VALUE IN HEALTH ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Do You Want to Hear the Bad News? The Value of Diagnostic Tests for Alzheimer's Disease

Axel Mühlbacher, PhD[1,2,*], *F. Reed Johnson, PhD*[3], *Jui-Chen Yang, MEM*[4], *Michael Happich, PhD*[5], *Mark Belger*[6]

[1]*Hochschule Neubrandenburg, IGM Institut Gesundheitsökonomie und Medizinmanagement, Neubrandenburg, Germany;* [2]*CHPIR Senior Research Fellow, Duke Global Health Institute, Duke University, Durham, NC, USA;* [3]*Senior Research Scholar, Duke Clinical Research Institute, Duke University, Durham, NV, USA;* [4]*Pacific Economic Research, LLC, Bellevue, WA, USA;* [5]*Lilly Deutschland GmbH, Bad Homburg, Germany;* [6]*Eli Lilly and Company Limited, Windlesham, Surrey, UK*

ABSTRACT

Objective: The diagnosis of Alzheimer's disease (AD) remains difficult. Lack of diagnostic certainty or possible distress related to a positive result from diagnostic testing could limit the application of new testing technologies. The objective of this paper is to quantify respondents' preferences for obtaining AD diagnostic tests and to estimate the perceived value of AD test information. **Methods:** Discrete-choice experiment and contingent-valuation questions were administered to respondents in Germany and the United Kingdom. Choice data were analyzed by using random-parameters logit. A probit model characterized respondents who were not willing to take a test. **Results:** Most respondents indicated a positive value for AD diagnostic test information. Respondents who indicated an interest in testing preferred brain imaging without the use of radioactive markers. German respondents had relatively lower money-equivalent values for test features compared with respondents in the United Kingdom. **Conclusions:** Respondents preferred less invasive diagnostic procedures and tests with higher accuracy and expressed a willingness to pay up to €700 to receive a less invasive test with the highest accuracy.

Keywords: discrete-choice experiment (DCE), Alzheimer's disease (AD), diagnostic test information, money-equivalent value (MEV).

Introduction

Alzheimer's disease (AD) is the most common cause of dementia in older adults and affects approximately 7.3 million people in Europe. The worldwide cost of AD care is estimated to be €440 billion [1]. Symptoms of AD include memory deterioration, progressive difficulty with language and the ability to communicate, declining ability to perform routine tasks, and personality and mood changes.

Despite the existence of standardized medical criteria [2,3], clinical diagnosis remains difficult. With the use of standardized criteria, diagnosis is approximately 81% sensitive and 70% specific compared with the gold standard—pathology at autopsy [4]. The accuracy levels have not changed since the publication of the criteria in 2001; Beach et al. [5] reported in 2012 that concordance of clinical diagnosis and pathologic findings is still around 71% for both sensitivity and specificity in patients with a probable clinical AD diagnosis.

No general tool exists for making a definitive diagnosis of AD. There is a new approach, however, to detecting preclinical AD in older adults. The blood test predicted AD within a 2- to 3-year timeframe, with over 90% accuracy [6]. Several other biomarkers accompany the progression of AD and are hypothesized to develop at different stages of the disease. It is a widely accepted assumption that an initiating event in the development of AD is the accumulation of β-amyloid plaques in the brain, which are targeted by using diagnostic tests such as β-amyloid positron emission tomography (PET) tracer or indirectly by using cerebrospinal fluid. After a time, neuronal dysfunction and neurodegeneration with brain atrophy follow. Fluorodeoxyglucose positron emission tomography (FDG-PET) and magnetic resonance imaging (MRI) are used to detect these developments [7].

Biomarker tests vary in their invasiveness. MRI, for example, requires patients to lie in a machine that takes images of the brain. PET tests also take pictures of the brain, but as a prerequirement, a radioactive marker or dye is injected into blood in order to monitor tracer activities in the brain. To collect samples of cerebrospinal fluid, a needle is inserted into the spine to remove the fluid. The tests vary in reliability and accuracy, with sensitivity and specificity ranging from 65% to 95% [4,8].

There is an unmet need for improved test precision. The proposal by Dubois and colleagues [9] to include pathologically linked biomarkers of AD in research diagnostic criteria aims to reduce the frequency of false-positive diagnosis and increase the

Conflict of interest: Lilly Deutschland GmbH provided the funding for this research.

[*]*Address correspondence to:* Axel Mühlbacher, Hochschule Neubrandenburg, IGM Institut Gesundheitsökonomie und Medizinmanagement, Brodaer Straße 2, 17033 Neubrandenburg, Germany.

E-mail: muehlbacher@hs-nb.de

1098-3015$36.00 – see front matter Copyright © 2016, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2015.10.011.

validity of a clinical diagnosis of AD, especially at its earliest symptomatic stage. Studies by Elson [10], Mattsson et al. [11], and Turnbull et al. [12] suggest that most patients are keen to know the cause of their symptoms, even if the cause is AD. The clinical value of a diagnostic test, however, depends heavily on the potential for treating the disease. Currently, there is no effective treatment for AD. In the absence of treatment benefits with regard to mortality, morbidity, or quality of life, the value of diagnostic tests remains questionable.

The aim is to quantify the preferences of older adults without memory problems for AD test technologies and the perceived value of diagnostic information by applying discrete-choice experiment (DCE) and contingent-valuation (CV) questions [13–15].

