavioral bases of age differences in perceptions of trust. *Proc Natl Acad Sci U S A* 2012; 109:20848–52.
- [51] Haddow LJ, Robinson AJ. A case of a false positive result on a home HIV test kit obtained on the internet. *Sex Transm Infect* 2005; 81:359.
- [52] McMullan M. Patients using the Internet to obtain health information: How this affects the patient–health professional relationship. *Patient Educ Couns* 2006;63:24–8.
- [53] Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the internet on health care and the physician-patient relationship: National U.S. survey among 1050 U.S. physicians. *J Med Internet Res* 2003;5:e17.
- [54] Kortum P, Edwards C, Richards-Kortum R. The impact of inaccurate Internet health information in a secondary school learning environment. *J Med Internet Res* 2008;10:e17.
- [55] Muthukumarasamy S, Osmani Z, Sharpe A, England RJ. Quality of information available on the world wide web for patients undergoing thyroidectomy: Review. *J Laryngol Otol* 2012;126:116–9.