Materials and Methods

Study Designs

The survey presented respondents with a series of DCE questions (see Appendix A in Supplemental Materials at: http://dx.doi.org/10.1016/j.jval.2015.10.011 [16–26]). The levels and analytical variable names used for each attribute are summarized in Appendix A (in Supplemental Materials at: http://dx.doi.org/10.1016/j.jval.2015.10.011). Each question included two virtual diagnostic test profiles and a no-test option. Each profile was defined by three attributes based on a review of relevant literature, consultation with clinical experts, and face-to-face interviews with target populations in Germany and the United Kingdom. The attributes were diagnostic test type; test precision (false-positive or false-negative test results); and test cost. Pretest interviews indicated that respondents had difficulty understanding the distinction between false-negative results and false-positive results and therefore had difficulty evaluating both types of errors in the same DCE question. The presentation of diagnostic test imprecision was improved, and the questions were divided into two sets of questions: one that defined imprecision as the rate of false positives, and another that defined imprecision as the rate of false negatives. The sequence of the two groups of questions was randomly assigned.

The attribute levels were chosen to encompass the range of levels described in the literature as well as the range over which respondents were willing to accept tradeoffs in pretest interviews. Diagnostic test cost was included as an attribute in the choice questions to estimate the money-equivalent value (MEV), also called "willingness to pay" (WTP), for improvements in the levels of the diagnostic test attributes. The range of costs was chosen based on the likely limit of respondents' willingness to accept tradeoffs between cost and different levels of the diagnostic test attributes. We expected a majority of respondents to reject the highest cost shown, even when the associated test attributes were all set at the most desirable levels.

In addition to the DCE question format, CV questions were included to capture the value of diagnostic test information assuming that treatment is available to patients. Respondents were asked to consider the following scenario: 1) "If the test says you do have AD, you would get a treatment that would keep your memory problems from getting worse for 1 year"; 2) "If the test says you do not have AD, you would get a treatment that would keep your memory problems from getting worse for 2 months"; and 3) "The test uses brain imaging with radioactive marker."

Using a double-bounded, dichotomous-choice format (see Appendix B in Supplemental Materials at: http://dx.doi.org/10.1016/j.jval.2015.10.011 for the CV question format), respondents were then asked what they would pay for this diagnostic test. The resulting responses made it possible to identify two types of respondents among those who always chose the no-test option in the DCE questions: 1) respondents who would be interested in additional testing and had a zero or positive MEV for diagnostic information; and 2) respondents who would not be interested in additional testing and had a negative MEV for diagnostic information.

While developing the surveys, we used a robust forward-backward translation. Both surveys were then pretested with a sample of respondents aged 60 years or more in face-to-face interviews in the United Kingdom (15 pretests) and Germany (18 pretests). The survey also collected data on demographic and socioeconomic characteristics, health history, experience with AD diagnostic tests, and experience in caring for or knowing someone with AD (see Appendix C in Supplemental Materials at: http://dx.doi.org/10.1016/j.jval.2015.10.011 for the English version of the survey). The Research Triangle Institute's Office of Research Protection and Ethics approved the study design.

Experimental Design

The experimental design optimized the assignment of the three attributes and their levels into pairs of virtual test profiles. A fractional factorial experimental design with 36 choice questions was constructed using a D-optimal algorithm in the Statistical Analysis System (SAS Institute, Cary, NC) [13,27,28]. Choice questions were divided into six versions of six questions each. For each pair, respondents were asked to indicate whether they would choose one of the two virtual tests or the no-test option. Figure 1 presents an example DCE question for the false-negative error (Appendix D in Supplemental Materials at: http://dx.doi.org/10.1016/j.jval.2015.10.011 contains the German version). Each respondent was given 12 questions from the design (six questions with incidence of false-positive errors and six questions with incidence of false-negative errors). For the assigned version, each respondent also was randomly assigned to answering either the false-positive choice questions first or the false-negative choice questions first.

Data Collection

Sample-size power calculations represent a technical challenge in DCEs. Based on the existing DCE literature [29,30], with a sample of 300 to 350 respondents, the accuracy of estimation is generally good but representativeness is limited [13,31,32].

The hypothesis that numerous respondents might not be interested in taking the AD diagnosis test led to the decision to target a bigger sample size. Sample sizes of 800 for each country provided better representativeness and allowed for sufficient power for analyzing the subsample that had a positive value for diagnostic test information.

The German and UK respondents were recruited from GfK's Custom Research online panels. To qualify for inclusion in this study, respondents had to be able to read and understand German for the German sample or English for the UK sample; be aged 60 years or older; not have memory problems or dementia; and not currently take a prescription medication to treat AD.

Data Analysis

The preference data were analyzed by using random-parameters logit models with NLOGIT 4.0 (Econometric Software Inc., Plainview, NY). Random-parameters logit avoids potential estimation bias from unobserved taste heterogeneity among respondents by estimating a distribution of tastes for each preference parameter [33,34]. Taste distributions were assumed to be normal except for cost, which was estimated as a nonrandom variable [33]. Other distributional assumptions yielded implausible estimates.

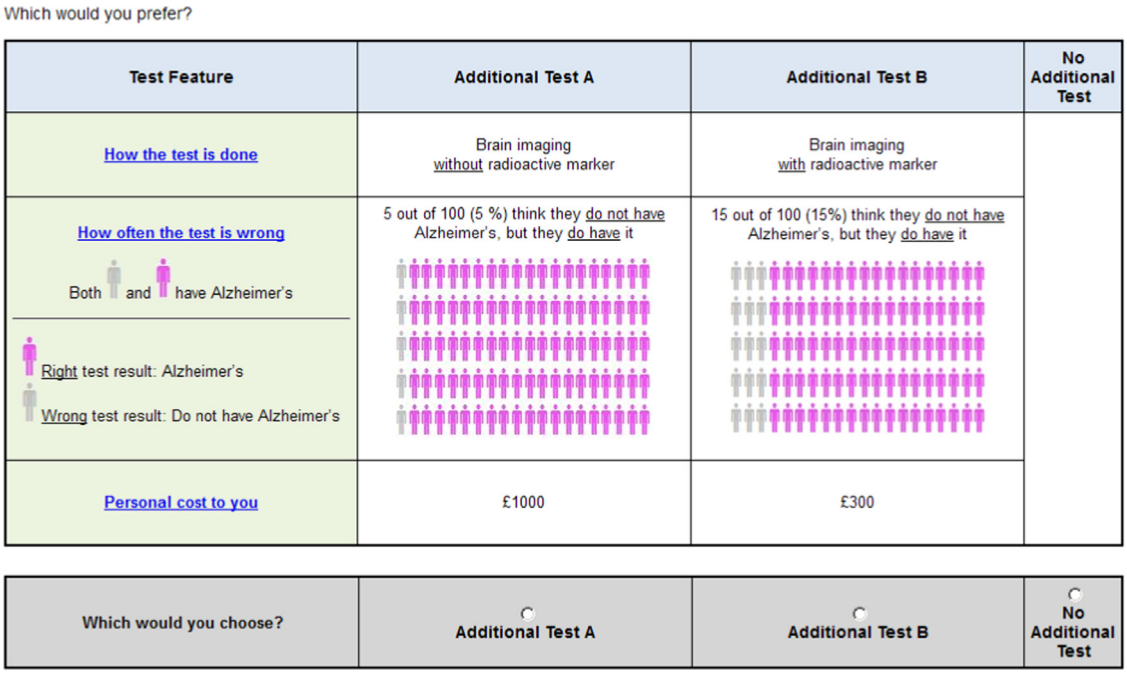

Figure 1 – Example DCE Question, UK Survey, False-Negative Version. DCE, discrete-choice experiment.

The dependent variable was the preferred profile in each choice task. The independent variables were levels of each of the attributes in each profile. All attributes (see Appendix A in Supplemental Materials at: http://dx.doi.org/10.1016/j.jval.2015.10.011) were included in the model as effects-coded variables except personal cost and the no-test option. After specification testing, personal cost was specified as a linear numeric variable. The no-test option was specified as a dummy variable.

Effects coding normalizes the mean level effect at zero for each attribute, rather than at the omitted level as in dummy coding [35]. The procedure produces a parameter estimate for all level categories. Standard errors were calculated by using the Krinsky-Robb method [36]. The resulting log-odds estimates relative to the mean effect indicate the relative strength of preference for each level. Positive parameter estimates indicate choice probabilities larger than the mean effect, whereas negative parameter estimates indicate choice probabilities smaller than the mean effect. For ease of comparison and to eliminate scale differences between the two countries, the model results are rescaled from 0 to 10, using linear transformation of β-coefficients from 0 (the least preferred level) to 10 (the most preferred level).

The relative importance of an attribute *conditional on the range of levels presented* is determined by the relative difference between the most important and the least important levels. Thus, the mean relative importance score for each factor is interpreted as the relative value of an improvement from the worst level to the best level.

The preference-score index has no meaning except in relative comparisons. Rescaling of parameter estimates by using a continuous attribute as a numeraire converts parameter estimates into a more intuitive measure of preference intensity. In particular, relative importance is indicated by the MEV of utility differences. For the specified linear-additive indirect utility function, MEV for test characteristics is the expected mean compensating-surplus welfare measure for respondents with an interest in the specific test attribute. Ex ante MEV accounts for the probability of being "in the market" for a particular test and thus is the value of having the test in the individual's choice set. The estimates were calculated by using the standard random-utility log-sum formula [37]:

$$\mathrm{MEV} = \frac{\ln(e^{U_i} + e^{U_{OPTOUT}}) - \ln(e^{U_{OPTOUT}})}{-\partial U/\partial \mathrm{Cost}}$$

where U_i is the utility of a test of specified type and precision, U_{OPTOUT} is the utility of not testing, and the denominator is the constant marginal utility of income.

Given the importance of obtaining a valid estimate of the marginal utility of income, a validity test of sensitivity to scope was included to test the hypothesis that respondents paid attention to absolute cost levels and did not interpret the cost levels simply as "low," "medium," and "high." To test this hypothesis, each respondent was given only four cost levels. Half of the German respondents were given €240, €360, €600, and €960 per diagnostic test and the other half €240, €360, €600, and €1200 per diagnostic test. Half of the UK respondents were given £200, £300, £500, and £800 per diagnostic test and the other half £200, £300, £500, and £1000 per diagnostic test. Attention to absolute costs is indicated by greater sensitivity to cost in the arm with the wider range of costs.*

In this study, some respondents chose the no-test alternative in all the DCE questions and did not express an interest in the diagnostic test described in the CV question even at zero cost. The respondents provided no usable information for quantifying preferences for diagnostic test features and thus were dropped from the choice model. We estimated a probit model to investigate the effects of respondents' demographic characteristics, experiences, and perceptions of AD and its tests on interest in test information. Missing data on personal characteristics were replaced with mean values of respondents who provided the data.

The preference data included respondents who picked "Additional Test A" or "Additional Test B" for at least one of the DCE

*The coefficients were plotted against the price points to see if the cost attribute is linear. The R-squared values were between 0.89 and 0.93 for the German sample and between 0.93 and 0.98 for the UK sample.

questions, as well as respondents who always picked "No Additional Test" but stated an interest in an additional diagnostic test in the CV questions if the test were offered at low or zero cost.

We analyzed the choice data by specifying indirect utility U as:

$$U = \beta_{WITHOUT\_MARKER} \times WITHOUT\_MARKER + \beta_{WITH\_MARKER} \times WITH\_MARKER + \beta_{SPINAL\_TAP} \times SPINAL\_TAP + \beta_{FALSE\_5\%} \times FALSE\_5\% + \beta_{FALSE\_15\%} \times FALSE\_15\% + \beta_{FALSE\_30\%} \times FALSE\_30\% + \beta_{OPTOUT} \times OPTOUT + \beta_{COST} \times COST,$$

where OPTOUT is an indicator variable equal to one for the "No Additional Test" option and zero otherwise. β_{OPTOUT} is defined relative to the mean test utility, which is zero for all attributes except cost. We removed the effect of the mean cost by calculating no-test utility as the dummy-variable parameter minus the cost parameter times the mean cost. COST is a continuous variable set at the level shown in each question. We estimated the same choice model for each country separately. Specification tests supported combining false-positive and false-negative imprecision levels in a single attribute.[†]

Results

Sample Characteristics

In Germany, of the 6377 individuals invited to participate in the survey, 1630 responded, 876 were eligible and consented, and 815 completed the survey. In the United Kingdom, of the 7852 individuals invited to participate in the survey, 1580 responded, 860 were eligible and consented, and 800 completed the survey. A total of 281 respondents (167 German; 114 UK) chose the no-test alternative in all the DCE questions and did not express an interest in the diagnostic test described in the CV question even at zero cost and thus were dropped from the choice model. Another 33 respondents (17 German; 16 UK) were dropped from the choice model because they always chose "Additional Test A" or "Additional Test B" in all DCE questions, which indicated that they did not pay close attention to the questions given. The median reported completion time was 15 minutes. Summary statistics are reported in Table 1.

The German respondents had less experience of AD. About a third of the German respondents had a family member or friend with AD or any other kind of serious memory problem compared with about half of the UK respondents. In both countries, however, one-third of the respondents provided care for someone with AD or serious memory problems. Only 4.3% and 2.5% of the German and UK respondents, respectively, stated that they had taken an AD test. In addition, overall, 32% did not think that an AD test is useful if there is no cure and no way to slow disease progression (Germany 37%; UK 27%).

Contingent Valuation: Who Is Willing to Take an AD Test?

A double-bounded, dichotomous-choice CV question elicited the respondents' direct WTP for a brain imaging test using a radioactive marker with 15% imprecision. About one-third of the respondents (184 German; 130 UK) were not willing to pay for diagnostic information in the DCE question. About 53% of these respondents who were not willing to pay, however, indicated that they would be interested in taking a diagnostic test at low or zero cost in the CV question.

†Log-likelihood ratio tests suggested the false-positive/false-negative sequence (Germany: P value = 0.995; UK: P value = 0.765) and information treatment (i.e., availability of treatment for AD) (Germany: P value = 0.148; UK: P value = 0.105) had no statistically significant effect on preferences for AD diagnostic testing.

Choice-Model Estimates: What Are the Important Attributes of an AD Test?

Table 2 compares rescaled log-odds mean preference weights and standard deviations of the distribution of preference weights by country for the 631 German and the 670 UK respondents who indicated an interest in testing. In the DCE data, about 40% of the German respondents and 34% of the UK respondents chose the no-test alternative in every question.

The standard errors and deviations of these distributions reported in Table 2 are also depicted in Figure 2. The German respondents had greater taste heterogeneity than the UK respondents. That is, there was less consensus among German respondents about each diagnostic test feature than among the UK respondents. The only exception is the middle-level test imprecision of 15%, where both the German and the UK respondents were in strong agreement.

There were no statistically significant differences between the samples from the two countries in mean preference-parameter estimates for either test characteristics or cost. Test precision strongly influenced diagnostic test preferences and was significantly more important than test technology. For the test-type attribute, the German and UK respondents preferred brain imaging without radioactive markers to brain imaging with radioactive markers (Germany: $P < 0.01$, UK: $P < 0.01$); brain imaging without radioactive markers to spinal tap (Germany: $P < 0.01$; UK: $P < 0.01$); and brain imaging with radioactive markers to spinal tap (Germany: $P < 0.01$; UK: $P < 0.01$).

Although mean parameters were similar, all standard-deviation measures of taste heterogeneity for the German respondents were significantly larger than the estimates for UK respondents except for 15% imprecision.

The Value of Diagnostic Information: What Is the MEV of AD Tests?

The preference weights were used to calculate ex ante mean MEV for AD test profiles. Although parameter weights between the two countries were statistically similar, the strong nonlinearity in the ex ante MEV ratio formula resulted in larger differences in those values. Table 3 presents MEV estimates for all combinations of test technology and test precision. Each MEV estimate indicates the predicted value of having the diagnostic test as an option in the individual's choice set. The most valued diagnostic test was brain imaging without radioactive markers with best test precision and with estimated MEVs of €342 among German respondents and €704 among UK respondents. The least valued diagnostic test was spinal tap, with worst test precision and estimated MEVs of €49 and €37 among German and UK respondents, respectively. In general, German respondents were willing to pay less than UK respondents for improvements in test type and test precision. All of the German MEV estimates except one were less than the UK estimates and five of the nine differences were statistically significant.

Respondent Characteristics: Who Does Not Want to Know Their AD Status?

Table 4 summarizes the effect of personal characteristics on the likelihood of having an interest in the diagnostic test information. Respondents who were more likely to have an interest in an additional test included the UK respondents ($P = 0.034$), male respondents ($P = 0.095$), respondents who thought an AD test is useful even though there is no cure ($P < 0.001$), respondents who thought a false-negative result would be devastating ($P < 0.001$), respondents who thought their memory problems probably would be much or somewhat less serious than AD if an additional test

Table 1 – Descriptive statistics.

Variable	Overall N = 1301	German respondents n = 631	UK respondents n = 670	P value*
Male	56.5%	62.9%	50.5%	< 0.001
Mean age in years (SD)	65.8 (4.6)	65.4 (4.5)	66.2 (4.7)	0.002
Has a college degree or higher	26.3%	31.4%	21.6%	< 0.001
Current total annual household income (before any deductions for income tax, national insurance, etc.) (£ in UK and € in Germany)				
Less than £20,000 / Less than €24,000	38.5%	36.2%	40.6%	0.200
£20,000-£39,999 / €24,000-€47,999	36.8%	37.5%	36.2%	
£40,000 or more / €48,000 or more	12.9%	14.1%	11.7%	
Do not know or missing	11.8%	12.2%	11.5%	
Has had a family member or friend with AD or any other kind of serious memory problem	41.4%	32.8%	49.4%	< 0.001
Has provided care for someone with AD or serious memory problems	35.5%	35.8%	35.4%	0.925
Has had a test for memory problems or AD	3.4%	4.3%	2.5%	0.082

Note. Percentages displayed exclude missing values.
AD, Alzheimer's disease; SD, standard deviation.
* P values indicate the statistical significance of differences between the UK and German respondents and are reported based on chi-squared tests for categorical variables and Student's *t* tests for continuous variables.

indicated that the reason for their memory problem was not AD ($P < 0.001$), and respondents who thought it somewhat or very likely that researchers will find an effective treatment for AD in the next 5 years ($P = 0.011$). Respondents, however, would be less likely to have interest in an additional test if they reported no existing medical conditions ($P = 0.085$), previously had a test for memory problems or AD ($P = 0.120$), or were not concerned about developing AD in the future ($P < 0.001$).

Discussion

On the basis of the literature and the pretest results, we expected to identify two different groups: (1) those who did not want to know their AD status, and (2) those who did want to know and were willing to accept the tradeoffs between diagnostic test attributes and personal costs. Results from qualitative questions, the CV data, and the DCE data confirmed this expectation. The questions that were answered are described below.

Who Was Willing to Take an AD Test?

The CV questions showed that a large percentage of respondents were not willing to take a diagnostic test or did not state any interest if asked about their WTP, perhaps for fear of receiving bad news with little prospect of effective treatment. Depending on the question format and the nationality of the population, 33% to 40% were not interested in taking the AD diagnostic test. The result is consistent with findings that a diagnosis of AD is often made late in the disease continuum [38]. Indeed, many people with dementia never receive a formal diagnosis at all [38]. Studies have found that stigma can be an important influence on delays in the recognition and diagnosis of dementia in primary care, with some patients refusing to be assessed and caregivers also being in denial or preferring not to know the diagnosis [39,40].

Table 2 – Germany and United Kingdom rescaled choice-model estimates.

Attribute	Level	Germany (N = 631)		UK (N = 670)	
		Coefficient	Standard error	Coefficient	Standard error
Rescaled Mean Preference Weights					
Test type	Brain imaging without radioactive marker	9.0946	0.8870	8.2469	0.5039
	Brain imaging with radioactive marker	5.2005	0.7208	5.5920	0.4591
	Spinal tap	1.7451	1.0396	1.8255	0.6390
Test imprecision	5%	10.0000	1.1092	10.0000	0.6602
	15%	6.0402	0.6115	5.6643	0.3515
	30%	0.0000	1.1552	0.0000	0.6958
No test	1, if no test was chosen; otherwise, 0	1.9769	2.2501	2.5488	1.2517
Test cost	A continuous variable	5.3259	0.0017	5.2085	0.0011
Rescaled Standard Deviations of Preference-Weight Distributions					
Test type	Brain imaging without radioactive marker	15.9881	1.1047	9.4408	0.5775
	Brain imaging with radioactive marker	12.0153	0.8877	8.6638	0.5198
	Spinal tap	28.0035	1.5260	18.1046	0.7222
Test imprecision	5%	22.7278	1.1867	14.5778	0.7860
	15%	0.5828	1.0265	0.6308	0.5293
	30%	23.3106	1.4581	15.2086	0.9482
No test	1, if no test was chosen; otherwise, 0	70.3128	3.3571	40.2972	1.8009

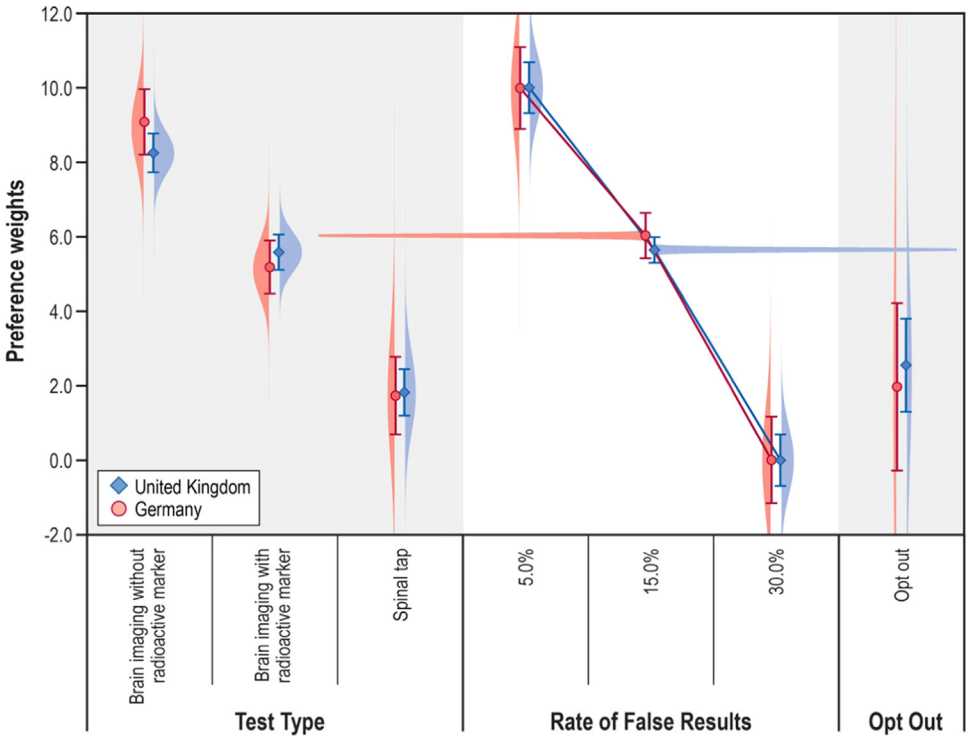

Figure 2 – Germany and UK Rescaled Choice-Model Estimates. Illustrates rescaled log-odds mean preference weights and standard deviations of the distribution of preference weights by country for the 631 German and 670 UK respondents. The circles and diamonds in the figure represent the relative preference weights for attribute levels for Germany and the UK, respectively. The vertical bars around each relative preference weight indicate the 95% confidence interval. If the confidence intervals between levels of a single attribute do not overlap, the preference weights are statistically significantly different from each other at the 5% level. The distributions around each preference-weight estimate indicate the estimated taste heterogeneity among respondents.

We found that the likelihood of rejecting AD diagnostic information was, indeed, correlated with various attitude and health-history variables. The more aware and concerned people were, the more likely they were to choose to take and pay for an AD test. Given the importance of diagnostic test precision, it is likely that more accurate, less invasive tests, as well as the availability of better treatments, would also improve interest in AD diagnostic information. Because of the respondents who did not state any preference, the choice-model estimates and the MEVs analyzed data only from respondents who were interested in taking the AD test.

What Were the Important Attributes of an AD Test?

Respondents who had an interest in taking an AD test preferred the less invasive diagnostic procedures. In particular, the major-

Table 3 – Money-equivalent values relative to no test (selected comparisons).

Test Technology	Test Imprecision	MEV in €*,†		P value
		Germany (N = 631)	UK (N = 670)	
		DCE ESTIMATES		
Spinal tap	False = 30%	49 (30, 68)	37 (23, 51)	0.33
	False = 15%	104 (72, 136)	126 (90, 162)	0.38
	False = 5%	164 (117, 211)	284 (211, 357)	< 0.01
Brain imaging with radioactive marker	False = 30%	76 (49, 103)	85 (59, 111)	0.63
	False = 15%	155 (113, 197)	257 (201, 314)	< 0.01
	False = 5%	237 (178, 295)	506 (410, 601)	< 0.01
Brain imaging without radioactive marker	False = 30%	121 (80, 163)	147 (104, 189)	0.41
	False = 15%	235 (175, 295)	397 (319, 475)	< 0.01
	False = 5%	342 (265, 419)	704 (590, 818)	< 0.01

MEV, money-equivalent value; DCE, discrete-choice experiment.
* At the time of the study, the exchange rate was £1 = €1.26.
† Numbers in parentheses are 95% confidence intervals.

Table 4 – Likelihood of having interest in an additional test for given respondents' characteristics.

Respondents' characteristics	Effect on likelihood of having interest in an additional test for AD Pooled samples (N = 1582)
UK resident	3.99% (0.31%, 7.66%)*
Male	3.07% (−0.53%, 6.68%)†
Respondent's age	−0.15% (−0.51%, 0.21%)
Married	1.78% (−1.86%, 5.42%)
Reports no existing medical conditions	−3.87% (−8.50%, 0.77%)†
Has had a test for memory problems or AD	−8.70% (−21.28%, 3.88%)‡
Provides care for someone with AD	3.40% (−1.35%, 8.14%)
Thinks that an AD test is useful even though there is no cure and no way to slow the progression of AD	19.53% (15.49%, 23.57%)§
Not at all concerned about developing AD in the future	−8.50% (−12.20%, −4.80%)§
Thinks that a false-negative would be the worse result	7.69% (4.17%, 11.20%)§
Thinks that memory problems probably would be much or somewhat less serious than AD if an additional test said the reason for the memory problem was not AD	6.48% (2.93%, 10.03%)§
Thinks that it is very likely or somewhat likely researchers will find an effective treatment for AD in the next 5 years	4.91% (1.13%, 8.69%)*

Note. Values in parentheses are the 95% confidence intervals.
AD, Alzheimer disease; UK, the United Kingdom.
* Significant at 5%.
† Significant at 10%.
‡ Significant at 15%.
§ Significant at 1%.

ity of respondents in Germany and the United Kingdom preferred brain imaging without radioactive markers, followed by brain imaging with radioactive markers, and finally spinal tap. The result is consistent with a study in older patients, in which lumbar puncture was cited as the strongest disincentive for participation in clinical research [41]. Taste heterogeneity was larger for the German cohort than for the UK cohort and was particularly large for no-test utility.

How Much Were Respondents Willing to Pay?

Respondents preferred the tests with higher accuracy and indicated significantly higher monetary values for more accurate tests. The mean MEV for the least invasive, most accurate test was €342 and €704 for the German and UK samples, respectively. Advertised prices for brain scans in UK private clinics range from approximately €300 to €1500 equivalents. We found, however, large taste variation in preference parameters relative to cost. Although the preference for less invasive procedures might be related to patient convenience, higher WTP for more accurate diagnosis could have direct economic and public health consequences. The economic cost of false-positive diagnoses of dementia has been explored in a US study [42]. That study found that persons with false-positive dementia diagnoses made Medicare claims that for were significantly ($P < 0.001$) more expensive, costing $11,294 compared with $4,065 for true negatives.

Were There Differences between UK Respondents and German Respondents?

Despite the fact that the ordinal structure of the preferences did not vary between the UK and German populations, the different models discovered differences in taste heterogeneity, WTP, and person-specific attributes, which led to a rejection of the AD diagnostic testing. The differences could be related to the fact that the study populations differed by the percentage of respondents who were familiar with AD through a friend or family member and who had serious memory problems. Moreover, the percentage of male respondents varied, whereas age, annual household income, and the percentage of respondents who provided care to someone with AD did not vary dramatically. Both populations had a low percentage of respondents who had tested for AD before. Choice-model estimates for the attributes of test type and test precision did not vary much between the two populations. The test cost coefficients varied to a great extent. Also, the standard deviations were different, leading to the conclusion of great differences in taste heterogeneity. That is, there was less consensus among the German respondents about each test feature than among the UK respondents. The fact that test cost coefficients were different could be analyzed in more detail by estimating the MEVs. The UK respondents were willing to pay twice as much for the most preferred AD testing option compared with the German respondents (€704 versus €342). Finally, the likelihood of having (no) interest in an additional test for AD was found to be greatly influenced by nationality.

No treatments are currently available to cure or stop the progression of AD; however, interventions are offered to support and improve the lives of people with AD, including pharmaceutical treatments and nonpharmaceutical interventions [38,43]. Our study indicates that people are willing to pay for diagnostic clarity, even in the absence of viable treatment options [15]. A study by Kopits and colleagues [44] studied WTP for AD genetic testing and found that genetic risk information was valuable, even if such information has no therapeutic value. Similarly, Green and colleagues [45] disclosed AD genotype information to asymptomatic adults and found psychological benefits to those with negative test results and only transient distress to those with positive results. The findings are consistent with previous research showing that people dislike ambiguous, uncertain situations and prefer certainty, even in the case of situations with negative consequences.

Our analysis had a number of limitations. Although the respondents were recruited at different ages above 60 years and from different regions in the consumer panels, the panels are not strictly representative of the general older populations in the two countries studied.

Dementia is a complex condition, and achieving a timely and accurate diagnosis is challenging. AD often develops slowly, overlaps with other comorbidities, and can include symptoms of cognitive impairment similar to those of other conditions [46]. Furthermore, the boundaries between subtypes of dementia are indistinct, and most patients with dementia have been found to have mixed forms [47]. The diagnostic complexities are difficult to reflect in a preference study where the number of addressed items needs to be balanced against information overload to respondents. Thus, respondents could have various expectations regarding their health state after receiving a positive or negative diagnosis of AD. We controlled for these expectations and found no significant differences.

Also, a number of well-known factors can lead to hypothetical bias in stated-preference surveys [48,49]. Our survey was designed to mitigate hypothetical bias by offering respondents choices among realistic tests using well-defined attributes. However, the respondents in this study made decisions independent of physicians' advice. The result is not reflective of real life, where diagnosis is mainly in the hands of physicians and treatment decisions are at best a collaborative effort of patients and physicians. Although not likely to be predictive of clinical outcomes, increased interest in patient-centric health care and shared decision helps shed light on patients' choices when not mediated strongly by clinical practice guidelines and reimbursement considerations.

Conclusions

Health care decision makers must prioritize health care interventions on the basis of the relationship between benefit to patients and costs imposed on society. Cultural and institutional differences can influence prioritization decisions. The findings of this study can help inform decisions on public needs and priorities and can raise awareness of public preferences for diagnostic information when therapeutic options are limited or nonexistent. Our findings should be interpreted in light of the current pharmacologic and behavioral interventions for dementia, which do not prevent or alter disease progression. The use of these interventions can only temporarily affect detrimental cognitive, functional, and behavioral outcomes [38,50]. To aid in reducing uncertainty in health policy making, a more targeted confirmatory study needs to be conducted to verify and expand on these results.

Acknowledgments

The authors would like to thank the respondents for their help and Lilly Deutschland GmbH for providing the funding for this research.

Supplemental materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2015.10.011 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Wimo A, Jönsson L, Bond J, et al. The worldwide economic impact of dementia 2010. Alzheimers Dement 2013;9:1–11.
[2] McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;34:939–44.
[3] American Psychiatric Association. Practice Guideline for the Treatment of Patients with Alzheimer's Disease and Other Dementias. Arlington, VA: American Psychiatric Association, 2007.
[4] Knopman DS, DeKosky ST, Cummings JL, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001;56:1143–53.
[5] Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. J Neuropathol Exp Neurol 2012;71:266–73.
[6] Mapstone M, Cheema AK, Fiandaca MS, et al. Plasma phospholipids identify antecedent memory impairment in older adults. Nat Med 2014;20:415–8.
[7] Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010;9:119–28.
[8] Bloudek LM, Spackman DE, Blankenburg M, Sullivan SD. Review and meta-analysis of biomarkers and diagnostic imaging in Alzheimer's disease. J Alzheimers Dis 2011;26:627–45.
[9] Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007;6:734–46.
[10] Elson P. Do older adults presenting with memory complaints wish to be told if later diagnosed with Alzheimer's disease? Int J Geriatr Psychiatry 2006;21:419–25.
[11] Mattsson N, Brax D, Zetterberg H. To know or not to know: ethical issues related to early diagnosis of Alzheimer's disease. Int J Alzheimers Dis 2010: pii:841941.
[12] Turnbull Q, Wolf AM, Holroyd S. Attitudes of elderly subjects toward "truth telling" for the diagnosis of Alzheimer's disease. J Geriatr Psychiatry Neurol 2003;16:90–3.
[13] Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011;14:403–13.
[14] Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16:3–13.
[15] Neumann PJ, Cohen JT, Hammitt JK, et al. Willingness-to-pay for predictive tests with no immediate treatment implications: a survey of US residents. Health Econ 2012;21:238–51.
[16] Clark CM, Schneider JA, Bedell BJ, et al. Use of florbetapir-PET for imaging β-amyloid pathology. JAMA 2011;305:275–83.
[17] Weiner MW, Veitch DP, Aisen PS, et al. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimers Dement 2011;1:1–67.
[18] Marshall DA, Johnson FR, Kulin NA, et al. How do physician assessments of patient preferences for colorectal cancer screening tests differ from actual preferences? Comparison in Canada and the United States using a stated-choice survey. Health Econ 2009;18:1420–39.
[19] Phillips KA, Maddala T, Johnson FR. Measuring preferences for health care interventions using conjoint analysis: an application to HIV testing. Health Serv Res 2002;37:1681–705.
[20] Marshall DA, Johnson FR, Phillips KA, et al. Measuring patient preferences for colorectal cancer screening using a choice-format survey. Value Health 2007;10:415–30.
[21] Peskind ER, Riekse R, Quinn JF, et al. Safety and acceptability of the research lumbar puncture. Alzheimer Dis Assoc Disord 2005;19:220–5.
[22] Blennow K, Wallin A, Häger O. Low frequency of post-lumbar puncture headache in demented patients. Acta Neurol Scand 1993;88:221–3.
[23] Peskind E, Nordberg A, Darreh-Shori T, Soininen H. Safety of lumbar puncture procedures in patients with Alzheimer's disease. Curr Alzheimer Res 2009;6:290–2.
[24] Wong DF, Rosenberg PB, Zhou Y, et al. In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (florbetapir [corrected] F 18). J Nucl Med 2010;51:913–20.
[25] Sassi F, McKee M. Do clinicians always maximize patient outcomes? A conjoint analysis of preferences for carotid artery testing. J Health Serv Res Policy 2008;13:61–6.
[26] Lewis SM, Cullinane FM, Carlin JB, Halliday JL. Women's and health professionals' preferences for prenatal testing for Down syndrome in Australia. Aust N Z J Obstet Gynaecol 2006;46:205–11.
[27] Kuhfeld WF. Marketing Research Methods in SAS: Experimental Design, Choice., Conjoint, and Graphical Techniques. Cary, NC: SAS Institute, Inc, 2010.
[28] Kuhfeld W, Tobias F, Garratt M. Efficient experimental design with marketing research applications. J Mark Res 1994;31:545–57.
[29] Orme BK. Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research. (2nd ed.). Madison, WI: Research Publishers LLC, 2010.
[30] Louviere JJ, Hensher DA, Swait JD. Stated Choice Methods: Analysis and Applications. New York: Cambridge University Press, 2000.
[31] Johnson FR, Yang J-C, Mohamed AF. In defense of imperfect experimental designs: statistical efficiency and measurement error in choice-format conjoint analysis. Proceedings of the Sawtooth Software Conference. March 2012. Available at: http://www.sawtoothsoftware.com/download/techpap/2012Proceedings.pdf. [Accessed December 31, 2013].

[32] Marshall D, Bridges JFP, Hauber AB, et al. Conjoint analysis applications in health—how are studies being designed and reported? An update on current practice in the published literature between 2005 and 2008. Patient 2010;3:249–56.
[33] Train K. Discrete Choice Methods with Simulation. Cambridge, UK: Cambridge University Press, 2003.
[34] Train K, Sonnier G. Mixed logit with bounded distributions of correlated partworths. In: Scarpa R, Alberini A,eds. Applications of Simulation Methods in Environmental and Resource Economics. Dordrecht (The Netherlands): Springer Publisher, 2005.
[35] Hensher DA, Rose JM, Greene WH. Applied Choice Analysis. Cambridge, UK: Cambridge University Press, 2005.
[36] Krinsky I, Robb A. On approximating the statistical properties of elasticities. Rev Econ Stat 1986;86:715–9.
[37] Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. Health Econ 2004;13:901–7.
[38] Alzheimer's Disease International. World Alzheimer report 2011: the benefits of early diagnosis and intervention. Available at: http://www.alz.co.uk/research/world-report-2011. [Accessed December 31, 2013].
[39] Vernooij-Dassen MJ, Moniz-Cook ED, Woods RT, et al. Factors affecting timely recognition and diagnosis of dementia across Europe: from awareness to stigma. Int J Geriatr Psychiatry 2005;20:377–86.
[40] Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. Alzheimer Dis Assoc Disord 2009;23:306–14.
[41] Marcantonio ER, Aneja J, Jones RN, et al. Maximizing clinical research participation in vulnerable older persons: identification of barriers and motivators. J Am Geriatr Soc 2008;56:1522–7.
[42] Taylor DH Jr, Østbye T, Langa KM, et al. The accuracy of Medicare claims as an epidemiological tool: the case of dementia revisited. J Alzheimers Dis 2009;17:807–15.
[43] Bond M, Rogers G, Peters J, et al. The effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease (review of Technology Appraisal No. 111): a systematic review and economic model. Health Technol Assess 2012;16:1–470.
[44] Kopits IM, Chen C, Roberts JS, et al. Willingness to pay for genetic testing for Alzheimer's disease: a measure of personal utility. Genet Test Mol Biomarkers 2011;15:871–5.
[45] Green RC, Roberts JS, Cupples LA, et al. REVEAL Study Group. Disclosure of APOE genotype for risk of Alzheimer's disease. N Engl J Med 2009;361:245–54.
[46] Piccini C, Bracco L, Amaducci L. Treatable and reversible dementias: an update. J Neurol Sci 1998;153:172–81.
[47] Neuropathology Group of the Medical Research Council Cognitive Function and Aging Study. Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. Lancet 2001;357:169–75.
[48 Smith RD. Contingent valuation in health care: does it matter how the "good" is described? Health Econ 2008;17:607–17.
[49] Hanley N, Ryan M, Wright R. Estimating the monetary value of health care: lessons from environmental economics. Health Econ 2003;12:3–16.
[50] Hort J, O'Brien JT, Gainotti G, et al. EFNS Scientist Panel on Dementia. EFNS guidelines for the diagnosis and management of Alzheimer's disease. Eur J Neurol 2010;17:1236–48.